Федеральное государственное бюджетное учреждение науки Научно-исследовательский институт системных исследований РАН

ТРУДЫ НИИСИ РАН

TOM 3 N 1

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ СЛОЖНЫХ ТЕХНИЧЕСКИХ СИСТЕМ:

ТЕОРЕТИЧЕСКИЕ И ПРИКЛАДНЫЕ АСПЕКТЫ

MOCKBA 2013

Редакционный совет НИИСИ РАН:

В.Б. Бетелин (председатель), О.М. Белоцерковский, Е.П. Велихов, В.А. Галатенко, В.Б. Демидович (отв. секретарь), П.П. Кольцов, Б.В. Крыжановский, А.Г. Кушниренко, А.Г. Мадера, М.В. Михайлюк, В.Я. Панченко, В.П. Платонов, В.Н. Решетников

Главный редактор журнала: В.Б. Бетелин

Научные редакторы тома:

П.П. Кольцов, Е.П. Волошина, А.Г. Мадера

Тематика номера:

Моделирование, технология, производство в микро- и нано- электронике, информационные технологии

Журнал публикует оригинальные статьи по следующим областям исследований: математическое и компьютерное моделирование, обработка изображений, визуализация, системный анализ, методы обработки сигналов, информационная безопасность, информационные технологии, высокопроизводительные вычисления, микро- и нано- электроника

The topic of the issue:

Simulation, technology, production of micro- and nano- electronics, information technology

The Journal publishes novel articles on the following research arias: mathematical and computer modeling, system analysis, image processing, visualization, signal processing, information security, information technologies, high-performance computing, micro- and nano- electronics

Заведующий редакцией: Ю.Н. Штейников

Издатель: Российская академия наук, НИИСИ РАН, 117218, Москва, Нахимовский проспект 36, к. 1

[©] Федеральное государственное бюджетное учреждение науки Научно-исследовательский институт системных исследований РАН, 2013 г.

2 января 2014 года исполняется 25 лет Научно-исследовательскому институту системных исследований Российской академии наук (НИИСИ РАН). В этой связи было решено приурочить к юбилейной для Института дате издание 3-го тома за 2013 год и 4-го тома за 2014 год журнала «Труды НИИСИ РАН».

Данный № 1 тома 3 журнала «Труды НИИСИ РАН» содержит работы сотрудников Института, объединенные в нескольких тематических рубриках. Помещенные в номере статьи покрывают различные направления исследований в Институте, по которым можно судить о тематической широте работ, проводимых в НИИСИ РАН.

Номер открывается серией статей, посвященных результатам исследований в моделировании, проектировании и производстве субмикронных СБИС. Представленные здесь работы охватывают практически весь цикл проектирования, технологии и производства, который проходит в процессе своего создания СБИС. К указанной тематике примыкают работы по математическому и компьютерному моделированию в микро- и нано- электронике, а также по проблемам, возникающим в сложных технических системах. В номере представлены также работы по информационным технологиям – тематике, традиционной для Института. Новой для журнала «Труды НИИСИ РАН» является рубрика «Из истории науки и техники», которая открывается статьей об истоках создания в нашей стране отечественной школы вычислительной математики и программирования.

Представляется, что как предлагаемый, так и последующие юбилейные номера журнала, вызовут несомненный интерес у специалистов.

Главный редактор журнала академик РАН

В.Б. БЕТЕЛИН

СОДЕРЖАНИЕ

I. ТЕХНОЛОГИЯ И ПРОИЗВОДСТВО СУБМИКРОННЫХ СБИС

Организация производства микроэлектронной продукции

Моделирование технологических процессов

Конструирование элементов СБИС

Технологические процессы изготовления СБИС

Операционный контроль в производстве СБИС

Статистический контроль технологических процессов

Анализ отказов и брака продукции в производстве СБИС

<i>А.А.Захарченко, М.В.Орешков, В.Ю.Троицкий.</i> Анализ дефектов интегральных микросхем с помощью фотонно-эмиссионной микроскопии
В.Ю.Троицкий, Т.В.Козырева, С.В.Седов. Анализ отказов и брака продукции в производстве микросхем технологического комплекса НИИСИ РАН
II. МОДЕЛИРОВАНИЕ В МИКРО- И НАНО- ЭЛЕКТРОНИКЕ
В.А.Горячев. Особенности германиевых технологий КМОПИС
<i>И.А.Люосева</i> . Компьютерное моделирование зависимости характеристик нанотранзисторов от топологии
<i>Н.В.Масальский</i> . Влияние высокой диэлектрической проницаемости подзатворного диэлектрика на характеристики КНИ КМОП нанотранзисторов
III. МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ
А.Г.Мадера, М.Ж.Акжолов, И.Г.Лебо. Моделирование развития процессов конвекции плюс теплопроводность в воздухе вблизи процессора
П.И.Кандалов, А.Г.Мадера. Компьютерное моделирование свободного конвективного теплообмена горизонтальных проволочных теплоотводов
<i>М.Л.Бахмутский</i> . Два алгоритма фильтрации, сглаживания и выделения тренда зашумленных больших временных рядов: Алгоритм 1

IV. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ: МОДЕЛИРОВАНИЕ И ПРИЛОЖЕНИЯ

V. ИЗ ИСТОРИИ НАУКИ И ТЕХНИКИ

В.Б.Демидович,	В.М.Тихомиров.	У	истоков	вычислительной	математики	И	
программирования: Ј	Іазарь Аронович Л	юсте	рник				132

Организация производства микроэлектронной продукции на технологическом комплексе НИИСИ РАН

С.И. Волков¹, Е.П. Волошина

1-кандидат технических наук

В статье рассмотрены вопросы организации производства микроэлектронной продукции на технологическом комплексе НИИСИ РАН. Описаны основные задачи, технологического комплекса. Определены характеристики продукции, подлежащей изготовлению на технологическом комплексе. Подробно рассмотрены проблемы, характерные для мелкосерийного производства, и организационные, производственные и технические решения, позволившие реализовать восемь базовых технологических процессов.

Введение

Технологический комплекс НИИСИ РАН был сформирован в 1998-2001 годах на территории НИЦ «Курчатовский институт». Задачами технологического комплекса были изначально определены:

 разработка, на базе последних достижений мировой микроэлектроники, новых конструктивнотехнологических решений микросхем, разрабатываемых НИИСИ РАН, и их реализация при изготовлении макетов и опытных образцов;

- передача и адаптация разработанных конструктивно-технологических решений на другие отечественные предприятия, изготавливающие микроэлектронную продукцию;

- изготовление микросхем по проектам НИИСИ РАН для использования в радиоэлектронной аппаратуре специального назначения (изделия «частного» или «узкоспециального» применения), разработанной как НИИСИ РАН, так и предприятиями-партнерами.

При реализации поставленных задач, перед НИИСИ РАН возникла еще одна проблема, которой первоначально не предавалось большого значения, однако, которая потребовала принятия целого набора комплексных решений, от административных до сугубо технических. Эта задача – определение формы организации производственного процесса на базе создаваемого технологического комплекса.

В первую очередь, были проведены оценки потребности в продукции технологического комплекса НИИСИ РАН. В результате которых были определены характеристики рынка сбыта продукции технологического комплекса:

 комплектуемая аппаратура представляет собой специализированные вычислительные комплексы, предназначенные для использования в сложных условиях и отличающиеся широкой номенклатурой потребляемых интегральных микросхем;

- комплектуемая аппаратура отличается, как правило, низкими показателями ремонтопригодности, при высоких требованиях к наработке на отказ (а во многих случаях – к наработке до отказа) и высокой «ценой отказа»;

- общий годовой объем потребления функционально сложных микросхем составляет от десятков до нескольких сотен изделий каждого типа, что на порядки величины меньше объемов потребления аналогичных изделий радиоэлектронной аппаратурой промышленного и бытового назначения;

- и, наконец, условия разработки комплектуемой аппаратуры выдвигают очень жесткие требования к длительности периода от формулирования технического задания на микросхему до получения первых образцов.

Исходя из указанных особенностей аппаратуры, были определены основные характеристики продукции, подлежащей изготовлению на технологическом комплексе, а именно:

- номенклатура производимой продукции должна включать изделия с функциональными характеристиками, ориентированными на применение в конкретных образцах радиоэлектронной аппаратуры специального назначения. Выпускаемые изделия не должны дублировать продукцию, соответствующую технологическим возможностям отечественных предприятий того времени, т.е. должны представлять собой современные СБИС – микропроцессоры, микроконтроллеры, оперативные и постоянные запоминающие устройства, цифро-аналоговые и аналого-цифровые преобразователи, устройства обработки ВЧ и СВЧ сигналов;

 конструктивное оформление изготавливаемых изделий должно соответствовать разнообразным требованиям и представлять собой полупроводниковые (бескорпусные и в корпусе) и гибридные микросхемы (в том числе микросборки и многокристальные модули);

- работы, проводимые на технологическом комплексе должны органично совмещать изготовление различных объектов (моделей, макетов, опытных образцов и др.) и изготовление продукции, предназначенной для комплектования аппаратуры;

- объемы производства конкретных изделий должны составлять от десятков до нескольких сотен изделий в год, что соответствует определению единичного или мелкосерийного производства, принятому в отечественной микроэлектронике;

Здесь необходимо отметить принципиальное отличие системы приоритетов принятых при мелкосерийном производстве от традиционно действующей в массовом (как отечественном, так и зарубежном) производстве изделий коммерческого и промышленного назначения.

Так для мелкосерийного производства приоритетными являются:

- проблема морального старения выпускаемых изделий (снижение спроса);

 проблема морального износа материалов и комплектующих изделий (исчезновение с рынка);

 проблема обеспечения максимальной способности производства к реконфигурированию;

 проблема оценки и обеспечения надежности изделий.

Остановимся на некоторых из них подробнее.

Проблема морального старения изделий

Узкий круг потребителей, проведя модернизацию изготавливаемой ими специализированной аппаратуры, закрывает рынок сбыта для ставшего ненужным для них комплектующего изделия. Поэтому способность минимизировать время удовлетворения потребностей разработчиков и изготовителей радиоэлектронной аппаратуры является едва ли не главным фактором, определяющим «живучесть» предприятия на этом рынке. Для массового же производства, ориентированного на широкий круг применения в различных видах аппаратуры, уход с рынка отдельных потребителей и даже целых их групп, оказывает незначительное влияние на общий уровень спроса. В этом случае изготовитель имеет значительный запас времени, чтобы учесть изменение тенденций на рынке и перепрофилировать свое производство.

Проблема морального износа материалов и комплектующих изделий

Доля микросхем для производства специализированной радиоэлектронной аппаратуры, составляет, в настоящее время, менее одного процента рынка. В связи с этим предприятия мелкосерийного производства полностью утратили возможность оказывать давление на мировой рынок сбыта микроэлектронных материалов и комплектующих изделий. Этот рынок полностью подчинен производителям массовой коммерческой и промышленной продукции и меняется в соответствии с их потребностями, игнорируя запросы единичных и мелкосерийных производств. Как следствие, предприятия мелкосерийного производства вынуждены приспосабливаться к изменениям или создавать запасы.

Проблема обеспечения максимальной способности производства к реконфигурированию

Чтобы оперативно реагировать на изменения потребностей комплектуемой аппаратуры производственный процесс и конструкция производимых в нем изделий требуют «нарезки» на отдельные «независимые» модули таким образом, чтобы реакция на эти изменения осуществлялась путем замены только одного (или ограниченного количества) «модулей» (например, блоков технологического процесса или отдельных элементов изделия), пусть даже с некоторым ущербом для функциональных и эксплуатационных характеристик изделий. Массовое же производство, наоборот, имеет возможность «затачивать» выделенные производственные мощности, конструктивные решения и технологию под одно конкретное изделие, добиваясь наиболее эффективного сочетания его функциональных и эксплуатационных характеристик, с одной стороны, и экономических показателей, с другой.

Проблема обеспечения и оценки надежности изделий

Длительность сроков активного существования специальной радиоэлектронной аппаратуры, комплектуемой изделиями единичного и мелкосерийного производства, при высокой, как правило, цене отказа такой аппаратуры и ее низкой ремонтопригодности, ставит перед изготовителем комплектующих изделий задачи, зачастую, неизвестные изготовителям изделий массового производства. В частности, объемы единичного и мелкосерийного производства слишком малы даже для того, чтобы отобрать достаточное количество изделий для проведения их контрольных испытаний. Проблема надежности изделий массового производства не может быть отнесена к разряду приоритетных по нескольким причинам. Прежде всего, радиоэлектронная аппаратура, комплектуемая изделиями массового производства, предъявляет, как правило, относительно невысокие требования к их надежности. Во-вторых, объемы массового производства позволяют, при недостатке понимания механизмов и первопричин отказов изделий, проводить разбраковку продукции и отбор изделий, отвечающих требованиям достаточных конструктивно-технологических запасов. И, наконец, объемы массового производства позволяют формировать репрезентативные выборки для проведения контрольных испытаний продукции (в том числе, разрушающих).

Напротив, в ряду приоритетных проблем при организации массового производства находятся:

- проблема захвата рынка;

- проблема удовлетворения непрерывно возрастающих требований потребителей;

проблема технологичности;

 проблема обеспечения привлекательности (сложности) проекта и его отработанности;

- проблема обеспечения контролепригодности изделий.

Проблема захвата рынка

Предприятие массового производства, вышедшее на рынок первым, длительное время занимает на нем лидирующее положение, даже при появлении на рынке изделий, обладающих лучшими функциональными и эксплуатационными характеристиками. Таким образом, для массового производства приоритетной задачей является сокращение, любой ценой, периода разработки продукции. Это условие является ключевым в системе приоритетов массового производства. Напротив, эта проблема не может быть отнесена к приоритетным в условиях единичного и мелкосерийного производства, изначально нацеливаемого на узкий сектор рынка, существующий в условиях минимальной конкуренции.

Проблема удовлетворения непрерывно возрастающих требований потребителей к функциональным и эксплуатационным характеристикам изделий

Следствием проблемы, рассмотренной выше, является необходимость непрерывного обновления технологических возможностей изготовителя и парка его технологического оснащения, в том числе, за счет отказа от «вчерашних» методов и средств производства. Хотя эта проблема существует и для мелкосерийного производства, однако, по сравнению с коммерческой аппаратурой она не может быть отнесена к разряду приоритетных.

Проблема технологичности

Технологичность изделий массового производства должна быть «встроена» в их проекты, и высокие значения выхода годных должны быть достигнуты немедленно за счет высокой робастности изделий по отношению к изменчивости процесса пусть и достигаемой путем снижения эксплуатационных характеристик (например, конструктивно-технологических запасов). Такой подход вполне допустим вследствие мягких условий применения изделий в бытовой, а зачастую, и в промышленной радиоэлектронной аппаратуре и «низкой цены» ее отказов. Напротив, при мелкосерийном производстве изделий при конфликте интересов обеспечения технологичности и надежности, проблема технологичности отходит на второй план.

Проблема обеспечения сложности проекта и его отработанности

Необходимость исключить недоработки сложного проекта находится в непримиримом противоречии с условием максимального сокращения времени выхода на рынок в условиях массового производства. Как правило, выбор делается в пользу минимального времени выхода на рынок пусть и «сырого», но сложного проекта. Такой подход тоже оправдан низкой «ценой отказа». Комплектование же радиоэлектронной аппаратуры с высокой «ценой отказа» требует от изготовителя мелкосерийной продукции тщательной отработки проекта изделия.

Проблема обеспечения контролепригодности изделий

Разработка эффективных способов обнаружения дефектных изделий (в том числе, программ тестирования) требует существенного времени и средств, что находятся в непримиримом противоречии с ключевым приоритетом массового производства – минимальным временем выхода новых изделий на рынок и минимизацией их себестоимости. Понятно, что и в этом случае выбор делается в ущерб обеспечению контролепригодности в пользу минимального времени выхода на рынок продукции, пусть и несколько «засоренной» дефектными изделиями. Это оправдано низкой «ценой отказа» изделий и комплектуемой ими радиоэлектронной аппаратуры. Создание же радиоэлектронной аппаратуры специального назначения с высокой «ценой отказа» требует принятия всех возможных мер по исключению из поставляемой продукции комплектующих изделий с дефектами, пусть даже с некоторым ущербом в части удлинения технологического цикла изготовления.

Таким образом, организацию процесса единичного и мелкосерийного производства изделий специального назначения на технологическом комплексе НИИСИ РАН пришлось проводить с учетом следующих соображений:

- необходимости минимизации сроков «от заказа до поставки» (то есть, фактически, минимизации итерационных циклов отработки проектных решений при разработке конкретных изделий, что в зарубежной микроэлектронике называют «успехом с первой попытки (first pass success)»);

 необходимости выполнения производственного плана в условиях расширения технологических возможностей производства, а значит, его периодической приостановки для реконфигурации;

- необходимости гармонизации производства серийной продукции с одновременным изготовлением макетов и опытных образцов, проводимым в рамках выполнения НИР и ОКР;

 необходимости обеспечения и контроля качества продукции в условиях «рваного» производственного цикла и малых объемов производства отдельных изделий;

- необходимости обеспечения стабильного функционирования производства в условиях отсутствия резервирования средств технологического оснащения при их конечной надежности и непрерывного морального износа штатных запасных частей и ремонтных материалов.

Указанные соображения были учтены при организации производства путем реализации целого комплекса организационных, производственных и технических решений.

Ключевым из этих решений, вокруг которого «вращаются» все остальные, является опора на огра-

ниченное количество базовых технологических процессов. При этом, под термином «базовый технологический процесс» всегда понималось неразрывное единство технологического маршрута, режимов технологических операций (или, в отдельных случаях - правил их определения), правил проектирования микросхем, подлежащих изготовлению в этом процессе (сопровождаемых, как правило, библиотекой стандартных элементов) и унифицированного инструмента для мониторинга этого процесса - так называемого, параметрического монитора. Все базовые технологические процессы, разработанные для реализации на технологическом комплексе НИИСИ РАН, нацелены на изготовление КМОП микросхем и опираются на общую для этих процессов, унифицированную совокупность технологических решений по формированию ключевых конструктивно-технологических блоков и элементов микросхем, включающую, в том числе, уникальные решения, не имеющие прямых аналогов на зарубежных микроэлектронных предприятиях.

Принятые технологические решения обеспечили очень высокую робастность разработанных процессов, и, как следствие, высокие конструктивно-технологические запасы изготавливаемых изделий, позволившие обеспечить устойчивое изготовление изделий с высокими показателями стойкости к воздействию внешних факторов, в том числе, с показателями радиационной стойкости, близкими к экстремальным. Устойчивость производства таких изделий обеспечена, кроме того, мероприятиями по поддержанию работоспособности оборудования технологического комплекса, позволившими, в том числе обеспечить значения показателей его надежности (наработки на отказ), в среднем, в 1,5 раза более высокие, чем справочные значения, опубликованные изготовителями соответствующих единиц этого технологического оборудования.

В результате реализации указанных организационных, производственных и технических решений в НИИСИ РАН к 2012 году сформирован технологический комплекс для изготовления СБИС, обеспеченный необходимой производственной инфраструктурой и реализующий, в нынешней конфигурации, восемь базовых технологических процессов:

 - базовые технологические процессы изготовления логических КМОП СБИС на основе эпитаксиальных структур с минимальным топологическим размером 0,5 мкм, 0,35 мкм и 0,25 мкм; - базовые технологические процессы изготовления логических КМОП СБИС на основе структур «кремний на изоляторе» с минимальным топологическим размером 0,5 мкм, 0,35 мкм и 0,25 мкм;

 - базовые технологические процессы изготовления КМОП СБИС смешанного сигнала на основе эпитаксиальных структур с минимальным топологическим размером 0,35 мкм и 0,25 мкм.

При этом правила проектирования, разработанные для базовых процессов и сопровождающие их библиотеки стандартных элементов, опираются только на проектные решения, проверенные (и зарекомендовавшие себя) на всех уровнях проектирования микросхем (от архитектурного и вплоть до топологического), в том числе, на решения, использованные для базовых технологических процессов, предшествующих по срокам внедрения.

В этих базовых процессах освоено и серийно выпускается более 20 типов микросхем (в том числе бескорпусные микросхемы и многокристальный модуль) до VIII уровня интеграции.

При этом для микросхем, разработанных в строгом соответствии с правилами проектирования для базового процесса на основе эпитаксиальных структур с минимальным топологическим размером 0,35 мкм достигнуто, независимо от их функционального назначения, не только высокое, но и гарантированно воспроизводимое значение уровня радиационной стойкости $3V_C$ по фактору $7H_7$, а для микросхем, разработанных для базовых процессов на основе структур «кремний на изоляторе» с минимальными топологическими размерами 0,5 мкм и 0,35 мкм – еще более высокое, но столь же устойчивое значение более $4V_C$.

Кроме того, опора на принцип минимизации конструктивно-технологических решений на основе ограниченного количества базовых процессов позволила освоить на том же технологическом комплексе, параллельно с изготовлением микросхем, новую сферу деятельности – работу в режиме «кремниевой мастерской» – foundry (изготовление, в базовых технологических процессах так называемых, пластин с кристаллами заказанных элементов по заказам сторонних предприятий-партнеров – изготовителей микросхем).

В этом режиме проведено освоение более 20 типов изделий (пластин с кристаллами заказанных элементов), обеспечивших, в качестве полуфабрикатов, серийное производство более 25 типов микросхем на предприятиях-партнерах.

Organization of Manufacture of Microelectronic Production with Technological Complex of SRISA

S.I. Volkov, E.P. Voloshina

Abstract. The article considers organization of manufacture of microelectronic production with technological complex of SRISA. Major goals of the technological complex are indicated. The basic characteristics of the products to be manufactured with the technological complex are described. Specific problems of small-scale manufacture and organizational, pr oduction and engineering decisions are in detail considered. This makes realizable to create eight base technological processes.

Особенности методов проектирования СБИС с учетом результатов моделирования технологического процесса

А.В. Амирханов¹, А.А. Гладких, А.А. Глушко², Е.П. Михальцов², И.А. Родионов², А.А. Столяров

1-кандидат физико-математических наук, 2 – кандидат технических наук.

В статье рассмотрены особенности методов проектирования СБИС и разработки технологического процесса их изготовления с применением средств конструктивно-технологического моделирования. Подробно описаны основные блоки проектирования, а также задачи, решаемые в рамках каждого из блоков. В заключительной части определен круг задач, которые предстоит решить в ближайшем будущем.

1. Введение

За прошедшие 10-15 лет минимальные размеры элементов транзисторов СБИС были уменьшены с 0,35 мкм до 22 нм, что соответствует нескольким различным технологическим шагам. При этом на каждом шаге развития технологии [1] растет количество элементов СБИС, количество технологических операций увеличивается, а сами они усложняются, повышается точность и воспроизводимость, учитываются и контролируются все более сложные физические эффекты.

Следует отметить, что на сегодня в полупроводниковой технологии очень сложно отделить технологию, технологическое оборудование и конструкцию элемента (прибора), поскольку технологическое оборудование разрабатывают под конкретную технологию и наоборот, а конструкцию выбирают с учетом особенностей техпроцесса, при этом технологические режимы зачастую выбирают под конкретное изделие. Таким образом, технологические процессы изготовления СБИС можно представить, как сложную систему, которой необходимо точно управлять, чтобы получить требуемый результат.

В данной ситуации методы проектирования СБИС и разработки новых технологических процессов и отработки технологических режимов, основанные только на экспериментальных подходах, неприменимы ни с экономической точки зрения, ни по критерию затраченного времени.

Наиболее правильным является использование приемов конструктивно-технологического (далее встречается термин конструкторско-технологическое) моделирования для исследования сложных физических явлений, отработки технологических режимов, а также оптимизации конструкции элементов СБИС с учетом влияния особенностей процесса их формирования.

Основные ограничения на конструкцию (размеры и форму) элементов субмикронных СБИС связаны с процессами проекционной литографии, результат которых в высокой степени зависит от топографии подложки (кремниевой пластины). Так как глубина резкости современных литографических процессов ограничена десятыми долями микрона, неплоскостность поверхности, на которой формируется фоторезистивная маска, должна быть такого же порядка. Это достигается за счет проведения процессов химико-механической планаризации (далее ХМП) поверхности перед литографическими процессами. Наиболее ярко взаимосвязь этих двух технологических процессов проявляется при формировании многоуровневой металлизации СБИС.

Однако, в конечном счете, результаты отдельных технологических операций, хотя и являются важными составляющими, но не дают представления о параметрах конечного продукта, т.е. самой СБИС. Поэтому для полноценного проектирования и разработки СБИС необходимо суммировать результаты всех технологических операций, учитывать их особенности и на выходе предсказывать электрические характеристики разрабатываемых приборов. С этой целью используются САПР техпроцессов изготовления СБИС и моделирования их элементов [1].

В НИИСИ РАН при проектировании СБИС применяются два крупных блока конструктивнотехнологического моделирования, непосредственно связанных между собой:

 моделирование процесса ХМП и финишная модификация топологии СБИС перед изготовлением фотошаблонов с учетом результатов такого моделирования;

 моделирование процесса проекционной литографии и топологической коррекции эффекта оптической близости и финишная модификация топологии СБИС перед изготовлением фотошаблонов при изготовлении СБИС с проектными нормами, меньшими длины волны источника экспонирующего излучения проекционной установки (степпера).

Кроме того, специалистами МГТУ им. Н.Э. Баумана [15] предложена методика трехмерного моделирования элементов СБИС, а также исследования принципов их функционирования и влияния на них внешних воздействующих факторов, в том числе, специальных. Эту методику предполагается использовать в НИИСИ РАН для прогноза электрических характеристик элементов СБИС с учетом результатов моделирования процессов XMП и проекционной литографии.

В работе рассмотрена новая модель XMП для технологического процесса многоуровневой металлизации при изготовлении СБИС. Проведен сравнительный анализ методов калибровки этой модели и оптимизации топологии СБИС с ее использованием. Разработаны алгоритмы финишной коррекции топологии СБИС с минимальными проектными нормами 0,35 и 0,25 мкм путем введения фиктивных структур заполнения.

Также в работе представлены модели технологических процессов изготовления СБИС по эпитаксиальной и КНИ-технологиям с проектными нормами 0,5, 0,35 и 0,25 мкм. Предложены методы калибровки, улучшения сходимости Ньютоновских итераций, формирования расчетной сетки конечных элементов, построения конструкции элементов КНИ КМОП СБИС произвольной формы. Полученные результаты используются при разработке новых субмикронных СБИС.

1. Моделирование процесса ХМП

Многоуровневая металлизация в субмикронных СБИС представляет собой несколько слоев металлических межсоединений разделенных изолирующими слоями, называемыми межслойным диэлектриком (далее МСД). Формирование многоуровневой металлизации может осуществляться с использованием субтрактивного варианта технологического процесса. В таком технологическом процессе первоначально на поверхность пластины наносят сплошной проводящий слой, в котором с помощью операций литографии и последующего травления открытых участков получают требуемый рисунок межсоединений.

После получения рисунка проводящего слоя на него осаждают изолирующий слой (обычно изолятором является диоксид кремния), причем в процессе осаждения над проводниками образуются локальные ступеньки (рельеф). Если не принять соответствующих мер, то формирование следующего проводящего слоя будет осуществляться на поверхности со значительным рельефом (десятые доли микрона), что недопустимо для литографических процессов формирования топологии слоя. Также возникают сложности при плазмохимическом травлении проводящего слоя на рельефе, что в конечном счете сказывается на выходе годных и надежности изделия.

Для удаления локальной ступеньки, возникающей при осаждении диоксида кремния, используют технологическую операцию ХМП, в силу особенностей которой, на поверхности кристалла СБИС, тем не менее, возможно сохранение рельефа поверхности (глобальной ступеньки).

С учетом особенностей используемого технологического процесса была предложена полиномиальная модель операции ХМП [2]. Данная модель позволила повысить точность расчетов, что было подтверждено результатами ее калибровки по экспериментальным данным. Для примера на рисунке 1 представлено сравнение двух моделей: линейной, разработанной в работе





Рисунок 1. Сравнение экспериментальных данных с результатами моделирования: (а) – линейная модель (б) – полиномиальная модель

[3], и предложенной полиномиальной модели операции ХМП.

Наличие точных моделей операции ХМП позволило проводить модификацию топологии СБИС с целью снижения общего разброса толщин (далее ОРТ) межслойного диоксида кремния.

Метод основан на том, что практически в каждой топологии СБИС существуют области, не занятые функциональными элементами, что, в частности, и вызывает разброс плотности заполнения слоя. Размещая в данных областях фиктивные структуры заполнения (СЗ), которые не изменяют электрическую схему устройства, можно добиться снижения разброса толщины МСД. Для определения расположения и параметров СЗ для каждой уникальной топологии СБИС целесообразно использовать моделирование операции ХМП. В работах [4, 5] был предложен алгоритм модификации топологии СБИС, основанный на способе, предложенном в работе [6] и разработан комплекс мо-



Рисунок 2. Блок схема блока моделирования операции ХМП

делирования операции ХМП. Блок схема комплекса приведена на рисунке 2.

При реализации алгоритма необходим многократный расчет локальной плотности заполнения. Ее вычисление является ресурсоемкой задачей, так как при прямом расчете, например, для 0,35 мкм технологии, может потребоваться выполнение порядка 7.75×10¹⁰ операций пересечения полигонов, представляющих топологию слоя СБИС, с квадратами расчета локальной плотности заполнения. Поэтому в работах [5, 7] был предложен алгоритм рекурсивного разбиения полигонов топологического слоя, позволяющий существенно сократить число операций пересечения. В этих работах показано, что количество операций пересечения R при использовании предложенного алгоритма снижается до К×N/4^r, где: N – число квадратов расчета локальной плотности, К – число неперекрывающихся и непересекающихся многоугольников топологического слоя, r – глубина рекурсии. Так при r = 5, число операций уменьшается в 1 000 раз, а для приведенного выше случая r ≈ 8 – в 65 000 раз (подробнее см. [7]).

С использованием предложенного алгоритма проведены модельные исследования по заполнению топологии различных типов СБИС. Для сравнения проведено моделирование топологий кристаллов СБИС со стандартным заполнением (свободные места топологии заполняются СЗ равного размера с равными расстояниями между ними) и с заполнением по разработанному алгоритму на основе моделирования. На рисунке 3 приведены результаты моделирования распределения МСД кристалла микропроцессора (МП) после операции ХМП до (слева) и после оптимизации с помощью предложенного алгоритма (справа).

Из приведенного рисунка и гистограммы хорошо видно качественное снижение разброса толщины межслойного диоксида кремния. Количественные оценки результатов работы алгоритма для 20 различных кристаллов СБИС приведены на рисунке 4. Кроме кристаллов МП и статических оперативных запоминающих устройств (далее СОЗУ, в данном случае статиче-



Рисунок 3. Результаты заполнения топологического слоя СБИС микропроцессора №4 (см. рисунок 4)

(а) – стандартное заполнение, (б) – на основе моделирования,
(в) – гистограмма распределения толщины МСД (черным – на основе моделирования, белым – стандартное заполнение)



Рисунок 4. Результат моделирования топологических слоев различных СБИС после их заполнения

ской памяти) проверка алгоритма проводилась и на топологии тестовых кристаллов (далее ТК). Топология такого типа кристаллов отличается наличием локальных (размером 30-40% от площади кристалла) высокоплотных областей и разреженной остальной площадью кристалла.

Из рисунка 4 видно, что при использовании разработанного алгоритма для большинства кристаллов величина ОРТ диоксида кремния значительно снизилась. Исключения составляет топология СОЗУ, которая содержит блоки памяти высокой плотности заполнения. Эти блоки занимают практически всю площадь кристалла (более 90%), поэтому оптимизация топологии структурами заполнения практически невозможна.

Таким образом, в среднем ОРТ диоксида кремния для топологии микропроцессоров можно снизить на 36% только за счет оптимизации размещения СЗ в топологических слоях СБИС. Максимальное снижение разброса толщины МСД составляет 57%.

За счет полученного снижения ОРТ, т.е. улучшения неплоскостности подложки для выше лежащего слоя, были созданы условия для проведения литографических процессов с соблюдением требований к воспроизводимости минимальных линейных размеров. Задачей дальнейшей работы было добиться воспроизведения формы топологического рисунка с размерами меньшими длины волны источника экспонирующего излучения проекционной установки. Результаты решения этой задачи описаны во второй части статьи.

2. Моделирование процесса проекционной литографии и топологической коррекции эффекта оптической близости

Моделирование процесса литографии необходимо рассматривать в двух направлениях (рисунок 5), задачи которых пересекаются только на первый взгляд [1]. К первому отнесем, так называемое, «процессное» (полное физическое) моделирование, при котором последовательно обсчитываются все физико-химические явления от нанесения резиста до проявления/травления. Второе – «конструкторско-технологи-



Рисунок 5. Классификация литографических САПР

ческое» моделирование, при котором применяется математически упрощенное описание всех процессов, оптимизированное по времени счета.

В «процессных» симуляторах на первом этапе моделируют процессы нанесения и сушки резиста, на втором – поглощение света резистом в каждой его точке с заданной точностью сетки, и на последнем этапе – процесс проявления резиста, который, с точки зрения математического описания, является одним из наиболее сложных.

Такие модели требуют досконального понимания всех физических явлений процесса литографии и экспериментальных результатов, используемых для калибровки [8, 9], обеспечивают требуемую точность в диапазоне процессных параметров, а также крайне ресурсоемки с точки зрения вычислительных мощностей и СОЗУ. Так для моделирования полного слоя кристалла СБИС (технологии 0,25 мкм) с сеткой 10 нм потребуется вычисление $5,4 \times 10^{13}$ значений с плавающей точкой, что возможно только на кластерных вычислителях с сотнями ЦП и терабайтами СОЗУ.

В «конструкторско-технологических» симуляторах применяют алгоритмы ускоренных вычислений [10], которые частично учитывают физику процессов, при этом обладая точностью, достаточной для передовых полупроводниковых технологий.

Калибровка таких моделей не требует глубоких знаний о технологии, модель калибруется на основе измерений размеров сотен (тысяч) различных топологических структур для выбранного технологического режима, а вычисление результатов процессов литографии и травления полного слоя (для технологии 0,25 мкм) может быть выполнено за несколько часов на 4-8-ядерном сервере с 32 Гб СОЗУ.

С использованием «процессных» симуляторов проводят оптимизацию технологических режимов и параметров топологической коррекции, что заметно удешевляет исследовательские работы и снижает загрузку оборудования. В свою очередь «конструкторско-технологические» симуляторы используют для внедрения методов проектирования СБИС с учетом возможностей технологического процесса, что позволяет поднять выход годных кристаллов СБИС и улучшить их характеристики (частота, потребление и т.п.).



Рисунок 6. Экспериментальное получение колеоательных кривых

2.1. «Процессный» симулятор

В рамках работы для сверхконтрастного резиста Ultra-i123 (США) определены ключевые параметры модели резиста (доза Е0, коэффициент преломления, константа скорости экспонирования (рисунок 6), максимальная и минимальная скорости проявления) и получены модели литографического процесса, которые удовлетворяют точностным требованиям в заданных диапазонах изменения процессных параметров.

По результатам моделирования установлены и экспериментально подтверждены режимы литографического процесса с проектными нормами 0,25 мкм (параметры стека, числовая апертура, параметры внеосевого освеще-



Рисунок 7. Экспериментальная и промоделированная структура с соотношением шина/зазор 0,24/0,24 мкм

ния, фокусное расстояние, доза экспонирования) с использованием проекционной установки ASML PAS5500/250C (длина волны источника 0,365 мкм).

Промоделированы колебательные кривые для мини-



Рисунок 8. Результат моделирования топологии ячейки СОЗУ

мального критического размера (0,24 мкм) в зависимости от литографических режимов. Экспериментально подтверждена возможность проработки групповых структур с соотношением шина/зазор равным 0,24/0,24 мкм и шагом 0,48 мкм (рисунок 7) на проекционной установке с длиной волны источника 0,365 мкм.

По результатам моделирования установлены и экспериментально подтверждены параметры фигур коррекции эффекта оптической близости [11]. На примере ячейки СОЗУ подтверждена эффективность их применения [12] (рисунок 8).



Рисунок 9. Ошибка моделирования при калибровке модели для слоя поликремниевых затворов



Рисунок 10. Сравнение результатов моделирования и эксперимента (слой поликремниевых затворов, резист)

Проведено исследование различных подходов к

2.2. «Конструкторско-технологический» симулятор

Разработаны методики и откалиброваны векторные оптические модели высокоапертурных проекционных систем, а также полуэмпирические модели про-



Рисунок 11. Верификация моделей

цессов литографии и травления для критических топологических слоев техпроцесса с проектными нормами 0,25 мкм (рисунки 9, 10).



нисунок 12. улучшение электрических характеристин Н-транзистора с откорректированной топологией



Рисунок 13. Приращение тока Id (эксперимент) откорректированных Н- и О-транзисторов по сравнению со стандартными в зависимости от ширины канала



Рисунок 14. Топология ячейки памяти (слева), SEM снимок резистивной маски без коррекции (в центре) и с коррекцией (справа)

разработке полуэмпирических моделей [10, 12, 13].

Предложены и проведены 3 стадии верификация разработанных моделей литографии и травления, включая верификацию по сложным топологическим структурам (рисунок 11).

Экспериментально доказана эффективность внедрения разработанных методов как на примере транзисторов различных типов (рисунки 12, 13), так и на примере элементов 0,25 мкм СОЗУ (рисунок 14).

Экспериментально доказано повышение тока Id откорректированных О- и Н-транзисторов за счет улучшения формы канала на величину от 5 до 17 % в зависимости от ширины канала.

Визуально сложно оценить улучшение воспроизведения топологии, однако для подтверждения достаточно отметить, что кристаллы СБИС СОЗУ (4 Мбит) с откорректированной топологией в среднем работают на частоте на 8% выше.

Разработанный комплекс методов позволил осуществить переход к меньшим проектным нормам при обеспечении требуемого уровня стабильности процесса литографии (процессного окна с глубиной фокусировки $\approx 0,6-0,8$ мкм при фиксированном значении диапазона дозы экспонирования $\approx 8-10\%$), повысить быстродействие за счет уменьшения размеров получаемых элементов (около 30%), повысить степень интеграции СБИС за счет уменьшения размеров топологических элементов (до 40%). Универсальность разработанных методик и используемых моделей позволяет применять их на других предприятиях аналогичного профиля.

После получения возможности достоверного предсказания топологии основных технологических слоев следующей крупной задачей стояла необходимость расчета электрических характеристик приборов на основании данных об их конструкции и технологии изготовления. Рассмотрение этой задачи приведено в третьей части статьи.

3. Модели и методы моделирования процесса формирования структур и их электрических характеристик

Все рассмотренные примеры приведены для одной из наиболее актуальных задач – трехмерного конструкторско-технологического моделирования КНИ- структур («кремний на изоляторе») СБИС с учетом особенностей технологического процесса их формирования [14]. В частности, при достижении размеров, меньших длины волны источника экспонирующего излучения степпера, необходимо учитывать искажения топологии элементов СБИС в процессе литографии [1].

В ходе работы предложены методы и алгоритмы трехмерного моделирования процесса формирования элементов СБИС и расчета их электрических характеристик с использованием САПР приборно-



Рисунок 15. Конфигурация канала и сетки при моделировании КНИ МОП-транзистора 1-го (сверху) и 2-го (снизу) типов из ячейки памяти СОЗУ с минимальными проектными нормами 0,25 мкм

технологического моделирования и результатов литографического моделирования [15, 16]. Предложен метод построения сетки конечных элементов с привязкой к топологии затвора и возможностям САПР по реализации такого вида сетки. Примеры построения сетки транзисторов с различной конфигурацией канала представлены на рисунках 15 и 16.

При этом литографическое моделирование выполняется для целого фрагмента СБИС, например, ячейки статической оперативной памяти. На всех рисунках показана конфигурация канала, причем синий цвет и его оттенки соответствуют акцепторному типу приме-



Рисунок 16. Конфигурация канала и сетки при моделировании КНИ МОП-транзистора 3-го типа из ячейки памяти СОЗУ с минимальными проектными нормами 0,25 мкм

си, красный цвет и его оттенки – донорному. Наиболее мелкой сетка является в области изменения концентрации и при смене типа примеси.

Трехмерное моделирование позволяет оценить и эквивалентную ширину канала транзисторов, составляющих ячейку памяти, требуемую при схемотехническом моделировании ячейки памяти. Разработан метод определения эквивалентной ширины канала на основе линейной аппроксимации зависимости тока насыщения транзистора от геометрической ширины канала.

О точности используемых моделей, методов и подходов можно судить на основе сопоставления статических электрических характеристик, полученных в







Рисунок 18. Семейство выходных характеристик р-канального КНИ МОП-транзистора

результате моделирования, и экспериментальных данных. Достигнуто расхождение между расчетными и экспериментальными характеристиками, не превышающее 10%. Эти результаты получены для КНИ МОП-транзисторов различных конструкций с учетом литографических искажений топологии затвора. На рисунках 17 и 18 представлены семейства расчетных статических выходных характеристики одного из ти-



Рисунок 19. К анализу результатов моделирования растекания токов при проектировании сложных элементов СБИС

пов КНИ МОП-транзисторов в сравнении с экспериментальными данными (точками – экспериментальные данные). На рисунках показаны 10%-е планки погрешностей.

САПР конструкторско-технологического моделирования позволил «заглянуть внутрь» исследуемого элемента СБИС, отслеживая изменения электрических характеристик при вариации параметров технологического процесса и геометрических параметров в топологии транзистора. Так, к примеру, на рисунке 19 представлены результаты моделирования распределения тока электронов в транзисторах с различной геометрией затвора (контур затвора выделен жирным).

На рисунках 19 а, б, в – распределение токов для идеализированных транзисторов с различной топологией затвора (в предположении абсолютно точного переноса топологии затворов). На рисунках 19 г, д, е – распределение токов транзисторов, форма топологии которых рассчитана с использованием литографических моделей: 19 г – проведена коррекция на основе моделей, 19 д – проведена коррекция на основе правил, на рисунке 19 е – без применения коррекции.

В результате установлено, что искажение топологии транзисторов в ходе литографического процесса существенно влияет на электрические характеристики КНИ МОП-транзисторов. Добавление «скоса» в топологию (геометрического перехода по углом в 45° из вертикальной части затвора в нижнюю горизонтальную часть), присутствующего в реальных структурах, дает по самым оптимистичным прогнозам почти 7%-е снижение эквивалентной ширины канала транзистора (на практике из-за общего искажения топологии снижение эквивалентной ширины канала еще больше).

Моделирование также позволило оценить разработанные методы коррекции эффекта оптической близости в топологии СБИС на основе моделей и правил. Их можно считать эквивалентными, поскольку электрические характеристики транзисторов, полученные этими методами, отличаются примерно на 0,5%. А учитывая



Рисунок 20. Структура КНИ МОП-транзистора с сильным донным легированием акцепторной примесью под областью легирования истока: 1 – область истока; 2 – область сильного донного легирования; 3 – область затвора; 4 – область стока; 5 – область кармана; 6 – скрытый окисел

кардинальную разницу во времени проведения процедуры коррекции (почти на порядок), для техпроцесса с проектными нормами 0,25 мкм был обосновано сделан выбор в пользу более простого метода коррекции на основе правил.

Также были промоделированы и новые конструкции КНИ МОП-транзисторов, обладающих повышенной стойкостью к воздействию специальных факторов. В их основе лежит сильное донное легирование акцепторной примеси под областью легирования истока донорной примесью (рисунок 20).

Таким образом, разработаны методы и подходы к моделированию процесса формирования КНИ МОПструктур, которые нашли широкое применение при проектировании новых конструкций транзисторов.

Следует отметить и несколько наиболее важных задач, связанных с моделированием КНИ МОПтранзисторов, которые еще предстоит решить совместно специалистам НИИСИ РАН и МГТУ им. Н.Э. Баумана. К ним относятся следующие задачи.

1. Моделирование эффектов пробоя и калибровка модели генерации носителей. Моделирование пробоя особенно важно для технологий так называемого «глубокого субмикрона» (минимальные проектные нормы 0,25 мкм), когда вследствие бокового легирования области истока и стока оказываются на расстоянии порядка 0,15 мкм, что накладывает существенные ограничения на напряжение питания транзисторов (не более 4 В из-за опасности пробоя транзистора). Наибольшей сложностью моделирования процесса пробоя является то, что подключение моделей ударной ионизации усложняет сходимость Ньютоновских итераций. В случае усложнения сходимости процесса счета предполагается выполнить ступенчатое приближение, изменяя скорость генерации носителей от нуля до ее значения при заданных граничных условиях.

2 Моделирование динамических характеристик КМОП-элементов. В простейшем случае расчет динамических параметров должен быть произведен на примере КМОП-инвертора. Однако эта задача сопряжена с существенными вычислительными затратами: для расчета должны использоваться мощные вычислительные 64-разрядные серверы с объемом оперативной памяти не менее 16 Гбайт.

3 Получение параметров SPICE-моделей транзисторов: определение эквивалентной длины и ширины канала. Дело в том, что полученная по разработанной методике эквивалентная ширина канала не может быть напрямую введена в SPICE-модель транзистора, поскольку в эту модель вводятся эмпирические коэффициенты, позволяющие рассчитать ширину канала транзистора с привязкой к его геометрическим размерам [16]. Аналогично необходимо разработать методику определения эквивалентной длины канала для последующего ввода эмпирических коэффициентов вычисления длины канала в SPICE-модель транзистора. Решение этой задачи особенно важно на этапе долгосрочного прогноза характеристик технологического процесса еще до изготовления тестового кристалла.

4 Получение параметров технологического процесса по желаемым электрическим характеристикам. Это позволит, например, добиться максимума порогового напряжения паразитного КНИ МОП-транзистора при отклонении параметров основного транзистора не более, чем на 10%. Реализация алгоритма параметрической оптимизации предполагает как использование стандартных средств САПР, так и дополнительно разработанных управляющих модулей на языке TCL. Эта задача частично решена в работе [17].

Решение перечисленных задач расширит представление о физических процессах, протекающих в КНИ МОП-структуре, и позволит всесторонне исследовать ранее спроектированные структуры. Кроме того, позволит максимальным образом использовать возможности моделирования при проектировании СБИС, что ускорит процесс проектирования, и позволит значительно уменьшить количество экспериментальных партий, путем заменены их менее дорогостоящим и более быстрым вычислительным экспериментом [18].

Литература

1. А.А. Глушко, И.А. Родионов, В.В Макарчук. Моделирование технологии изготовления субмикронных КМОП СБИС с помощью систем ТСАD. // Технология и конструирование в электронной аппаратуре, 2007, № 4, стр. 32-34.

2. А.В. Амирханов, А.А.Гладких, В.В.Макарчук, А.Г. Пшенников, В.А. Шахнов. Полиномиальная модель химико-механической планаризации в производстве субмикронных СБИС. // Вестник МГТУ им. Н. Э. Баумана Серия «Приборостроение», 2012, №2, стр. 20-36 – ISSN 0236-3933.

3. D.O. Ouma. Modeling of Chemical Mechanical Polishing for Dielectric Planarization. Dis. PhD in Electrical Engineering and Computer Science / D. O. Ouma, Massachusetts Institute of Technology – Massachusetts, 1998. – 228 p.

4. А.В. Амирханов, С.И. Волков, А.А. Гладких, С.В. Демин, И.А. Родионов, А.А. Столяров, А.Г. Пшенников. Модификация топологии СБИС с учетом технологических ограничений операции химико-механической планаризации. // «Математическое и компьютерное моделирование систем: теоретические и прикладные аспекты» – Сборник научных трудов НИИСИ РАН, 2011, т.1, стр. 4-10.

5. А.А. Гладких, В.В. Макарчук, В.М. Курейчик. Методики оптимального размещения dummy-структур. [Электронный ресурс] // Электронное научно-техническое издание: наука и образование. – Электрон. журн. – 2012. – №05. – Режим доступа: http:// technomag. edu.ru/doc/368628.html – ISSN 1994-0408.

6. R. Boone, D.F. Wong, R. Tian. Model-Based Dummy Feature Placement for Oxide Chemical-Mechanical Polishing Manufacturability. dac, pp.667-670, 37th Conference on Design Automation (DAC'00), 2000.

7. А.А. Гладких. Алгоритм расчета локальной плотности заполнения топологии субмикронных СБИС для оптимального размещение dummy-структур. // М.: Издательство МГТУ им. Баумана, Сборник трудов Третьей Всероссийской Школы-семинара студентов, аспирантов и молодых ученых по направлению «Нано-инженерия», 2010. – стр. 256-267. ISBN 978-5-7038-3453-4.

8. S.H. Thornton, C.A. Mack. Lithography Model Tuning: Matching Simulation to Experiment // Optical Microlithography IX, SPIE Vol. 2726 (1996).

9. A. Sekiguchi, C.A. Mack, Y. Minami, and T. Matsuzawa. Resist Metrology for Lithography Simulation, Part 2. Development Parameter Measurements // Metrology, Inspection, and Process Control for Microlithography X, SPIE Vol. 2725 (1996).

10. И.А. Родионов [и др.]. Исследование методов калибровки процессных ОРС моделей VT-5 с переменным порогом чувствительности // Микроэлектроника, 2010, т. 39, № 6, стр. 468-480.

11. И.А. Родионов Исследование влияния введения ОРС фигур на параметры полупроводниковых структур с размерами элементов 0,25 мкм. // 10-ая Молодежная международная научно-техническая конференция «Наукоемкие технологии и интеллектуальные системы 2008». 16 апреля 2008 г. – М.: издательство МГТУ им. Н.Э.Баумана, 2008, стр. 115-120.

12. С.В. Демин, А.В. Амирханов, Е.П. Михальцов, И.А. Родионов, Е.В. Тафинцева. Особенности топологического проектирования субмикронных КМОП СБИС с учетом литографических ограничений. // «Математическое и компьютерное моделирование систем: теоретические и прикладные аспекты» – Сборник научных трудов НИИСИ РАН, 2009, стр. 24-31.

13. И.А. Родионов, В.А. Шахнов. Метод расчета значений весовых коэффициентов топологических структур для калибровки литографических моделей // Вестник МГТУ им. Н.Э.Баумана. Секция: Приборостроение. 2010. Спец. выпуск: Наноинженерия. Стр. 150-160.

14. В.Ю. Киреев. Введение в технологии микроэлектроники и нанотехнологии. М.: ФГУП ЦНИИХМ, 2008. 428 с.

15. А.А. Глушко, В.А.Шахнов. Особенности трехмерного моделирования КНИ МОП-транзисторов с непрямым затвором // Микроэлектроника, 2012, т. 41, №2, стр. 83-89. 16. В.В. Денисенко. Компактные модели МОПтранзисторов для SPICE в микро- и наноэлектронике. М.: ФИЗМАТЛИТ, 2010. 408 с.

17. А.А. Глушко, А.С. Юлкин. Особенности оптимизации технологических режимов производства кни кмоп сбис по критерию стойкости к воздействию спецфакторов // «Математическое и компьютерное моделирование систем: теоретические и прикладные аспекты» – Сборник научных трудов НИИСИ РАН, 2013, т.3, №1, стр. ххх. (статья в этом сборнике)

18. Sentaurus Device User Guide. Mountain View (California USA): Synopsys, 2010. 994 p.

The features of VLS design for manufacturing techniques

A.V. Amirkhanov, A.A. Gladkikh, A.A. Glushko, E.P. Mikhaltsov, I.A. Rodionov, A.A. Stoliarov

Abstract. The features of VLS design and manufacturing process development using topology and technological simulations are considered. The basic technology simulations blocks for IC design and all the tasks inside these blocks are described. Finally a number of tasks which are coming are described.

Особенности оптимизации технологических режимов производства КНИ КМОП СБИС по критерию стойкости к воздействию спецфакторов

А.А. Глушко¹, А.С. Юлкин

1-кандидат технических наук

В статье рассмотрена задача повышения стойкости элементов КНИ КМОП СБИС к воздействию спецфакторов путем оптимизации режимов технологических операций при их производстве. Кратко представлены преимущества и недостатки встроенных средств современных систем TCAD. Предложен алгоритм оптимизации параметров технологического процесса с учетом его особенностей.

Введение

Одним из важнейших требований к микросхемам современного производства является высокая стойкость к возлействиям внешних дестабилизирующих факторов, в частности, стойкость к воздействию ионизирующего излучения. Наиболее стойкими к воздействию внешних факторов являются микросхемы, изготовленные по КНИ КМОПтехнологии. Критичными элементами микросхемы, от работоспособности которых в значительной степени зависит работоспособность всей микросхемы в целом, являются транзисторы. Известно, что на КНИ МОПтранзистор в процессе воздействия ионизирующего излучения влияют, главным образом, три фактора: генерация электронно-дырочных пар в объемной области кармана за счет бомбардировки тяжелыми ионами, утечки по боковому окислу и накопление положительного заряда в захороненном окисле. При этом последний фактор вносит существенные изменения в работу транзистора, независимо от его конструкции. Поэтому задача повышения стойкости транзистора к накопленному положительному заряду (определяемому накопленной дозой излучения) является актуальной [1].

Накопление заряда в скрытом окисле эквивалентно прикладываемому на подложку положительному напряжению, и при определенной величине этого накопленного заряда происходит образование канала в «донном» паразитного проводящего п-канальном МОП-транзисторе, в котором роль подзатворного диэлектрика играет заглубленный окисел, а роль затвора – подложка [2] (рисунок 1). Поэтому, чтобы повысить устойчивость параметров транзисторной структуры к воздействию ионизирующего излучения, необходимо обеспечить настолько высокое пороговое напряжение паразитного «донного» транзистора, насколько это возможно, сохранив при этом параметры основного транзистора: пороговое напряжение и ток насыщения. Причем допустимым является их отклонение не более чем на 10% от первоначальных значений. Таким образом, ставится задача оптимизации, условия которой интуитивно ясны, но её решение даже в численном виле является сложным в силу нелинейной зависимости электрических параметров транзисторных структур микросхемы от изменяемых параметров технологического процесса. Для решения этой задачи приборносовременные системы применяют технологического моделирования, известные под названием TCAD. Метод оптимизации, предложенный лля решения общей задачи нелинейного программирования в современных системах TCAD, последовательном квадратичном базируется на SQP-метод программировании (SQP). решает нелинейную задачу, используя итеративный подход. При этом на каждой итерации нелинейная функция аппроксимируется квадратичным полиномом, по которому и осуществляется поиск оптимального решения [3]. Главным недостатком этого метода является направленность на решение общей задачи оптимизации, причем процесс оптимизации заканчивается при попадании на границу области допустимых решений.

Однако лля решения задачи повышения порогового напряжения «донного» транзистора при сохранении характеристик основного транзистора, как будет показано ниже, возможно модифицировать алгоритм оптимизации. Основная идея предлагаемого алгоритма заключается в том, что попадание на границу области допустимых решений должно приводить к поиску параметров, слабо влияющих на целевую функцию, но при этом влияющих на параметры-ограничения, что позволило бы снова вернуться в центр области допустимых решений и продолжить процесс оптимизации и т. д.

Следует также отметить, что одиночный расчет электрических параметров транзистора трудоемкий (занимает около 1,5 часов для двумерной модели), поэтому разработку программы нужно разбить на два этапа:

1. разработка основного функционального узла программы, в котором возможна замена реальных расчетов параметров транзистора на расчет по некоторым аналитическим формулам. При этом структуры используемых массивов должны иметь такой же формат, как и массивов, формируемых по результатам вычисления электрофизических параметров (TCAD); 2. собственно разработка программы оптимизации процессов формирования структуры.

Анализ зависимостей параметров транзистора от концентрационного профиля кармана

Исследуя частную задачу оптимизации, возникающую при попытке повышения стойкости интегральной микросхемы к спецвоздействиям, можно заметить, что некоторые параметры технологического процесса напрямую влияют на целевую функцию.



Рисунок 1. Структура КНИ МОП-транзистора

Рассмотрим структуру МОП-транзистора, полученного на подложке со скрытым оксидным слоем (см. рисунок 1). Как известно, пороговое напряжение транзистора определяется толщиной подзатворного диэлектрика (и его материалом), материалом затвора и уровнем легирования подложки [4]. В данном случае толщина подзатворного диэлектрика и его материал определяются особенностями производства подложек со скрытым слоем окисла и не могут быть изменены. Материал «затвора», образованного объемом подложки, тоже определяется в процессе eë Следовательно, наиболее производства. путем повышения предпочтительным порогового «донного» транзистора напряжения является повышение концентрации акцепторной легирующей примеси вблизи заглубленного окисла.

На технологической линейке ОМТ НИИСИ РАН все операции легирования осуществляются путем ионной имплантации, причем имплантация карманов выполняется в несколько этапов с различными энергиями легирования. Поскольку большим энергиям имплантации соответствует больший разброс концентрации примеси по глубине подложки и большая средняя глубина проникновения ионов в подложку, операции ионного легирования с меньшими энергиями влияют только на свойства основного транзистора, в то время как операции ионной имплантации с большими энергиями оказывают влияние как на основной транзистор, так и на «донный».

Такая особенность данного технологического процесса позволяет модифицировать алгоритм оптимизации, а именно: варьировать те технологические режимы (энергия и доза ионных имплантаций с большими энергиями), которые в наибольшей степени влияют на свойства «донного» транзистора лля увеличения его порогового напряжения, корректируя другие технологические режимы (энергия и доза ионных имплантаций с меньшими энергиями) для поддержания параметров основного транзистора в пределах допустимых ограничений.

Описание предлагаемого алгорит-ма оптимизации

На рисунке 2 показала блок-схема модифицированного метода оптимизации.



Рисунок 2. Алгоритм модифицированного метода оптимизации

На первом этапе осуществляется ввод исходных допуск tcl-файл: ланных в на отклонение электрических параметров OT первоначальных значений, максимальное количество итераций и т.п. После этого запускается итерационный процесс, в котором с помощью средств TCAD вычисляется целевая функция, функции-ограничения в исходной точке, а также в нескольких пробных точках, находящихся в некоторой окрестности от исходной. Определяются коэффициенты влияния исходных параметров на функции-ограничения и целевую функцию, выбирается направление перемещения в следующую точку. Если значения функций в предполагаемой следующей точке удовлетворяют ограничениям, то осуществляется переход в эту точку, и цикл оптимизации повторяется. В противном случае уменьшается параметр скорости перехода по данному направлению, и попытка перехода повторяется.

Если скорость перехода стала меньше заданной, то запускается вторая часть алгоритма, связанная с попыткой возврата в центр. Здесь осуществляется вычисление коэффициентов влияния, и из них выбирается тот коэффициент, влияние которого на целевую функцию пренебрежимо мало. Итерации аналогично оптимизационному продолжаются процессу до тех пор, пока не выполнится одно из двух условий: а) значения параметров-ограничений попали в заданную наперед є-окрестность от первоначальных значений; в этом случае реализуется переход к новому циклу оптимизации; б) максимальное отношение коэффициентов влияния і-го параметра на і-ю порогового функцию не ниже значения C. одновременно і-й параметр должен оставаться в пределах допустимых значений (например, доза и энергия легирования должны быть больше нуля и не меньше некоторой заданной наперед величины); в этом случае процесс счета заканчивается и система выдает параметры технологического процесса.

Предлагаемая структура программного комплекса и основной функциональный узел оптимизации

Структура программного комплекса приведена на рисунке 3 и включает в себя основной функциональный узел оптимизации, дополнительный функциональный блок и программу приборнотехнологического моделирования. В последней выполняются лишь вычисления целевой функции и функций-ограничений из алгоритма, представленного на рисунке 2.



Рисунок 3. Предлагаемая структура программного комплекса

Для тестирования программы достаточно заменить второй узел блока (программа приборно-технологического моделирования) на дополнительный функциональный блок, позволяющий рассчитать необходимые параметры на основе виртуальных аналитических формул (тем самым позволяя основной функциональный тестировать модуль, избегая существенных временных затрат - более 95% всего времени работы программы – на расчет большого числа характеристик). При этом структура массива исходных данных ориентирована на импорт данных из TCAD, сама же целевая и функция и функции ограничения задаются непосредственно пользователем (на языке TCL) как функцииограничения.

Программный комплекс оптимизации параметров технологического процесса реализован на языке TCL. После успешно проведенного тестирования с помощью дополнительного функционального блока осуществлен запуск процесса оптимизации технологического процесса изготовления КНИ СБИС с проектными нормами 0,25 мкм. Процесс оптимизации еще не завершен вследствие существенной длительности расчета одной итерации, однако, по результатам первых 25 выполненных итераций (время счёта 1 неделя) можно сделать вывод о рекомендуемом на данный момент режиме двухступенчатого легирования кармана р-типа: доза $1,35 \times 10^{13}$ $1/cm^2$, энергия 52 кЭв, доза $2,94 \times 10^{12}$ $1/cm^2$, энергия 17 кЭв. В таблице 1 представлено изменение параметров относительно первоначальных значений.

таолица т	Таблица	1
-----------	---------	---

Наименование параметра	Первоначаль- ное значение	По результатам оптимизации (25 итераций)	Отклонение, %
Пороговое	0,607	0,609	0,4
напряжение	,	,	,
при			
напряжении			
сток-исток			
0,1 B, B			
Ток стока в	1,25×10 ⁻⁴	1,24×10 ⁻⁴	0,5
открытом			
транзисторе			
при			
напряжении			
сток-исток			
0,1 B, A			
Пороговое	0,793	0,785	1
напряжение			
при			
напряжении			
сток-исток			
2,5 B, B		,	
Ток стока в	5,00×10 ⁻⁴	4,49×10 ⁻⁴	1,1
открытом			
транзисторе			
при			
напряжении			
сток-исток			
2,5 B, A			
Пороговое	14	22	57
напряжение			
«донного»			
транзистора, В			

Как видно из таблицы 1, основные параметры изменились несущественно (не более 1,1%), при этом пороговое напряжение «донного» транзистора возросло в 1,57 раза.

Заключение

В отличие от алгоритмов, заложенных в систему TCAD, предложенный алгоритм позволяет продолжать процесс оптимизации при достижении границы области ограничений, корректируя электрические параметры основного транзистора за счет изменения дозы легирования при ионной имплантации с меньшей энергией.

Тестирование основного функционального узла оптимизации на простых примерах показало, что предложенный алгоритм позволяет решить поставленную задачу оптимизации. При этом структура массива исходных данных позволяет имитировать использование программы приборнотехнологического моделирования. Таким образом, можно сделать вывод, что первый этап разработки программы оптимизации успешно завершен.

В качестве замечания следует отметить, что второй этап разработки программы пока завершить не удалось: получены лишь промежуточные результаты по оптимизации технологического процесса с целью увеличения стойкости к дозовым эффектам. Тем не что менее. моделирование показывает, при предлагаемом режиме легирования кармана (доза 1,35×10¹³ 1/см², энергия 52 кЭв, доза 2,94×10¹² 1/см², энергия 17 кЭв) сдвиг параметров относительно первоначальных значений составляет не более 1,1%, при этом достигнуто увеличение порогового напряжения «донного» транзистора в 1,57 раза.

Литература

1. А.А. Глушко. Исследование радиационной стойкости КНИ-транзисторов с помощью системы приборно-технологического моделирования ТСАD // Наукоемкие технологии и интеллектуальные системы 2009: Сборник трудов 11-ой Молодежной международной научно-технической конференции. М., 2009. с. 192-195.

2. С.И. Волков, А.А. Глушко, С.А. Морозов. Исследование моделирование факторов, И радиационную ограничивающих стойкость КНИ НИИСИ PAH. // Труды Математическое и компьютерное моделирование систем: прикладные аспекты. Т.1. №1. Москва, Издательство «НИИСИ PAH», 2011. c. 51-56.

3. R. Fletcher. Practical Methods of Optimization, New York: John Wiley & Sons, 1987.

4. Г.И. Зебрев. Физические основы кремниевой наноэлектроники: Учебное пособие. – М.: МИФИ, 2008. – 288 с.

The features of producing process parameter optimization of SOI CMOS VLSI by criterion of resistance to special factors

A.A. Glushko, A.S. Yulkin

Abstract. The article considers the problem of increasing the resistance elements SOI CMOS VLSI to influence of special factors through optimization of process steps in production. Advantages and disadvantages of modern embedded systems TCAD are summarized. The algorithm of technological process parameters optimization is offered in view of its features.

КНИ МОП-структура с непланарным контактом к карману

С.И. Бабкин¹, С.И. Волков¹, А.А. Глушко¹

1 - кандидат технических наук

В работе представлена конструкция КНИ МОП-транзистора с непланарным контактом к карману. Проведено моделирование транзистора предлагаемой конструкции и его сравнение со стандартным транзистором. Показано, что применение транзистора представленной конструкции позволяет выиграть не только в геометрических размерах, но и в стойкости к воздействию ионизирующего излучения.

Введение

Особенностью современных технологий является постоянное снижение проектных норм, которое ужесточает требования к параметрам единичного транзистора. С одной стороны, увеличивается плотность упаковки элементов СБИС a. следовательно, возрастают паразитные емкости в елиничной структуре. С другой стороны ужесточаются требования по быстродействию.

Внедрение КНИ-технологий позволило существенно ослабить паразитные связи за счет изоляции активного слоя от общей подложки слоем захороненного окисла. Однако, КНИ МОП-транзисторы без дополнительных контактов к карману демонстрируют ряд нежелательных эффектов, в частности, подвержены кинк-эффекту [1]. Обеспечение надежного контакта к карману позволяет минимизировать его влияние, но при этом усложняется конструкция транзистора, в случае ширококанальных транзисторов контакт может получиться неравномерным, поэтому используют дополнительные контакты. Это приводит к неэффективному использованию площади кристалла, приходящейся на единичный транзисторный элемент. Необходимо отметить, что удельные потери площади возрастают по мере уменьшения проектных норм.

С другой стороны, к большинству КНИ СБИС предъявляются высокие требования по радиационной стойкости. Сравнение СБИС, изготовленных по обычной и КНИ КМОП технологии показывает, что последние имеют неоспоримое преимущество по стойкости

к импульсным воздействиям. Во-первых, вследствие того, что, основная масса электронно-дырочных пар генерируется при облучении не в рабочем слое кремния, а в подложке, которая электрически от него изолирована, во-вторых, полная диэлектрическая изоляция активных элементов практически исключает проявление тиристорного эффекта.

Иначе обстоит дело с эффектами, обусловленными дозовыми воздействиями. Под излучений действием ионизирующих в диэлектрических слоях микросхемы, в том числе, в подзатворном диэлектрике, захороненном окисле и окисле шелевой изоляции формируется

радиационно-встроенной заряд, влияющий на параметры транзисторов. Для малоразмерных КНИ технологий наибольшую опасность представляет заряд в захороненном окисле. Очевидно, что уровень стойкости тем выше, чем выше пороговое напряжение паразитного «донного» транзистора (см. рисунок 1). При глубоком легировании истока и стока, используемом в стандартных КНИ МОПтранзисторах, и типовом донном легировании карманов пороговое напряжение «донного» транзистора оказывается недостаточно высоким, чтобы обеспечить высокую стойкость к накопленной дозе радиации. В связи с этим поиск конструкций транзисторов, исключающих влияние заряда на параметры транзисторов представляется весьма актуальным.



Рисунок 1 – Структура стандартного транзистора: 1 – исток, 2 – затвор, 3 – основной канал, 4 – сток, 5 – карман, 6 – паразитный канал, 7 – подложка (затвор «донного» транзистора)

Предлагаемая конструкция n-канального транзистора представлена на рисунке 2 [2].



Рисунок 2 – Структура предлагаемого транзистора: 1 – мелкозаглубленный исток, 2 – затвор, 3 – основной канал, 4 – сток, 5 – карман, 6 – подложка (затвор «донного» транзистора), 7 – область сильного донного легирования акцепторной примесью.

В область предлагаемой конструкции высоколегированного контакта к карману располагается под областью истока и имеет ширину, равную ширине истока. Электрическое соединение истока контакта областей И к карману обеспечивается силицидом. сформированным в канавке, расположенной вдоль затвора в области имеющей глубину истока И равную ипи превышающую глубину легирования истока.

На рисунке 3 представлены расчетные электрические характеристики стандартного транзистора при различных дозах радиационного облучения. Соответствие плотности накопленного в окисле заряда и дозы облучения соответствует данным, приведенным в таблице 1 [3].

Таблица 1 - Соответствие величины заряда и дозы облучения.

N⁰	Плотность	Доза
BAX	накопленного заряда	радиационного
	в окисле, 1/см ²	облучения, кРад
1	0	0
2	$7,19 \times 10^{13}$	250
3	$1,44 \times 10^{14}$	495
4	$2,16 \times 10^{14}$	745
5	$2,88 \times 10^{14}$	990



характеристики стандартного транзистора при различных дозах радиационного излучения: 1 – 0 кРад, 2 – 250 кРад, 3 – 495 кРад, 4 – 745 кРад, 5 – 990 кРад

На рисунке 4 представлено рассчитанное распределение токов в стандартном транзисторе после воздействия ионизирующего излучения, достаточного для образования паразитного проводящего канала в донной области.



Рисунок 4 – Распределение токов в стандартном транзисторе после воздействия ионизирующего излучения через основной и паразитный (нижний) канал

На рисунке 5 представлены расчетные электрические характеристики транзистора предлагаемой конструкции при различных дозах радиационного излучения.

На рисунке 6 представлено рассчитанное распределение токов в транзисторе предлагаемой конструкции после воздействия ионизирующего излучения, достаточного для образования паразитного проводящего канала в донной области.

Как видно из рисунков 3 – 6, наличие сильнолегированной «донной» области, блокирует возникновение паразитного канала в транзисторе. Во-вторых, очевидно, что сильнолегированная область должна выравнивать



Рисунок 5 – Расчетные электрические характеристики транзистора предлагаемой конструкции при различных дозах радиационного излучения: 1 – 0 кРад, 2 – 250 кРад, 3 – 495 кРад, 4 – 745 кРад, 5 – 990 кРад



Рисунок 6 – Распределение токов в транзисторе предлагаемой конструкции после воздействия ионизирующего излучения через основной и паразитный (нижний) канал

потенциал в кармане транзистора, что позволит избавиться от необходимости обеспечения дополнительных контактов к карману в ширококанальных транзисторах. В третьих, контакт к сильнолегированной области не должен снижать ширину канала транзистора, так как находится за пределами канала.

Таким образом, моделирование показало существенное преимущество предлагаемой конструкции транзистора по сравнению со стандартной конструкцией.

Литература

1 K. Bernstein, N.J.Roher SOI Circuit design concept. Kluwer Academic Publisher. 220p Boston/Dordrecht/London 2000

2 С.И.Бабкин, С.И.Волков, А.А. Глушко. Транзистор со структурой метал-окиселполупроводник (МОП) на подложке кремний на изоляторе (КНИ). Заявка на изобретение №2011130939 приоритетом от 25.07.2011 г. положительное решение о выдаче патента 1.10.2012

3 J.R.Schwank et al. Silicon on insulator field effect transistor with improved body ties for rad-hard applications. Patent US 6268630 B1 31.07.2001

The SOI MOS structure with nonplanar contact to the body

S.I. Babkin, S.I.Volkov, A.A.Glushko

Abstract. In the work design of SOI MOS transistor with nonplanar contact to the body is submitted. Modelling the transistor of the offered design in comparison with the standard transistor is carried out. It is shown, that application of the transistor of the offered design allows to win not only in the geometrical sizes of a design, but also in resistance to ionizing radiation influence.

Структура на основе пленок аморфного кремния для элемента программирования, интегрированного в технологию КМОП БИС

С.И. Бабкин¹, А.С. Новоселов

1 -кандидат технических наук

Исследованы электрофизические параметры конденсаторов, использующих в качестве диэлектрика тонкопленочную структуру на основе пленок аморфного кремния и двуокиси кремния с толщиной 120-400 Å и 30-35 Å соответственно. Структура формируется с использованием оборудования и процессов, которые интегрируются в технологический цикл создания многоуровневой металлизации КМОП БИС. Показано, что на основе исследованной структуры могут быть сформированы однократно программируемые элементы переключения типа antifuse с параметрами $R_{off} > 10^{10}$ Ом и $R_{on} <30$ Ом стабильными при температурах до 140^{0} С.

1. Введение

Программируемые логические интегральные схемы (ПЛИС), включающие в себя многочисленные функциональные блоки, изготовленные по КМОП технологии, уже давно находят применение при проектировании сложных электронных систем [1]. Необходимым условием для изготовления ПЛИС является наличие программируемого коммутирующего устройства, соединяющее составные части ПЛИС. В качестве такого устройства может служить однократно программируемый элемент (antifuse), сопротивление которого в выключенном состоянии Roff имеет большое, а во включенном состоянии Ron соответственно малое значения [2]. Элемент antifuse, металлизации расположенный между шинами многоуровневой разводки (antifuse Me to Me) имеет ряд преимуществ:

- малое значение R_{on} <100 Ом;

- низкое значение удельной емкости < $2\phi \Phi/MKM^{2}$;

- стабильность параметров при эксплуатации изделия в экстремальных условиях.

Применение antifuse Me to Me на основе пленок аморфного кремния позволило компании Actel создать семейство ПЛИС для военно-космических систем [3, 4].

К antifuse недостаткам данного относится необходимость использования специализированного оборудования, также сложность а получения относительно невысоких напряжений переключения (до 10 В) при одновременном обеспечении значений R_{off} > 10⁹ Ом. Для выполнения этих условий предлагается использовать многослойные структуры, состоящие из тонких пленок аморфного кремния, оксида и оксинитрида кремния [5].

В данной работе исследуются параметры структур на основе пленок аморфного кремния, предназначенных для формирования antifuse типа Ме to Me, изготовленных с использованием оборудования, применяемого в технологии многоуровневой металлизации КМОП БИС.

2. Методика проведения эксперимента

Пленки аморфного гидрогенезированного кремния сложного состава α -Si_xN_yH_z (далее SiHN) осаждались в реакторе WxZ, размещенном на платформе установки Endura PVD 5500 ф. Applied Materials [6]. Для осаждения пленок SiO₂ использовался процесс химического осаждения из газовой смеси тетраэтоксисилана и озона [7] на установке DCVD Centura ф. Applied Materials.

Тонкопленочная структура SiHN-SiO₂-SiHN исследовалась с помощью тестового конденсатора, конструкция которого представлена на рисунке 1. Роль обкладок выполняют межуровневое соединение на основе вольфрама (W) и шина металлизации верхнего уровня (M3). Указанная выше структура, обозначенная как A-Si, является диэлектриком конденсатора. ВАХ конденсаторов исследовались на измерительном комплексе Agilent Technologies B1500



Рис. 1 Конструкция конденсатора на основе пленок SiHN-SiO2.

3. Основные результаты

На рисунке 2 представлена зависимость напряжения пробоя (V_{br}) конденсатора, измеренная по

уровню тока 1 мА, от суммарной толщины SiHN в структуре SiHN-SiO₂-SiHN 1 – соответствует максимальному значению ($V_{br max}$), а 2 – минимальному значению($V_{br min}$) напряжения пробоя.

В целом зависимость имеет линейный характер. Наблюдаемая точка перегиба, связана с изменением толщины SiO₂ в структуре, что соответствует возможным колебаниям кратковременного процесса осаждения SiO₂ в начальной стадии. Получаемая 30-35 Å является толщина SiO₂ минимально реализуемых условиях. возможной в Разброс V_{br}=V_{br max}-V_{br min} не превышает 1,0 В по пластине, что в целом соответствует требованиям к элементам antifuse Следует отметить, что значение V_{br}, фиксируемое по току 1 мА, практически не отличается от V_{br}, фиксируемому по току 10 мА, что подтверждает лавинообразный характер пробоя.



Рис.2 Зависимость напряжения пробоя конденсатора от толщины SiHN.

Известно, что сопротивление antifuse после пробоя (Ron) зависит от величины пропускаемого тока, которая обычно составляет 5-15 мА [8]. В нашем случае использовался импульс тока амплитудой 10 мА и длительностью 1 мс. После этого измерялось сопротивление конденсатора Ron Результаты измерений Ron для структур с различной толщиной SiHN представлены на рисунке 3, где зависимости 1 и 2 относятся к максимальным и минимальным значениям Ron соответственно.

Из рисунка 3 следует, что абсолютное значение и разброс величины R_{on} уменьшаются с уменьшением толщины SiHN, что связано с формированием более короткой проводящей нити (filament) в результате пробоя [8]. В исследуемом диапазоне толщин R_{on} не превышает 70 Ом, а при толщинах SiHN<220 Å выходит на величины $R_{on} \sim 15-25$ Ом, что соответствует лучшим известным значениям для antifuse.



Рис.3 Зависимость сопротивления конденсатора после пробоя Ron от толщины SiHN.

Как отмечалось, для antifuse наряду с малым значением $R_{\rm on}$ необходимо обеспечить значительную величину сопротивления в выключенном состоянии $R_{\rm off}$. На рисунке 4 представлены зависимости от толщины SiHN максимального – 1 и минимального – 2 значения тока утечки через конденсатор при прикладываемом напряжении 3,3 В. Для толщин больше 220 Å зафиксированы минимальные значения абсолютной величины тока утечки и ее разброса, что соответствует значению $R_{\rm off} \sim 6 \cdot 10^{10}$ Ом. При меньших толщинах SiHN ток утечки может достигать $5 \cdot 10^{-10}$ А, что соответствует значению $R_{\rm off} \sim 6 \cdot 10^9$ Ом и также удовлетворяет критериям antifuse.



SiHN.

Для оценки стабильности структуры antifuse исследовалась зависимость тока утечки конденсатора от температуры для различных толщин SiHN при 3,3 B. приложенном напряжении Результаты представлены на рисунке 5. Зависимости соотносятся к толщине SiHN как: 1-120 Å, 2-160 Å, 3-220 Å, 4-280 Å, Наблюдается 5-400 Å, соответственно. резкое увеличение тока утечки конденсатора с ростом температуры при толщине SiHN 120 Å. В остальных случаях увеличение тока утечки незначительно.



Рис. 5 Зависимость тока утечки конденсатора от температуры при различных толщинах SiHN.

4. Выводы

Значения напряжений пробоя конденсаторов 6,5-7,5 В со структурой диэлектрика SiHN-SiO2-SiHN могут быть получены при толщинах SiHN и SiO2 ~220 Å и ~35 Å соответственно. При этом ток утечки конденсаторов при напряжении 3,3 В не превышает значений 2 10-10 Å и сохраняется при температурах до 140°С. Сопротивление такого конденсатора после воздействия импульса тока амплитудой 10мА и длительностью 1 мс менее 30 Ом. Полученные результаты позволяют формировать однократно программируемые элементы переключения типа antifuse с параметрами Vbr~6,5-7,5 B, $R_{\rm off} > 10^{10} \mbox{ Om}$ и $R_{on} < 30 \text{ Om}.$

Литература

- В. Майская. Программируемые логические схемы. Наступает эпоха перемен. Электроника: Наука, Технология, Бизнес 4/2004 с.10-19.
- Elodie Ebrard, BrunoAllard, PhilippeCandelier, PatriceWaltz. «Review of fuse and antifuse solutions for advanced standard CMOS technologies». Microelectronics Journal 40 (2009) 1755–1765.
- Е. Котельников. «Программируемая логика Actel». Электронные компоненты 2010 №9 с.88-93.
- С. Карпов. «Асtel: Новые технологии, передовые решения». Электроника: Наука, Технология, Бизнес 7/2007 с.78-79.
- 5. J. Michail Hart et al. «Multilayer amorphous silicon antifuse». Патент США №5726484 10/03 1998
- С.И. Бабкин, А.С. Трохин, А.С. Новоселов. «Процесс осаждения пленок аморфного кремния, интегрированный в технологию многоуровневой металлизации КМОП БИС». Электроннаятехника Серия 2. Полупроводниковые приборы 2013 Вып.1 (230) с.17-21.
- Handbook of Chemical Vapor Deposition (CVD) Principles, Technology, and Applications Second Edition by Hugh O. Pierson 1998 Norwich, New York, U.S.A.
- N. Vasudevan et al. «ON-state reliability of amorphous-silicon antifuses» Journal of Applied Physics Vol 84, № 11 1998 pp.6440-6447.

The structure based on the amorphous silicon films for programming element, integrated into the CMOS VLSI technology

S.I. Babkin, A.S. Novoselov

Abstract. The electrophysical parameters of capacitors with dielectric thin film structures based on amorphous silicon and silicon dioxide with the thickness of 120-400 Å and 30-35 Å, respectively are investigated. The structures were formed using the equipment and processes that are integrated into the technological cycle of the multilevel metallization VLSI CMOS. It is shown that one-time programmable antifuse type switching elements can be formed on the basis of the investigated structures with parameters $R_{off} > 10^{10}$ ohm and $R_{on} < 30$ ohm stable at temperatures up to 140^{0} C.

Особенности формирования МІМ конденсатора с электродом на основе пленок вольфрама, осажденных из газовой фазы

С.И.Бабкин¹, И.В.Волков, А.С.Новоселов, М.В.Орешков¹, Е.Э.Самоха, С.В.Седов

1 – кандидат технических наук

Исследованы особенности формирования МІМ конденсаторов с структурой W–TIN–SiO₂–TiN–Ti–Al–Ti в составе многоуровневой металлизации КМОП БИС. Показано, что конденсаторы с удельной емкостью 0,8–1,5 фФ/мкм² характеризуются величиной утечки \leq 3,5 фА/(пФ·В). Исследованы технологические факторы, влияющие на состояние поверхности нижнего электрода и на характеристики надежности МІМ конденсаторов.

1. Введение

Конденсатор со структурой металл-изоляторметалл (MIM) является одним из основных в ряду пассивных элементов современных аналоговых КМОП БИС. Конструкция МІМ конденсатора интегрирована в процесс формирования многоуровневой металлизации КМОП БИС, что предполагает использование традиционных процессов для данного типа технологии. Для «алюминиевой металлизации», использующей в качестве материала коммутационных шин тонкопленочную структуру титан-алюминий-титаннитрид титана (Ti-Al-Ti-TiN), наиболее традиционной является конструкция MIM конденсатора вертикального типа, представленная на рис. 1(а). В этом случае роль нижнего электрода (обкладки) конденсатора выполняет шина металлизации ME3. В качестве диэлектрика используются пленки SiO₂, Si₃N₄ толщиной 400-600 Å комбинация. полученные методом или их плазмохимического осаждения (PECVD) [1,2,3]. В качестве верхнего электрода используются дополнительные слои TiN-Ti-Al-Ti-TiN или одного TiN, которые соединяются с шиной металлизации ME4 посредством вольфрамовых межуровневых соединений VIA3. В случае одного TiN верхний электрод характеризуется поверхностным сопротивлением $R_{s} > 10 O_{M/\Box}$ что может ухудшать частотные характеристики схемы. Для достижения значений Rs ~ 0.2 Om/□ необходимо использовать пленку алюминия толщиной ~0,25 мкм, что в сочетании со слоями Ti-TiN дает суммарную толщину ~0,4-0,45 мкм. Рассматриваемый вариант изготовления MIM конденсатора имеет ряд недостатков:

- усложнение процесса формирования шин металлизации, выполняемого после травления верхнего электрода. Это связано с возникновением рельефа порядка 0,4–0,45 мкм, а также наличием диэлектрика на поверхности металла;

- сложность процесса травления верхнего электрода конденсатора, что обусловлено возможностью повреждения металла нижнего электрода, наличием продуктов травления на поверхности нижнего металла, а также проявление изотропной составляющей, приводящей к изменению топологического размера;

- воздействие плазмохимического процесса травления переходного отверстия (VIA) на сформированную структуру МІМ конденсатора.

Альтернативный конструктивно-технологический вариант МІМ конденсатора представлен на рис. 1(б) [5]. В качестве верхнего электрода конденсатора используется слой вольфрама, заполняющий отверстие в межуровневой изоляции (IMD). Толщина вольфрама составляет ~1,0 мкм, что соответствует значению Rs ~ 0,1 Ом/□. Формирование МІМ конденсатора совмещено с формированием межуровневых соединений (VIA), т.е. используется цикл осаждения структуры Ti-TiN-W с последующей химикомеханической полировкой (СМР).

В этом варианте диэлектрик МІМ конденсатора осаждается на поверхность TiN, которая подвергалась воздействию фторсодержащей и кислородной плазмы при анизотропном травлении SiO2 и последующем удалении фоторезистивной маски. Имеющиеся данные показывают, что состояние поверхности [3,4] электрода (рельеф, тип металла) оказывают влияние на токи утечек МІМ конденсатора. В качестве материала верхнего электрода предпочтительно использовать наиболее стабильные с точки зрения взаимодействия с ТіN. В этом случае получается SiO₂ пленки симметричная MIM структура. Отсюда следует необходимость осаждения структуры TiN-W без подслоя титана, что снижает адгезию к окислу и, как следствие, усложняет процесс СМР вольфрама. Кроме того, при использовании пленок вольфрама толщиной ~1 мкм существенно возрастают механические также могут влиять на напряжения, которые электрофизические параметры конденсаторов.

С учетом выше изложенного, в данной работе ставились следующие задачи:

- исследовать технологический маршрут блока операций формирования МІМ конденсатора с использованием пленок вольфрама, осажденных из газовой фазы в качестве материала для верхнего электрода;

W (400)»;

2.1.13. химико-механическая полировка вольфрама (CMPW);

2.1.14. химическая обработка;

2.1.15. формирование фоторезистивной маски (ФРМ);



в

части

Рис. 1 Конструктивно-технологические варианты МІМ конденсаторов в составе 4-х уровневой металлизации

микрорельефа и обработки в кислородной плазме на электрофизические параметры MIM конденсатора.

возможность

на основе металлических пленок Ti, Al, TiN, W;

электрода

технологию создания многоуровневой металлизации

- исследовать влияние состояния поверхности металла

MIM

его

структуры

интеграции

В

2. Методика подготовки образцов и проведения исследований

2.1 Подготовка образцов

- исслеловать

нижнего

Тестовые структуры с МІМ конденсаторами изготавливались по следующему маршруту:

изготавливались по следующему маршруту: 2.1.1. исходная кремниевая пластина р-типа Ø150 мм;

2.1.2. химическая обработка;

2.1.3. осаждение структуры металлов «Ті (20 нм) – Al (500) – Ті (20) – Ті N (80)»;

2.1.4. измерение микрорельефа поверхности металла;

2.1.5. плазмохимическое осаждение SiO₂ толщиной 1 мкм из газовой смеси тетраэтоксисилана и кислорода (PETEOS);

2.1.6. формирование фоторезистивной маски (ФРМ);

2.1.7. анизотропное плазмохимическое травление окисла до металла и плазмохимическое удаление ФРМ; 2.1.8. измерение микрорельефа поверхности металла;

2.1.9. химическая обработка;

2.1.10. обработка в кислородной плазме (O_2 +He₂) части пластин;

2.1.11. плазмохимическое осаждение SiO₂ различной толщины из газовой смеси тетраэтоксисилана и кислорода (PETEOS);

2.1.12. осаждение структуры металлов «TiN (25) -

2.1.16. анизотропное плазмохимическое травление окисла до металла и плазмохимическое удаление ΦРМ; 2.1.17. химическая обработка;

2.1.18. осаждение структуры металлов «Ті (20) – TiN (70) – W (500)»;

2.1.19. химико-механическая полировка вольфрама (CMPW);

2.1.20. измерение микрорельефа поверхности металла;

2.1.21. химическая обработка;

2.1.22. осаждение пленки алюминия толщиной 1 мкм на обратную сторону пластины.

Металлические пленки осаждались на установке ENDURA PVD 5500 (AMAT). Для получения пленок вольфрама использовался процесс химического осаждения из газовой фазы WF₆+H₂ при температуре 450^oC. Пленки Ti, Al и TiN наносились с использованием методов магнетронного и реактивного магнетронного распыления соответствующих мишеней.

Анизотропное травление окисла через ФРМ (2.1.6) с последующим ее удалением проводилось на установках ОХІDE ЕТСН СЕNTURA (AMAT) в фторсодержащей плазме CF_4+CHF_3+Ar и RAPID STRIP в реакторе с удаленной плазмой CF_4+O_2 .

Операция обработки поверхности в кислородной плазме (2.1.9) проводилась в том же реакторе установки DCVD CENTURA (AMAT), что и операция осаждения окисла, при температуре подложки 400^оС и мощности плазмы 500 Вт. Поперечное сечение изготовленной по данному маршруту тестовой



Рис. 2. Сечение тестовой структуры а) и фрагмент топологии с областями технологического и параметрического контроля тестовых структур б)

структуры для исследования МІМ конденсаторов представлено на рис. 2(а).

2.2 Методы исследования

Оценка микрорельефа поверхности металла проводилось на профилометре Dektak V200 SL в соответствии с методикой, изложенной в работе [6], в 5 областях пластины. Фиксировался параметр Ra, характеризующий среднее арифметическое значение отклонений рельефа от среднего уровня. Измерения Ra проводились в непосредственной близости от исследуемых тестовых структур C рис. 2(б).

Вольтамперные (ВАХ) и вольтфарадные (ВФХ) характеристики полученных МІМ конденсаторов снимались на измерительном комплексе Agilent B1500. Площадь конденсаторов составляла 9800 и 3025 мкм². Для исследования сечения тестовых структур использовался растровый электронный микроскоп LEO 1455 VP. Толщина окисла с локальностью 10 мкм контролировалась на спектроскопическом эллипсометре APECS 3020.

3. Основные результаты

3.1 Интеграция в технологию многоуровневой металлизации

Сечения тестовой структуры, после различных стадий формирования представлены на рис. 3.

Из рис. 3(а) следует, что после операции СМРW толщина вольфрама, составившая 340 нм на дне отверстия размером 6x6 мкм, незначительно отличается от номинального значения - 400 нм. При отверстий 100х100 мкм этом для больших наблюдаются аналогичные соотношения. Следует также отметить, что отсутствие титана в качестве адгезионного слоя не повлияло на процесс CMPW. Эрозия окисла при выбранном режиме CMPW (степень переполировки 50% при давлении 4 psi) составила ~80 нм. Помимо указанных параметров процесса CMPW на величину эрозии окисла существенно влияет топологическая плотность заполнения поверхности структурами типа «отверстий в окисле», которые заполняются вольфрамом. При большой плотности в момент переполировки слоя вольфрама резко возрастает давление полирующей подушки на окисел и, как следствие, увеличивается скорость эрозии [7]. Фактор эрозии для данной топологии необходимо учитывать И компенсировать толщину при формировании межуровневой изоляции (IMD).

Второй цикл осаждения и СМРW целесообразно совмещать с циклом создания межуровневых соединений (VIA), который включает в себя формирование ФРМ переходных отверстий, их травление и заполнение вольфрамом (оп. 2.1.15– 2.1.18). Наиболее критичным является влияние незаполненного отверстия МІМ структуры на процесс формирования ФРМ переходных отверстий. На рис. 3(б) и (в) представлены сечения тестовых структур различного размера с нанесенными

фоторезистом (PHR) и антиотражающим покрытием (BARC) толшинами. соответствующими с формированию ФРМ для VIA3. Из рис. 3(б) следует, что номинальная суммарная толщина PHR+BARC ~0,7 мкм для отверстий с размерами 6х6 мкм восстанавливается на расстоянии ~1 мкм от края отверстий отверстия. Для большого размера 100х100 мкм это расстояние увеличивается до 10 мкм. что это расстояние Очевидно, зависит от планаризующих свойств PHR и BARC. На практике большая емкость MIM конденсаторов набирается за счет матрицы конденсаторов размерами ~10x10 мкм [5]. На рис. 3(в) представлено сечение тестовой структуры после второго цикла осаждения Ti-TiN-W и СМРW, совмещенного с формированием VIA3. Таким образом, изготовлен тестовый элемент, показанный на рис. 2(а), который в дальнейшем использовался для исследования электрофизических параметров MIM конденсаторов. Следует отметить, что структура



а) оп. 2.1.14



б) оп. 2.1.15



в) оп. 2.1.15



г) оп. 2.1.19



характеризуется планарной поверхностью вольфрама (W2) со значением Ra = 10÷30 Å (таблица 1), что не должно вызывать затруднений при формировании шин металлизации верхнего уровня.

3.2 Электрофизические характеристики МІМ конденсаторов.

В таблице 1 приведены значения параметра Ra для поверхности металла нижнего электрода MIM конденсатора после различных стадий его формирования оп. 2.1.3 и 2.1.16. Из нее следует, что перед осаждением рабочего диэлектрика оп. 2.1.11 микрорельеф поверхности может иметь существенный больший разброс по сравнению с исходным состоянием после осаждения металлов.

Таблица 1. Параметры микрорельефа поверхности металлов.

Операция, после которой проводилось измерение (см. раздел 2.1)	Значения Ra, Å
Осаждение структуры металлов Ті(20нм)- Al(500)-Ti(20)-TiN(80);(2.1.3)	20–55
Анизотропное плазмохимическое травление окисла до металла и плазмохимическое удаление ФРМ (2.1.16)	20–210
Химико-механическая полировка вольфрама (2.1.19)	10–30

Влияние микрорельефа на электрофизические характеристики оценивалось несколькими способами:

 а) по величине напряжения, близкого к пробивному, при котором через конденсатор площадью 9800 мкм² проходил ток 1 нА;

б) сравнение и анализ ВАХ при проведении мгновенного пробоя в диапазоне температур 25÷150°С;

в) оценка надежности по результатам ускоренных испытаний на времязависимый пробой t_{bd} при постоянном токе через структуру 25 пА различной полярности;

г) сравнение ВФХ.

При этом помимо микрорельефа исследовалось



Рис. 4. Зависимость напряжения, фиксируемого при токе утечки 1 нА через МІМ конденсатор, от величины Ra микрорельефа поверхности нижнего электрода

влияние способа обработки поверхности нижнего электрода: с необработанной (N) и дополнительно обработанной (NO) в кислородной плазме.

Результаты измерения напряжения при токе 1 нА представлены на рис. 4. Можно отметить, что не наблюдается четкой корреляции между величиной Ra и напряжением для выбранного уровня тока утечки для всего исследуемого диапазона значений удельной емкости конденсаторов (1.1, 1.22, 1.52 фФ/мкм²) в зависимости от варианта обработки поверхности.



Рис. 5. Зависимость напряжения, фиксируемого при токе утечки 1нА через МІМ конденсатор, от величины удельной емкости.

На рис. 5 показаны зависимости напряжений, фиксируемых для тока утечки через конденсатор 1 нА для обработанной (NO) и необработанной (N) поверхностей электрода, характеризуемой величиной Ra микрорельефа ~100 Å, от значения удельной емкости.

Из представленных зависимостей следует, что ток 1 нА достигается при больших напряжениях для электродов, обработанных в кислородной плазме. Причем эта разница тем заметнее, чем выше удельная емкость конденсатора, т.е. чем тоньше рабочий диэлектрик МІМ конденсатора.

Для МІМ конденсаторов часто используется величина, интегрально связывающая ток утечки при данном напряжении и емкости и имеющая размерность $\phi A/(n\Phi \cdot B)$. Согласно [8], для технологического уровня до 0,13 мкм ее значение должно быть ≤ 7 . В нашем случае, для напряжения 3,3 В и удельных емкостей в диапазоне 0,8–1,5 $\phi \Phi/мкм^2$ ее значение не превышало 3,5.

Исследование зависимости ВАХ от температуры не выявило повышения токов утечки структур в пределах чувствительности измерительного оборудования в диапазоне напряжения от 0 до 15 В. Характерная особенность температурной зависимости заключается в параллельном сдвиге ВАХ в сторону меньших напряженностей электрического поля.

Отличительными особенностями ВАХ при проведении мгновенного пробоя стали:

 а) большой разброс напряжений пробоя по пластине, и даже между соседними структурами;

б) низкая и плохо воспроизводимая плотность предпробойного тока в диапазоне от 10^{-6} до 10^{-4} A/cm², особенно для отрицательной полярности;

в) невозможность выявить влияние микрорельефа и дополнительной обработки структур. На рис. 6 приведены графики в осях Вейбулла кумулятивной вероятности времени до пробоя t_{bd} для разных вариантов МІМ структур. При построении графиков кумулятивной вероятности учитывались только структуры, для которых зафиксировано падение абсолютного значения напряжения ниже 15 В в процессе воздействия в течение 1000 сек., при этом, из-за переходного процесса в начале стресса, пробои в течение первых 14 сек. также не учитывались. Количество структур, отсортированных по данным признакам, приведено в таблице 2.





На рис. 6 и в таблице 2 знак + и – соответствуют вариантам подачи на верхний электрод напряжения положительной и отрицательной полярности, а (N) и (NO) – варианты необработанной и обработанной дополнительно в кислородной плазме поверхности нижнего электрода, соответственно. Анализируя выше характеристики не имеют принципиальных отличий в зависимости от Ra и с достаточной точностью совпадают (см. –(N), Ra \approx 95 Å и –(N), Ra \approx 25 Å). Аналогичная зависимость наблюдается для варианта (NO) (см. –(NO), Ra \approx 60 Å).

- при увеличении Ra > 60 Å и положительном напряжении в независимости от вариантов (N) или (NO), и независимо от Ra и вариантов (N) или (NO) для отрицательного напряжения происходит активная деградация времени до пробоя, заключающаяся в появлении на распределении кумулятивной вероятности крутого участка при малых временах tbd (до 50 сек.).

Можно предположить, что поверхность верхнего электрода (TiN), контактирующая с диэлектриком конденсатора, в какой-то степени повторяет развитый микрорельеф нижнего электрода. При этом «верхний» TiN не подвергается обработке в кислородной плазме в технологическом цикле (см. оп. 2.1.7), а также при (оп. 2.1.10). дополнительной обработке При отрицательной полярности непассивированная поверхность TiN является более активным инжектором заряда в диэлектрик. Незначительное отличие характеристик для образцов (N) и (NO) при положительной полярности, вероятно, связано с тем фактом, что поверхность «нижнего» ТіN в достаточной степени пассивируется в кислородной плазме уже в результате других операций технологического цикла.

При положительной полярности воздействия наилучшие по среднему времени t_{bd} результаты получены для наименьших значений Ra у структур типа (N). Как следует из таблицы 2 около 30% этих структур выдержало воздействие в течение 1000 сек. Однако, для всех остальных структур присутствует огромный разброс по величине t_{bd} . Так, при небольшой величине среднего времени по графику кумулятивной

вариант структуры	14 сек.< <i>t_{bd}</i> < 1000 сек.	$t_{bd} > 1000$ сек.	<i>t_{bd}</i> < 14 сек.
–(NO), Ra≈210 Å	23	14	1
–(NO), Ra≈60 Å	22	13	3
+(NO) , Ra≈60 Å	24	11	3
+(NO) , Ra≈210 Å	36	0	2
–(N) , Ra≈95 Å	30	4	4
–(N) , Ra≈25 Å	26	9	3
–(N) , Ra≈25 Å	27	5	6
+(N) , Ra≈25 Å	26	10	2
+(N) , Ra≈95 Å	37	0	1
+(N) , Ra≈25 Å	26	12	0

Таблица 2. Распределение структур по группам, в различных интервалах полученного значения tbd

приведенные данные, можно сделать следующие выводы:

- для варианта (N) большие значения Ra при положительной полярности напряжения приводят к заметному уменьшению времени до пробоя (см. +(N), Ra \approx 95 Å и +(N), Ra \approx 25 Å). В этом случае заряд инжектируется в диэлектрик со стороны нижнего электрода. При отрицательных напряжениях

вероятности, имеется множество структур с t_{bd} , превышающим 1000 сек., либо пробившихся по прошествии первых 14 сек. (см. таблицу 2). Это указывает на возможные проблемы с надежностью предлагаемых структур и требует проведения целенаправленных исследований.

Точность измерения ВФХ не позволила выявить никаких отличий по однородности величины емкости

от напряжения для структур в зависимости от микрорельефа или наличия дополнительной обработки в кислородной плазме. Для определения коэффициента зависимости емкости от напряжения требуется её измерение с точностью не хуже 0,05%, что в имеющейся конфигурации измерительного оборудования достичь не представлялось возможным.

Заключение

Проведенные исследования показали возможность формирования МІМ конденсаторов со структурой: «верхний электрод (W-TIN) - SiO₂ - нижний электрод (TiN-Ti-Al-Ti)», с применением пленок вольфрама, с процессов, полученных использованием традиционных для создания межуровневых межсоединений в многоуровневой металлизации КМОП БИС. В этом варианте верхний и нижний электрод характеризуются величинами поверхностного сопротивления Rs < 0,2 Ом/ \Box и Rs < 0,06 Ом/ \Box , соответственно.

Для интеграции исследуемой структуры в процесс формирования многоуровневой металлизации требуются дополнительные операции фотолитографии, окисла плазмохимического травления И цикла и химико-механической осаждения полировки структуры TiN-W. Дополнительная операция химикомеханической полировки вольфрама приводит к увеличению эрозии окисла межуровневой изоляции, что необходимо учитывать при формировании многоуровневой металлизации.

При использовании в качестве диэлектрика окисла, полученного методом плазмохимического осаждения могут быть изготовлены MIM-(PETEOS), 10000мкм² конденсаторы площадью ЛО характеризуемые удельной емкостью 0,8-1,5 фФ/мкм² и величиной утечки $\leq 3,5 \ \phi A/(\pi \Phi \cdot B)$. С учетом условий формирования МІМ-структуры в отверстии в окисле межуровневой изоляции, а также фактора эрозии окисла данная структура наиболее эффективна в условиях малой топологической плотности MIM конденсаторов с площадью элемента емкости $< 100 \text{ MKm}^2$.

Среди факторов, связанных с особенностями формирования и определяющих состояние поверхности нижнего электрода, следует выделить микрорельеф поверхности. Его величина является определяющей по степени влияния на результаты ускоренных испытаний на время зависимый пробой *t*_{bd} MIM-конденсаторов.

Литература

1. A. Kar-Roy, C. Hu, M. Racanelli, C.A. Compton, P. Kempf, G. Jolly, P.N. Sherman, J. Zheng, Z. Zhang, and A. Yin. High density metal insulator metal capacitor using PECVD nitride for mixed signal and RF circuits. Int. Interconnect Tech. Conference, pp. 245–247, 1999.

2. M.C. King, C.F. Chang, H.J. Lin, and Albert Chin. Comparison of MiM Performance with Various Electrodes and Dieletric in Cu Dual Damascene of CMOS MS/RF Technology. J. Electrochem. Soc. 2006 153(12): pp G1032–G1034.

3. Ng, Chit-Hwei Hwei et al. Characterization and comparison of PECVD silicon nitride and silicon oxynitride dielectric for MIM capacitor. IEEE Electron Device Letters 2003 Volume: 24, Issue: 8 pp 506–508.

4. A.W. Groenland and R.A.M. Wolters, and A.Y. Kovalgin, and J. Schmitz On the leakage problem of MIM capacitors due to improper etching of titanium nitride. In Proceedings of STW. ICT Conference 2010, 18–19 Nov 2010, Veldhoven, The Netherlands. pp. 89–92. Technology Foundation STW. ISBN 978-90-73461-67-3.

5. Eun-Seok Shin et al. Capacitor and capacitor array. Patent № US 7560796 B2 Jul. 14. 2009.

6. С.И. Бабкин, В.Ю. Киреев, Т.В. Козырева, О.Л. Голикова, Е.П. Михальцов, А.С. Трохин. Возможности оценки качества систем металлизации интегральных микросхем на основе алюминиевых сплавов различными методами. Известия вузов. Электроника №5 2003 с.7–22.

7. M.R. Oliver. Chemical Mechanical Planarization of Semiconductor Materials. Springer–Verlag, Berlin Heidelberg, New York, 2004, p. 407.

8. The international Technology Roadmap for Semiconductors, Semiconductor Industry Association, 2002.

The features of formation MIM capacitor with an electrode on the basis of the chemical vapor deposition tungsten films

S.I. Babkin, I.V. Volkov, A.S. Novoselov, M.V. Oreshkov, E.E. Samokha, S.V. Sedov

Abstract. The features of MIM capacitor formation with structure W–TiN–SiO₂–TiN–Ti–Al–Ti as a part of CMOS process multilevel metallization are investigated. Devices with specific capacity of 0,8–1,5 fF/ μ m² have been characterized by volume of a leakage \leq 3,5 fA/(pF·V). The technology factors influencing the bottom electrode surface condition and the reliability of such MIM capacitors are investigated.

Особенности операционного контроля в условиях мелкосерийного производства кристаллов микросхем

В.Ю. Троицкий¹, М.В. Орешков², Т.В.Козырева

1- кандидат физико-математических наук, 2 – кандидат технических наук

В статье рассмотрена организация операционного контроля технологического процесса изготовления субмикронных интегральных микросхем в условиях мелкосерийного и прерывистого производства на базе технологического комплекса НИИСИ РАН. Приведены описания методик контроля процессов изготовления кристаллов микросхем с субмикронными проектными нормами, позволяющих определить влияние входных параметров отдельных операций технологического процесса на выходные характеристики сформированной структуры. Описываются принципы организации комплексной системы операционного контроля, обеспечивающей своевременную обнаружение дефектных пластин на стадии кристального производства. Предложена интеграция методов квалификации технологического процесса и оценки конструктивно-технологических запасов в систему операционного контроля в условиях мелкосерийного производства для обеспечения указанных требований.

1. Введение

Мелкосерийное производство в силу своей специфики неизбежно диктует специфичные требования к системе операционного контроля, зачастую сильно отличающиеся от требований массового серийного производства. Первое и главное отличие вытекает из характера производства и заключается в ограниченных объемах выпуска продукции при широкой номенклатуре выпускаемых изделий. Кроме того, при проведении технологических операций в процессе производства задействовано, как правило, оборудование для индивидуальной, а не групповой обработки пластин. Так, в кристальном производстве на базе технологического комплекса НИИСИ РАН индивидуальная обработка пластин осуществляется на всем обрабатывающем и измерительном оборудовании, за исключением установки ионной имплантации.

Результатом такого вида обработки может стать снижение воспроизводимости параметров формируемых структур от пластины к пластине не только между партиями в рамках одного базового процесса, но и между пластинами внутри одной партии. Таким образом, возникает необходимость контроля резульпроведения каждой структурообразующей тата операции на каждой пластине из партии. Такой сплошной контроль, в отличие от выборочного контроля при массовом производстве, возможен только при применении неразрушающих методов контроля, обладающих достаточной локальностью для использования их на пластинах с топологическим рисунком. Однако набор методов, отвечающих этим требованиям, достаточно ограничен и не может обеспечить контроль всех необходимых параметров формируемых на пластине структур.

В связи с этим, при организации операционного контроля в условиях индивидуальной обработки пластин и широкой номенклатуры изделий возникла необходимость организовать оптимальную систему операционного контроля, обеспечивающую достаточный объем и эффективность контроля, и позволяяющую гарантировать высокий выход годных и надежность изделий при минимально возможных затратах на ее реализацию.

Кроме этого, в рамках управляемой технологии результаты операционного контроля должны давать информацию о требуемой коррекции технологического процесса, т.е. обеспечивать эффективную обратную связь между входными и выходными параметрами процесса.

С учетом сказанного, система операционного контроля в кристальном производстве на базе технологического комплекса НИИСИ РАН была построена на шести нижеследующих принципах.

1. Использование самых современных физических методов для контроля выходных характеристик технологических процессов.

2. Применение комбинированной системы контроля технологического процесса, включающей:

 периодическую оценку технологической готовности (аттестация) технологического оборудования на привносимую дефектность;

периодическую аттестацию технологических
процессов по выходным параметрам;

 – контроль выходных характеристик технологических процессов непосредственно на рабочих пластинах;

 – контроль выходных характеристик технологических процессов на пластине-спутнике при обработке рабочих партий.

3. Использование различных методов контроля одной и той же характеристики, исходя из ее критичности, трудоемкости метода, его информативности и точности, что достигается, в том числе, сочетанием прямых и косвенных измерений. При этом, предпочтительно применение неразрушающих методов контроля на рабочих пластинах.

4. Использование расширенного диагностического контроля электрофизических параметров параметрического монитора кристалла СБИС для своевременной отбраковки потенциально ненадежной
продукции и снижения затрат на ее дальнейшую обработку и испытания.

5. Установление технологически и статистически обоснованных нормативных границ контролируемых величин с учетом всех основных влияющих на них факторов.

6. Применение схемы двуступенчатого операционного контроля для уменьшения объемов контроля внутри одной партии.

2. Выходные характеристики

технологических процессов и

методы их контроля

В базовых технологических процессах, разработанных в рамках возможностей технологического комплекса НИИСИ РАН, можно выделить ряд блоков формирующих операций, выходные характеристики которых требуют постоянного контроля. Структура операционного контроля кристального производства с привязкой к блокам технологических операций и указанием контролируемых характеристик приведена в таблице 1.

Таблица 1. Номенклатура выходных характеристик технологических процессов, подлежащих операционному контролю в кристальном производстве.

Наименование процесса	Выходные характеристики
1. Фотолитография	
1.1 Нанесение адгезионного слоя, фоторезиста и антиотражающего покрытия	Толщина пленки и равномерность толщины по пластине
1.2 Экспонирование	_
1.3 Проявление	Минимальные контролируемые размеры и их разброс по пластине
	Точность совмещения слоев
2. Физическое и химическое осаждение из газовой фазы диэлектрических и металлических слоев, осаждение кремния и поликремния	Толщина пленки и равномерность толщины по пластине
	Коэффициент отражения (для металлических пленок)
	Напряжение в пленке
	Концентрация и однородность по пластине легирующей примеси в пленке ФСС (бора, фосфора)
	Микрорельеф поверхности пленки (для металлических пленок)
	Поверхностное сопротивление (для металлических пленок)
	Заполнение рельефа поверхности пластины
3. Плазменное травление, плазменное удаление фоторезиста, (для ионной чистки перед нанесением Ті контролируется скорость и равномерность травления)	Скорость, равномерность и селективность травления
	Профиль
	Плазменные повреждения
	Величина подтравливания нижележащего окисла
	Минимальные контролируемые размеры и их разброс по пластине
4. Ионная имплантация	Равномерность ионного легирования по пластине
	Поверхностное сопротивление (после отжига) и его равномерность по пластине
	Концентрация и профиль распределения имплантированной примеси (до и после отжига)
5. Быстрая термическая обработка	Равномерность толщины пленки или удельного поверхностного сопротивления по пластине
	Механические напряжения в рабочих слоях
	Привнесенная дефектность
 Кимическая отмывка и травление в жидких реагентах. 	Скорость, неравномерность и селективность травления
	Шероховатость поверхности
	Привнесенная дефектность
	Остаточная концентрация металлов на поверхности пластины
7. Химико-механическая планаризация	Шероховатость поверхности
	Толщина остаточной пленки и ее равномерность по пластине
	Неудаляемые дефекты (плотность дефектов)
	Профиль края пластины
*Для всех процессов	Привносимая оборудованием дефектность
8. Контроль пластины после формирова	ния первого уровня металлизации и перед разделением на кристаллы
8.1 Контроль параметров тестовых структур параметрического монитора	Согласно требованиям к параметрическому монитору и применяемым методикам контроля
8.2 Контроль функционирования перед разделением на кристаллы	Согласно ТУ на микросхему

Выходные характеристики технологических процессов

В разработанную систему операционного контроля технологического процесса изготовления кристаллов микросхем были включены как методы, использовавшиеся ранее в российской микроэлектронной промышленности при контроле процессов предыдущих поколений (0,8 мкм и более), так и принципиально новые методы, используемые в настоящее время ведущими мировыми производителями микросхем. Все новые методы были впоследствии предложены для внесения в систему стандартов российской микроэлектронной отрасли в качестве дополнения к ОСТ 11 14.1012-99 [1].

Многие характеристики допускают контроль с использованием различных методов. Факторами, влияющими на выбор конкретного метода контроля являются оперативность, локальность, информативность, трудоемкость, необходимость разрушения образца, погрешность измерения в сравнении с настроенностью контролируемого процесса.

Некоторые физические параметры являются выходными характеристиками многих процессов и контролируется с применением одних и тех же методов. Такими параметрами являются:

- толщина слоя и равномерность толщины слоя по пластине;

- критичные (линейные) размеры элементов;

- поверхностное сопротивление слоя;

- рельеф поверхности слоя;

- привносимая дефектность по результатам аттестаций оборудования;

- дефектность топологического рисунка на рабочих пластинах.

Кроме этого выходными характеристиками отдельных процессов являются:

точность совмещения слоев;

- напряжение в пленке;

- заполнение рельефа поверхности пластины;

- доза и равномерность ионного легирования

- распределение имплантированной примеси по глубине;

- остаточная поверхностная концентрация металлов;

- концентрация и однородность легирующей примеси в легированных стеклах;

- скорость и селективность травления.

В целом же качество изготовления кристаллов оценивается по финишным выходным характеристикам:

параметрам тестовых структур параметрического монитора;

- электрическим параметрам (например, токи потребления) и функциональным тестам.

Методы операционного контроля

В настоящее время для контроля выходных характеристик технологических процессов в кристальном производстве технологического комплекса НИИСИ РАН используются следующие методы:

- метод спектральной эллипсометрии;

- метод оптической спектрофотометрии;

- четырехзондовый метод измерения поверхностного сопротивления;

- метод фотоакустических измерений (другое название – метод генерации термических волн);

- метод растровой электронной микроскопии, включая методы изготовления поперечных шлифов (сечений) контролируемых структур;

- метод контактной профилометрии;

- метод регистрации рассеянного поверхностью света;

- метод дифракционной микроскопии;

- метод ИК-спектрофотометрии;

- метод вторичной ионной масс-спектроскопии;

 метод рентгено-флуоресцентной спектроскопии в геометрии полного внешнего отражения первичного рентгеновского пучка;

- электрофизические методы измерения параметров тестовых структур параметрического монитора;

- контроль электрических характеристик и функциональный контроль.

Контроль выходных характеристик процессов 1. Контроль толщин слоёв.

Все контролируемые слои делятся на два типа: оптически прозрачные и оптически непрозрачные. Метод контроля зависит от типа слоя.

Для измерения толщины *оптически прозрачных* слоев, к которым в процессах технологического комплекса НИИСИ РАН относятся фоторезист, оксид кремния, поликремний, тонкие, менее 100 нм, слои титана и нитрида титана и равномерности толщины слоя по пластине применяются методы спектроскопической эллипсометрии и спектрофотометрии.

Метод спектральной эллипсометрии основан на облучении исследуемого объекта фокусированным поляризованным световым пучком в широком спектральном диапазоне и регистрации изменения параметров поляризации отраженного от исследуемого объекта света в диапазоне длин волн от 320 нм до 800 нм.

Метод оптической спектрофотометрии основан на облучении исследуемого объекта фокусированным световым пучком в диапазоне длин волн от 190 нм до 800 нм и регистрации спектра интенсивности оптического излучения, отраженного от исследуемого объекта. Оба метода являются бесконтактными и способны измерять многослойные структуры (эллипсометр – до 9 слоев, спектрофотометр – до 5 слоев). Метод спектроскопической эллипсометрии обладает большей чувствительностью, но меньшей локальностью по сравнению со спектрофотометрией (минимальный размер светового пятна эллипсометра составляет 50×50 мкм, спектрофотометра – 4,5 мкм).

Эллипсометрия используется для измерения толщин сверхтонких слоев (менее 100 нм) и многослойных структур, а спектрофотометрия преимущественно для аттестационных измерений и измерений оптических характеристик металлических слоев (напр. коэффициентов отражения).

Для измерения *толщины оптически непрозрачных* слоев (алюминий, вольфрам, силицид титана)

используются четырехзондовый метод и метод генерации термических волн. На пластинах спутниках применяется без топологического рисунка автоматизированный четырехзондовый метод измерения удельного поверхностного сопротивления. Четырехзондовый метод основан на измерении разности потенциалов между двумя зондами при пропускании тока через два других внешних зонда. Зонды контактируют с поверхностью слоя. расположены в линию на равных расстояниях друг от друга. Полученное значение сопротивления пересчитывается в значение толщины слоя или концентрации активной примеси полупроводникового слоя.

Термоволновой метод является прямым и позволяет контролировать толщину металлических слоев с помощью измерения параметров термоакустических волн, сгенерированных в исследуемом образце модулированным лазерным облучением. Детектируемым откликом является комплексный коэффициент отражения света от образца, несущий информацию о свойствах его приповерхностной области. Термоволновой метод является бесконтактным и высоко локальным (минимальный размер пятна лазера 6 мкм), что позволяет использовать его на рабочих пластинах.

Оба этих метода применяются также для контроля *дозы и однородности ионной имплантации*.

Термоволновой метод имеет более высокую чувствительность по дозе (от 10^{10} см⁻²) и не требует обязательного активационного отжига, что позволяет проводить контроль на рабочих пластинах и максимально исключить факторы, влияющие на результат операционного контроля (дополнительный отжиг, химическая отмывка). При дозах легирования выше 10^{13} см⁻² может применяться четырехзондовый метод, требующий активационного термического отжига.

2. Контроль критичных (линейных) размеров элементов топологии проводится с помощью метода растровой электронной микроскопии. Метод основан на принципе сканирования поверхности пластины пучком электронов, регистрации пространственного распределения вторичных или отраженных электронов и анализа полученного распределения с учетом скорости сканирования. По интенсивности рассеяния также можно косвенно судить о профиле исследуемого объекта.

3. Рельеф поверхности слоя контролируется с контактной профилометрии, помощью метода заключающегося В регистрации сигнала, пропорционального вертикальному перемещению датчика с алмазной иглой. Датчик расположен перпендикулярно контролируемой поверхности и может перемещаться вдоль этой поверхности. Полученный профиль также может использоваться для определения геометрии пластины, по которой можно контролировать механические напряжения в формируемых слоях.

4. Привносимая оборудованием дефектность.

Для контроля используется метод регистрации рассеянного поверхностью света. Метод заключается в сканирующем освещении рабочей поверхности пластины с помощью лазерного луча и регистрации интенсивности обратно рассеянного в некоторый телесный угол света. Интенсивность рассеянного света зависит от размера поверхностного дефекта. Метод также используется для контроля *микрорельефа* поверхности. 5. Точность совмещения слоев.

При экспонировании элементов с минимальным топологическим размером 0,35 мкм И выше допускается визуально-оптический контроль. При этом величина рассовмещения слоев определяется путем измерения размеров фигур совмещения (например, «квадрат в квадрате»). При переходе к меньшим размерам элементов топологии требования к точности совмещения слоев возрастают и данный вид контроля не может быть использован. Для технологии 0,25 мкм в ОМТ был разработан метод автоматизированного контроля совмещения слоев, основанный на принципах дифракционной микроскопии. Контроль проводится по специальной метке в четырех диаметрально противоположных модулях на пластине. Точность метода составляет 5 нм.

6. Заполнение рельефа поверхности пластины контролируется с помощью исследования поперечного сечения контролируемой структуры в растровом Для формирования электронном микроскопе. сечения поперечного травление используется (FIB) фокусированным ионным пучком или прецизионный механический скол влоль кристаллографической плоскости кристалла. При использовании FIB метод позволяет контролировать структур параметры поперечного сечения без разрушения пластины в любой ее точке с высокой локальностью. В случае механического скола добиться высокой локальности без использования специализированного оборудования практически невозможно, в ряде случаев требуется дополнительное декорирование полученного поперечного сечения в растворе травителя.

7. Концентрация и однородность примеси в легированных стеклах. Метод ИК – спектрофотометрии используется для определения содержания фосфора в пленках фосфоросиликатного стекла (ФСС). Метод основан на сравнении ИК-спектров поглощения измеряемых и контрольных пластин по видам дифференциальной кривой, которая, при прочих равных условиях, зависит от различия в составе ФСС. Кроме того может использоваться метод рентгеноспектрального анализа для определения концентрации различных элементов в слое. Метод основан на регистрации интенсивности рентгеновского характеристического излучения элемента, возбуждаемого в образце электронным зондом.

8. Определение скорости и равномерности травления на пластинах без топологического рисунка сводится к измерению толщины слоя до и после травления с последующим вычислением скорости на основании данных о времени травления. Для определения скорости травления на пластинах с топологическим рисунком измеряется глубина канавок (между полосками фоторезиста) до и после травления, для чего применяется метод профилометрии. Метод контактный и достаточно трудоемкий, но получаемый результат не зависит от оптических свойств материала.

10. Профиль травления определяется либо методом контактной профилометрии, либо методами, используемыми для исследования заполнения рельефа поверхности пластины, описанными выше.

11. Плазменные повреждения и иные дефекты топологического рисунка могут быть зафиксированы визуально-оптическим методом с помощью оптического микроскопа в режиме светлого и темного поля, однако такой вид контроля подвержен воздействию человеческого фактора, не позволяет зафиксировать дефекты менее 0,5 мкм и не обеспечивает контроль всей площади пластины. Более эффективным является использование автоматизированного метода поиска дефектов, основанного на последовательном освещении всей поверхности пластины световым пучком, падающим перпендикулярно к поверхности (светлое поле) или под малым углом к поверхности (темное поле) и последующем сравнении получаемых изображений соседних кристаллов с помощью специализированного программного обеспечения.

После локализации дефектов, проводится их анализ и классификация с помощью оптической или растровой электронной микроскопии.

На сегодняшний день данный метод автоматизированного контроля в составе технологического комплекса НИИСИ РАН отсутствует. В ближайшее время запланирована закупка установки автоматизированного оптического контроля дефектности топологического рисунка на полупроводниковых пластинах ЭМ-6429Б, производства «КБТЭМ ОМО» НПО «Планар», Беларусь. Установка позволяет осуществлять 100%-й контроль поверхности пластины в автоматическом режиме с разрешением, достаточным для контроля технологического процесса с проектными нормами 0,25 мкм.

12. Распределение примеси по глубине после ионного легирования контролируется с помощью вторичной ионной масс-спектроскопии на этапе технологического Метол разработки процесса. регистрации заключается В многоканальным детектором отношения массы к заряду (однозначно характеризующего химический элемент и его изотоп) вторичных ионов, образованных в результате распылении поверхностного слоя образца первичными пучком. В результате для ионным кажлого химического элемента строится зависимость количества ионов от времени распыления, то есть профиль распределения примеси по глубинные при известной скорости травления.

13. Остаточная концентрация металлов на поверхности пластины контролируется с помощью рентгено-флуоресцентной спектроскопии в геометрии внешнего отражения первичного полного рентгеновского пучка. Метод основан на сборе и анализе спектра характеристического рентгеновского полученного после возбуждения излучения, поверхностного слоя образца (40-80Å) рентгеновским пучком и регистрации интенсивности излучения с определённой энергией. В результате может быть определен количественный элементный состав на поверхности пластины.

Недостатком метода является его недостаточная чувствительность (более 10¹¹ см⁻²) к ионам натрия и калия, обусловливающих термополевую нестабильность подзатворной системы МОП-транзисторов. Поэтому для контроля подвижных ионов планируется использовать электрофизические методы, такие как метод динамических вольтамперных характеристик (ДВАХ), термостимулированных ионных токов (ТСИТ) или по сдвигу вольтфарадной характеристики (ВФХ).

14. Контроль параметрического монитора.

Отдельным видом комплексного операционного контроля является контроль качества выполнения технологических операций по финишным выходным характеристикам на основании электрофизических измерений. Такие измерения проводятся на тестовых структурах, входящих в состав параметрического монитора, расположенного на линиях реза между кристаллами на пластине. Основными контролируемыми параметрами являются нижеследующие.

1. Параметры транзисторных структур:

• пороговое напряжение (*V_T*) для всех видов транзисторов;

• линейная крутизна (g_m) для всех видов транзисторов;

• эффективная длина канала (L_{eff}) для транзисторов с минимальной длиной канала каждого типа;

• ток открытого транзистора (*I*_{on}) на репрезентативной выборке транзисторов из комплекта;

• ток закрытого транзистора (I_{off}) на репрезентативной выборке транзисторов из комплекта;

• время задержки на тестовой структуре в виде функциональной цепочки, которая позволяет измерить время задержки распространения сигнала;

• полевая утечка между смежными транзисторами, находящимися на минимальном расстоянии друг от друга, при максимально допустимом напряжении.

2. Параметры, характеризующие качество металлизированной разводки:

• поверхностные (слоевые) сопротивления всех проводящих слоев;

• контактные сопротивления всех межуровневых контактов.

Контроль параметрического монитора в технологическом маршруте изготовления кристаллов микросхем проводится дважды – первый раз после формирования первого уровня металлизации для своевременной отбраковки потенциально ненадежной продукции и снижения затрат на ее дальнейшую обработку и испытания, второй раз после формирования последнего, верхнего уровня металлизации для контроля качества проведения всего технологического процесса в целом.

Кроме этого, с помощью электрофизических измерений оцениваются надежностные характеристики изготавливаемых изделий. Такие измерения проводятся на структурах быстрой оценки надежности, позволяяющих оперативно оценить развитие конкретного известного механизма отказа, основными из которых являются:

- деградация параметров транзисторов под влиянием горячих носителей;

- эффект электромиграции;
- времязависимый пробой диэлектрика;

- нестабильность порогового напряжения р-канальных МОП-транзисторов при повышенной температуре;

- деградация контактного сопротивления;

- диффузия в затворе.

Планы контроля и способы своевременной отбраковки

С целью уменьшения объема контроля в пределах одной партии была введена двухступенчатая схема операционного контроля, предусматривающая первичный контроль первой и последней пластины из партии с применением более узких контрольных границ (КГ), по сравнению с нормативными границами (НГ). Контрольные границы строятся по результатам статистического контроля процесса. В случае, если результат контроля не удовлетворяет установленным КГ, проводится контроль всех пластин из партии. Партия бракуется при непопадании в НГ для данного базового технологического процесса результатов контроля хотя бы одной пластины из партии. Такой способ контроля эффективен для отработанного техпроцесса, обладающего высокой точностью и настроенностью. На стадии отработки техпроцесса в мелкосерийном производстве и при индивидуальной обработке пластин необходим контроль всех пластин из партии.

3. Оценка метрологических характеристик контрольноизмерительного оборудования

Под метрологическими характеристиками средства измерения понимают характеристики, влияющие на результат измерений и на его погрешность.

Точность и стабильность результатов измерений необходима для обеспечения правильности принимаемых решений на всех уровнях управления и на всех этапах изготовления и испытаний микроэлектронных изделий. Получение недостоверной информации по результатам измерений приводит к снижению качества продукции и росту общих затрат на ее изготовление.

Общепринятым критерием достоверности контроля выходной характеристики процесса является отношение точности измерения характеристики к ее допуску (precision to tolerance ratio) P/T. Показателем хорошего уровня достоверности является выполнение отношения P/T < 0,1, приемлемого уровня – P/T < 0,3. Допуск на каждую характеристику устанавливается в соответствии с технологическими требованиями, определяется точность измерения аппаратными возможностями измерительного оборудования сочетании с внешними источниками изменчивости, такими как температура и влажность воздуха,

субъективный фактор (оператор, выполняющий измерение), и другие.

На технологическом комплексе НИИСИ РАН, в соответствии с отечественными и зарубежными рекомендациями по оценке возможностей измерительного оборудования, принята процедура периодической оценки погрешности измерений. Для расчета погрешности измерений на установке применяются нижеследующие статистические параметры.

1. Статическая повторяемость σ_{SR} (Static Repeatability)- стандартное отклонение значений от среднего, вычисляемое по N измерениям.

Этот параметр характеризует стабильность аппаратного комплекса за время измерений.

$$X^{SR} = \frac{\sum_{i=1}^{i=N} X_i}{N} \sigma_{SR} = \sqrt{\frac{\sum_{i=1}^{i=N} (X^{SR} - X_i)^2}{N - 1}}$$

2. Динамическая повторяемость σ_{DR} (Dynamic Repeatability) – стандартное отклонение значений от среднего, вычисляемое по M измерениям с промежуточной загрузкой-выгрузкой (или изменением положения) измеряемой пластины. Характеризует стабильность измерительной системы в совокупности с загрузчиком за время проведения серии измерений.

$$X^{DR} = \frac{\sum_{j=1}^{j=M} X_j^{SR}}{M}, \ \sigma_{DR} = \sqrt{\frac{\sum_{j=1}^{j=M} (X^{DR} - X_j^{SR})^2}{M - 1}}$$

 Стабильность σ_S (Stability)- параметр, характеризующий стабильность измеряемой величины в течение длительного времени.

Ежедневно, в течение длительного промежутка времени (L дней), проводятся $N \times M$ измерений подряд по алгоритму, описанному в п.п. 1 и 2. Стабильность $\sigma_{\rm S}$ рассчитывается по формуле:

$$X^{S} = \frac{\sum_{k=1}^{L} X_{k}^{DR}}{L}, \sigma_{S} = \sqrt{\frac{\sum_{k=1}^{L} (X^{S} - X_{k}^{DR})^{2}}{L - 1}},$$

4. Воспроизводимость σ_R (Reproducibility) – суммарная вариация (отклонение) параметров измерительной системы, включающая в себя вклад статической и динамической повторяемости:

$$\sigma_{R} = \sqrt{\sigma_{SR}^{2} + \sigma_{S}^{2}}$$

5. Погрешность измерений *P* (Precision) определяется как:

$$P = 6 \sigma_R (\pm 3\sigma_R)$$

4. Заключение

В рамках технологического комплекса НИИСИ РАН разработана и реализована комплексная система операционного контроля кристального производства, включающая в себя результаты контрольных измерений рабочих пластин, пластин спутников, а также результаты аттестаций технологического оборудования. 42

современных методов контроля, ряд из которых был впервые реализован в российском микроэлектронном производстве, на сочетании прямых и косвенных измерений, на использовании двухуровневой схемы контроля. Все это позволяет своевременно отбраковывать потенциально ненадежную продукцию и снизить затраты на ее изготовление и испытания.

Тем не менее, при переходе к изготовлению изделий по проектным нормам 0,25 мкм выявился ряд недостатков используемой системы операционного контроля, над устранением которых сейчас ведутся работы по следующим направлениям:

- расширение набора и параметров тестовых структур параметрического монитора для выявления влияния короткоканальных эффектов на характеристики транзисторов и построения обратной связи между измеряемыми параметрами и режимами технологических процессов. - разработка процессов автоматизированного контроля дефектности топологического рисунка различных технологических слоев на базе планируемой к закупке установки ЭМ-6429Б, НПО «Планар», позволяющих осуществлять 100%-й контроль поверхности пластины в автоматическом режиме с разрешением 0,25 мкм.

- корректировка КГ и НГ на электрофизические параметры в соответствии со схемотехническими требованиями и технологическими возможностями.

Литература

1. ОСТ 11 14.1012–99. Микросхемы интегральные. Технические требования к технологическому процессу. Система и методы операционного контроля. 22 ЦНИИИ МО, 1999

Process monitoring specificities for low-volume manufacturing of integrated circuit chips

V.J. Troitsky, M.V. Oreshkov, T.V. Kozyreva

Abstract. This work is focused on organization of process monitoring of integrated circuits (IC) manufacturing under low-volume and interrupted conditions at technological complex of SRISA RAS. The descriptions for IC manufacturing process monitoring methods are given. Such methods have to make it possible to determine the influence of input parameters on monitored characteristics of given test structures. The organization principles of complex process monitoring system are discussed that ensures separation of defective wafers in advance during chip production. An integration of qualification and reliability methods is supposed as part of process monitoring system of low-volume production for ensuring stated requirements.

Методика контроля точности совмещения фотолитографических слоев при изготовлении микросхем с проектными нормами 0,25 мкм и менее

А.А Захарченко¹, Н.С. Кукина, Е.П. Михальцов², А.А.Столяров, В.Ю. Троицкий¹

1 кандидат физико-математических наук, 2-кандидат технических наук

В статье описывается методика контроля точности совмещения топологических слоев с помощью дифракционных меток и ее применение в НИИСИ РАН при изготовлении микросхем с проектными нормами 0.25 мкм. Методика позволяет осуществлять контроль рассовмещения слоев с погрешностью 10 нм без дополнительной модернизации или закупки нового оборудования. Проведено испытание методики путем сравнения полученных результатов с результатами измерений на установке CD SEM.

1. Введение

При уменьшении проектных норм микроэлектронных изделий все более сложным становится контроль совмешения фотолитографических слоев В технологическом процессе изготовления этих изделий. Так, если для микросхем с проектными нормами 0.35 мкм и более вполне эффективен контроль совмещения слоев с помощью оптической микроскопии с использованием шкалы «нониус» или фигур «квадрат в квадрате» (погрешность таких измерений порядка 100 нм), то при уменьшении размеров элементов до 0.25 мкм и менее такой контроль уже неприменим, так как не может обеспечить требуемой погрешности измерений (<50 нм).

На зарубежных микроэлектронных предприятиях для контроля рассовмещения фотолитографических слоев в технологическом процессе изготовления субмикронных микросхем используется специализированное И весьма дорогостоящее оборудование. Таким образом, при переходе к изготовлению изделий с проектными нормами 0.25 мкм и менее перед ОМТ НИИСИ стоял выбор: пойти по пути дооснащения фабрики специализированной установкой контроля совмещения слоев, либо разработать альтернативный метод контроля и специальное программное обеспечение (ПО), которые могут быть реализованы на имеющемся оборудовании.

Было принято решение идти по второму пути, в результате была разработана и реализована методика контроля точности совмещения слоев с помощью дифракционных меток, имеющая погрешность измерений 10 нм.

2. Описание методики

В [1, 2] описаны дифракционные методы, наиболее распространенные в настоящее время в мире для контроля процессов изготовления субмикронных изделий с проектными нормами 90 нм и менее. Методы основаны на анализе отраженного от многоуровневых дифракционных решеток света, свойства которого зависят от сдвига уровней решетки друг относительно друга.

Недостатком таких методов является сложность расчетов, необходимость калибровки, зависимость от условий регистрации и т.п. Вместе с тем существует группа дифракционных методов, которые позволяют обойти указанные недостатки. Речь идет о методах контроля целого набора решеток с различным контролируемым сдвигом в расположении решеток верхнего и нижнего уровня [3, 4]. Достоинством этих методов является высокая устойчивость к изменениям условий измерения, отсутствие необходимости построения сложной математической молели формирования изображения И необходимости калибровки. Другими словами вся аппаратная часть метода. обеспечивающая высокую точность измерений, переносится на сами структуры. Именно эти методы и легли в основу разработанной методики контроля совмещения слоев при изготовлении микросхем.

Фигуры контроля, разработанные в ОМТ НИИСИ РАН, представляют собой набор накладывающихся друг на друга N дифракционных решеток (N = 11 или 13). В нижележащем контролируемом слое все решетки одинаковые (базовая фигура), а в лежащем сверху слое решетки последовательно сдвинуты на контролируемый сдвиг шагом в 50 нм следующим образом (рис.1):

- центральные решетки в обоих слоях совпадают (сдвиг 0);

- следующие слева от центральной решетки сдвинуты на 50, 100, 150, 200, 250 нм влево;

- следующие справа от центральной решетки сдвинуты на 50, 100, 150, 200, 250 нм вправо.

2.1 Алгоритм измерений и расчеты

При наблюдении изготовленных дифракционных структур в оптическом микроскопе при небольшом увеличении происходит дифракция света на объективе, которая зависит от сдвига нижней и верхней решетки относительно друг друга и длины волны света. Поэтому решетки при заданных условиях наблюдения (апертура объектива достаточно мала) отображаются как области, заполненные одним цветом, каждая своим. При нулевом рассовмещении цвета решеток будут распределены симметрично относительно центра. Если рассовмещение отлично от нуля, центр цвета будет симметрии сдвигаться на соответствующую рассовмещению величину. Следовательно, задача измерения рассовмещения сводится к поиску центра цветовой симметрии структуры.

Для корректных измерений необходимо обеспечить такие условия регистрации изображений, при которых достигается хороший цветовой контраст, и элементы дифракционных решеток не разрешаются микроскопом. Практика показала, что для этого подходит режим светлого поля инспекционной станции Leica INS3000 при увеличении объектива 10х (2-ой объектив) с максимальной апертурной диафрагмой. Соответствующие изображения для слоев **ACTIVE-POLY** показаны на рис. 2.

Для измерения центра цветовой симметрии цветовая информация записывается в виде набора интенсивности красного, зеленого и синего цветов: R_i, G_i, B_i , i = 1..N, где N – число решеток в наборе контроля рассовмещения двух слоев.

Поиск центра симметрии производится методом наименьших квадратов путем минимизации отличия измерений каждой компоненты цвета от их Фурьеаппроксимации с четными коэффициентами (невязки): min $\eta^2(\Delta)$,

$$\eta^{2}(\Delta) = \begin{bmatrix} \sum_{i=1}^{N} \left(R_{i} - \sum_{j=0}^{M} r_{j} \cos \frac{2\pi}{P} (j + \Delta) \right)^{2} + \\ \sum_{i=1}^{N} \left(G_{i} - \sum_{j=0}^{M} g_{j} \cos \frac{2\pi}{P} (j + \Delta) \right)^{2} + \\ \sum_{i=1}^{N} \left(B_{i} - \sum_{j=0}^{M} b_{j} \cos \frac{2\pi}{P} (j + \Delta) \right)^{2} \end{bmatrix}, \quad (1)$$

где $\eta^2(\Delta)$ - невязка, Р – период решетки (1 мкм для партии 2850), М – количество коэффицентов ряда

COM'



0024

Расчеты



Фурье (исключая нулевой), 1 < M < N, Δ рассовмещение слоев, r_i, g_i, b_i - коэффиценты ряда Фурье для аппроксимации R_i, G_i, B_i соответственно, которые получаются решением следующих задач для каждой компоненты цвета:

$$\min_{r_1..r_M} \sum_{i=1}^{N} \left(R_i - \sum_{j=0}^{M} r_j \cos \frac{2\pi}{P} (j+\Delta) \right)^2$$
(2)

Данный метод вытекает из того факта, что любая четная непрерывная функция раскладывается в ряд Фурье, в котором присутствуют только косинусы. Если же функция нечетна, то в ряде Фурье присутствуют только коэффициенты при синусах.

Для решения задачи (2) используется псевдообратная матрица для матрицы:

$$A_{\Delta} = \begin{bmatrix} 1 & \cos\frac{2\pi}{P}(\Delta_{1} + \Delta) & \dots & \cos\frac{2\pi}{P}M(\Delta_{1} + \Delta) \\ 1 & \cos\frac{2\pi}{P}(\Delta_{2} + \Delta) & \dots & \cos\frac{2\pi}{P}M(\Delta_{2} + \Delta) \\ 1 & \dots & \dots & \dots \\ 1 & & & & \dots \\ 1 & \cos\frac{2\pi}{P}(\Delta_{N} + \Delta) & & \cos\frac{2\pi}{P}M(\Delta_{N} + \Delta) \end{bmatrix}_{,(3)}$$

где $\Delta_i, i = 1..N$ - известные сдвиги решеток, заложенные в проект.

Невязка из (1) при этом может быть записана как $\eta^{2}(\Delta) = \left\| R - A_{\Delta}A_{\Delta}^{+}R \right\|^{2} + \left\| G - A_{\Delta}A_{\Delta}^{+}G \right\|^{2} + \left\| B - A_{\Delta}A_{\Delta}^{+}B \right\|^{2}$

где R, G, B - результаты измерений, представленные в виде векторов в N-мерном евклидовом пространстве, A_{Λ}^+ - псевдообратный оператор (матрица) к A_{Λ} .

На рисунке 3 представлены графики, иллюстрирующие процесс решения задачи (1) для

M=3.

По краям диапазона невязка будет стремиться к минимуму, обусловленному тем фактом, что любую функцию за областью ее определения можно продолжить четной функцией. Поэтому диапазон корректных измерений меньше заданного диапазона смещения меток. Необходимо оставить вне диапазона смещения меток. Необходимо оставить вне диапазона поиска минимума хотя бы по два-три измерения с каждой стороны. Установлено, что на практике для N=11 и диапазона контролируемого сдвига [-0.25;0.25] мкм можно измерить рассовмещение в диапазоне [-0.18;0.18] мкм. Причем, если величина рассовмещения точно совпадает с краем диапазона поиска минимума, то результату доверять не следует.

2.2. Выбор числа М

Если число М велико, возможна ошибочная аппроксимация, которая аппроксимирует даже шумы, что приводит к неверным результатам. Так при М=N для любого смещения не из набора известных сдвигов решеток возможно подобрать аппроксимацию, доставляющую 0 для невязки (1).

Если М мало, то возможна слишком грубая аппроксимация, что приводит к потери точности. Реально число М должно соответствовать физическим принципам формирования изображения. Для шага 50 нм и средней длины волны целесообразно выбрать М=3.

2.3. Погрешность измерений

В рамках указанной математической модели минимальная невязка (1) объясняется шумом цвета. Будем считать, что шум каждой компоненты цвета



Рис. 3. Графики, иллюстрирующие процесс решения задачи (1). Цвета соответствуют компонентам R,G,B. Кругами обозначены результаты измерений, пунктирными линиями – результаты наилучшей аппроксимации $A_{\Lambda}A_{\Lambda}^{+}R$, $A_{\Delta}A_{\Lambda}^{+}G$,

 $A_{\Delta}A_{\Delta}^{+}B$. Черным пунктиром показан корень из невязки в (1), умноженный на коэффициент $\frac{20}{\sqrt{3N}}$



Рис. 4. Графики зависимости рассовмещения слоев Active – Poly от порядкового номера кристалла. Синий – измерения на фоторезистивной маске на INS3000, зеленый – измерения после травления на INS3000, фиолетовый – измерения после травления на INM300, красный – измерения на NanoSEM 3D после травления.

является случайной величиной с нормальным распределением $v \sim N(0, \sigma^2)$, а σ^2 не зависит от цвета. Тогда невязка, отнесенная к σ^2 , имеет распределение хи-квадрат с 3N степенями свободы, близко что к нормальному распределению $\eta^2(\Delta)$ Это $\sim N(3N,6N)[5].$ означает, что с σ^2 вероятностью около 95% невязка будет меньше чем $\sigma^2(3N+2\sqrt{6N})$

На практике отклонение результатов измерений от модели объясняется не только шумом, но и несовершенством модели, а распределение шума может быть не нормальное. Поэтому целесообразно оценить параметр σ^2 исходя из величины минимальной невязки η^{2}_{\min} . Наилучшей его оценкой

будет
$$\sigma = \sqrt{\frac{\eta^2_{\min}}{3N}}$$

Таким образом, погрешность измерений следует определить как половину интервала значений Δ , для которых:

$$\left(\frac{\eta(\Delta)}{\eta_{\min}}\right)^2 < 1 + 2\sqrt{\frac{2}{3N}} \tag{4}$$

На рисунке 3 вертикальными линиями показан интервал измерений рассовмещения с учетом погрешности, рассчитанный по формуле (4).

3. Результаты

Для осуществления точного расчета рассовмещения слоев с использованием данной методики было разработано программное обеспечение, позволяющее



Рис. 5. График зависимости полученного рассовмещения слоев Poly, Cont от порядкового номера кристалла. Синий – измерения на фоторезистивной маске, красный – измерения после травления.



Рис. 6. Общий вид базовой фигуры контроля совмещения

оператору в полуавтоматическом режиме выбрать нужную фигуру контроля и получить значение рассовмещения слоев в области расположения фигуры с указанием погрешности измерений. Время расчета при каждом измерении составляет менее 1 сек.

Были получены результаты, однозначно указывающие на применимость разработанной методики для контроля совмещения слоев с требуемой точностью.

Так, были измерены рассовмещения слоев Active -Poly с помощью инспекционной станции Leica INS3000 И микроскопа Leica INM300 на фоторезистивной маске и после процессов травления. Полученные результаты были проверены на измерительном растровом электронном микроскопе NanoSEM 3D (CD SEM), погрешность измерений которого менее 3.5 нм. Измерения были проведены на трех пластинах в 11 кристаллах, расположенных по горизонтальному диаметру пластины. Все результаты совпали с учетом найденной погрешности измерений, составляющей не более 8 нм (рис. 4,5).

На основании проведенных исследований были выработаны требования к фигурам контроля рассовмещения слоев.

Фигура контроля должна состоять из двух групп решеток, каждая из которых должна иметь вертикальный размер не менее 30 мкм. Расположение элементов в решетках разных групп взаимно ортогонально. В базовой фигуре (рис. 6) все решетки в одной группе одинаковы.

В фигуре со сдвигом все решетки, кроме центральных, последовательно сдвинуты на 50 нм относительно соседних решеток. Решетки первой группы, расположенные слева от центральной, сдвигаются влево, а расположенные справа – вправо. Решетки второй группы, расположенные слева от центральной сдвигаются вниз, а расположенные справа – вверх. Центральные решетки совпадают с базовыми.

Для каждых двух контролируемых слоев необходимо впечатывать базовую фигуру в нижний слой и фигуру со сдвигом в верхний слой. Корректным является контроль рассовмещения только двух слоев. Недопустима ситуация, когда друг на друга накладываются три и более фигур (слоев), т.к. каждая из них имеет свое рассовмещение, что приводит к искажению волновой картины.

Период решетки для контроля всех слоев, кроме слоев МЕТх-СОΝТ, должен составлять 1 мкм с соотношением выступа - впадины 1:1. Для слоев МЕТх-СОNТ необходимо использовать решетку с периодом 2 мкм с соотношением выступа - впадины 1:2

Количество элементов решеток в фигуре должно быть не менее N=11.

4. Выводы

Разработана методика инструментального контроля рассовмещения фотолитографических слоев с использованием дифракционных структур и метод расчета погрешности измерений при использовании данной методики.

Методика позволяет на имеющемся оборудовании осуществлять контроль рассовмещения слоев с погрешностью 10 нм, вполне достаточной для изготовления микросхем с проектными нормами 0.25 мкм и ниже.

Проведено испытание методики путем сравнения полученных результатов с результатами измерений на установке NanoSEM-3D (CD SEM). Результаты совпали, с учетом погрешности измерений, составившей 8 нм.

Разработано программное обеспечение, позволяющее в полуавтоматическом режиме проводить измерения рассовмещения слоев по всей поверхности пластины на установке Leica INS 3000 в рамках операционного контроля технологических процессов изготовления микросхем.

Литература

1 J. Bischoff, R. Brunner, J. Bauer and U. Haak. LIGHT DIFFRACTION BASED OVERLAY MEASUREMENT Proceedings SPIE Vol. 4344-28 (2001).

2 A. Braun. "Diffraction metrology measures overlay down to 45nm", Semiconductor International, January, 2003 3 P. Heimann. "The Color-Box alignment vernier: a sensitive lithographic alignment vernier read at low magnification." Optical Engineering, vol. 29, No. 7, Jul. 1990, pp. 828-836

4 US-Patent 7,230,703 B2

5 А.А. Боровков. Математическая статистика. Оценка параметров. Проверка гипотез. М.:Наука, 1984.

The photolithography overlay precision control technique in 0.25 μm and less IC fabrication process

A.A. Zakharchenko, N.S. Kukina, E.P. Mikhaltsov, A.A. Stolyarov, V.U. Troitskiy

Abstract. The developed precision technique for circuit pattern overlay control, based on the diffraction marks and its application to 0.25 μ m IC fabrication process at SRISA RAS microelectronic division are described. This technique being realized on the inspection station Leica INS 3000 makes it possible to control pattern overlay with 10 nm accuracy without any additional new equipment. The technique has been verified by comparison with the CD SEM measurement results.

Метод контроля профиля эффективной концентрации примеси вдоль канала субмикронных МОП-транзисторов в рамках сток-истоковой инженерии

М.В. Орешков¹, В.Ю. Троицкий²

1-кандидат технических наук, 2- кандидат физико-математических наук

В работе предлагается метод оперативной оценки профиля концентрации примеси вдоль канала транзистора. Метод основан на принципе зарядовой накачки и эффективен при решении задач стокистоковой инженерии и при проведении анализа брака в кристальном производстве неразрушающими электрофизическими методами. Достоинством предлагаемого метода является минимизация воздействия на зарядовое состояние подзатворной системы транзистора и простота реализации.

Введение

В производственной практике отделения микротехнологии НИИСИ РАН приходится решать ряд характерных для субмикронной КМОП технологии сток-истоковой инженерии, задач связанных с оптимизацией режимов отжига, ионной имплантации, геометрических характеристик спейсера, исследованием эффектов горячих носителей, с целью повышения надежности и улучшения характеристик активных элементов интегральных схем. В связи с этим возникает необходимость оперативно оценить характер изменения профиля легирования вдоль канала транзистора. Ключевым электрофизическим методом, который позволяет решать обозначенную задачу применительно к субмикронной технологии без принципиальных ограничений к минимальным размерам элементов, является метод зарядовой накачки [1]. Существует ряд модификаций этого метода, различающихся способом воздействия на МОП-транзистор (МОПТ) для измерения тока зарядовой накачки (I_{cp}) и последующей обработкой полученных зависимостей этого тока от изменяемого параметра. Чаще всего инженерами ставится задача нахождения профиля распределения встроенного заряда и плотности поверхностных состояний (ПС) вдоль канала МОПТ, при этом профиль концентрации примеси не анализируется, либо получен разрушающими физическими методами или рассчитан аналитически [2,3]. Для извлечения профиля легирования, как правило, используется подача импульсов напряжения на затвор с переменной амплитудой относительно постоянного уровня напряжения [4]. Такой способ подачи напряжения требует использования специализированного импульсов, а также базируется на генератора извлечении эквивалентного порогового напряжения. Это требует использования достаточно большой амплитуды импульсов и может приводить к изменению исходного зарядового состояния, что приводит к ошибке в определении концентрации примеси и остальных параметров. В данной работе предлагается метод, предусматривающий использование импульсов постоянной амплитуды, симметрично накладываемых на напряжение базового уровня через RC-цепь, и последующая экстракция величины напряжения, эквивалентной напряжению плоских зон для области канала транзистора. Такой способ не требует специализированного генератора импульсов, прост в реализации и обработке результатов и не требует подачи высоких напряжений, способных существенно исказить зарядовое состояние подзатворной структуры транзистора.

Экспериментальные образцы и измерительное оборудование

В качестве экспериментальных образцов для отработки методики были использованы пи рканальные MOIIT. входящие в состав параметрического монитора И тестового технологий характеристического кристалла C топологическими нормами (ТН) 0.35 и 0.25 мкм, изготовленные на опытно-экспериментальном участке отделения микротехнологий НИИСИ PAH. Транзисторы имели симметричную структуру со слаболегированными областями стока и истока (LDD). Фактическая длина затворов для 0.25 мкм ТН составила 0.18, 0.22 и 0.26 мкм, ширина канала 10 мкм. Для TH 0.35 мкм транзисторы имели топологическую длину канала 0.35 и 2.0 мкм и ширину 20 мкм.

Схема подключения МОПТ приведена на рис. 1 и соответствует классическому методу зарядовой накачки. І_{ср} измеряется пикоамперметром постоянного тока, подключенным к стоку и истоку МОПТ, а не к карману, как в классической реализации метода, с целью минимизации тока утечки через подзатворный диэлектрик. Для экстракции профиля распределения примеси измеряется зависимость тока зарядовой накачки от напряжения базового уровня постоянного напряжения (V_b) , на которое через **RC-**пепь накладывается переменное напряжение от генератора импульсов.

Предлагаемый метод был реализован на оборудовании кафедры «Полупроводниковой электроники» ФГБОУ НИУ «МЭИ». Подробное описание измерительного комплекса приведено в [5].



Рис. 1. Схема снятия зависимости $I_{cp}(V_b)$

Описание метода

Суть метода заключается в построении профиля распределения напряжения вдоль канала МОПТ, соответствующего определенной концентрации носителей заряда (H3) того типа, который является основным для области канала транзистора. Для *п*-МОПТ строится профиль напряжения для дырок, для р-МОПТ – для электронов, т. е. по аналогии с МДПконденсатором - это будет являться некоторым подобием напряжения плоских зон в отношении центральной части канала МОПТ. В простейшем случае, пороговая концентрация НЗ, необходимая для обеспечения достаточного потока НЗ для заполнения состояний в запрещенной зоне при зарядовой накачке, будет задаваться соотношением (1), получаемом из теории рекомбинации Шокли-Рида-Холла:

$$n_{th} = \frac{1}{\sigma_n \tau v_t}$$
для электронов;

$$p_{th} = \frac{1}{\sigma_p \tau v_t}$$
для дырок, (1)

где σ_n и σ_p – эффективное сечение захвата для электронов и дырок, соответственно; τ – время ответа ПС, приравниваемое к периоду подаваемых на затвор импульсов напряжения, v_t – тепловая скорость H3.

Рассмотрим алгоритм метода и его ключевые применительно особенности субмикронным к транзисторам на примере 0.35 мкм *n*-МОПТ. На рис. 2 приведена типичная зависимость I_{cp} от напряжения базового уровня V_b при частоте импульсов f = 500 кГц субмикронного транзистора с достаточно для «затянутым» профилем легирования сток-истоковых областей, обусловленного LDD-технологией. Зависимость получена традиционным способом, однако из-за сильно неоднородного профиля концентрации вдоль канала, на графике практически отсутствует пологий участок с максимальным значением тока.



Рис. 2. Зависимость I_{cp} от V_b для 0.35 мкм МОПТ

Первое ограничение для субмикронных МОПТ при экстракции профиля концентрации связано с необходимостью приложения высокой напряженности электрического поля к подзатворному диэлектрику, что достигается использованием достаточно большой амплитуды импульсов. Критерием правильного выбора амплитуды является наличие пологого участка кривой вблизи максимума тока зарядовой накачки I_{cn}^{max}, что будет свидетельствовать об участии в I_{cp} всей площади под затвором транзистора. В нашем случае зависимость на рис. 2 получена при амплитуде импульсов $V_A = 3.65$ В. При такой амплитуде максимальное напряжение на затворе достигает величины порядка 3.85 В, которое, при толщине подзатворного диэлектрика 7 нм, не способно привести к заметной деградации МОПТ.

При построении профиля концентрации на исходном транзисторе (без стрессовых воздействий), принимаются следующие ключевые допущения:

1. Плотность ПС эффективный И заряд, приведенный к нижней границе раздела, считаются канала. Данное всей длине однородными по отсутствием обеспечиваться требование должно антенных эффектов на этапах проведения технологических операций, а также проведением заключительного отжига В формер-газе (H₂), обеспечивающего как можно более однородную плотность ПС и минимальную величину захваченного заряда в подзатворном диэлектрике, как было отмечено в работе [6].

2. Выбранная амплитуда импульсов обеспечивает участие в токе зарядовой накачки всей площади под затвором МОПТ при I_{cp}^{max} .

При построении профиля с использованием указанного вида H3 (для *n*-МОПТ — дырок) необходимо использовать часть кривой, расположенной в интервале V_b от напряжения, соответствующего минимальной производной dI_{cp}/dV_b (указано справа на графике), до напряжения, при котором достигается максимум тока зарядовой накачки. В этом случае, имея в виду принятые допущения, координатная зависимость эквивалентного напряжения V_{fb} можно вычислить по формулам (2):

$$x(V_b) = \frac{L_g}{2} \left[1 - \frac{I_{cp}(V_b)}{I_{cp}^{\max}} \right];$$

$$V_{fb} = V_b - V_A / 2,$$
(2)

где L_g — фактическая топологическая длина затвора данного МОПТ, например, измеренная с помощью растрового электронного микроскопа (РЭМ). Т.о. для каждого значения V_b находятся соответствующие значения x и V_{fb} . Получаемый профиль напряжения для зависимости $I_{cp}(V_b)$ (рис. 2) приведен на рис. 3.



Рис. 3. Зависимость эквивалентного напряжения V_{fb} от расстояния от края затвора

При определении эффективной концентрации примеси канал МОПТ представляется в виде МДП-конденсаторов с обеими обкладками из полупроводника. Поверхностный потенциал верхней обкладки (затвора) ψ_s^{gale} находится из уравнения электронейтральности:

 $Q_{sc}^{sub}\left(\psi_{p0}\right) + Q_{it}\left(\psi_{p0}\right) + Q_{sc}^{gate}\left(\psi_{s}^{gate}\right) + Q_{eff}^{sub} + Q_{eff}^{gate} = 0, (3)$ где $Q_{e\!f\!f}^{sub}$ и $Q_{e\!f\!f}^{gate}$ – эффективный заряд, приведенный к нижней верхней границе диэлектрика, И Q_{sc}^{gate} — приведенный соответственно; заряд В полупроводнике затвора от верхней границы раздела при ψ_{s}^{gate} ; Q_{sc}^{sub} – приведенный заряд в полупроводнике от нижней границы при значении поверхностного потенциала ψ_{p0} , соответствующего «пороговой» концентрации p_0 , вычисленной по формуле (1); Q_{it} – заряд ПС при $\psi_s = \psi_{p0}$. Подставляя в формулу (3) значение ψ_{p0} для определенной концентрации примеси, численным методом подбирается соответствующее значение поверхностного потенциала ψ_s^{gate} . Величина полупроводнике вычисляется заряда В по стандартному соотношению для простейшей МДПструктуры, полученному из решения уравнения Пуассона:

$$Q_{sc}(\Psi_s) = \sqrt{2qn_i L_i} [\lambda (\exp(-\beta \Psi_s) + \beta \Psi_s - 1) + \lambda^{-1} (\exp(\beta \Psi_s) - \beta \Psi_s - 1)]^{1/2},$$
(4)

где $\beta = \varphi_t^{-1} = q/kT$, q – элементарный заряд, k – константа Больцмана, T – абсолютная температура, $\lambda = n_i/n_0$ – характеристика легирования

полупроводника, n_i и n_0 – собственная и равновесная концентрация НЗ в полупроводнике, соответственно.

Большая плотность ПС может привести к ощутимому сдвигу каждой точки профиля напряжения и его искажению, поэтому при расчете уравнения электронейтральности и напряжения на затворе необходимо учитывать Q_{it} . С этой целью по кривой зарядовой накачки также определяется средняя плотность ПС D_{it} из классической формулы (5) для тока зарядовой накачки [1]:

$$I_{cp} = 2Af q D_{it} kT \ln\left(\nu_t n_i \sqrt{\sigma_n \sigma_p} \frac{\left|V_{th} - V_{fb}\right|}{V_A} \sqrt{t_r t_f}\right), \quad (5)$$

где A – площадь под затвором МОПТ, участвующая в I_{cp} ; V_{th} – эквивалентное пороговое напряжение для зарядовой накачки; t_r и t_f – время нарастания и спада затворных импульсов.

Значение порогового напряжения V_{th} условно определяется по точке, соответствующей половине максимальной производной dI_{cp}/dV_b слева на графике плюс половина амплитуды импульсов, а напряжения плоских зон V_{fb} — в аналогичной точке в правой части кривой минус половина амплитуды импульсов (см. рис. 2). Данный способ определения значений V_{th} и V_{tb} дает наиболее адекватный результат при сравнении с теми же значениями для МДП-структуры и передаточной ВАХ МОПТ. С учетом Q_{it} и разности работ выхода Φ_{ms} , напряжение на затворе, соответствующее достижению необходимой концентрации носителей заряда p_0 (n_0), складывается из следующих нескольких компонент в соответствии с формулой (6):

$$V_{g}^{n(p)}\left(N\right) = \frac{1}{C_{ox}} \left[-Q_{sc}^{sub}\left(\Psi_{n_{th}}(p_{th})\right) - Q_{eff}^{sub} -qD_{it}\left(\varphi_{b} - \Psi_{n_{th}}(p_{th})\right)\right] + \Psi_{n_{th}}(p_{th}) - \Psi_{s}^{gate} - \Phi_{ms},$$
(6)

где C_{ox} – удельная емкость подзатворного $\varphi_{h} = \varphi_{t} \ln(\lambda)$ диэлектрика; объемный электростатический потенциал. Следует иметь в виду, что практически все величины, входящие в формулу (6), кроме C_{ox} , D_{it} и Q_{eff}^{sub} , являются функциями концентрации, но в формуле для простоты представления это не показано. Величина эффективного заряда по возможности должна МДП-конденсаторе, вычисляться на имеющем аналогичную МОПТ конструкцию, по величине напряжения плоских зон или с использованием метода наименьших квадратов. Поверхностный потенциал, соответствующий «пороговой» концентрации n_{th}(p_{th}), вычисляется по формулам Шокли:

$$\Psi_{n_{th}} = \varphi_t \ln(n_{th} / n_0); \Psi_{p_{th}} = -\varphi_t \ln(p_{th} / p_0).$$
(7)

Разность работ выхода Φ_{ms} между материалами подложки и затвора вычисляется как разница между объемными электростатическими потенциалами, один из которых является функцией искомой концентрации, а второй определяется степенью легирования поликремниевого затвора МОПТ.

Для корректного расчета необходимо определить координату *p*-*n*-перехода для правильного учета знаков

заряда в полупроводниках каждой из обкладок конденсатора. Для этого рассчитывается значение $V_g^{n(p)}$ при концентрации, равной собственной для полупроводника при температуре измерения, т.е. в случае полной компенсации примеси; и на графике $V_{fb}(\mathbf{x})$ фиксируется соответствующая координата (см. рис. 3).

В целом, алгоритм нахождения профиля эффективной концентрации примеси выглядит следующим образом:

1. По высокочастотной вольтфарадной характеристике для МОП-конденсатора, имеющему структуру, аналогичную исследуемому МОПТ, и расположенному на пластине вблизи от него, определяется величина напряжения плоских зон V_{fb} , которая разделяется на компоненты, обусловленные влиянием захваченного (фиксированного) заряда в диэлектрике Q_{eff}^{sub} и разностью работ выхода Φ_{ms} .

2. Из зависимости $I_{cp}(V_b)$ извлекается профиль эквивалентного напряжения плоских зон $V_{fb}(\mathbf{x})$.

3. Определяется координата *p*-*n*-перехода.

4. По формулам (3,4,6,7) для каждой точки зависимости $V_{fb}(x)$ с требуемым шагом по координате (обычно целесообразно выбирать в диапазоне от 1 нм до 5 нм) численным методом рассчитывается концентрация примеси, соответствующая данному напряжению, при этом учитывается предполагаемый тип проводимости в соответствии с текущей координатой относительно *p*–*n*-перехода.

Процедура получения профиля эффективной концентрации примеси была автоматизирована на языке MATLAB, что позволяет в интерактивном режиме оперативно провести расчет профиля и проанализировать влияние входных параметров на результат вычисления. Расчёт профиля занимает не более 10 сек. на компьютере среднего уровня.

Анализ результатов

Рассчитанный из зависимости на рис. 2 профиль эффективной концентрации примеси приведен на рис. 4. На графике для сравнения также приведены профили для МОПТ с другой длиной канала и технологическими отличиями. Следует отметить особенности полученного некоторые профиля. Поскольку профиль получается на основе той части кривой зарядовой накачки, где преобладает ток со всей площади канала, а также используется метод при постоянной амплитуде импульсов, то правая часть зависимости на рис. 3.39 не имеет явно выраженной «полки», соответствующей концентрации примеси в канале МОПТ. В этом случае целесообразно по достижении определенного уровня концентрации отбрасывать оставшуюся часть зависимости. Данный уровень концентрации можно условно определить с помощью метода определения эффективной концентрации в канале MOIT. примеси использующего зависимость порогового напряжения от обратного смещения на подложке.

В ходе разработки метода было проанализировано влияние следующих входных параметров, знание

которых требуется с той или иной точностью при расчете профиля:

а) толщина подзатворного оксида *t*_{ox};

б) геометрические размеры затвора транзистора – длина L_G и ширина W_G ;

в) значение эффективного сдвига подпороговой ВАХ V_{ot} за счет влияния встроенного заряда в диэлектрике;

г) параметры сигнала на затворе – частота f, амплитуда V_A , фронты нарастания t_r и спада t_f импульса.



Рис. 4. Зависимости эффективной концентрации примеси вдоль канала МОПТ, имеющих различную структуру и длину канала.

Неточность задания толщины подзатворного окисла достаточно слабо влияет на качественную картину получаемого профиля эффективной концентрации примеси, в частности, на вычисленную координату *p*-*n*-перехода, поскольку напряженность электрического поля вблизи этой координаты невелика при достижении порогового значения концентрации H3.

Влияние ширины W_G МОПТ на профиль эффективной концентрации примеси пренебрежимо мало, т.к. оно осуществляется только через пересчет средней плотности ПС. Длина L_G пропорционально деформирует координатную зависимость в соответствии с формулой (2). Т.к. метод основывается на охвате всей площади под затвором МОПТ, то для корректного определения координатной зависимости необходимо определять реальную длину затвора транзистора с помощью прямых методов измерения (например, с помощью измерительного РЭМ) на том же объекте.

Точность задания амплитуды импульсов одинаково влияет в первую очередь на рассчитанную величину концентрации в каждой точке по координате, зависимость напряжения ОТ расстояния т.к. претерпевает параллельный сдвиг по вертикальной оси. Это, в свою очередь, также отражается и на вычислении координаты *р-п*-перехода, поскольку положение по координатной оси одного и того же напряжения на затворе, характеризующего положение *p*-*n*-перехода, будет различаться в наибольшей степени. В целом, чем больше значение рассчитанной эффективной концентрации примеси, тем меньше сказывается вариация амплитуды на вид зависимости.

Несколько хуже дело обстоит с величиной V_{ot}, поскольку при расчете эта переменная фигурирует

сразу в двух местах: в качестве постоянной добавки в формуле для напряжения (6) и после пересчета на Q_{aff}^{sub}

в уравнении электронейтральности (3). Так, изменение V_{ot} всего на 0.1 В приводит к значительному изменению как координаты *p*-*n*-перехода на величину порядка 10 нм, так и абсолютного значения эффективной концентрации примеси более чем в 2 раза.

Влияние уровня легирования поликремниевого затвора и плотности ПС на нижней границе относительно невелико. Олнако пятикратное уменьшение **у**величение D_{it} И двукратное концентрации примеси в затворе до 10¹⁹ см⁻³ приводят практически к аналогичным изменениям: координата *p*-*n*-перехода изменяется в пределах 3 нм, а уровень концентрации в среднем на 50%. Из этого следует, что в обоих случаях необходимость использования более сложной модели является обоснованной, поскольку время расчета одного профиля в среднем составляет 10 секунд и не является лимитирующим фактором, а точность экстракции профиля при этом возрастает.

Полученные закономерности остаются в силе и для случая профилирования по методам, аналогичным [4], в которых применяется специализированный генератор импульсов, а в основе используется профиль напряжения $V_t(x)$. При этом алгоритм расчета остается аналогичным, меняются местами лишь типы H3 и соответствующие им потенциалы, формулы для которых приведены выше.

Кроме того было установлено, что получаемые концентрационные профили хорошо коррелируют с эффективной длиной канала, определенной по методу [7], наиболее применимому к МОПТ с *LDD*-структурой. Это обстоятельство подтверждает работоспособность методики и её незаменимость при проведении контрольно-диагностических операций при отработке технологических процессов сток-истоковой инженерии и анализе брака кристального производства.

Заключение

Резюмируя сказанное нужно подчеркнуть, что для получения наиболее точных и информативных результатов при применении предлагаемого метода следует обеспечить следующие ключевые требования:

1. Минимизировать воздействие антенных эффектов на анализируемый МОПТ.

2. Произвести отжиг в формер-газе для нейтрализации встроенного заряда и создания более однородной плотности ПС вдоль канала МОПТ.

3. Определить фактическую длину канала (по затвору) анализируемого МОПТ с помощью измерительного РЭМ.

4. С максимальной точностью определить величину приведенного к нижней границе встроенного заряда в подзатворном диэлектрике.

5. Использовать высокочувствительное оборудование, позволяющее регистрировать относительное изменение I_{cp} от 10^{-13} А при неизменности метрологических характеристик в процессе измерения одной зависимости $I_{cp}(V_b)$.

Приведенные результаты подтверждают применимость данного метода для решения задач стокистоковой инженерии для разработки и совершенствования технологических процессов в ОМТ НИИСИ РАН.

Литература

1. G. Groeseneken, H.E. Maes, N. Beltrán, R.F. de Keersmaecker. A reliable approach to charge-pumping measurements in MOS transistors // IEEE Trans. ED. 1984. Vol. 31. No 1. P. 42–53.

2. R.G.-H. Lee, J.-S. Su, S.S. Chung. A New Method for Characterizing the Spatial Distributions of Interface States and Oxide-Trapped Charges in LDD n-MOSFET's // IEEE Trans. ED. 1996. Vol. 43. No 1. P. 81—89.

3. S. Mahapatra, C.D. Parikh, J. Vasi, V. Ramgopal Rao, C.R. Viswanathan. A direct charge pumping technique for spatial profiling of hot-carrier induced interface and oxide traps in MOSFET's // Solid-State Electronics. 1999. Vol. 43. P. 915—922.

4. C. Chen, T.-P. Ma. Direct Lateral Profiling of Hot-Carrier-Induced Oxide Charge and Interface Traps in Thin Gate MOSFET's // IEEE Trans. ED. 1998. Vol. 45. No 2. P. 512—520.

5. М.В. Орешков, В.С. Солдатов. Измерительный комплекс электрофизических методов для углубленного анализа субмикронной КМОП технологии // Вестник МЭИ. 2012. № 1. с. 102-107.

6. M. Tsuchiaki, H. Hara, T. Morimoto, H. A Iwai. New charge pumping method for determining the spatial distribution of hot-carrier-induced fixed charge on p-MOSFET's // IEEE Trans. ED. 1993. Vol. 40. No 10. P. 1768—1778.

7. J.Y.-C. Sun, M.R. Wordeman, S.E. Laux. On the accuracy of channel length characterization of LDD MOSFET's // IEEE Trans. ED. 1986. Vol. 33 P. 1556.

The evaluation of concentration profile along the submicron MOSFET channel within the scope of source-drain engineering

M.V. Oreshkov, V.Yu. Troitskiy

Abstract. Method for quick evaluation of concentration profile along the MOSFET channel is suggested. This method can be used within the scope of source-drain engineering and for failure analysis as nondestructive investigation. The main advantages of developed method are minimal degradation of transistor's undergate system due to measurement process and simplicity of implementation.

Реализация основных принципов контроля стабильности базовых технологических процессов в условиях мелкосерийного производства

Т.М. Антонова, Е.С. Темников

В мелкосерийном и прерывистом производстве, в силу недостаточности статистических данных, статистический контроль процесса изготовления каждого изделия в отдельности становится невозможным. В то же время, именно в мелкосерийном и прерывистом производстве проблема подтверждения стабильности производственного процесса стоит особенно остро. Оптимальным выходом из положения является переход от статистического контроля процессов изготовления отдельных изделий к статическому контролю базовых процессов.

1. Введение

При внедрении методов статистического управления в мелкосерийном и прерывистом производстве следует учитывать его особенности, резко ограничивающие возможности традиционных подходов:

- недостаток статистических данных: для накопления количества партий, необходимого для построения контрольных границ, может потребоваться несколько лет;

- продолжительное время отклика: если одна партия изделия производится раз в несколько месяцев, произошедший выход из состояния статистической управляемости может быть зафиксирован с запозданием, исключающим возможность реализации корректирующих мероприятий;

- невозможность контроля большого числа изделий: в мелкосерийном производстве многие изделия производят лишь в количестве нескольких партий, и построение контрольной карты для них невозможно.

Одним из важнейших условий эффективного применения статистического контроля процесса, как использования контрольных карт, так и оценки точности и настроенности процесса является наличие достаточного количества статистических данных. Для расчета статистически значимых контрольных границ на контрольной карте требуется, по меньшей мере, десять партий изделий. Столько же требуется и для расчета статистически значимых значений коэффициентов точности и настроенности процессов.

В мелкосерийном и прерывистом производстве многие изделия изготавливаются в количестве трехпяти партий за все время их изготовления. Если же для отдельных, наиболее востребованных изделий объем выпуска и превышает десять партий, то эти партии бывают выпущены в течение огромного количества времени - в течение года или нескольких лет. И в течение всего этого времени, пока не будет изготовлено десять партий, статистический контроль остается невозможным для этих изделий.

В то же время, недостаточный объем выпуска каждого изделия в отдельности делает особенно важной проверку стабильности производственного процесса. Ведь если нельзя напрямую убедиться в стабильности свойств выпускаемой продукции, то необходимо быть уверенными, что, по крайней мере, эта продукция изготавливается в стабильном производственном процессе.

2. Метод статистического контроля в мелкосерийном и прерывистом производстве, используемый в НИИСИ РАН

Основные подходы к статистическому контролю процесса в мелкосерийном и прерывистом производстве изложены в действующей редакции Приложении Л к ОСТ 11 14.1011, в целом вполне корректно, но нуждаются в конкретизации и дополнении. Правда, изложенные также в Приложении Л методы обработки результатов изготовления партий различного объема и расчета толерантных границ имеют очень отдаленное отношение к этому вопросу.

Для производства продукции на базе технологического комплекса в НИИСИ РАН разработана и несколько лет успешно используется на практике методика статистического контроля в мелкосерийном и прерывистом производстве. На основании опыта работы были подготовлены материалы, которые, в числе прочих, включены в предложения по корректировке ОСТ 11 14.1011 в рамках СЧ ОКР «Кузбассовец».

Основным приемом, используемым при статистическом контроле в мелкосерийном и прерывистом производстве, является объединение на одной контрольной карте результатов контроля нескольких изделий, условия изготовления, с использованием одних и тех же средств технологического оснащения, и свойства которых считают идентичными.

Ключевым принципом признания идентичности изделий, с точки зрения статистического управления процессами, является принцип конструктивнотехнологического подобия. Результаты операционного контроля и испытаний изделий, объединенных признаками конструктивно-технологического подобия, рассматривают на контрольной карте, как правило, совместно, как результаты контроля процесса изготовления одного и того же изделия .

При статистическом управлении процессами обработки пластин в качестве основного признака конструктивно-технологического подобия используют признак изготовления изделий в одном базовом технологическом процессе.

В случаях, при анализе характеристик, присущих конкретному типу микросхемы, например, определяемых топологией кристалла, используют дополнительный признак конструктивно-технологического подобия – соответствие проектов микросхем одним и тем же правилам проектирования.

В некоторых случаях, имеется возможность объединять на контрольной карте результаты операционного контроля изделий, изготавливаемых в различных базовых технологических процессах (если условия проведения контролируемой операции и контролируемые свойства изделий идентичны, то есть, результаты контроля статистически однородны).

При статистическом управлении процессами сборки микросхем в качестве признаков конструктивнотехнологического подобия, в зависимости от особенностей контролируемой характеристики используют (в различных сочетаниях):

- степень «одинаковости» конструкции корпуса – использование корпуса одного и того же типа, подтипа, типоразмера и (или) исполнения по ГОСТ Р 54844-2011 ;

- изготовление корпуса одним и тем же изготовителем, с использованием одних и тех же конструктивных материалов и (или) с применением одних и тех же конструктивно-технологических решений,

- применение в технологическом процессе одних и тех же конструктивно-технологических решений, режимов и (или) материалов (в том числе, материалов и методов монтажа межсоединений и кристалла и (или) одних и тех же методов герметизации).

При статистическом управлении отбраковочными испытаниями микросхем, в некоторых случаях учитывают дополнительный признак конструктивнотехнологического подобия – одинаковую степень интеграции.

Примечание – Например, объединение на одной контрольной карте результатов испытания прочности монтажа кристалла на сдвиг возможно для изделий в корпусах одного типа, изготовленных одним изготовителем из одних и тех же материалов, с одним видом используемого клея и одним размером кристалла.

В любом случае, объединение на одной контрольной карте результатов операционного контроля (и отбраковочных испытаний) различных изделий возможно при условии статистической однородности результатов контроля для всех изделий, если:

- различия в среднем значении характеристики для различных изделий статистически незначимы;

 различия в изменчивости характеристики по пластине для различных изделий статистически незначимы;

- различия в изменчивости характеристики по партии для различных изделий статистически незначимы; - различия в изменчивости характеристики от партии к партии для различных изделий статистически незначимы.

При использовании одной контрольной карты для контроля одной характеристики нескольких изделий, объединенных либо по принципу их изготовления в одном базовом технологическом процессе, либо по другим критериям конструктивно-технологического подобия, понимают, что полученный сигнал о выходе данной характеристики из состояния статистического контроля следует в равной степени относить ко всем изделиям, контролируемым на данной контрольной карте. И до выяснения причин выхода из-под контроля и выполнения корректирующих мероприятий изготовление всей номенклатуры изделий, контролируемых на данной контрольной карте, в необходимых случаях, приостанавливают. Корректирующие мероприятия, если они требуются, также проводят в отношении всех изделий, контролируемых совместно на этой контрольной карте.

3. Особенности оценки точности и настроенности процесса в мелкосерийном и прерывистом производстве

Подход к оценке точности и настроенности процесса в мелкосерийном и прерывистом производстве вытекает из метода контроля нескольких, объединенных по принципу конструктивно-технологического подобия изделий.

Оценка точности и настроенности процесса проводится путем сопоставления границ естественной изменчивости процесса и нормативных границ для отдельных характеристик этого процесса.

Контрольные границы, построенные для одной из характеристик группы таких изделий, будут отражать границы естественной изменчивости этой характеристики для каждого изделий, контролируемых совместно. Если объединение этих изделий на одной контрольной карте проведено правомерно, то границы естественной изменчивости будут идентичны для всех изделий.

Что же касается нормативных границ, то они, строго говоря, могут быть различны для одной характеристики разных изделий, контролируемых совместно (как правило, в случае, если границы заданы не исходя из особенностей процесса, а исходя из требований к конечному продукту, например, функциональному назначению изделий).

Таким образом, в противоположность результатам контроля стабильности процесса, распространяемым одновременно на все изделия, контролируемые совместно, результаты оценки точности и настроенности процесса следует распространять на каждое изделие в отдельности. При этом, возможны (и встречаются в реальном производстве!) такие ситуации, когда для одних изделий результаты оценки точности и настроенности процесса будут удовлетворительными, а для других нет. Это означает, что одни типы контролируемых изделий могут изготавливаться в данном технологическом процессе, имея хорошие конструктивнотехнологические запасы, а для других этот процесс не подходит (иными словами, технические решения, принятие при проектировании этих изделий, в той или иной степени не соответствуют возможностям процесса, используемого для их изготовления).

4. Пример применения статистического контроля в мелкосерийном и прерывистом производстве

На рисунке 1 представлена контрольная карта для толщины окисла щелевой изоляции после операции химико-механической планаризации в базовом технологическом процессе изготовления КМОП микросхем с минимальным топологическим размером 0,5 мкм. На ней представлена информация о партиях пяти различных изделий, проходящих обработку в этом базовом ному изделию – TVR. Изделие TVR отличается от остальных изделий, представленных на данной контрольной карте, иной плотностью заполнения, из-за чего, несмотря на одинаковые режимы обработки, толщина окисла щелевой изоляции оказывается другой. Несмотря на то, что толщина окисла и для изделий TVR, и для остальных изделий на данной карте удовлетворяет установленным нормативным ограничениям, разница между ними является статистически значимой. Следовательно, на контрольной карте для толщины окисла щелевой изоляции нельзя объединять результаты контроля изделия TVR с результатами контроля остальных изделий, проходящих обработку в этом базовом технологическом процессе.

На рисунке 2 представлена контрольная карта для толщины окисла щелевой изоляции после той же технологической операции того же базового процесса, для всей номенклатуры изделий, проходящих обработку в этом процессе, за исключением изделия TVR. На этой контрольной карте видно, что толщина окисла щелевой изоляции находится в состоянии статистического



Рис. 1. Контрольная карта для толщины окисла щелевой изоляции после планаризации в базовом технологическом процессе КМОП с топологическими нормами 0,5 мкм, для всех изделий, изготавливаемых в базовом процессе. Контрольные границы отображены сплошной линией, нормативные границы – пунктиром.

процессе.

Контрольная карта демонстрирует, что значения толщины в партиях 203, 223, 239, 240, 244, 253, 254, 259, 277, 293, 296 резко отличаются от значений толщины в других партиях изделий. Партии 203, 223, 239, 240, 244, 253, 254, 259, 277, 293, 296 принадлежат одконтроля. За пределами контрольных границ находятся значения толщины только в партии 261. Анализ причины выхода за пределы границ показал, что измерения характеристик этой партии проводились с использованием другого оборудования.



Рис. 2. Контрольная карта для толщины окисла щелевой изоляции после планаризации в базовом технологическом процессе КМОП с топологическими нормами 0,5 мкм, для всех изделий, изготавливаемых в базовом процессе, кроме изделия TVR. Контрольные границы отображены сплошной линией, нормативные границы – пунктиром.

На рисунке 3 представлена контрольная карта для толщины окисла щелевой изоляции после той же операции того же базового процесса, составленная только для изделия TVR. На этой контрольной карте видно, что толщина окисла щелевой изоляции для изделий TVR также находится в состоянии статистического



Рис. 3. Контрольная карта для толщины окисла щелевой изоляции после планаризации в базовом технологическом процессе КМОП с топологическими нормами 0,5 мкм, только для изделий TVR. Контрольные границы отображены сплошной линией, нормативные границы – пунктиром.

контроля.

Таким образом, метод статистического контроля технологических процессов с помощью контрольных

карт, даже в условиях мелкосерийного и прерывистого производства, вполне эффективен, при условии корректного учета особенностей его применения.

The implementation of the basic principles for control of stability of basic technological processes in small-scale production

T.M. Antonova, E.S. Temnikov

Abstract: In small-scale and discontinuous production, in virtue of insufficient statistical data, statistical control for process of manufacture of each individual product becomes impossible. At the same time, in the small-scale and discontinuous manufacture requirement of stability of the production process is essentially actually. The optimal solution for this situation is transaction from statistical process of control for manufacture of individual products to static control of basic processes.

Применение контрольных карт иерархических процессов в технологических процессах изготовления СБИС с высоким числом уровней изменчивости

Т.М. Антонова

В статье рассмотрено применение контрольных карт иерархических процессов в микроэлектронном производстве изготовления СБИС с высоким числом уровней изменчивости.

1. Введение

Стандартные контрольные карты Шухарта, повсеместно применяемые для статистического контроля процесса, не всегда адекватно отражают особенности этого процесса. В частности, они неприменимы для иерархических процессов, т.е. процессов с несколькими уровнями изменчивости. В микроэлектронном производстве, особенно в кристальном производстве, большинство процессов является как раз иерархическими, что требует особого подхода к построению контрольных карт. Применение же стандартных контрольных карт Шухарта приведет к заведомо неадекватным результатам.

1. Ограничения использования стандартных контрольных карт Шухарта

Для статистического контроля процесса, как правило, используются контрольные карты Шухарта средних значений и среднеквадратических отклонений (или размахов). Эти типы контрольных карт используются столь широко, что очень часто перед их использованием даже не проводится анализ адекватности этих типов контрольных карт для конкретного процесса. А ведь очень часто данные типы контрольных карт неприменимы, и их использование приводит к совершенно неадекватному результату.

Метод работы контрольных карт Шухарта средних значений и среднеквадратических отклонений (размахов) опирается на то, что разброс значений от выборки к выборке вызван теми же причинами, что и их разброс внутри выборки, и рассчитывается исходя из значения последнего. В случае, когда появляются особые причины изменчивости, отсутствующие внутри выборки и проявляющиеся от выборки к выборке, связь разброса по выборке и от выборки к выборке нарушается, что приводит к выходу за контрольные границы и трактуется как выход из-под контроля.

Однако, далеко не всегда, даже в нормальном и контролируемом состоянии процесса, разброс значений от выборки к выборке вызван теми же причинами, что и их разброс внутри выборки. Очень часто особенности процесса таковы, что изменчивость значений внутри выборки определяется одними факторами, а изменчивость от выборки к выборке – совсем другими. Такие процессы можно назвать иерархическими процессами.

Для иерархических процессов контрольные карты Шухарта средних значений и среднеквадратических отклонений (размахов) абсолютно неприменимы, их использование приведет к совершенно неадекватным результатам.

2. Иерархический характер процессов в микроэлектронике

В микроэлектронике практически все процессы являются именно иерархическими.

При партионной обработке кремниевых пластин изменчивость любых характеристик пластин (например, поверхностных сопротивлений слоев металлизации, толщины диэлектрических слоев, линейных размеров элементов, электрофизических характеристик и т.д.) внутри партии обуславливается следующими факторами:

- различием исходных свойств кремниевых пластин;

- особенностями работы оборудования (при одновременной обработке – распределением условий обработки внутри камеры, при последовательной – способностью оборудования к воспроизведению условий обработки, наличием или отсустствием т.н. «эффекта первой пластины» и т.п.).

Изменчивость тех же характеристик от партии к партии пластин определяется уже значительно большим количеством факторов:

- различием исходных свойств кремниевых пластин;

- воспроизводимостью оборудования;

- изменением свойств используемых материалов по мере их старения или смены на новую партию;

- изменением состояния оборудования после плановых и внеплановых технических обслуживаний, замены деталей и т.п.;

- изменением состояния окружающей среды;

- человеческим фактором.

Помимо изменчивости характеристик кремниевых пластин внутри партии и от партии к партии, еще имеется изменчивость характеристик по пластине. Как правило, в микроэлектронике каждому процессу обработки пластин соответствует свой типичный профиль распределения характеристик по пластине (например, «выпуклый» профиль, когда значение характеристики в центре пластины выше, чем на периферии, либо значение характеристики может быть выше с одного края пластины и ниже – с другого края, и т.п.).

Изменчивость характеристик по пластине определяется, в первую очередь:

 - физическими особенностями процессов обработки пластин;

- привносимой дефектностью.

Понятно, в данном случае изменчивость характеристик кремниевых пластин по пластине, от пластины к пластине внутри партии и от партии к партии являются принципиально различными и не могут зависеть друг от друга. Как правило, изменчивость от партии к партии превышает изменчивость внутри партии, а изменчивость внутри партии значительно превышает изменчивость по пластине, хотя это и не обязательно.

Поэтому, в микроэлектронике контрольные карты Шухарта средних значений и среднеквадратических отклонений (размахов), как правило, неприменимы. Результат некорректного применения контрольных карт Шухарта средних значений и среднеквадратических отклонений для иерархического процесса – толщины диэлектрического слоя на поверхности кремниевых платин – представлен на рисунке 1 (На этом рисунке и на всех последующих – контрольные границы отображены сплошной линией, нормативные границы



Рисунок 1. Контрольная карта Шухарта средних значений и среднеквадратических отклонений для иерархичекого процесса.

- пунктиром).

При использовании контрольной карты Шухарта средних значений контрольные границы рассчитываются на основе незначительного разброса по партии, а разница между точками на контрольной карте (то есть между партиями) существенно превышает этот разброс, то есть большинство точек находится далеко за пределами контрольных границ (см. рисунок 1). Тем не менее, на контрольной карте видно, что процесс в достаточной степени однороден, и на самом деле находится в состоянии статистического контроля, а контрольные границы построены явно неадекватно.

3. Контрольные карты для иерархических процессов

Иерархические процессы распространены очень широко. В каждом сложном производстве иерархических процессов даже больше, чем процессов с одним уровнем изменчивости. Для иерархических процессов обычные типы контрольных карт неприменимы – для специалистов по статистическому контролю это давно известный факт. Более того, были сформулированы общие подходы к тому, как адаптировать контрольные карты для иерархических процессов [1]. Для иерархических процессов в двумя уровнями изменчивости была разработана трехсторонняя контрольная карта [2], [3]. Однако, конкретных рекомендаций по построению контрольных карт для иерархических процессов с тремя и более уровнями изменчивости до сих пор сформулировано не было.

Если изменчивость значений внутри партии и от партии к партии отличны и независимы друг от друга, они не могут сравниваться друг с другом. Это значит, что контрольные границы на контрольной карте для средних по партии значений не могут рассчитываться на основе разброса (размаха или среднеквадратического отклонения) внутри партии. Для иерархического процесса контрольные границы на контрольной карте для средних по партии значений должны рассчитываться на основе разброса от партии к партии.

Сделать это позволяет контрольная карта Шухарта индивидуальных значений и скользящих размахов (XmR). Если для средних по партии значений построить контрольную карту Шухарта индивидуальных значений и скользящих размахов (XmR), трактуя средние по партии значения как индивидуальные, тогда контрольные границы будут рассчитаны на основе скользящего размаха, т.е. размаха между соседними партиями. Таким образом, подобное будет сравниваться с подобным: разброс между партиями с разбросом между партиями.

Результат применения контрольных карт Шухарта индивидуальных значений и скользящих размахов для иерархического процесса представлен на рисунке 2.



Рисунок 2. Контрольная карта Шухарта индивидуальных значений и скользящих размахов для иерархического процесса.

На этом рисунке представлен тот же самый процесс, что и на рисунке 1 – толщина диэлектрического слоя на поверхности кремниевых платин, но контрольная карта выбрана другая. В результате видно, что контрольные границы рассчитаны вполне корректно, и процесс находится в состоянии статистического контроля, как мы и рассчитывали.

4. Трехсторонние контрольные карты

Контрольная карта Шухарта индивидуальных значений и скользящих размахов (XmR), где индивидуальными значениями считают средние по партии значения, подходит для контроля иерархического процесса, но она обладает существенным недостатком: на этой карте никак не отображается изменчивость внутри партии. А эта изменчивость также очень важна и тоже должна контролироваться.

Чтобы контролировать не только средние по партии значения, но и разброс по партии для иерархического процесса, было предложено использовать две пары карт XmR:

- на первой паре индивидуальными значениями считают средние по партиям значения,

- на второй паре индивидуальными значениями считают среднеквадратические отклонения по партиям.

Таким образом, первая пара контрольных карт контролирует средний уровень и его изменение от партии к партии, а вторая пара — разброс по партии и его изменение от партии к партии.

Однако, контролировать для каждой характеристики по две пары контрольных карт несколько затруднительно. В то же время, четвертая карта – карта скользящих размахов выборочных среднеквадратических отклонений – дает не так много информации, т.к. при выходе значения скользящего размаха за контрольные границы увеличение расстояния между соседними точками будет заметно и на карте индивидуальных значений для выборочных среднеквадратических отклонений. Поэтому многие авторы рекомендуют исключить из рассмотрения четвертую карту, а карту выборочных среднеквадратических отклонений (трактуемых как индивидуальные значения) поместить на одном рисунке с парой карт для средних выборочных значений.

В итоге будет получена тройка контрольных карт:

- средние по партиям значения как индивидуальные значения;

 скользящие размахи средних по партиям значений;

- среднеквадратические отклонения по партиям как индивидуальные значения.

Устоявшейся терминологии для обозначения такой тройной контрольной карты в настоящий момент нет. В различных зарубежных источниках она упоминается как «3D chart» (Lyle-Kearsley Systems, Inc.) или «three-way chart» (D.J. Wheeler). В русском языке термин этот термин введен при переводе книги D.J. Wheeler [2] как «трехсторонняя контрольная карта». Дабы избежать путаницы в терминах, предлагаем именно этот термин и использовать.

Результат применения трехсторонней контрольной карты для иерархического процесса представлен на рисунке 3.



На этом рисунке представлен тот же самый процесс, что и на рисунках 1 и 2 – толщина диэлектрического слоя на поверхности кремниевых платин. Первые две контрольные карты повторяют рисунок 2, но добавлена также третья карта – для среднеквадратических отклонений по партиям.

5. Иерархические процессы с тремя уровнями изменчивости

Трехсторонние карты позволяют контролировать процессы лишь с двумя уровнями изменчивости.

Для контроля процессов с тремя уровнями изменчивости можно использовать две двухуровневых карты. На первой трехсторонней карте выборкой будет являться совокупность значений в одной партии, и она будет состоять из следующих контрольных карт:

- карта индивидуальных значений для средних по партии значений;

- карта скользящих размахов для средних по партии значений;

 карта индивидуальных значений для среднеквадратических отклонений по партии средних по пластине значений.

На второй трехсторонней карте выборкой будет являться совокупность значений на одной пластина, и она будет состоять из следующих контрольных карт:

- карта индивидуальных значений для средних по пластине значений;

- карта скользящих размахов для средних по пластине значений;

- карта индивидуальных значений для средне-квадратических отклонений по пластине.

Однако при этом со второй трехсторонней картой могут возникнуть сложности.

Как мы уже отмечали ранее, изменчивость процесса внутри партии и от партии к партии различны и независимы между собой. Обычно, разброс от партии к партии превышает разброс по партии. Поэтому, на контрольной карте, где выборкой является совокупность значений на одной пластине, точки будут группироваться по партиям: точки, соответствующие пластинам из одной партии, будут лежать относительно близко друг к другу, а группы точек, соответствующих пластинам из различных партий, будут несколько отстоять друг от друга (см. рисунок 4).



Рисунок 5. Трехсторонняя контрольная карта для толщины окисла щелевой изоляции. Каждая точка соответствует значению характеристики в одной партии.

На рисунке 4 и 5 изображены контрольные карты для толщины окисла щелевой изоляции после процесса планаризации. На рисунке 4 представлена трехсторонняя контрольная карта от пластины к пластине (т.е. каждая точка соответствует значению характеристики на одной пластине), на рисунке 5 – трехсторонняя контрольная карта от партии к партии (т.е. каждая точка соответствует значению характеристики в одной партии).

Можно ли считать такой процесс находящимся в состоянии статистического контроля? С одной стороны, видно, что множество точек находится за пределами контрольных границ. Кроме этого, мы видим повторяющиеся паттерны. Все эти признаки, вообще говоря, свидетельствуют о том, что процесс не находится в состоянии статистического контроля.

С другой стороны, такая особенность структурирования данных является «встроенной» особенностью процесса и не может быть устранена. При этом, процесс остается предсказуемым, и вполне соответствует предъявляемым к нем требованиям. Это значит, что представленный процесс все же следует считать контролируемым. Это подтверждается трехсторонней контрольной картой для этого же процесса, где выборкой будет является совокупность значений в одной партии (рисунок 5).

В таком случае, контрольная карта на рисунке 4 свидетельствует о том, что для иерархических процессов с тремя и более уровнями изменчивости стандартные контрольные карты не подходят. Неоднородность данных на контрольной карте для иерархического процесса приводит к неправильной оценке границ естественной изменчивости процесса.

6. Особенности расчета контрольных границ для иерархических процессов с тремя уровнями изменчивости на стандартной контрольной карте Шухарта

Рассмотрим для примера процесс с тремя уровнями изменчивости: продукция изготавливается партиями, каждая партия, в свою очередь, состоит из трех пластин, на пластинах в нескольких точках измеряется контролируемая характеристика.

Для такого процесса, на контрольной карте индивидуальных значений, где выборкой является совокупность точек на одной пластине, точки будут располагаться тройками (т.к. в каждой партии три пластины). На карте скользящих размахов в этом случае расположение точек будет иметь следующую конфигурацию: два малых размаха (внутри партии), один большой (новая партия), два малых, один большой, и т.д. Следовательно, средний скользящий размах будет складываться:

- из разброса пластин внутри партии (меньшего) с весовым коэффициентом 2;

- из разброса пластин от партии к партии (большего) с весовым коэффициентом 1.

То есть, средний скользящий размах окажется больше среднего размаха по партии, но меньше сред-

него размаха от партии к партии, причем больший вклад в его значение внесет первый, то есть меньший размах, а именно:

$$\overline{mR}_{o \delta u \mu} \approx \frac{2}{3} \overline{mR}_{en} + \frac{1}{3} \overline{mR}_{Mn}, \quad (1)$$

 \overline{mR}_{obu} – средний скользящий размах от пластины к пластине;

*mR*_{вп} – средний скользящий размах от пластины к пластине внутри партии;

*mR*_{мп} – средний скользящий размах от партии к партии (между партиями).

$$\overline{mR} \approx \frac{\sigma}{1.128}$$

Учитывая, что 1.128 , для значения контрольных границ на контрольной карте Шухарта индивидуальных значений мы получим:

$$BK\Gamma \approx \overline{X} + \frac{2\sigma_{_{\thetaH}} + \sigma_{_{MR}}}{\sqrt{3}}$$
$$HK\Gamma \approx \overline{X} - \frac{2\sigma_{_{\thetaH}} + \sigma_{_{MR}}}{\sqrt{3}}, \quad (2)$$

где:

 $\sigma_{_{en}}$ – среднеквадратическое отклонение от пластины к пластине внутри партии;

σ_{мп} – среднеквадратическое отклонение от партии к партии (между партиями).

 X_{-} среднее значение наблюдаемой характеристики.

Такие значения контрольных границ будут получены, если для иерархического процесса использовать стандартные контрольные карты Шухарта индивидуальных значений и скользящих размахов, либо трехстороннюю карту, которая, в свою очередь, состоит из контрольных карт Шухарта индивидуальных значений и скользящих размахов.

Как видно из рисунка 4, и как это подтверждает приведенный расчет, полученные таким образом значения контрольных границ оказываются значительно заужены и не отображают естественной изменчивости процесса.

7. Ширина границ естественной изменчивости для иерархических процессов с тремя уровнями изменчивости

Чтобы оценить корректность расчета контрольных границ, необходимо определить границы естественной изменчивости иерархического процесса.

Разброс средних по пластине значений, принадлежащих различным партиям, включает в себя как разброс средних по пластине значений внутри партии, так и разброс от партии к партии. Метод расчета общего разброса от пластины к пластине будет зависеть от того, насколько связаны между собой разброс внутри партии и разброс от партии к партии. Некоторая связь между этими величинами обычно имеется, но она, как правило, не очень значительна. Если считать разброс по партии и от партии к партии полностью независимыми, то значение общего среднеквадратического отклонения составит:

$$\sigma_{o \delta u \mu} = \sqrt{\sigma_{en}^2 + \sigma_{Mn}^2}, \qquad (3)$$

Полученное значение $\sigma_{o \delta u}$, независимо от зна-

чений σ_{sn} и σ_{Mn} , будет больше любого из них, как среднеквадратического отклонения по партии, так и среднеквадратического отклонения от партии к партии.

Между тем, оценка значения $\sigma_{o \delta u u}$, некорректно рассчитанная из среднего скользящего размаха от пластины к пластине (1), была больше минимального из

значений $\sigma_{_{6n}}$ и $\sigma_{_{Mn}}$, но меньше максимального.

Границы естественной изменчивости будут находиться на уровне:

$$B\Gamma E H = \overline{X} + \frac{\sqrt{\sigma_{en}^{2} + \sigma_{Mn}^{2}}}{\sqrt{3}}$$
$$H\Gamma E H = \overline{X} - \frac{\sqrt{\sigma_{en}^{2} + \sigma_{Mn}^{2}}}{\sqrt{3}}$$
(4)

что как минимум в 1.4 раза шире, чем контрольные границы, рассчитанные для контрольной карты Шухарта индивидуальных значений и скользящих размахов по формуле (2). Причем, разница между контрольными границами и границами естественной изменчивости будет тем больше, чем больше разница между разбросом по партии и разбросом от партии к партии.

Именно поэтому на рисунке 4, в то время как процесс находится под контролем, на контрольной карте Шухарта индивидуальных значений и скользящих размахов множество точек оказалось за пределами контрольных границ.

8. Корректный метод расчета контрольных границ для иерархических процессов

Корректные значения контрольных границ следует рассчитывать таким образом, чтобы они были близки к значениям границ естественной изменчивости.

Следовательно, при расчете контрольных границ для иерархического процесса следует учесть все уровни изменчивости процесса, представленные на контрольной карте, и оценить их вклад в общую изменчивость процесса. Изменчивость процесса внутри партии оценивается средним среднеквадратическим отклонением по партии:

$$\hat{\sigma}_{\scriptscriptstyle 6n} = \frac{\overline{S_{\scriptscriptstyle 6n}}}{c_4(n)} = \frac{\sum_{j=1}^N S_{\scriptscriptstyle 6n,j}}{N}, \qquad (5)$$

где:

 $S_{\scriptscriptstyle BN}$ – среднее выборочное среднеквадратическое отклонение от пластины к пластине по партии;

S_{вп,j} – выборочное среднеквадратическое отклонение от пластины к пластине по j-той партии;

j = 1...N – номер партии;

N – количество партий, по которым рассчитываются контрольные границы;

n – количество пластин в партии;

с4 – коэффициент, значение которого определяется из таблицы в зависимости от количества пластин в партии.

Данное значение определяет положение средней линии на третьей карте трехсторонней контрольной карты, где выборкой является совокупность значений в одной партии (т.е. на контрольной карте для среднеквадратических отклонений по партии).

Изменчивость процесса от партии к партии оценивается через значение скользящего размаха от партии к партии:

$$\hat{\sigma}_{Mn} = \frac{\overline{mR_{Mn}}}{1.128} = \frac{\sum_{j=2}^{N} \left| \overline{X}_{j} - \overline{X}_{j-1} \right|}{1.128 \cdot (N-1)}, \quad (6)$$

где:

$$mR_{_{Mn}}$$

*иис*_{*мп*} – среднее значение скользящего размаха от партии к партии;

X_j – среднее по партии значение измеряемого параметра в j-той партии.

Данное значение определяет положение средней линии на второй карте (карте скользящих размахов) трехсторонней контрольной карты, где выборкой является совокупность значений в одной партии (т.е. на контрольной карте для среднеквадратических отклонений по партии).

Исходя из этого, мы можем рассчитать положение контрольных границ для всех карт трехсторонней контрольной карты, где выборкой является совокупность значений на одной пластине.

Контрольные границы на карте индивидуальных значений (выборкой является совокупность значений на одной пластине):

$$BK\Gamma = \overline{\overline{X}} + 3 \cdot \sqrt{\left(\frac{\overline{S_{en}}}{c_4(n)}\right)^2 + \left(\frac{\overline{mR_{Mn}}}{1.128}\right)^2},$$
$$HK\Gamma = \overline{\overline{X}} - 3 \cdot \sqrt{\left(\frac{\overline{S_{en}}}{c_4(n)}\right)^2 + \left(\frac{\overline{mR_{Mn}}}{1.128}\right)^2},$$
$$IJ\Pi = \overline{\overline{X}}.$$
(7)

где:

X – среднее значение измеряемого параметра по всем партиям, по которым рассчитываются контрольные границы.

Контрольные границы на карте скользящих размахов (выборкой является совокупность значений на одной пластине):

$$BK\Gamma = 3.686 \cdot \sqrt{\left(\frac{\overline{S_{en}}}{c_4(n)}\right)^2 + \left(\frac{\overline{mR_{Mn}}}{1.128}\right)^2},$$
$$HK\Gamma = 0,$$
$$I \downarrow \Pi = 1.128 \cdot \sqrt{\left(\frac{\overline{S_{en}}}{c_4(n)}\right)^2 + \left(\frac{\overline{mR_{Mn}}}{1.128}\right)^2}.$$
(8)

Контрольные границы на карте среднеквадратических отклонений по пластине: (выборкой является совокупность значений на одной пластине):

$$BK\Gamma = \overline{\overline{s}} + 3 \cdot \sqrt{\left(\frac{\overline{S}s}{c_4(n)}\right)^2 + \left(\frac{\overline{mR_s}}{1.128}\right)^2},$$
$$HK\Gamma = \overline{\overline{s}} - 3 \cdot \sqrt{\left(\frac{\overline{S}s}{c_4(n)}\right)^2 + \left(\frac{\overline{mR_s}}{1.128}\right)^2},$$
$$IJ\Pi = \overline{\overline{s}}.$$
(9)

где:

s – среднеквадратическое отклонение по пластине измеряемого параметра (sij), усредненное по всем пластинам (по i = 1...n) всех партий, по которым определяются контрольные границы (по j = 1...N);

 S_s – среднеквадратическое отклонение по партии среднеквадратических отклонений по пластине измеряемого параметра (SSj), усредненное по партиям, по которым определяются контрольные границы (по j = 1...N);

 mR_s – скользящий размах от партии к партии средних по партии значений среднеквадратических отклонений по пластине, усредненный по партиям, по которым определяются контрольные границы (по j = 2...N).

Значения параметров для расчета контрольных границ на карте среднеквадратических отклонений рассчитываются следующим образом:

$$\overline{s_{j}} = \frac{\sum_{i=1}^{n} s_{ij}}{n}, \quad \overline{s} = \frac{\sum_{j=1}^{N} \overline{s_{j}}}{N}, \quad (10)$$

$$S_{sj} = \sqrt{\frac{\sum_{i=1}^{n} (s_{ij} - S_{j})^{2}}{n-1}}, \quad \overline{S_{s}} = \frac{\sum_{j=1}^{N} S_{sj}}{N}, \quad (11)$$

$$\overline{mR_{s}} = \frac{\sum_{j=2}^{N} |\overline{s_{j}} - \overline{s_{j-1}}|}{N-1}, \quad (12)$$

где:

i = 1...n – номер пластины в партии;

sij – среднеквадратическое отклонение по пластине измеряемого параметра на i-той пластине в j-той партии;

 S_{j} – среднеквадратическое отклонение по пластине измеряемого параметра, усредненное по всем пластинам (по i = 1...n) j-той партии;

s – среднеквадратическое отклонение по пластине измеряемого параметра, усредненное по всем пластинам (по i = 1...n) всех партий, по которым определяются контрольные границы (по j = 1...N);

SSj – среднеквадратическое отклонение по партии среднеквадратических отклонений по пластине измеряемого параметра в j-той партии;

 S_{s} – среднеквадратическое отклонение по партии среднеквадратических отклонений по пластине измеряемого параметра, усредненное по всем партиям (по j = 1...N);





Рисунок 6. Контрольная карта для иерархических процессов для толщины окисла щелевой изоляции.

партии значений среднеквадратических отклонений по пластине.

На рисунке 6 представлена трехсторонняя контрольная карта, где выборкой является совокупность значений на одной пластине, а контрольные границы рассчитаны по формулам (7) – (9). Каждая точка соответствует значению характеристики на одной пластине. Положения точек на контрольной карте на рисунке 6 те же, что и на рисунке 4. Однако, видно, что теперь контрольные границы рассчитаны корректно, и процесс находится в состоянии статистического контроля, что и соответствует действительности.

Выводы

1. Контрольные карты для иерархических процессов не применимы для процессов с мультимодальным распределением контролируемой характеристики, а также для процессов с автокорреляцией или трендом.

2. Необходимо учитывать, что из всех критериев выхода из-под контроля можно использовать лишь основной – выход точки за пределы контрольных границ. Для оценки иерархического процесса неприемлемы такие критерии выхода из-под контроля, как:

- две из трех расположенных подряд точек за пределами границ на уровне 2σ по одну сторону от центральной линии;

- четыре из пяти расположенных подряд точек за пределами границ на уровне 1σ по одну сторону от центральной линии;

- девять точек подряд по одну сторону от центральной линии.

3. Объем данных для расчета контрольных границ должен позволять удовлетворительно оценивать изменчивость на обоих уровнях: по партии и от партии к партии. Это значит, что контрольные границы для иерархического процесса необходимо рассчитывать на основе результатов наблюдений не менее 10 партий, независимо от количества платин в одной партии.

Литература

1. W. H. Woodall, E. V. Thomas. Statistical Process Control With Several Components of Common Cause Variability // IIE Transactions, 27, 757-764 (1995).

2. Д. Уиллер, Д. Чамберс. Статистическое управление процессами. Оптимизация бизнеса с использованием контрольных карт Шухарта. Пер. с англ., – М., Альпина Бизнес, 2009. – 406 с.

3. D. J. Wheeler; D. S. Chambers, Understanding statistical process control (2 ed.). Knoxville, Tennessee: SPC Press., 1992. – 409 p.

Application of control cards of hierarchical processes in technological processes of manufacturing VLS with high number of levels of variability

T.M. Antonova

Abstract. Application of control cards of hierarchical processes in microelectronic manufacture of manufacturing VLS with high number of levels of variability is considered in article.

Анализ дефектов интегральных микросхем с помощью фотонно-эмиссионной микроскопии

А.А. Захарченко¹, М.В. Орешков², В.Ю. Троицкий¹

1- кандидат физико-математических наук, 2-кандидат технических наук

Методика анализа дефектов субмикронных интегральных микросхем с помощью регистрации фотонной эмиссии в последние годы широко используется на ведущих зарубежных полупроводниковых предприятиях при отладке проектов кристаллов микросхем и в серийном производстве [1,2]. Суть методики заключается в регистрации и анализе слабого электромагнитного излучения оптического диапазона от кристалла работающей микросхемы, которое позволяет выявить на ней дефектные области. В настоящее время выпускаются аналитические установки, предназначенные для тестирования микросхем по данной методике (наиболее известным производителем является японская фирма Hamamatsu), но они достаточно дороги и их импорт в Россию весьма затруднен в силу различных экспортных ограничений в странах производителей (Япония, США). В данной статье описана программно-аппаратная реализация методики регистрации фотонной эмиссии и первые результаты ее применения на опытно-экспериментальном участке изготовления микросхем с субмикронными размерами элементов в ОМТ НИИСИ РАН.

1. Введение

Основные принципы возникновения фотонной эмиссии в кристалле интегральной микросхемы основаны на испускании фотонов при переходе носителей с избыточной энергией на уровни с меньшей энергией [3]. Так, в прямо-смещенном р-п-переходе большое количество дырок И электронов межзонной рекомбинируют механизму по рекомбинации с испусканием фотонов с энергией порядка ширины запрещенной зоны или более. Ток утечки в обратносмещенном p-n-переходе также создает фотонную эмиссию за счет релаксации прошедших переход электронов, в обратном направлении и набравших достаточную энергию. Доминирующий механизм эмиссии лля короткоканального КМОП-транзистора представляет

собой релаксацию энергии высоко-энергетичных («горячих») носителей рекомбинацию и сгенерированных ими избыточных носителей в области отсечки канала. Кроме того, фотонная эмиссия наблюдается при протекании значимого тока через подзатворный диэлектрик как В результате некатастрофического пробоя, так и в случае локально повышенной напряженности электрического поля.

Спектр фотонной эмиссии находится в основном в ближнем ИК диапазоне, соответствующем длине волны с энергией порядка ширины запрещенной зоны кремния (1.1 эВ – 1.12 мкм). Кроме того, кремний практически полностью прозрачен в ближнем ИК диапазоне, что позволяет наблюдать эмиссию с обратной стороны кристалла микросхемы, где она не экранируется слоями металлизированной разводки. Интенсивность фотонной эмиссии достаточно низкая,



Рис. 1. Общая схема установки фотонно-эмиссионной микроскопии

поэтому для ее регистрации необходимо использовать высокочувствительную аппаратуру.

2. Оборудование

В ОМТ НИСИИ РАН была сконструирована установка фотонно-эмиссионной микроскопии (рис. 1) на базе инфракрасного микроскопа Leica INM100IR (Германия) с видеокамерой XEVA 320 фирмы Xenics (Бельгия). Микроскоп имеет возможность ИК подсветки объекта как сверху через объектив, так и наблюдения объекта на снизу, для просвет. Максимальное увеличение объектива микроскопа составляет 100х, минимальное – 2.5х. Матрица видеокамеры изготовлена на базе соединения InGaAs, обладает чувствительностью в диапазоне от 900 до 1700 нм и имеет разрешение 320х240 при размере сенсора 9.6х7.68 мм. В процессе измерений температура матрицы видеокамеры поддерживалась на уровне 240 К с помощью системы охлаждения на элементах Пельтье с водяным теплосъемом. Типичные времена экспозиции видеокамеры для наблюдения фотонной эмиссии составляют от 10 до 100 секунд в зависимости от интенсивности эмиссии.

Для снятия фотонно-эмиссионного изображения всей поверхности кристалла микросхемы использовался моторизованный столик фирмы Marzhauser (Германия), точность позиционирования которого составляет 2.5 мкм. Было разработано программное обеспечение, позволяющее автоматически получать фотонно-эмиссионные сигналы от различных частей кристалла и накладывать их на оптическое изображение кристалла и на топологический проект.

В качестве исследуемых объектов были выбраны микросхемы статической оперативной памяти (СОЗУ). изготовленные по КМОП технологии с минимальными топологическими размерами 0.35 мкм и 0.25 мкм. Перед анализом кристалл микросхемы крепился на плату с отверстием, совпадающим по форме и размеру кристаллом, необходимые для проведения с исследований выводы кристалла разваривались на внешние выводы платы. Микросхемы испытывались в статическом режиме, при котором на микросхему подается лишь питание. При нахождении КМОП СОЗУ микросхем в статическом состоянии фотонная эмиссия должна практически отсутствовать в силу крайне малого тока потребления КМОП элементов в статическом состоянии. Если же в микросхеме присутствуют дефекты, вызывающие повышенные токи утечки или потребления ее элементами, то они могут приводить к возникновению фотонной эмиссии. В данном случае, ток статического потребления исследуемых микросхем составлял до 100 мА, что на порядок превышает значение данного параметра для



Рис. 2. Сигнал фотонной эмиссии от дефектного конденсатора микросхемы, наложенный на ее оптическое ИК изображение (слева) и на проект (справа). Объектив 100х, выдержка 100с



Рис 3. Сигнал фотонной эмиссии от дефектной ячейки памяти микросхемы, наложенный на ее оптическое ИК изображение (слева) и на проект (справа). Объектив 100х, выдержка 50с.



Рис 4. Иллюстрация возможных причин возникновения фотонной эмиссии в исследуемых микросхемах СОЗУ

полностью работоспособных микросхем (20-30 мкА).

3. Результаты и обсуждения

На рисунках 2 И 3 представлены зарегистрированные на описанной выше установке сигналы фотонной эмиссии от кристаллов дефектных микросхем, наложенные на оптические изображения этих же кристаллов и результат совмещения сигнала фотонной эмиссии с топологическим проектом. На проекте зеленым цветом изображен поликремний, красным – активная область (карманы активных элементов), синим и оранжевым контуром - области имплантации n- и p-типа, желтыми точками – первый слой металлизированной разводки, оранжевыми квадратами - контакты нижележащих структур к первому слою металлизированной разводки, белыми квадратами – контакты между первым и вторым слоем металлизированной разводки, белыми точками второй слой металлизированной разводки.

Совместный анализ полученных изображений позволил установить предположительную причину возникновения фотоэмиссии из кристаллов данных микросхем. Для случая, показанного на рис. 2, свечение может быть вызвано локальным повышением напряженности электрического поля в подзатворном диэлектрике, вызванное уменьшением его толщины на границе раздела активной области и щелевой изоляции вследствие аномального «проседания» (уменьшения высоты) изолирующего окисла (рис 4а). В результате повышается вероятность возникновения в этой области токов утечки через подзатворный диэлектрик, в том числе и за счет появления дополнительных ловушечных уровней, увеличивающих протекающий ток по прыжковому механизму (аналогичному *SILC*току [4]).

рис. 3 может Эмиссия на быть вызвана повышенным током утечки обратно смещённого стокового *p*-*n*-перехода *p*-канального транзистора в месте, где осуществляется контакт истока к карману при помощи n+-имплантации (квадратная область на рис. 3, отмеченная пунктиром правее места свечения). утечки Повышенный ток вызван сужением поликремниевого затвора месте свечения, В усугубленного наличием ступеньки при переходе поликремниевой шины с активной области на область разделительной изоляции, И, как следствие, недопустимым сближением *n*+ и *p*+-областей (рис. 4б). Следует иметь в виду, что указанное на рис.3 место свечения может не точно совпадать с фактическим изза погрешности позиционирования, сопоставимой с диаметром пятна свечения.



Рис 5. Сигнал фотонной эмиссии от дефектной части микросхемы СОЗУ, наложенный на ее изображение (слева) и на проект (справа). Объектив 63х IR, выдержка 50 с.



Рис 6. Сигнал фотонной эмиссии от схемы защиты СОЗУ 0.25 мкм, наложенный на ее изображение (слева) и на проект (справа). Объектив 20х, выдержка 50 с. locsal – маска локального силицидирования

На рис. 5 представлен сигнал фотонной эмиссии от *p*-канальных транзисторов в двух рядом расположенных инверторах. Форма и расположение сигнала указывает на то, что данные транзисторы нормально открыты и их стоки закорочены на землю. Это, скорее всего, указывает на наличие общей закоротоки по металлу, так как выходы инверторов идут параллельно и расположены достаточно близко к шине «земля».

На рис 6. представлен сигнал фотонной эмисии микросхемы СОЗУ, изготовленной по технологии 0.25 мкм в области схемы защиты по питанию. Выяснилось, что свечение исходит из некоторых мест контакта сильнолегированной и слаболегированной областей противоположного типа легирования (n+ - p и p+ - n) в кольцевых диодах. Область базы этих диодов защищена от образования силицида титана при

формировании контактов к сильнолегированным областям по технологии локального силицидирования (LOCSAL).

Для выявления возможного источника свечения было изготовлено поперечное сечение области кристалла, где оно было зарегистрировано. Для визуализации областей легирования было проведено декорирование полученного шлифа спецмальным травителем. В результате выяснилось, что области сильного легирования n+ и p+ недопустимо сблизились (рис. 7), что вызвало повышенной ток утечки обратно смещенного диода, который и привел к появлению фотонной эмиссии.

В целом для установления более детальной причины появления фотонной эмиссии и природы дефекта необходимо проводить дальнейшее исследование области свечения с помощью



Рис 7. Изображение декорированного поперечного шлифа диодов в РЭМ.

изготовления пренизионных шлифов, травления фокусированным ионным пучком, изолирование дефектного места и напыления контактов для электрофизического анализа и т.п. Будет крайне полезным получить изображение фотонной эмиссии при работе микросхемы не в статическом, а в динамическом режиме, причем с использованием алгоритма тестирования, дающего сбой в конкретной Но такой микросхеме. анализ сопряжен с трудностями, дополнительными связанными с оснастки изготовлением и полключением ЛЛЯ тестирования совместно с регистрацией фотоэмиссионного изображения. Также заметим, что наличие закороток по металлу трудно детектируется с помощью данной методики, так как присутствие такого дефекта в одном месте может приводить к фотонной эмиссии совсем в другом месте, а может и вообще не приводить к какому-либо излучению. Для выявления закороток необходимо использовать другие методики. такие как метол индуцированного оптическим изменения сопротивления лучом (OBIRCH) или термо-индуцированного метод изменения напряжения (TIVA) [5].

4. Выводы

Разработана и сконструирована установка для анализа дефектов субмикронных микросхем с применением фотонно-эмиссионной микроскопии. Представлены первые результаты применения данной методики, продемонстрировано одно из главных ее Методика может использоваться как в качестве эффективного инструмента при анализе брака и отказов субмикронных микросхем, позволяющего максимально локализовать область поиска дефекта, так и при проектировании микросхем с целью отладки проекта.

Литература

1. C.F. Hawkins, J.M..Soden, E.I.Cole Jr., E.S. Snyder. The use of light emission in failure analysis of CMOS ICs // Proceedings of the International Symposium for Testing and Failure Analysis, ISTFA'90, Los Angeles, U.S.A., 29 Oct.-2 Nov. 1990, pp. 55-67

2. J. Kolzer, C. Boit, A.Dallmann, G. Deboy, J. Otto, D. Weinmann. Quantitative Emission Microscopy // J. Appl. Phys., Vol 70, No 11, 1992, pp 23-41

3. A Review of Near Infrared Photon Emission Microscopy and Spectroscopy / J.C.H. Phang, D.S.H. Chan, S.L. Tan, W.B. Len et al. // IPFA 2005, Singapore, pp. 275-281

4. J.H. Stathis. Percolation models for gate oxide breakdown // J. Appl. Phys. 1999. Vol. 86. pp. 5757—5766.

5. F. Beaudoin, R. Desplats, P. Perdu, C. Boit Principles of Thermal Laser Stimulation Techniques // Microelectronics Failure Analysis Desk Reference (Fifth Edition), 2004, pp. 418-425

IC failure analysis using photon emission microscopy

A.A. Zakharchenko, M.V. Oreshkov, V.Yu. Troitskiy

Abstract. The photon emission ICs failure analysis technique is widely used in the leading semiconductor fabs all over the world for design debug and serial production monitoring [1, 2]. The main idea of this technique is registration and analysis of week optical emission from the working IC crystal that makes it possible to localize defect regions on it. State of the art analytical equipment for ICs testing using this technique is produced now by several vendors (one of the most known is Hamamatsu), but such equipment is extremely expensive and falls under export limitations in vendor countries (Japan, USA).

The hardware-software complex for photon emission analysis, developed at SRISA microelectronics division, is described as well as the first results of submicron ICs analysis.

Анализ отказов и брака продукции в производстве микросхем технологического комплекса НИИСИ РАН

В.Ю. Троицкий¹, Т.В. Козырева, С.В. Седов

1-кандидат физико-математических наук

В статье описана процедура анализа брака и отказов микросхем, разработанная и применяемая в производстве микросхем, организованном на базе технологического комплекса НИИСИ РАН. Определена роль группы анализа брака и отказов в процессе микроэлектронного производства. Приведена блок-схема процесса принятия решения по каждому событию отказа, описаны документы, разрабатываемые в процессе анализа. Дается краткая справка о методах, применяемых для анализа брака за рубежом, и тех, которые доступны в технологическом комплексе НИИСИ РАН. В конце статьи приведены конкретные примеры результатов анализа отказов и брака, проведенного в процессе производства микросхем в НИИСИ РАН.

1. Введение

Процедура анализа отказов микросхем при эксплуатации и брака при их изготовлении является неотъемлемой частью современного микроэлектронного производства.

Анализ дефектов, приводящих к отказам интегральных микросхем в эксплуатации, позволяет в зависимости от найденной причины отказа совершенствовать проект микросхем или режимы технологических процессов их изготовления с целью повышения надежности микросхем, а также снижения финансовых и временных потерь со стороны изготовителя и потребителя.

Уменьшение проектных норм изготовления микросхем, приводящее к миниатюризации их элементов и усложнению конструкции кристаллов, корпусов и изделий в целом предъявляет все более жесткие требования к качеству их проектирования и изготовления и делает все более критичной задачу выявления и анализа всех параметров, влияющих на надежность функционирования субмикронных интегральных микросхем.

Обеспечение качества и ожидаемых рабочих характеристик микросхем субмикронного диапазона становятся все более сложными задачами, требуя увеличения роли аналитических исследований на всех этапах жизненного цикла изделий, что позволит своевременно разрешать потенциальные проблемы и поддерживать непрерывные процессы усовершенствований на этапах проектирования и изготовления микросхем.

С другой стороны, повышение роли аналитических исследований диктуется динамикой развития микро-электронного рынка и отрасли в целом, отвечая задачам сокращения длительности жизненных циклов на стадии изготовления микросхем и их выведения на рынок и максимального снижения издержек их производства.

На ведущих предприятиях-изготовителях микросхем эффективная деятельность по анализу микросхем на всех стадиях их жизненного цикла дает возможность не только избежать дополнительных издержек, но даже успешно реализовывать планы по сокращению производственных затрат. Общее сокращение затрат полностью оправдало использование групп анализа дефектов, брака и отказов микросхем и высоко подняло роль процесса анализа как вида деятельности, увеличивающей прибыль производителя [1].

В этой связи изменились роль и задачи подразделений, исторически занимавшихся в основном анализом отказов готовых изделий. Группы анализа инициировать организацию лолжны разработчиков. взаимодействия группами с технологов и с подразделе-ниями, занимающимися контролем техпроцесса и испытаниями изделий, позволяющих добиться своевременного обеспечения всех этих производ-ственных подразделений достоверными данными И необходимой для разработки корректирующих мероприятий информацией.

Современное микроэлектронное производство требует вовлечения аналитических подразделений в процессы выявления и объединения симптомов неисправностей элементов, узлов и микросхем в целом на всех этапах их жизненного цикла, что позволяет сфокусировать усилия по анализу для достижения целей усовершенствования процессов на всех этапах изготовления микросхем.

2. Процедура анализа

Процедура анализа отказов и брака интегральных микросхем представляет собой последовательность действий, предпринимаемых для выяснения причин возникновения конкретного брака/отказа (дефекта изготовления) схемы и устранения этих причин.

В производстве микросхем, организованном на базе технологического комплекса НИИСИ РАН, разработана стандартная процедура анализа отказов микросхем, вовлекающая все необходимые подразделения. На рисунке 1 представлена укрупненная блок-схема процедуры анализа отказов изделий


Рис. 1. Укрупненная блок-схема процедуры анализа отказов.

при эксплуатации. В случае анализа производственного брака процедура начинается с блока «Изолятор брака».

3. Документирование процедуры анализа

Анализ брака и отказов – организационно и методологически сложная процедура, вовлекаю-щая в совместную работу большое число людей из разных подразделений или даже организаций, с высокой степенью ответственности за достовер-ность конечного результата. В связи с этим особенно важно надежно фиксировать не только итоговые результаты, но и все исходные и промежуточные данные, планы, решения. Для этой цели в производстве микросхем, организованном на базе технологического комплекса НИИСИ РАН, в процессе реализации процедуры анализа отказов и брака разрабатывается ряд документов, необходимых и достаточных для фиксации исходных данных, произведенных действий и полученных результатов.

1. Паспорт несоответствующего изделия (ПНИ). Оформляется на группу однородных изделий, направляемых для проведения анализа. В ПНИ включают имеющуюся исходную информацию, необходимую для правильной организации анализа и получения его качественных результатов. Паспорт представляет собой формуляр для заполнения, содержащий вопросы о цели проведения анализа, первичных параметрах отказавшего изделия, его истории от отгрузки до обнаружения отказа, форме проявления и способе регистрации отказа, контактных лицах у потребителя, а также любых других полезных для проведения анализа данных. Он реализован в бумажном (для заполнения на территории потребителя) и в электронном (для заполнения на территории ОМТ, а также для последующей структуризации и анализа полученных данных) виде.

2. Программа анализа. Если ситуационный анализ не дает достаточной информации для заключения о причинах отказа (брака), разрабатывается программа анализа, которая может включать при необходимости дополни-тельные проверки функционирования и физический анализ. В случае отказа при эксплуатации анализ проводится по утвержденной типовой программе, в случае отказа на испытаниях или анализа брака производства разрабатывается индивидуальная программа.

3. **Протокол анализа**. По результатам проведения ситуационного анализа и реализации программы анализа составляется протокол анализа, содержащий описание произведенных действий, выводы о причинах отказа (брака) и рекомендации по его устранению и/или предотвращению.

4. **Программа корректирующих мероприятий**. По данным в протоколе анализа выводам и рекомендациям ответственное подразделение разрабатывает и реализует программу корректирующих мероприятий.

корректирующим 5. Отчет no мероприятиям. В отчете по корректирующим мероприятиям содержится описание всех проведенных корректирующих воздействий и анализ эффективности. В случае недостаточной ИХ эффективности проведенных корректирующих мероприятий разрабатывается новая программа корректирующих мероприятий.

лействие. Кажлое связанное с анализом забракованного (отказавшего) изделия (включая прием изделия на анализ, разработку программы анализа, выполнение действия из нее, составление протокола анализа), разработкой и внедрением корректирующих мероприятий, а также каждый документ, разработанный в процессе анализа, регистрируется в специально разработанной на производстве НИИСИ РАН электронной базе данных. Это позволяет отслеживать стадию анализа и каждого несоответ-ствующего местонахождение изделия, а также иметь быстрый доступ к любому документу, составленному в ходе выполнения анализа, или приобщенного к процедуре анализа, даже проводя параллельно анализ нескольких разных изделий.

4. Методы физического анализа

Набор методов, применяемых при физическом анализе брака (отказов) изделий в производстве микросхем, организованном на базе технологического комплекса НИИСИ РАН, зависит от этапа жизненного цикла интегральной микросхемы, на которой он был обнаружен. При анализе брака микросхемы на этапе кристального производства может применяться весь арсенал методов операционного контроля [2]. При анализе брака сборочного производства в основном применяется визуальный контроль, электронная микроскопия, анализ элементного состава методом рентге-новского микроанализа.

С наибольшими трудностями приходится сталкиваться установлении причины при производственного отказа готовой микросхемы, если он обусловлен скрытым дефектом. Это связано с тем, что сначала необходимо выявление местоположения и характера проявления дефекта, и только потом целесообразно проводить физические исследования. Чем точнее удается определить местоположение и характер дефекта, тем эффективнее последующие физические исследования. По мере все более глубокого перехода в область субмикронных размеров и вызванного этим усложнения конструкции интегральных микросхем и увеличения степени их интеграции задача поиска элементов микросхемы, ответственных за ее полную или частичную существенно неработоспособность, также усложняется.

Невозможно провести анализ отказа микросхемы и выявить причину его возникновения за экономически целесообразное время, а иногда и в принципе, без сужения области поиска с миллионов элементов, расположенных на одном кристалле, до нескольких сотен, а лучше десятков штук. За рубежом для этих целей применяют два взаимодополняющих подхода:

1. проектирование интегральных микросхем таким образом, чтобы по результатам функционального тестирования было возможно определить место расположения критического дефекта;

2. применение специализированного оборудования, реализующего один или несколько методов, основанных на возбуждении оптического или электрического эффекта в области нарушения функционирования микросхемы [1].

В производстве микросхем, организованном на базе технологического комплекса НИИСИ РАН, ведутся работы по внедрению одного из этих методов в собственной программно-аппаратной реализации [3].

Для визуализации дефектного элемента методом электронной микроскопии необходима предварительная подготовка образца. Она может включать локальную шлифовку корпуса или обратной стороны кристалла, резку и/или утонение кристалла, последовательное удаление вышележащих слоев микросхемы. Все эти работы успешно проводятся в производстве микросхем, организованном на базе технологического комплекса НИИСИ РАН, при анализе брака.

Для электрической характеризации отдельного дефектного элемента ИМС необходимо формировать контактные площадки к нему при помощи локального травления и напыления металла сфокусированным ионным пучком (FIB). В производстве на базе технологического комплекса НИИСИ РАН ведутся некоторые экспериментальные работы в этом направлении с привлечением установок локального ионного травления сторонних организаций.

5. Результаты анализа отказов

За период с 2008 по 2013 годы по претензиям от потребителей был проведен анализ более, чем 100 микросхем, однако для большей их части при контрольной проверке отказ был признан (рисунок 2а). Больше всего эксплуатационным производственных отказов микросхем происхо-дило причине несовершенства проекта по (рисунок 2б).





Рис. 2. Сводная диаграмма результатов анализа отказов изделий при эксплуатации.

Для каждого случая отказа с выясненной причиной были внедрены корректирующие мероприятия: по ошибкам оператора - вне-очередные инструктажи, по несовершенству проекта – доработка программы и аппаратуры тестирования, технической документации, по низкому качеству материалов поставщика. корректирующие замена Bce мероприятия были признаны достаточно эффективными и позволили существенно снизить долю брака в продукции, поставляемой потребителю.

Несколько примеров частных результатов анализа брака и отказов приведены ниже.

1. Микросхема отказала после проведения испытаний на воздействие одиночных ударов. Наблюдались отсутствие контактирования и короткие замыкания отдельных выводов. Контроль вскрытого корпуса микросхемы в сканирующем электронном микроскопе позволил обнаружить трещину на дне монтажного колодца (рисунок 3а). С помощью оптического микроскопа была обнаружена трещина и на нижней стороне корпуса (рисунок 3б).





Рис. 3. Трещина, вызвавшая отсутствие кон-тактирования, обнаруженная: а) на дне мон-тажного колодца, б) на нижней стороне корпуса.

Анализ истории проведения испытаний выявил нарушение режима испытаний, устранение которого позволило избежать такого отказа в следующих испытаниях.



Рис. 4. Касание проволочным выводом дорожки металлизации.

Визуальный осмотр контактных площадок на кристалле выявил касание проволочным выводом дорожки металлизации, которое могло привести к короткому замыканию (рисунок 4).

2. Микросхема отказала при кратковре-менных испытаниях на безотказность. Исследование в оптическом микроскопе микросхем в корпусах вскрытыми крышками позволило обнаружить многочисленные расплав-ления шин питания и земли в области контактных площадок (рисунок 5).



Рис. 5. Оплавление шин питания микросхемы.

Причиной отказа явилось резкое повышение питающего напряжения в ходе испытаний.

3. На визуальном контроле готовых микросхем были обнаружены зазоры между контактными площадками внешних выводов микросхем и керамическим корпусом. Сравнительный анализ микросхемы в сборе и нового корпуса, показал, что подобный отслоения имеются и на новом корпусе и не связаны с обработкой (рисунок 6).



Рис. 6. Вид нового корпуса: а) отслоение между внешним выводом и керамикой



 б)
 Рис. 6. Вид нового корпуса: б) след после удаления расслоившегося вывода

4. Партия изделий была забракована после формирования второго слоя металлизации из-за повреждений металлических шин, обнаруженных на операции визуального контроля (рисунок 7).



Рис. 7. Повреждения металлических шин.

В результате анализа скола методом электронной микроскопии было обнаружено нарушение внутренней структуры металлических межсоединений: затекание вольфрама в нижележащий слой алюминия (рисунок 8).

Коррекция режимов операций блока формирования металлизации позволила в дальнейшем избежать разрушения барьерного слоя на поверхности металлических шин, приводящего к их эрозии.





Рис. 8. Затекание вольфрама в нижележащий слой алюминия.

Заключение

Анализ отказов и брака является неотъемлемой частью современного микроэлектронного производства. Его ролью является как удовлетворение претензий потребителей, так и постоянное повышение качества продукции и эффективности производства. В производстве микросхем, организованном на базе PAH. технологического комплекса НИИСИ разработана и внедрена цельная процедура анализа, включающая регламент взаимодействия подразделений, документирование всех производимых действий и результатов от исходных данных до оценки эффективности корректирующих мероприятий. Она позволяет во многих случаях успешно определять и корректировать причину возникновения отказа или брака изделия. Лля эффективности повышения локализации и обнаружения дефекта, приводящего к отказу готового изделия, необходима реализация методов локализации дефектов, в частности. с привлечением специализированного оборудова-ния, а также более тесная кооперация с разра-ботчиками проекта изделия.

Литература

1. Failure analysis of integrated circuits. Tool and techniques. Edited by L.C. Wagner, Kluwer Academic Publishers, Boston / Dordrecht / London, 2000.

2. В.Ю Троицкий., М.В. Орешков, Т.В.Козырева. Особенности операционного контроля в условиях мелкосерийного кристального производства. (В этом сборнике)

3. А.А. Захарченко, М.В. Орешков,

В.Ю. Троицкий. Анализ дефектов интегральных микросхем с помощью фотонно-эмиссионой микроскопии. (В этом сборнике).

Failure analysis by integrated circuits (IC) manufacturing at technological complex of SRISA RAS

V.J. Troitsky, T.V. Kozyreva, S.V. Sedov

Abstract. Failure analysis procedure developed and applied by integrated circuits (IC) production at technological complex of SRISA RAS is stated in the work. The role of failure analysis group by IC manufacturing is determined. The flow block of decision-making process for every failure event is set out. The documents are described that are developed by analysis process. An information is given about methods that are used for failure analysis abroad and at the (IC) manufacturing at technological complex of SRISA RAS. Some case studies from the manufacturing are printed in the end of the paper.

Особенности германиевых технологий КМОП ИС

В. А. Горячев

кандидат физико-математических наук

Для проектирования высокопроизводительных КНИ КМОП микропроцессоров с низкой мощностью рассмотрены отдельные особенности применения германия в конструкционных элементах ИС. Даётся представление об усовершенствованиях в технологических процессах создания микро- и наноконструкций: планарных и нитевидных нанокристаллов кремния с гетеропереходами Ge-Si. Отмечаются характеристики физических эффектов в различных технологиях. Делается вывод о возможности производства Ge-Si КМОП ИС, совместимого с традиционными технологиями микроэлектроники.

Введение

Проблема повышения скорости обработки данных – ключевая в развитии информационных технологий (IT). Сегодня требуются скорости 10 Гбит/с и выше. Но медные проводники достигли физического предела скорости передачи информации до ~ 300 Мбит/с.

Второй важнейшей проблемой является потребление электроэнергии. На сегодняшний день в разных областях технологических задач требуются компьютеры уровня 1–5 Тфлопс (10^{12} операций с плавающей запятой в секунду). В ближайшие годы потребление энергии суперкомпьютером экзафлопной мощности вырастет более чем на два порядка и превысит гигантское значение ~ 10^5 кВт. Наибольшая часть потребляемой энергии при этом расходуется на нагрев проводников и рассеивается в окружающее пространство.

Альтернатива медным проводам - оптические соединения. По прогнозам, чтобы преодолеть «медный суперкомпьютер в 10 петафлопс тупик» (на сегодняшний день достигнутая мощность составляет 2,57 петафлопса, и выдает её китайский Tianhe-1A) должен содержать 5 млн оптических каналов, а ЭВМ мощностью 1 экзафлопс – более 10⁹ каналов с пропускной способностью по 100 Гбит/с каждый [1]. Но современные каналы бытового интернета предлагают в лучшем случае 100 Мбит/с, то есть в тысячу раз меньше.

С уменьшением размеров транзисторов в чипах производительность вычислительных систем постоянно повышается, однако без модификации конструкции этих элементов, которая оставалась практически неизменной в течение последних 40 лет, индустрия может встретиться с большими IT сложностями. Один из методов разрешения проблемы - это включение германия (Ge) в технологический процесс. Подобное включение является уже знакомой технологией почти для всех основных производителей чипов, поскольку добавка Ge увеличивает скорость работы кремниевых микросхем и может сочетать как улучшение проводимости медных соединений, так и новую систему беспроводной передачи данных. По сравнению с кремнием, у германия более низкое напряжение pn-перехода (0.1V÷0.3V против 0.6V÷0.7V). Это делает германий более экономичным и в плане энергозатрат.

В отличие от других материалов, германий легко существующему процессу присоединять к производства кремниевых чипов. Подобную схему разработали ученые из Национального Тайваньского Университета, что позволит передавать информацию в 100 раз быстрее, чем по протоколу Wi - Fi, сообщает Cnews, ссылаясь на Taiwan News [2]. Размер созданного тайваньскими учеными чипа составляет всего 0,5 х 0,5 мм – это в 10 раз меньше современных микросхем беспроводной связи. При этом стоит чип в 10 раз дешевле и потребляет меньше электричества. Спустя некоторое время такой чип, как утверждают создатели, будет стоить ~ \$1, и это сделает его привлекательным для широкого круга применения.

Корпорация IBM совместно с Калифорнийским университетом в Беркли ведут разработку аналогичных устройств, однако ученые из тайваньского университета совершили настоящий прорыв – ни одна существующая разработка не может предложить подобную скорость и стоимость.

1. Физические особенности технологических процессов

Внедрение Ge сопряжено со значительными технологическими изменениями И усовершенствованиями, особенно при проектировании больших интегральных микросхем. Проблема состоит в том, что при создании таких структур необходимо использование процессов самоорганизации нанокристаллов (СНК) германия. Однако подобные процессы, как правило, включают четыре стадии: а) термическое окисление пластины для получения слоя туннельного окисла толщиной $h = 3 \div 8$ HM, разделяющего слой СНК и канал МОП-транзистора; б) последовательное осаждение методом молекулярнолучевой эпитаксии (МЛЭ) тонких слоев Ge ($h = 0.5 \div 0.9$ нм) и слоев Si $(h = 5 \div 30$ нм); в) термическое оксидирование верхнего слоя Si ЛО его соприкосновения с туннельным оксидом; г) быстрый термический отжиг структуры в инертной среде для формирования слоя нанокристаллов Ge [3].

В целом процедура формирования структур содержит обычные, отработанные технологические оксидирования, методы осаждения И отжига использовать позволяющие традиционное технологическое оборудование для производства электронных приборов. Но самоорганизация нанокристаллов недетерминированный процесс, _ результаты часто оказываются которого отличающимися друг от друга.

B более поздней попытке отечественных исследователей из ИФП СО РАН моделирования Монте-Карло проанализировано метолом формирование гетероструктур Si-Ge на основе нитевидных нанокристаллов (ННК), выращенных по механизму пар-жидкость-кристалл [4]. Найдено, что при росте аксиальных гетеропереходов (ГП) также невозможно получить атомарно однородную или гладкую гетерограницу Si-Ge, что связано с постепенным изменением состава капли катализатора при переключении потоков. Исследована зависимость состава переходного слоя G_{ex}Si_{1-x} от соотношения потоков и длительностей осаждения германия и кремния. Установлено, что ширина аксиальных гетеропереходов Si-Ge зависит от диаметра ННК. В адсорбционном режиме роста она линейно зависит от диаметра, а в диффузионном ширина ГП убывает с увеличением диаметра. Это означает, что при уменьшении толщины нитевидных нанокристаллов (ННК) резкие гетеропереходы в системе Ge-Si оптических волокон можно получить только в методе внутреннего осаждения CVD, когда несущественна диффузионная составляющая роста. Такой метод заключается в осаждении стеклообразующего и легирующего окислов SiO2 и GeO2 соответственно и характеризуется малыми потерями. Перспективным является использование ННК Ge и Si в качестве химических и биологических сенсоров, туннельных диодов, элементов памяти, одноэлектонных транзисторов. Одним из этапов создания электронных приборов на основе ННК является формирование *p* – *n* переходов и гетеропереходов Ge-Si. В Si-Ge нанопроволоках можно создавать аксиальные и радиальные переходы. Ha германиевых нанопроволоках, покрытых кремниевой оболочкой, создаются вертикальные транзисторы с более высокой подвижностью носителей. Формирование аксиальных гетеропереходов в ННК не требует дополнительных технологических операций в отличие от отмеченных выше планарных структур, так как гетеропереходы формируются непосредственно в процессе роста. Важно, что гетероструктуры Si-Ge также совместимы с традиционной кремниевой технологией.

2. Замена электрических сигналов оптической связью

Для преодоления возникающих трудностей в период 2010 – 2011 г.г. главной идеей зарубежных специалистов по созданию высокопроизводительных ИС становится использование оптических соединений [5].

Примерно в тот же период времени представители компании IBM сообшили о значительном продвижении на пути замены электрических сигналов оптической связью. Их новый сверхпроизводительный фотодетектор может быть встроен в кремниевые чипы. Устройство, которое сами создатели называют «нанофотонным лавинным фотоприёмником» (nanophotonic avalanche photodetector), - ключевой компонент будущих оптических процессоров [6]. Уже сейчас этот потоковый фотодетектор по величине полосы пропускания в 340 ГГц почти влвое недавний образец от превосходит опытный университета Калифорнии в Дэвисе (http://www.ucdavis.edu/) [7].

В кремне-германиевом чипе IBM используются возможности лавинного фотоэффекта, умножающего первоначальные сигналы во много раз. Лавинный эффект происходит в масштабе нескольких десятков нанометров, позволяя, как утверждают представители IBM, снижать уровень шума на 50 – 70% в сравнении с другими прототипами подобных детекторов.

Следует отметить, что исследователи из Лаборатории технологий микросистем Массачусетского технологического института (MIT's Microsystems Technology Laboratories) не только усовершенствовали германиевые интегральные схемы, но и продемонстрировали первый лазер, полученный на основе германия. Такой лазер может генерировать световые волны с длиной, необходимой лпя осуществления оптической связи. Кроме того, это



Рис. 1a). Микрофотография задней части чипа IBM Holey Optochip с фотоприемниками и лазерами фотопередатчиками, которые видны в отверстиях подложки.

первый германиевый лазер с непрямыми межзонными переходами, работает при комнатной температуре [8].

Последнее время корпорация Intel проводит эксперименты с микросхемами на полностью обедненных КНИ-подложках, с 3D-схемами со сквозными отверстиями через кремний (Through-Silicon-Via, TSV) для оптических волоконных связей и р-канальными полевыми транзисторами на полупроводниковых соединениях A₃B₅ с Ge-каналом. 3D-схема со сквозными отверстиями для оптических волоконных связей изображена на рисунке 1 а).

Стандартный КМОП чип, выполненный по технологии 90 нм, представлен ниже на рисунке 1 б).

Похожие устройства появятся в продаже в ближайшие два-три года.

3. Полевые транзисторы с германиевым каналом

МОП транзисторы на кремниевых соединениях представляют собой приборы n-типа, но для КМОП



Рис. 16). Стандартный 90-нм КМОП-чип с отверстиями в подложке, заимствованный с сайта http://www.flcd.ru/news/ibm_holey_optochip_visokoskorost noi_transiver/

схем необходимы и транзисторы р-типа, для чего и создается Ge-канал. На IEDM 2010 Intel сообщила о разработке р-канального полевого транзистора на основе структуры с квантовыми ямами (QWFET) и с германиевым каналом, который предназначен для маломощных А₃В₅ КМОП микросхем [9]. Транзистор выполнен с затвором HfO₂/TiN, самосовмещенными областями истока/стока, полученными имплантацией ионов бора и изолирующим слоем фосфора, подавляющим параллельную проводимость буферного слоя GeSi. При толщине оксида 145 нм подвижность носителей (дырок) равна 770 см²/В·с при их плотности 5·10¹²см². Это самое высокое значение подвижности для оксидов толщиной менее 400 нм, в четыре раза превышающее этот показатель для совершенных структур с напряженным кремнием. Отношение токов транзистора в состояниях ON/OFF оказалось $> 10^2$. Транзисторы создавались по технологии начального формирования затвора (gate first) при t = 350°C.

Экспериментальная структура нового германиевого транзистора, изготовленного ИЗ распространенных (часто применяемых) полупроводниковых материалов, может стать тем элементом конструкции, что совместно с быстрыми позволит существенно увеличить связями производительность микропроцессоров [10].

Таким образом, чипы на основе новых транзисторов можно выпускать с помошью существующих технологических процессов и на существующем оборудовании, не подвергая оборудование существенным переделкам, чего нельзя сказать о транзисторах из более экзотических материалов, к примеру, из углеродных нанотрубок или графена.

уже говорилось, Как новый германиевый транзистор является транзистором р-типа, заряд внутри которого переносится за счет движения дырок, образованных в кристаллической решетке материала с помощью примесей. Для повышения скорости работы транзистора, т.е. создания более высокой частоты переключения, некоторые элементы германиевого транзистора подвергаются своего рода механическому напряжению. Это механическое напряжение деформирует кристаллическую решетку германия, сближает атомы в узлах решетки, что позволяет электронным дыркам передвигаться в материале с большей скоростью.

Для того, что бы создать механическое напряжение в Ge, исследователи вырастили слой германия сверху «бутерброда» из слоев нескольких видов кремния. Атомы германия естественным путем пытаются упорядочиться относительно атомов кремния в нижних слоях, что приводит к смещению атомов германия и возникновению механической напряженности «Эти "напряженные" слои все время кристалла. "сломаться". стремятся Но авторы разработали несколько способов, позволяющих выращивать механически напряженные слои, сохраняющие свои свойства длительное время без возникновения какихлибо дефектов» - рассказывает Джеймс Теэрэни (James Teherani), исследователь ИЗ Массачусетского технологического института [10].

Вторым компонентом, обеспечивающим высокую скорость работы нового, германиевого полевого транзистора, является его топологическая структура с каналами, окруженными с трёх сторон затвором.



Рис. 2. Структура квантовой модели транзистора из [11]

Собственно канал германиевого транзистора находится выше плоскости самого чипа. Такую конструкцию можно сравнить с поездом, стоящим на рельсах. Все каналы «обернуты» затвором, подобно изображенному на рисунке 2, заимствованному из работы [11]. Похожая структура может обеспечить возможность надежного управления потоком электронных дырок через Ge-канал транзистора.

Работа исследователей ИЗ Массачусетского технологического института по созданию высокоскоростного германиевого транзистора была проведена по заказу Управления перспективных исследовательских программ Пентагона DARPA и при поддержке компании Semiconductor Research Corporation.

В работе исследованы закономерности формирования МОП-структур, содержащих плотные массивы самоорганизованных нанокристаллов Ge, локализованных внутри слоя SiO₂ на расстоянии ~ 3,9 \pm 1,4 нм от Si-подложки и имеющих высокую однородность по размерам (4,5 \pm 1,8 нм). Эта процедура включает последовательное создание туннельного оксида, МЛЭ-выращивание слоев Ge (0,7 - 0,9 нм) и Si (10 - 20 нм), оксидирование верхнего слоя Si и отжиг в инертной среде для получения самоорганизованных Ge-нанокристаллов.

Заключение

Таким образом, ближайшие годы обещают существенные изменения как технологии, так и качества характеристик элементов наноэлектроники. «Прорыв в интеграции КМОП схем и нанофотоники обеспечить увеличение может беспрецедентное функций И производительности чипов через повсеместное использование оптических соединений с низкой мощностью - между стойками, модулями, чипами или даже внутри самих чипов» - сказал доктор Юрий Власов, управляющий подразделением Silicon Nanophotonics в IBM Research. С добавлением всего нескольких производственных модулей в цикл стандартного КМОП производства новые технологии позволят получать множество нанофотонных компонентов, таких как модуляторы, германиевые фотодетекторы, ультракомпактные мультиплексоры и др., пригодные для интеграции с аналоговыми и цифровыми цепями КМОП ИС. Реализации подобных работ по отечественной электронно-компонентной базе невозможна без создания собственных, специализирован-ных САПР и производственного оборудования.

Разрабатываемый подход можно считать переходным от традиционной технологии СБИС к будущей технологии наноэлектронных схем. При этом появляется возможность изготовления не только одноэлектронных ячеек памяти, но и одноэлектронных

транзисторов, выполняющих логические функции. Предполагается, что наноструктуры позволят преодолеть фундаментальные ограничения технологии микроэлектроники (ограничения в размерах, связанных фотолитографией; проблемы с формирования мелких переходов и т. д.), а также совершенно новые полупроводниковые созлать приборы и ранее неизвестные возможности в формировании трехмерных интегральных схем. На основе предложенных структур с кристаллами германия могут быть разработаны схемы памяти и одноэлектронные транзисторы с улучшенными характеристиками, такими как скорость записи и информации, считывания длительность энергонезависимого хранения информации, плотности элементов на схеме и высокой радиационной устойчивостью.

Литература

1. http://books.ifmo.ru/file/pdf/228.pdf

2. http://www.caravan.ru/about/press/white-paper/58/

3. П.И. Гайдук, А.Н. Ларсен, П. Норманд, А. Клавари Формирование нанокристаллов Ge в слоях Si0₂ при быстрой термической обработке слоев Si/Ge // Вестник БГУ. Сер. 1, 2007, № 2, стр. 46-50

4. А.Г. Настовьяк, И.Г. Неизвестный, Н.Л. Шварц Моделирование роста нитевидных нанокристаллов кремния с гетеропереходами Ge-Si // Поверхность. Рентгеновские, синхротронные и нейтронные исследования, 2011, № 9, стр. 1-9

5. <u>http://www.3dnews.ru/news/ibm-ocherednoy-shag-</u> k-sistemam-s-opticheskimi-soedineniyami

6. http://www.membrana.ru/particle/3753

7. <u>http://www.ferra.ru/click/news/http://www.ibm.com/.http://domino.research.ibm.com/comm/research_projects.nsf/pages/photonics.apd.html</u>

8. <u>http://studyamerica.ru/?action=article&ar_id=378</u>

9. М. Гольцова. «IDEM 2010 Новые процессы, новые материалы» Электроника НТБ № 1, 2011, стр. 124-134

10. http://www.nanonewsnet.ru/news/2013/

11.<u>http://www.nanosoft.ru/articles/ftian_proc_2007_v</u> yurkov_2.doc

Features of the ICs CMOS Germanic technologies

V.A. Goryachev

Abstract. For design of highly productive SOI CMOS microprocessors with low power separate features applications of germanium in the ICs constructional elements are considered. An idea is given of enhancements in technological processes micro and nanoconstructions: planar and filamentary silicons nanocrystals with Ge-Si heterojunctions. Characteristics of physical effects in different technologies are marked. The output about possibility of production Ge-Si CMOS ICs, compatible with traditional technologies of microelectronics is drawn

Компьютерное моделирование зависимости характеристик нанотранзисторов от топологии

И.А. Люосева

В статье рассматривается два эффекта влияния элементов топологии на проектирование аналоговых схем – эффект близости кармана и длина области диффузии. Предлагается методика компьютерного моделирования этих эффектов

При проектировании аналоговых схем по технологии 20 нм становится актуальной проблема, связанная с появлением физических эффектов, зависящих от расположения элементов схемы относительно друг друга (layout dependent effects – LDE). Но и для технологии 28 нм и даже для 65 нм следует учитывать влияние топологии на характеристики прибора.

Было рассмотрено влияние двух LDE на характеристики прибора.

Первый – WPE (well proximity effect) – «эффект близости кармана» представлен на рис. 1. Для технологических норм 90 нм и менее на характеристики МОП-транзистора начинают влиять соседние элементы топологии ИС, не принадлежащие данному транзистору. Пороговое напряжение и усиление транзистора меняются в зависимости от того расстояния на которое карман окружает транзистор (PMOS) или отстоит от него (NMOS). В результате рабочая точка транзистора может быть сдвинута на 20-30%, что приводит к потере работоспособности аналоговых ИС.



Рис. 1. WPE (эффект близости кармана)

Второй – LOD-эффект (length of the OD region effect) – учитывает изменение порогового напряжения и тока стока при изменении расстояния от затвора транзистора до краев области диффузии (sa и sb на рис. 2). При приближении активной области транзистора к границе изоляции LOD-эффект усиливается и оказывает заметное влияние на характеристики приборов.

Моделирование проводилось на САПР Cadence. Использовались КМОП транзисторы из технологической библиотеки фирмы TSMC, изготовленные по технологии 65 нм. Методика моделирования представлена на рис. 3.





Рис. 3. Методика моделирования LDE

Для определения влияния WPE и LOD эффектов определялась зависимость тока стока от напряжения на затворе при постоянном напряжении на стоке и зависимость тока стока от напряжения на стоке при постоянном напряжении на затворе транзисторов. При этом для WPE эффекта брали по четыре р- и пканальных транзистора, размещенных на различных расстояниях от краев кармана. В результате моделирования получалась зависимость порогового напряжения от расстояния sc следующего вида (рис. 4).

Изменение тока стока при изменении расстояния sc в четыре раза составило менее 4%.



1.5

2

2.5



sc (um)

Рис. 4. Зависимость порогового напряжения от расстояния до кармана

Для LOD эффекта брали по восемь p- и пканальных транзисторов, соединенных «гребенкой», с различными расстояниями от затворов до краев диффузии. Изменение тока стока в зависимости от этого расстояния представлено на рис. 5. При изменении sa или sb примерно в пять раз изменение Id составило менее 2%.



1

Рис. 5. Изменение тока стока в зависимости от расстояния до границ области диффузии

0.5

Таким образом результаты моделирования говорят о безусловном наличии WPE и LOD эффектов в технологии 65 нм, но не о сильном их влиянии на характеристики приборов (особенно по сравнению с технологиями 20 и 28 нм).

Литература

1.4

1.2

0.8

0.6

0.2

-0.4

0

∆Id

(%)

1. http://electronicdesign.com/article/eda/lde-standsmart-device-74614

2. http://www. springsoft.com/ assets/files / Newsletters / 2012/Poster_LDE Aware Design Solution 20120504.pdf

3. P. G. Drennan, M. L. Kniffin, D. R. Locascio. Implications of Proximity Effects for Analog Design.-// IEEE Custom Integrated Circuits Conference (CICC). – 10-13 Sept. 2006

Computer modeling of the nanotransistor characteristics from the topology

I.A. Lyuoseva

Abstract. This paper addresses two layout dependent effects, well proximity and length of the OD region, as they impact to analog circuit design. Comuter simulation methodology of layout dependent effects has been proposed.

Влияние высокой диэлектрической проницаемости подзатворного диэлектрика на характеристики КНИ КМОП нанотранзисторов

Н.В. Масальский

кандидат физико-математических наук

На основе численных решений анализируются характеристики ультра коротких двух затворных КНИ нанотранзисторов с тонкой нелегированной рабочей областью для случая применения в качестве подзатворного диэлектрика затвора материалов с высоким значением диэлектрической проницаемости. Основное внимание уделяется анализу степени проявления коротко-канальных эффектов и эффектов размерного квантования в не масштабируемых характеристиках транзисторов.

Введение

Одной из основных проблем при переходе к новым топологическим нормам и новым технологиям масштабирование является физических характеристик, в частности, топологических размеров, используемых электронных устройств [1-4]. Согласно последней редакции объединенной программы International technology roadmap for semiconductor (ITRS) очень [5]. тонкий подзатворный диэлектрик с эквивалентной окисной толщиной (ЕОТ) меньше чем 1 нм требуется для реализации всех КМОП приборных технологий таких High performance (HP) как высокая производительность, Low standby power (LSTP) - с потребляемой мощностью в режиме низкой ожидания и Low operation power (LOP) - низкой операционной мощностью. Более тонкая окись позволяет снизить ток утечки между затвором и подложкой, ограничить потребляемую мощность и улучшить управление в ON и OFF состояниях [1, 3]. Чтобы получить требуемый ЕОТ, необходимо использовать изолятор затвора с высокой диэлектрической проницаемостью. Например, оксид гафния HfO₂, который является перспективным кандидатом, так как ОН обладает высокой диэлектрической постоянной $\varepsilon = 25\varepsilon_0$ где ε_0 абсолютная диэлектрическая проницаемость, И хорошей тепловой устойчивостью в контакте с кремниевой пленкой Одновременно [6]. осуществляется масштабирование к более низким значениям и других топологических параметров транзисторов, приводящее к более высокой как производительности, так и степени интеграции электронных устройств. Следовательно, возрастает актуальность поисковых исследований и разработки с учетом квантовых эффектов и технологических требований физических моделей компонентов нанотранзисторных СБИС, что определяет, прежде всего, возможности проектирования микросхем с низким уровнем напряжением питания И уменьшенным значением потребляемой мощности на единицу площади [4, 7].

Одним из перспективных активным компонентом нанотранзисторных СБИС является двух затворный КНИ нанотранзистор с ультра тонкой нелегированной рабочей областью (рис. 1) [3, 6].



Рис. 1 Схема транзистора

Для данного сорта транзисторов, наряду с коротко-канальными эффектами (ККЭ), возникают и другие, специфические эффекты размерного квантования (ЭРК), приводящие, в частности, к сдвигу порогового напряжения [1, 8]. Поэтому в необходимо данной области одновременно ККЭ рассматривать как в двухмерном (2D) приближении, так и ЭРК в тонком рабочем слое кремния. Количественные результаты в общем быть могут получены лишь случае при использовании численных методов моделирования. В данной работе используется подход, который основывается на последовательном решении уравнений Шредингера, Пуассона и токовых уравнений [4, 9, 10]. Используемая модель для транспорта носителей следует из квантового дрейфодиффузионного метода, где носители находятся в локальном равновесие, которое характеризуется локальным уровнем Ферми.

Моделирование КНИ нанотранзисторов на физическом уровне важно как для понимания

физических процессов в нем, так и для оптимизации их конструкции для достижения оптимальных характеристик, а также оптимизации технологического маршрута. Тем не менее, при КНИ устройств с целью масштабирование повышения степени интеграции и быстродействия без одновременного ухудшения надежности и тепловых потерь, существует ряд факторов, которые не масштабируются ни с физическими размерами, ни с приложенным напряжением [2]. Исследование их поведения в присутствии в качестве подзатворного диэлектрика материала с высокой диэлектрической проницаемостью является целью настоящей работы.

1. Основные положения

Основная причина немасштабируемости ряда параметров состоит в наличии теплового потенциала kT/q, (где k – константа Больцмана, T – температура, q – заряд электрона), а также ширины запрещенной зоны E_g в кремнии [1, 2]. Первое из двух названных обстоятельств является существенным лпя масштабируемости объяснения отсутствия подпорогового режима, т.е. пороговое напряжение U_{th} не может уменьшаться подобно другим параметрам. Второе приводит к отсутствию масштабируемости встроенного потенциала, глубины обедненного слоя и ККЭ.

Вследствие экспоненциальной зависимости подпорогового тока І_{ОFF} пороговое напряжение не может быть существенно уменьшено без значительного повышения тока в состоянии OFF. На самом деле, даже если уровень U_{th} поддерживать постоянным, ток I_{OFF} будет все же возрастать в κ_M раз (к_м – коэффициент масштабирования) по мере того, как физические размеры будут уменьшаться в км раз [11]. Это накладывает серьезные ограничения на то, насколько низким быть может пороговое напряжение. Ограничения на величину U_{th} в свою очередь приводят к нижнему пределу для напряжения питания U_{dd}, т.к. временные задержки быстро возрастают с ростом отношения U_{th}/U_{dd} , в частности, когда оно превышает 0.3 [7].

2. Влияние эффектов размерного квантования

Для ультратонких рабочих слоев кремния структурные ограничения, наряду с поперечной компонентой напряженности электрического поля, приводят к перенормировке энергетического спектра [12]. Проявляется это прежде всего в изменении характера зонного закона дисперсии – вместо зависимости энергии носителей от квазиимпульса $E(\mathbf{p})$, например, в зоне проводимости возникает ряд квазинепрерывных зависимостей в подзонах, в результате чего возрастает пороговое напряжение, ухудшается затворная емкость и токовые характеристики транзистора [8-10].

На рис. 2 представлены результаты моделирования, где для двух типов подзатворных

Электронная

плотность, х10¹⁸ см⁻³

SiO₂ и HfO₂ и для двух различных окислов смещения затворов напряжений изображены энергетическая диаграмма зоны проводимости и распределение электронной плотности вдоль затвора в центре рабочей области (x=0). Граница зоны проводимости для SiO₂ и HfO₂ маркирована 'CB' и квантовая электронная плотность 'qD'. Зона проводимости в рабочей области ниже, чем уровень Ферми. Уровень Ферми почти плоский и равен -0.35 эВ для SiO₂, и для HfO₂ -0.27 эВ.

Энергия, эВ



Рис. 2а. Случай SiO₂



Рис. 2б. Случай HfO₂

Для различных напряжений смещения затвора как показано на диаграммах электронной плотности, электронная плотность HfO_2 значительно отличается от SiO₂. Она варьируется для SiO₂ от 5.5×10^{18} см⁻³ при U_g =0.6 В до 0.89×10^{18} см⁻³ при U_g =0 В, соответственно, и для HfO_2 от 25.4×10^{18} см⁻³ при U_g =0 В, соответственно.

На рис. 3 графически представлены для двух случаев подзатворного окисла и различных ЕОТ распределения электронной плотности вдоль канала в центре рабочей области при смещении стока 50 мВ.



Рис. За. Случай SiO₂, где 1 - t_{Si} =5 нм, EOT=1.5 нм; 2 - t_{Si} =3 нм, EOT=1.5 нм; 3 - t_{Si} =5 нм, EOT=1 нм; 4 - t_{Si} =3 нм, EOT=1 нм; 5 - t_{Si} =5 нм, EOT=0.8 нм; 6 - t_{Si} =3 нм, EOT=0.8 нм

Электронная плотность, x10¹⁸ см⁻³



Рис. 36. Случай HfO_2 где 1 - t_{Si} =5 нм, EOT=1.5 нм; 2 - t_{Si} =3 нм, EOT=1.5 нм; 3 - t_{Si} =5 нм, EOT=1 нм; 4 - t_{Si} =3 нм, EOT=1 нм; 5 - t_{Si} =5 нм, EOT=0.8 нм; 6 - t_{Si} =3 нм, EOT=0.8 нм

Хорошо видно, что электронная плотность достигает самых высоких значений в середине канала, а не на границах Si/SiO₂ и Si/HfO₂, где она фактически нулевая, потому что из-за барьеров SiO₂ и HfO₂, волновая функция имеет тенденцию стремиться к нулю на границах Si/SiO₂ и Si/HfO₂.

Для смещения затворного напряжения 0.5 В как показано на диаграммах электронной плотности, электронная плотность HfO_2 существенно отличается от SiO₂ и достигает максимального значения при $t_{Si} = 3$ нм. Электронная плотность SiO₂ для различных EOT варьируется от 6.7×10^{18} см⁻³ до 24.5×10^{18} см⁻³ для $t_{Si} = 5$ нм и от 16.2×10^{18} см⁻³ до 34.2×10^{18} см⁻³ для $t_{Si} = 3$ нм; электронная плотность HfO_2 для различных EOT варьируется от 32.0×10^{18} см⁻³ до 45.6×10^{18} см⁻³ для $t_{Si} = 5$ нм и от 34.0×10^{18} см⁻³ до 80.5×10^{18} см⁻³ для $t_{Si} = 3$ нм.

На рис. 4 представлены зависимости токов гипотетического «длинного» (по сравнению с выше рассмотренным, считая с что в исследуемом образце ККЭ полностью подавлены) транзистора n-типа в состояниях *ON* и *OFF* с учетом ЭРК и без него.



Рис.4. Зависимости токов I_{ON} (нечетные кривые) и I_{OFF} (четные кривые) без учета ЭРК (кривые 1 и 2) и с учетом ЭРК (кривые 3 и 4)

Без учета размерного квантования зависимость $I_{OFF}(t_{Si})$ определяется двухмерными эффектами, в то время как постоянное значение $I_{ON}(t_{Si})$ говорит о преобладающем вкладе одномерных эффектов, при которых пороговое напряжение не зависит от толщины t_{Si} . Проявление ЭРК для транзисторов ртипа будут выше, чем для электронных благодаря меньшему значению отношения масс для дырок [8, 12]. Со сдвигом порогового напряжения, обусловленного ЭРК, и определяемого выражением

$$\Delta U_{ih}^{QM} = \frac{s}{kT/q\ln(10)} \left(\frac{A}{(m/m_0)t_{Si}^2}\right), \quad \text{где s} \quad -$$

обратная величина крутизны ВАХ, m/m_0 – отношение эффективной массы носителей в направлении структурного ограничения к массе свободного электрона (т.е. 0,92 для электронов и 0,29 для дырок в направлении <100> Si), A – численная константа [8], ток транзистора в закрытом состоянии *OFF* теперь зависит, в дополнение к ККЭ, и от ЭРК. Вообще, ток нанотранзистора КНИ с ультратонким слоем Si в закрытом состоянии *OFF* всецело будет зависеть от ККЭ и ЭРК, каждый из которых зависит от толщины рабочего слоя.

Результаты моделирования, представленные на рис. 5, иллюстрируют зависимость уровня подпороговой крутизны от значения диэлектрической проницаемости.

S, мB/дек.



Рис. 5. Зависимость S(EOT), где 1 - $t_{Si} = 5$ нм, SiO₂; 2 - $t_{Si} = 3$ нм, SiO₂; 3 - $t_{Si} = 5$ нм, HfO₂; 4 - $t_{Si} = 3$ нм, HfO₂. Здесь и ниже геометрическими фигурками отмечены результаты расчетов, выполненных при помощи программы Nextnano, которые заимствованы из [10]

Для SiO₂ подпороговый наклон ухудшается почти на 400% при изменении толщины окисла от 0.8 нм до 5 нм для величины рабочей области 5 нм. В то время как для HfO2 подпороговый наклон ухудшается только 148 % для того же случая. Данные моделирования показывают также, что и при сокращении рабочей области до 3 нм те же тенденции сохраняются как для SiO₂ так и для HfO₂. Подпороговый наклон достигает минимального значения 90 мВ/дек при ЕОТ=1.5 нм. Можно сделать вывод, что подпороговая характеристика КНИ с подзатворным диэлектриком на основе HfO₂, значительно лучше, чем на SiO2, но остается далеким от рекомендации ITRS, поскольку величина подпорогового наклона не должена быть выше, чем 80 мВ/дек.

Из результатов, представленных на рис. 4, также следует, что транзисторы с толщиной рабочего слоя, масштабированной вплоть до значений 2 нм, могут полезными низковольтных оказаться для приложений. В противоположность этому. сравнительно малое возрастание в токе I_{ON} для толщин t_{Si}, больших 7 нм, означает, что лучшая производительность не может быть достигнута за счет увеличения толщины tsi. Для данного случая и номинального значения $t_{Si} = 7$ нм из рис. 4 следует, что 15%-ное возрастание t_{Si} будет приемлемо для возрастания тока I_{OFF} на один порядок.

В дополнение к отмеченным зависимостям с учетом ЭРК токи I_{OFF} и I_{ON} также зависят от «подвижности» носителей, которая для ультратонких слоев Si уменьшается с уменьшением толщины t_{Si} [1, 3, 9]. Данная деградация подвижности объясняется возрастанием скорости фононного рассеяния в пространственно ограниченных тонких слоях, а также рассеянием на поверхностных неоднородностях нерегулярностях. Данную И деградацию подвижности следует также отнести, как и проявление эффектов поперечного размерного квантования, к особенностям структурных ограничений носителей. Следует отметить, что, несмотря на деградацию, подвижность носителей в тонкослойных КНИ транзисторах оказывается выше, чем в объемных КМОП устройствах с таким же значением U_{th} [6, 7].

3. Физические ограничения, связанные с эффектом туннелирования

На границе области масштабирования КНИ технологии, необходимо учитывать ряд физических ограничений, возникающих из-за роста туннельных токов или токов утечки при переходе к устройствам ультра малых размеров [1-3]. Отметим два основных источника токов утечки: утечка в подзатворнам диэлектрике и утечка транзистора в состоянии OFF. Каждый ИЗ указанных источников играет определяющую роль при уменьшении характерных размеров. По отношению к другим источникам токов утечки ток утечки в подзатворном диэлектрике при масштабировании подзатворного диэлектрика связан с ограничениями на возможную толшину диэлектрического слоя (заметим, что данное значение превышает мономолекулярную величину 0,28 нм). При достижении толщины диэлектрика в несколько межатомных слоев затворный ток утечки будет доминирующим в состоянии OFF. При этом следует подчеркнуть, что ток утечки затвора незначителен по сравнению с током в состоянии ON.

На рис. 6 приведены численно рассчитанные зависимости прямого туннельного тока затвора (I_g) как функция напряжения на затворах (U_g) для двух типов подзатворного диэлектрика при U_{dd}=0.5 В. Величина исследуемого тока затвора планомерно возрастает при сокращении физической толщины окисла t_{ox}. А его уровень становится намного выше при переходе от SiO₂ к HfO₂. Для параметра t_{ox}, варьирующегося в диапазоне значений от 5 нм до 1 нм, максимальный ток затвора увеличивается от 9x10⁻⁹ A/м для HfO₂. При нулевом смещении затворов ток затвора увеличивается от 9x10⁻¹⁰ A/м до 4x10⁻⁴ A/м для SiO₂ и от 5x10⁻⁶ A/м для HfO₂.

На рис. 7 представлены зависимости $I_g(U_g)$ для $t_{Si} = 3$ нм и $U_{dd} = 0.5$ В. Аналогичные тенденции как для случая $t_{Si} = 5$ нм наблюдаются и в данных зависимостях.



Рис. 6. Зависимость тока затвора I_g от напряжения U_g для t_{Si} =5 нм, где левый - SiO₂, правый - HfO₂



Рис. 7. Зависимость тока затвора I_g от напряжения U_g для t_{Si} =3 нм, где левый - SiO₂ и правый HfO₂

Для физической толщины t_{ox} , варьирующаяся от 5 нм до 1 нм, максимальный туннельный ток увеличивается от 1×10^{-7} А/м до 8×10^{-2} А/м для SiO_2 и от 6×10^{-8} А/м до 1 А/м для HfO_2 . С точки зрения эквивалентной окисной толщины отмечаем более низкий ток утечки для HfO_2 по сравнению с SiO_2 . Учитывая, что ЕОТ для HfO_2 составляет только 18% от своей физической толщины, можно сделать вывод, что HfO_2 - перспективный кандидат по сравнению с SiO_2 для подавления тока утечки.

Из результатов моделирования можно сделать вывод, что, как для окиси кремния так и для окиси гафния, прямой туннельный ток затвора в диапазонах исследуемых толщин рабочей области и подзатворного диэлектрика, принимает минимальное значение на верхней границе обоих диапазонов толщин. Уменьшение размеров толщин приводит к нелинейному росту туннельного тока. При этом сохраняется возможность корректировать ток утечки, варьируя толщину рабочей области.

Заключение

В работе проанализированы не масштабируемые характеристики ультра коротких двух затворных КНИ нанотранзисторов с тонкой нелегированной рабочей областью для случая применения в качестве подзатворного диэлектрика затвора материалов с высоким значением диэлектрической проницаемости. Исследована степень влияния ряда технологических параметров и оксидов кремния и гафния на уровень проявления ККЭ и ЭРК на границе области масштабирования. Оксид гафния HfO2 значительно эффективен как для снижения уровня тока I_{off} и для максимизации отношения I_{on} /I_{off}.

Прогнозируется сильная зависимость времени отклика транзистора от напряжения питания в зависимости от толщины рабочей области. Большее значение задержки для толщины 3 нм ожидается изза более низкого значения тока I_{ON} в результате проявления ЭРК, которые более выражены для низких значений напряжения питания. Поэтому изменение толщины t_{Si} позволяет контролировать не только проявление ККЭ (а, следовательно, и тока I_{OFF}) в транзисторах, но и благодаря ЭРК, как

значение тока I_{ON} так и производительность. Для номинального значения толщины кремниевого слоя 5 нм приемлемые значения вариации параметра t_{Si} для контроля за током I_{OFF} в действительности больше, чем для контроля за производительностью И значение тока ION оказывается меньше. Однако для t_{Si} менее 5 нм проявление ЭРК благодаря структурным ограничениям становятся значительным и для диапазона напряжений питания меньше 1 В ухудшение скорости работы и ее чувствительность к толщине рабочей области уже недопустимы.

Литература

1. D.J. Frank, R.H. Dennard, E. Nowak. Device Scaling Limits and Their Application Dependencies // *Proc. IEEE.*- 89(3), 2001. p. 259-287

2. G. Roy, A. R. Brown, F. Adamu-lema, S. Roy, M. Asenov. Simulation study of individual and combined sources of intrinsic parameter fluctuations in conventional nano-Mosfet's *// IEEE Trans Electron Devices.*- No. 12(53), 2006. p. 3063-3070

3. F. Boeuf, M. Sellier, A. Farcy, T. Skotnicki An Evaluation of the CMOS Technology Roadmap From the Point of View of Variability, Interconnects, and Power Dissipation // *IEEE Trans Electron Devices.*- No. 6(55), 2008. p. 1433 – 1440

4. Н. В. Масальский. Масштабирование характеристик двух затворных КНИ нанотранзисторов // Труды НИИСИ.- №1(1), 2011. с. 16-19

5. International technology roadmap for semiconductor 2010 edition. Available from: (http://public.itrs.net)

6. T. Ernst, C. Tinella, C. Raynaud, S. Cristoloveanu. Fringing fields in sub-0.1 μ m fully depleted SOI MOSFETs: optimization of the device architecture // *Solid-State Electronics.*- Vol. 46, 2002. p. 373-378

7. B. Nikolic. Design in the Power-Limited Scaling Regime // *IEEE Trans Electron Devices.*-No. 1(55), 2008. p. 71 – 83

8. R. Difrenza, P. Llinares, G. Ghibaudo. The impact of short channel and quantum effects on MOSFET transistor mismatch // *Solid-State Electronics.*- Vol. 47, 2003. p. 1161-1167

9. D. Munteanu, J.-L. Autran, X. Loussier, S. Harrison, R. Cerutti, T. Skotnicki. Quantum short channel compact modeling of drain-current in Double-gate MOSFET // *Solid-State Electron.* – Vol. 50, 2006. p. 680-688

10. S. Birner, T. Zibold, T. Andlauer, T. Kubis, M. Sabathil, A. Trellakis, P. Vogl Nextnano. General Purpose 3-D Simulations // *IEEE Trans Electron Devices.*- No. 9(54), 2007. p. 1932-1938

11. Н. В. Масальский. Вопросы масштабирования характеристик КМОП СБИС // Успехи современной радиоэлектроники.- № 7, 2009. с. 3-27

12. M. Gilbert, D. Ferry. Efficient quantum threedimensional modeling fully depleted ballistic silicon-oninsulator MOSFET // *J. Applied Physics.*- Vol. 95, 2004. p. 7954-7961

The Impact of high dielectric permittivity of subgate dielectric on SOI CMOS nanotransistor characteristics

N.V. Masalsky

On the basis of numerical decisions characteristics of ultra short two gate SOI CMOS nanotransistors with thin undoped work area for an application case as subgate gate dielectric of materials with high value of dielectric permittivity are analyzed. The main attention is paid to the analysis of a level of manifestation shortly - channel effects and effects of sized quantization in not scalable of transistor characteristics.

Моделирование развития процессов «конвекция плюс теплопроводность» в воздухе вблизи процессора

А.Г. Мадера¹, М.Ж. Акжолов², И.Г. Лебо³

1 - доктор техн. наук. 2 - кандидат физ.-мат. наук, 3 - доктор физ.-мат. наук

(Работа выполнена при поддержке гранта РФФИ № 12-07-00076-а)

Дано описание физико-математической модели и двумерной программы «ENS_T» (Equation of Navies-Stocks with thermo conductivity) для численного решения уравнений Навье-Стокса с учетом теплопроводности. Представлены результаты численных расчетов нагрева и конвекции воздуха вблизи электронных устройств.

Введение

Вблизи поверхности электронного устройства с воздухом происходит передача тепла в газ, формируется переходной пограничный слой. Скорость отвода энергии будет существенно зависеть от конвективных процессов в газе и, возможно, теплопроводности. Влияние этого процесса возрастает с уменьшением масштаба неоднородности как ~ k^2 , где $k = 2\pi/L$ - волновое число, а L – размер неоднородности. Поэтому, в разрабатываемой авторами статьи программе был развит алгоритм расчета уравнения теплопроводности совместно с уравнениями газовой динамики и создана новая программа «ENS_T» (Equation of Navies-Stocks with thermo conductivity).

В настоящей статье представлены предварительные результаты компьютерного моделирования конвективных процессов в воздухе вблизи поверхности электронных устройств, выделяющих тепло.

1. Постановка задачи

Газодинамические процессы в воздухе вблизи нагревающей поверхности описываются системой уравнений Навье-Стокса (1):

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u}{\partial x} + \frac{\partial \rho v}{\partial y} = 0,$$

$$\frac{\partial \rho u}{\partial t} + \frac{\partial \rho u^{2}}{\partial x} + \frac{\partial \rho u v}{\partial y} + \frac{\partial \rho}{\partial x} =$$

$$\frac{\partial}{\partial t} \left\{ \mu \left\{ \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right\} + 2\mu \frac{\partial v}{\partial y} \right\} + \frac{\partial}{\partial y} \left\{ \mu \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right) \right\}$$

$$\frac{\partial \rho v}{\partial t} + \frac{\partial \rho u v}{\partial x} + \frac{\partial \rho v^{2}}{\partial y} + \frac{\partial p}{\partial y} + \rho g_{0} = \frac{\partial \partial v}{\partial y} \left\{ \mu A \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) + 2\mu \frac{\partial v}{\partial y} \right\} + \left\{ \mu \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right) \right\}$$

$$\frac{\partial \rho E}{\partial t} + \frac{\partial (\rho u E)}{\partial x} + \frac{\partial (\rho v E)}{\partial y} + \frac{\partial (\rho u)}{\partial x} + \frac{\partial (\rho v)}{\partial y} + \rho g_{0} v = \frac{\partial \partial v}{\partial x} \left\{ \mu A \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) + \mu \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right) + \mu \frac{\partial \partial v}{\partial x} \left(u^{2} + B \cdot J \right) \right\} + \frac{\partial \partial v}{\partial y} \left\{ \mu A \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) + \mu \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right) + \mu \frac{\partial \partial v}{\partial y} \left(v^{2} + B \cdot J \right) \right\} + \frac{\partial \partial v}{\partial x} \left\{ \kappa I \left(\frac{\partial T}{\partial x} \right) \right\} + \frac{\partial \partial v}{\partial y} \left\{ \kappa I \left(\frac{\partial T}{\partial y} \right) \right\}$$
(1)

где W = (u, v) - вектор скорости, J - удельная внутренняя энергия, Е- полная удельная энергия, ρ - плотность газа, А, В – коэффициенты вязкости. Для

идеального газа
$$A = \frac{2}{3}$$
, $B = \frac{\chi^2}{C_p}$, χ , C_p -

коэффициент температуропроводности и теплоёмкость газа g_0 - ускорение свободного падения, направленное строго вдоль оси *OY*, *T* - температура, κ - коэффициент

теплопроводности,
$$\chi = \frac{\kappa}{C_p \cdot \rho}$$

Система уравнений (1) замыкается уравнением состояния (2)

 $P = (\gamma - 1)\rho J = (\gamma - 1)\rho [E - (u^2 + v^2)/2]$ (2). γ – показатель адиабаты, *p* – давление.

Для решения уравнений (1-2) использовался метод крупных частиц, описанный в [1].

Постановка задачи была взята из [2], но расчеты были сделаны с учетом теплопроводности. Коэффициент

теплопроводности воздуха $\kappa = 0.026 \text{ Bt/(град·м)}$, $C_p =$ 1004.8 Дж/(кг град). Поскольку в постановке задачи в [2] имелась зеркальная симметрия относительно прямой, проходящей через центр области параллельно оси 0У, то в представленных ниже расчетах была взята левая половина счетной области. Размер этой области $L_x \times L_y = 0.09 \times 0.18$ м. На нижней границе расчетной области у правого края помещался тепло элемент, имеющий ширину 0.02 м. Он выделял мощность W= 5 Начальная плотность воздуха 1.225 кг/м³, а Bt. температура 300 град.К. В предыдущем варианте на верхней границе было задано условие второго рода, то есть W_{up} = 0. В рассматриваемом варианте задано условие первого рода Т_{ир} = 300 град. К Таким образом, в новом варианте часть тепловой энергии покидает систему. (В предыдущем варианте размер тепло элемента был задан 0.038 м, в новом – 0.04 м).

2. Результаты расчетов

В расчетах использовалась сетка 45х90 ячеек. Были выполнены расчеты с учетом влияния теплопроводности и гравитации (вариант а). Результаты расчетов сравнивались с данными, полученными без учета теплопроводности (вариант b).

На представленных ниже рисунках по оси абсцисс отложены номера ячеек счетной области.

На рис.1 даны распределения температуры на моменты времени t = 1 сек для вариантов (а) и (b).



Рис.1. Распределения температуры вдоль оси 0Y при x=0.09 м на момент времени t = 1 сек для вариантов (а) и (b).



Рис.2. Распределения температуры вдоль оси ОУ при x=0.09 м на момент времени t = 3.96 сек для вариантов (а) и (b).



Рис.3. Распределения плотности при x=0.09 м на момент времени 1 и 3.96 с в варианте b).

На рис.3, 4 показаны распределения плотности вдоль оси 0У при x=0.09 м на моменты времени t = 1 сек - - - (1) и 3.96 сек • • • (2) для вариантов (b) и (a).



Рис. 4. Распределения плотностей вдоль оси 0У при x=0.09 м на момент времени t = 1сек - - - (1) и 3.96 сек • • • (2) для вариантов (а).

На рис.5, 6 показаны поля температуры в области на моменты времени t = 1 (а) и 3.96 сек (b) в вариантах с учетом и без учета теплопроводности.









Рис.5. Поля температуры в варианте с учетом теплопроводности на моменты времени t = 1 (a) и 3.96 сек.



Рис.6. Поля температуры в варианте без учета теплопроводности при t = 1 сек (а) и при t = 3.96 сек.

На следующих рисунках (рис.7а,b,c) показано развитие «вихрей» в газе ($rot \vec{V}$)на различные моменты времени (a) t=0,0004c, b) t=0,03c, c) t=1,0c).





7c) t=1,00 c

Рис.7. Развитие поля вихрей на моменты времени 0.0004, 0.03 и 1 с.

В начальные моменты времени вихри формируются вблизи тепло элемента. Звуковые волны, отразившись от верхней и боковых стенок, двигаются в противоположном направлении и «интерферируют» с новыми зародившимися волнами. В результате, формируются вихревые структуры. К моменту времени t = 1 с поле вихрей распространилось по всей области счета.

3. Основные результаты

1. Разработана физико-математическая модель и создана двумерная программа «ENS_T» для решения уравнений газовой динамики с учетом вязкости и теплопроводности.

2. Проведены расчеты, моделирующие конвекцию и диссипацию энергии вблизи нагревающей поверхности электронного устройства.

3. Сделано сравнение с результатами расчетов, выполненных по программе «ENS_G» и показано, что на масштабах неоднородностей порядка 1 см и мощности теплоэлемента порядка 10 Вт процесс теплопроводности не дает существенного вклада, хотя из-за оттока тепла в варианте а) температура несколько ниже, чем в варианте b). С развитием вихревых течений, роль теплопроводности может возрасти.

4. Получены предварительные результаты по зарождению вихревых течений вблизи электронного нагревательного устройства в воздухе.

Литература

1. О.М. Белоцерковский, Ю.М. Давыдов. Метод крупных частиц в газовой динамике. // Москва, Наука, 1982

2. М.Ж. Акжолов, П.И. Кандалов, И.Г. Лебо, А.Г. Мадера. Компьютерное моделирование конвективных процессов в воздушной среде вблизи электронных устройств. Труды НИИСИ РАН, т.1, №2, 44-46, (2011)

The modeling of convection and thermo conductivity processes in the air near processing element

A.G. Madera, M.J. Akjolov, I.G. Lebo

Abstract. The description of physical-mathematical model and 2D code «ENS_T» (Equation of Navies-Stocks with thermo conductivity) for computer simulation of Navies-Stocks equations with allowance for thermo conductivity. We have demonstrated the results of Air heating and convection simulations near of electronic devises.

Компьютерное моделирование свободного конвективного теплообмена горизонтальных проволочных теплоотводов

П.И. Кандалов, А.Г. Мадера¹

1 – доктор технических наук, профессор

Работа выполнена при поддержке гранта РФФИ № 12-07-00076-а

Приведены результаты вычислительного эксперимента свободного конвективного теплообмена горизонтальной проволоки миллиметрового и субмиллиметрового диаметра с воздушной средой при числах Рэлея $10^{-8} < Ra < 10^2$. Получены численная и аппроксимирующая ее аналитическая зависимость в виде уравнения подобия Nu = f(Ra). Сравнение полученной аппроксимирующей аналитической зависимости с известными из литературных источников экспериментальными данными, показало их хорошее согласие.

При разработке и создании радиоэлектронных систем (РЭС) различного назначения получают распространение теплоотводы от электронных компонентов, выполненные из проволочных элементов миллиметрового и субмиллиметрового диаметра. Проволочные элементы используются в конструкциях петельнопроволочных радиаторов, миниатюрных теплообменников, в электронных модулях и других электронных компонентах. Теплоотдающая поверхность одного проволочного элемента при малых диаметрах чрезвычайно мала, однако, в случае высокой концентрации проволочных элементов в малом объеме, суммарная площадь теплообмена может достигать чрезвычайно высоких значений. В условиях жестких требований, предъявляемых к плотности упаковки РЭС, использование принудительной конвекции не всегда возможно, поэтому теплообмен путем свободной естественной конвекции, в условиях недостатка конструктивного пространства РЭС остается актуальным.

Для проведения расчетов проволочных теплообменников необходимо располагать методами вычисления коэффициента теплоотдачи с поверхности проволок различных диаметров в среду в условиях свободного конвективного теплообмена. Существующие на сегодняшний день аналитические зависимости [1 – 6] для расчета коэффициента теплоотдачи горизонтальных цилиндров, не могут быть использованы в практике теплового проектирования РЭС, поскольку либо охватывают сравнительно узкий диапазон чисел Рэлея $(Ra) Ra > 10^{-3} [5, 6]$, либо составляются из нескольких эмпирических аппроксимаций, сшиваемых между собой, причем без указания точности полученных зависимостей [1, 2], либо получены экстраполяцией экспериментальных данных за пределы, в которых эксперимент не проводился [1 – 4]. Вместе с тем при конструировании РЭС используются проволочные элементы, у которых число Рэлея изменяется в диапазоне 10-8 $< Ra < 10^2$. Для указанного диапазона отсутствуют надежные аналитические зависимости, пригодные как для практических расчетов коэффициента теплоотдачи проволочных элементов.

В работе была поставлена задача проведения исследований свободно конвективного теплообмена в воздушной среде (Pr = 0,72) вокруг горизонтальной проволоки с миллиметровым и субмиллиметровым диаметром и числом Рэлея (Ra), изменяющимся в диапазоне $10^{-8} < Ra < 10^2$. В качестве инструмента исследования использовался метод вычислительного эксперимента на компьютере с помощью системы автоматизации вычислительного эксперимента в области тепломассообмена и динамики жидкости ANES [8]. Выбор в качестве инструмента исследования метода вычислительного эксперимента обусловливается тем, что проведение натурных экспериментов в условиях свободного конвективного теплообмена чрезвычайно тонких проволочек с диаметрами, вплоть до субмиллиметровых, и аномально низкими значениями числа Рэлея, наталкиваются на практике серьезные трудно преодолимые препятствия. По результатам проведенного компьютерного моделирования получена зависимость между числом Нуссельта (Nu) и числом Рэлея в виде уравнения подобия Nu = f(Ra), или Nu = f(Gr Pr), которая с высокой точностью (менее 1%) была аппроксимирована аналитической формулой, удобной для применения при проведении практических расчетов тепловых режимов РЭС.

Тепловая модель объекта моделирования представляет собой бесконечно протяженный горизонтальный цилиндр того же диаметра, что и моделируемая проволока, расположенный вдоль оси z. На поверхности проволоки задана фиксированная температура, не изменяющаяся вдоль координаты z, поэтому поля температуры и скорости среды являются двумерными и изменяются только в плоскости поперечного сечения проволоки, с координатами x и y. Математическая модель, описывающая стационарный свободно конвективный теплообмен с поверхности проволоки в среду относительно переменных температурного поля T' и

95

поля давления в среде *p*', записанная в векторной форме имеет следующий вид [9]:

- уравнение Навье-Стокса

$$(\mathbf{v}\nabla)\mathbf{v} = -\nabla \frac{p'}{\rho} - g\,\beta\,T' + v\,\Delta\mathbf{v}\,,\qquad(1)$$

- уравнение переноса энергии в среде

$$a\nabla T' = a\Delta T', \qquad (2)$$

$$div \mathbf{v} = 0, \qquad (3)$$

 граничные условия: на поверхности проволоки задана нулевая скорость среды и фиксированная температура,

где $T = T_a + T'$, $p = p_0 + p'$, T_a и p_0 – средние значения температуры и давления в среде, от которых отсчитываются неравномерности температуры T' и давления p'; $\mathbf{v} = \mathbf{v}(x, y)$ – двумерное поле скорости среды; $v, a = \lambda/\rho c, \lambda, \rho, c, \beta, - \phi$ изические характеристики среды: кинематическая вязкость, температуропроводность, теплопроводность, плотность, удельная теплоемкость, температурный коэффициент расширения; g – ускорение свободного падения.

Решение математической модели (1) – (3) проводился с помощью автоматизированной системы компьютерного моделирования тепломассообмена и динамики жидкости ANES [8], путем непосредственного численного решения уравнений на компьютере. На внешней границе цилиндрической расчетной области, так же как и на границе проволоки, задавались фиксированные температуры с заданным перепадом (ΔT) между ними равным $\Delta T = 50$ °С, 100°С и 150°С. При проведении вычислений учитывалось изменение теплофизических характеристик среды от температуры. Расчетная область разбивалась в цилиндрической системе координат на контрольные объемы, причем в радиальном направлении располагалось 200 контрольных объемов с логарифмическим сгущением сетки к границе проволоки, а в направлении угловой координаты – 120 контрольных объемов. Уравнения математической модели (1) – (3) решались напрямую численными методами на компьютере без использования гипотезы о наличии пограничного слоя вокруг проволоки.

Результаты компьютерного моделирования (рис. 1, сплошная кривая с жирными точками) по моделированию числа Нуссельта $Nu = \alpha \cdot d/\lambda$ (α – коэффициент теплоотдачи; d – диаметр проволоки) в зависимости от числа Рэлея $Ra = \beta g d^3 \Delta T/va$, представленные в логарифмическом (логарифм десятичный) масштабе, показывают, что характер зависимости lg(Nu) от lg(Ra) существенно отличается от прямолинейного. Наилучшее приближение, имеющее точность меньшую 0,1%, дается следующей аналитической зависимостью (рис. 1, пунктирная линия возле сплошной кривой, соответствующей компьютерному моделированию):

$$lg(Nu) = 0,0066(lgRa)^{2} + 0,1266lg(Ra) - 0,0428.$$
 (4)



Рис 1. Зависимость числа Nu от числа Рэлея Ra, построенная в логарифмическом масштабе при $\Delta T = 100^{\circ}$ С: данные эксперимента [1, 2] – кружки; данные компьютерного моделирования, выполненного в настоящей статье – сплошная линия с жирными точками; аппроксимация полученных в статье данных посредством формулы (5) или (6) – пунктирная линия; существующие зависимости из литературных источников [5 – 7] – пунктирная линия с квадратиками

После несложных преобразований получим окончательную аналитическую зависимость числа Nu от числа Ra во всем диапазоне изменения числа Рэлея $10^{-8} < Ra < 10^2$:

$$Nu = 0.90615 \cdot Ra^{0,1266} \cdot Ra^{0,0066 \lg(Ra)}.$$
 (5)

Принимая во внимание соотношение между числами Рэлея, Грасгофа (Gr) и Прандтля (Pr), а именно Ra = Gr Pr, можно записать формулу (5) в виде функциональной зависимости Nu = f(Gr Pr):

$$Nu = 0.90615 \cdot (GrPr)^{0.1266} \cdot (GrPr)^{0.0066 \lg (GrPr)}.$$
 (6)

Таким образом, свободно конвективный теплообмен горизонтально расположенной проволоки со средой в диапазоне изменения числа Рэлея $10^{-8} < Ra < 10^2$ описывается уравнением подобия (5) или (6). Вычислив по формулам (5) или (6) число Нуссельта, далее из равенства $Nu = \alpha \cdot d/\lambda$ может быть найдено значение коэффициента теплоотдачи α с поверхности проволоки в среду. Для сравнения на рис. 1 нанесены также экспериментальные данные (отмечены кружками), приведенные в работах [1, 2].

Сравнение полученной в статье зависимости (5) или (6), построенной на рис. 1 (пунктирная линия), с экспериментальными данными (рис. 1, показаны кружками), приведенными в литературе [1, 2], показывает достаточно хорошее их совпадение, имеющее максимальную относительную погрешность не превышающую 16%.

Анализ полученной в настоящей работе зависимости коэффициента теплоотдачи α от диаметра проволоки *d* показывает, что с уменьшением диаметра, коэффициент теплоотдачи быстро возрастает, $\alpha \sim 1/d^{0,62-0,06lg} d$, что следует из формулы (6) и выражений для чисел Грасгофа $Gr = \beta g d^3 \Delta T/v^2$ и Нуссельта. Так, например, для свободного конвективного теплообмена в воздушной среде (Pr = 0,72) и диаметра проволоки d = 0,1мм ($Ra = 3,35 \cdot 10^{-3}$) коэффициент теплоотдачи составит $\alpha = 152$ Вт/м²К, но уже при d = 0,01мм ($Ra = 3,35 \cdot 10^{-6}$) получим $\alpha = 965$ Вт/м²К. Наблюдаемое при компьютерном моделировании явление роста коэффициента теплоотдачи с поверхности проволоки в среду с уменьшением ее диаметра объясняться тем, что с уменьшением диаметра проволоки, во-первых, возрастает плотность теплового потока q с ее поверхности в среду, что приводит к росту градиента температуры и, в результате – к более интенсивному конвективному переносу микрообъемов среды, прилегающих к поверхности проволоки, а, вовторых, концентрация поверхности проволоки, а, вовторых, концентрация поверхностных центров адгезии инородных частиц, всегда присутствующих в среде, уменьшается, так что образующийся на поверхности проволоки инородный слой убывает, а с ним и потери теплового потока с поверхности нагретой проволоки в среду.

Литература

- С.У. Чарчилл. Свободная конвекция около погруженных в жидкость тел / В кн. Справ. по теплообменникам: в 2 т. Т.1. – М.: Энергоатомиздат, 1987
- S.W.Churchill, H.S. Chu. Correlating equations for laminar and turbulent free convection from a horizontal cylinder Original Research Article // Int. J. of Heat and Mass Transfer, 1975, Vol. 18, P. 1049-1053
- 3. S. Nakai, T. Okazaki. Heat transfer from a horizontal circular wire at small Reynolds nd Grasgoff numbers I. Pure convection // Int. J. of Heat and Mass Transfer, 1975, Vol. 18, P. 387 396
- 4. D.C. Collis, M.J. Williams. Free convection of heat from fine wires // Aust. Res. Lab. Aero., Note 140, 1954
- 5. М.А. Михеев, И.М. Михеева. Основы теплопередачи. – М.: Бастед, 2010
- 6. Г.Н. Дульнев. Тепло- и массообмен в радиоэлектронной аппаратуре. – М. Высш. шк. 1984
- 7. С.С. Кутателадзе. Основы теории теплообмена. М.: Атомиздат, 1979
- В.И. Артемов, А.Г. Муров, В.И. Шиков, Г.Г. Яньков. Система автоматизации численного эксперимента ANES: идеология и архитектура - // Препринт №8-247.- М.: ИВТ АН СССР, 1988.
- 9. Л.Д. Ландау, Е.М. Лифшиц. Гидродинамика. М.: Наука, 1992

Computational modeling of free convective heat transfer from a horizontal wire at a small Rayleigh numbers

P.I. Kandalov, A.G. Madera

Abstract. The results of computer simulation of free convective heat transfer from a horizontal wire both millimeter and sub-millimeter diameter at a small Rayleigh numbers ($10^{-8} < \text{Ra} < 10^2$), as well as the numerical and approximating its analytical dependence by the similarity equation Nu = f (Ra), are obtained. Comparing approximating analytic dependence with known from of literary sources the experimental data are in good agreement. The advantage of the proposed dependence in comparison with existing ones, are shown.

Два алгоритма фильтрации, сглаживания и выделения тренда зашумленных больших временных рядов: Алгоритм 1

М.Л. Бахмутский

Кандидат физико-математических наук

Рассмотрены особенности применения сингулярного спектрального анализа для выделения тренда и сглаживания больших, порядка десятков тысяч значений, временных рядов. Для решения этой задачи, предлагается сочетать сингулярный спектральный анализ с применением к исходному временному ряду (апгоритм 1) или траекторной матрице (алгоритм 2) вейвлет-пакетного разложения. Этот подход иллюстрируется на примерах временных рядов из десятков тысяч значений давления, полученных в ходе экспериментов по гидродинамическому прослушиванию нефтяных скважин на месторождениях, а также некоторых численных экспериментов.

1. Сингулярный спектральный анализ

При зашумпенных изучении сильно нестационарных временных рядов в настоящее время начинает достаточно широко применяться метод сингулярного спектрального анализа (SSA) [1]. Метод SSA основан на построении по исходному временному ряду т.н. траекторной матрицы и ее аппроксимации при помощи сингулярного разложения матриц. Поэтому его можно рассматривать как исследование временного ряда хорошо известным в статистике методом главных компонент (другое название метод естественных функций). ортогональных Для построения траекторной матрицы, временной ряд $\{x_1, ..., x_M\}$ последовательно, со сдвигом на один отсчет, проходится окном размером т-отсчетов и строится множество векторов задержки, размерности т $\dot{\omega}_i = (x_i, x_{i+1}, \dots, x_{i+m-1})^T$ [2], которые составляют траекторную матрицу размерностью m*n

$$A = \begin{pmatrix} x_1 & x_2 & x_3 & & x_n \\ x_2 & x_3 & x_4 & \dots & x_{n+1} \\ x_3 & x_4 & x_5 & & x_{n+2} \\ \vdots & & \ddots & \vdots \\ x_m & x_{m+1} & x_{m+2} & \dots & x_N \end{pmatrix}$$
(1)

Здесь $X_1, X_2, ..., X_N$ – значения ряда в моменты времени t=1,2,...,N.

$$n=N-m+1, m \leq \left[\frac{N+1}{2}\right].$$

Для траекторной матрицы находится ее сингулярное разложение, т.е. матрица А представляется в виде :

 $A=U\sum V^T$, где U –матрица размерности m*n с ортонормальными столбцами, U ={ u_p }, V={ v_p }матрица размерности n*n с ортонормальными столбцами и $\sum = diag (\sigma_p)$ – матрица размерности m*n с неотрицательными диагональными элементами (сингулярными числами). Матрица аппроксимируется при помощи г - троек { δ_p, u_p, v_p }, отвечающих г наибольшим сингулярным числам .Иными словами матрица A заменяется матрицей \tilde{A} , где:

 $\widetilde{\boldsymbol{A}} = \boldsymbol{\sigma}_1 \ast \boldsymbol{E}_1 + \boldsymbol{\sigma}_2 \ast \boldsymbol{E}_2 + \dots + \boldsymbol{\sigma}_r \ast \boldsymbol{E}_r \,, \quad (2)$

 $E_{i} = u_{i} \times v_{i}^{T}$ – внешнее произведение i-го столбца матрицы U и соответствующего столбца матрицы V. E_{i} –простая матрица ранга 1 с единичной нормой. Временной ряд восстанавливается путем диагонального усреднения матрицы \tilde{A} , т.е. матрицы в которой оставлены только г гладких компонент.

$$\widetilde{\widetilde{X}_{j}}^{=} \begin{cases} \frac{1}{j} * \sum_{i=1}^{j} \widetilde{a}_{i,j-i+1, -1 \leq j \leq m} \\ \frac{1}{m} \sum_{i=1}^{m} \widetilde{a}_{i,j-i+1, -m} & m \leq j \leq n \\ \frac{1}{N-j+1} \sum_{i=1}^{N-j+1} \widetilde{a}_{i+j-n,n-i+1, -m \leq j \leq N} \end{cases}$$

$$(3)$$

Если параметр m выбран таким, что m *«п* то тогда можно решить спектральную задачу для матрицы

$$C^{A*}A^{T} = U^{*}(\Sigma^{*}\Sigma^{T})^{*}U^{T}.$$

Матрица U является матрицей собственных векторов матрицы C, а сингулярные числа матрицы A являются положительными корнями собственных чисел матрицы

С.Умножив скалярно матрицу U^T на A , получим матрицу главных компонент

$$Y = U^{T} * A = \begin{pmatrix} Y_{1} \\ Y_{2} \\ \vdots \\ Y_{m} \end{pmatrix}, r = 1, 2, ..., m$$
строки

длины n

Тогда $\tilde{A} = \sum_{k=1}^{r} u_i \times Y_i$ и временной ряд восстанавливается по формуле (3). Этот подход

довольно широко используется в задачах нахождения периодических зависимостей во временных рядах наблюдений, фильтрации шумов и сглаживания временных рядов, т.к. использование сингулярного (спектрального) разложения матрицы позволяет выделить наиболее значимые составляющие ряда и отсеять случайные возмущения. Однако при непосредственном применении этого подхода к рядов обработке очень больших наблюдений возникают значительные трудности. В изложенном методе SSA есть два существенных параметра. Эти параметры – размерность окна, т.е. размерность траекторной матрицы и количество сингулярных троек главных используемых (компонент), лля аппроксимации траекторной матрицы. Количество используемых главных компонент, в нашей задаче в крайнем случае можно определять интерактивно. Выбор длины скользящего окна, т.е. размерности траекторной матрицы определяется основной идеей метода SSA, т.е. соображением о том, что если исходный временной ряд имел какую-то структуру, то и его отрезки (столбцы траекторной матрицы) имеют такую же структуру. Если стоит задача выделения составляющих с известным периодом, то размер окна равен или кратен этому периоду. Поэтому в задаче о выделении глобального тренда и сглаживании ряда рекомендуемый размер окна равен максимально возможному и примерно равен половине длины ряда. При ЭТОМ выбор размера движущегося окна оказывается очень существенным. Результаты численного эксперимента, иллюстрирующие влияние выбора размера окна приведены на рис.2 и рис.3. На рис.1 изображен временной ряд, полученный в ходе реального эксперимента по гидродинамическому прослушиванию скважин.



Рис.1 Исходный временной ряд 130816 значений давления, измереных с интервалом 1 минута



На рис 2. и рис.3 приведены результаты сглаживания (белая линия) исходного временного ряда (черная

линия) с использованием окон в 100 и 1000 отсчетов соответственно.

Рис.2 Исходный ряд (черная линия) и его сглаживание(светлая линия) методом SSA с движущимся окном в 100 отсчетов



Рис.3 Исходный ряд и его сглаживание методом SSA с движущимся окном в 1000 отсчетов



Рис. 4. Сглаживание исходного ряда методом SSA (сплошная черная линия - окно 100 отсчетов, сплошная светлая – окно 500 отсчетов, светлые точки – 5000 отсчетов)

Ha рис.4 приведено сопоставление результатов сглаживания (тренд) для различных окон (100,500 и 5000 отсчетов). На рис.5 приведены результаты отсчетов нахождения тренда при выборе окна в 2500 и 5000. Из этих расчетов следует, что для надежного и корректного нахождения тренда временного ряда в общем случае, необходимо использовать траекторную матрицу с максимально возможной длиной столбца, равной примерно половине длины ряда. Заметим, к тому же, что вычисление сингулярного разложения порядка 5000*5000 матрицы размерности на компьютере с CPU Intel Core i5-750 (2.67GHz) требует недопустимо большое время счета. Это вполне понятно, так как вычисление SVD - разложения матрицы размерности m*n требует порядка

$$4*n^2 * m - \frac{4}{3} * n^3 + O(n^2)$$
 операций [3].

Отсюда следует, что непосредственно применять метод SSA к выделению тренда и сглаживанию больших временных рядов не удается, т.к. мы не можем вычислить сингулярное разложение матриц размерности порядка 65000*65000 элементов. Во всяком случае, для компьютеров традиционной последовательной архитектуры. Использовать же часто применяемый алгоритм усреднения в каком *либо* окне с последующей аппроксимацией полиномом по методу наименьших квадратов также невозможно. Легко видеть (рис.1),что полезная аномалия часто меньше амплитуды шума и неудачный выбор окна или степени полинома недопустимо исказит искомый



Рис.5. Сглаженный тренд по методу SSA - сплошная линия - окно 2500 отсчетов, пунктир - окно 5000 отсчетов

полезный сигнал. Алгоритма же правильного выбора небольшого размера окна и степени полинома нет. Поэтому необходимо применить метод SSA выбирая максимально возможный размер окна. Для того чтобы преодолеть трудности связанные с огромным размером траекторной матрицы предлагается сочетать метод SSA и применение вейвлет-пакетов к исходному временному ряду.

2. Первый алгоритм обобщения метода SSA на случай больших временных рядов.

Пусть у нас имеется большой, порядка десятков тысяч отсчетов, временной ряд длины N. Будем полагать, без потери общности, что N= $2^k * N_1$.Пусть N=2*M. Вместо временного ряда

X={ $x_1, x_2, x_3, x_4, \dots, x_{2*M-1}, x_{2*M}$ } рассмотрим ряд Z, состоящий из двух временных рядов S и D. Z={S|D}, где S={ s_1, s_2, \dots, s_M } и D={ d_1, d_2, \dots, d_M }.

Переход от исходного ряда X к ряду Z происходит следующим образом:

$$S_{i} = \frac{x_{2*i-1} + x_{2*i}}{\sqrt[2]{2}}; \ d_{i} = \frac{x_{2*i-1} - x_{2*i-1}}{\sqrt[2]{2}}$$

Обратное преобразование от S и D к X имеет следующий вид:

$$x_{2*i-1} = \frac{s_i + d_i}{\sqrt[2]{2}}; \quad x_{2*i} = \frac{s_i - d_i}{\sqrt[2]{2}},$$

i= 1, 2, ..., M (5)

Компоненты ряда S представляют собой локально сглаженные, усредненные значения, а компоненты D – уточняющие коэффициенты для элементов ряда X.

Фактически используя (4), мы раскладываем ряд Х по вейвлетам Хаара [4].

Преобразования (4), (5) являются взаимнооднозначными и, как легко убедиться непосредственно, сохраняющими нормы. Т.е. если

$$||X||^{2} = \sum_{i=1}^{2*M} (x_{i})^{2},$$

$$||S||^{2} = \sum_{i=1}^{M} (s_{i})^{2}, ||D||^{2} =$$

$$\sum_{i=1}^{M} (d_{i})^{2}, TO$$

$\|X\|^2 = \|S\|^2 + \|D\|^2.$

Таким образом пренебрежение какой либо величиной в преобразованном временном ряде означает (после обратного преобразования) пренебрежение такой же малой величиной в исходном. На следующем этапе применим введенное нами преобразование к рядам S и D T.e. от ряда Ζ перейлем к ряду $Z^{(2)} = (SS)$ DSISD **DD**). ПО ТЕМ ЖЕ

правилам, что и при переходе от ряда Х к ряду Z.

$$ss_i = \frac{s_{2*i-1} + s_{2*i}}{\sqrt[2]{2}}; ds_i = \frac{s_{2*i-1} - S_{2*i}}{\sqrt[2]{2}}; sd_i = \frac{d_{2*i-1} + d_{2*i}}{\sqrt[2]{2}}; dd_i = \frac{d_{2*i-1} - d_{2*i}}{\sqrt[2]{2}}$$

И обратный переход:

$$s_{2*i-1} = \frac{ss_i + ds_i}{\sqrt[2]{2}}; \ s_{2*i} = \frac{ss_i - ds_i}{\sqrt[2]{2}}; \ d_{2*i-1} = \frac{sd_i + dd_i}{\sqrt[2]{2}}; \ d_{2*i} = \frac{sd_i - dd_i}{\sqrt[2]{2}},$$

здесь i=1, 2, ..., M_1 , $M_1 = \frac{M}{2} = \frac{N}{4}$.Все

то, что говорилось о переходе от ряда X к ряду Z, справедливо и для перехода от ряда Z к ряду $Z^{(2)}$.

Ряд SS является временным рядом локально сглаженных, усредненных значений элементов ряда S, а ряд DS состоит из уточняющих коэффициентов. Аналогичное справедливо и относительно рядов SD и DD. При этом ряд SS описывает самую гладкую часть исходного временного ряда. Сохраняя взаимно – однозначное соответствие, мы перешли от исходного временного ряда к ряду,состоящему из двух временных рядов (гладкой части и поправок), а затем к ряду состоящему из 4-х временных рядов (гладкой части и поправок различной гладкости). Последовательно, k раз, применяя этот прием, называемый вейвлет- пакетным разложением[4],

получаем временной ряд состоящий из 2^k временных рядов, каждый из которых имеет длину N_1 . Для каждого из этих 2^k временных рядов, построим свою траекторную матрицу (1) размерностью m*n, где

$$m = \left[\frac{(N_1+1)}{2}\right], n = N_1 - m + 1$$
. Вычислим

сингулярные разложения этих матриц. Заметим, что эта часть алгоритма может быть эффективна выполнена как на традиционных последовательных так и на параллельных системах. Анализируя сингулярное разложение траекторной матрицы наиболее гладкой части, определяем количество сингулярных троек (или главных компонент) описывающих тренд и, тем самим определяем тот порог ниже которого пренебрегаем сингулярными числами всех 2^k траекторных матриц. После этого формируем аппроксимации (2) траекторных матриц, по формулам (3) восстанавливаем временные ряды, и, последовательно применяя (5) получаем сглаженную версию исходного временного ряда. Обратимся к первому нашему примеру. Исходный ряд состоит из 130816 отсчетов,

т.е. N=130816, k=8, **N**₁=511. На 8 – м уровне вейвлет – пакетного разложения мы имеем 256 временных ряда длиной 511 значений, каждый из которых соответствует (как гладкая часть и поправочные коэффициенты) всему исходному временному ряду. На рис.6 приведен график гладкой части вейвлет разложения, на рис.7 следующий временной ряд поправочных коэффициентов.Легко видеть резкое влияние шума на этом графике, хотя закономерности исходного графика(рис.1) также просматриваются. Заметим, что рис.6 внешне похож на один из графиков на рис.4 (соответсвующему окну в 100 отсчетов), что неудивительно поскольку последовательное применение первой формулы преобразования (4) можно рассматривать как своеобразное усреднение с окномв 256 отсчетов.



Рис.6. Самая гладкая часть 8 - го уровня вейвлет - преобразования исходного временного ряда



Рис.7 Следующий после самой гладкой части массив поправочных коэффициентов 8 - го уровня вейвлет-преобразования

На рис.8 приведен график значений 20 максимальных сингулярных чисел

На рис.9 приведены графики первых 20 сингулярных чисел траекторных матриц 2 – го, 3 –го, 4 –го и 256ряда вейвлет-коэффициентов.



Рис.8 Сингулярные числа траекторных матриц 1-го (сплошная линия) и 2 – го (пунктир) временного ряда вейвлеткоэффициентов



Рис.9. Первые 20 сингулярных чисел траекторных матриц 2 –го (ромб), 3 –го (круг), 4 – го (треугольник), 256 – го (пунктир и квадрат) ряда вейвлет – коэффициентов

Предположим, что в разложении (2) мы решили ограничиться только первым членом.Тогда, отбрасывая все члены в разложении (2) для всех рядов коэффициентов, сингулярные числа которых меньше использованного и k раз проделав обратное вейвлет преобразование получим тренд (светлая линия) на (2) рис.10.Добавляя в разложение матрицы последовательно повышаем новые члены, МЫ

детальность воспроизведения временного ряда. На рис. 11 приведен график первого ряда вейвлет – коэффициентов, восстановленного с использованием 10 первых членов в (2). А на рис.12, где приведен точно тот же график, но восстановленный с помощью 11 первых членов (2), ясно видна появляющаяся шумовая компонента.



Рис. 10. Исходный ряд (черная линия) и тренд, найденый с использованием 1-й сингулярной тройки



Рис.11. Тренд первого ряда вейвлет - коэффициентов. Построен при использовании 10 членов разложения (2)



Рис.12. Тренд первого ряда вейвлет - коэффициентов. Построен при использовании 11 членов разложения (2)

Таким образом, для выделения тренда исходного временного ряда, восстанавливаем первый ряд вейвлет – коэффициентов при помощи 10 членов разложения (2), во всех других рядах вейвлет – коэффициентов в разложениях типа (2) оставляем только те члены, сингулярные числа которых не меньше 10 – го сингулярного числа из SVD –разложения первого ряда вейвлет – коэффициентов. По формуле (3) восстанавливаем ряды и, проделав обратое вейвлет – преобразование получим сглаженный тренд. На рис 13 приведен этот тренд.



Рис.13. Исходный ряд (черная линия) и тренд, найденый с использованием 10 сингулярных троек(10 членов аппроксимации (2))



Рис.14. Исходный ряд (черная линия) и тренд, найденый с использованием 11 сингулярных троек.Видно появление шумовой компоненты

Рис. 14 (11 членов в аппроксимации матрицы) и рис.15 (50 членов в аппроксимации) иллюстрируют тот факт, что дальнейшее повышение числа членов в (2), ведет к повышению точности аппроксимации временного ряда и все большему появлению шумовых компонент.

Сравним стоимость SVD разложения траекторной матрицы исходного ряда, в предположении равенства длины движущегося окна половине длины ряда, и 2^k траекторных матриц предлагаемого алгоритма.



Рис.15. Исходный ряд (черная линия) и тренд, найденый с использованием 50 сингулярных троек. Шумовая компонента усиливается

Стоимость SVD разложения матрицы размерности m*n примерно равна

$$4 * n^2 * m - \frac{4}{3} * n^3 + O(n^2)$$
 операций [3].

Если длина движущегося окна примерно равна половине ряда, то $n \approx m \approx 2^{k-1} * N_1$ и стоимость SVD разложения

$$pprox rac{2^{2k}}{3} * N_1^3 + O(2^{2k-2} * N_1^2)$$
 операций. Для

нашего алгоритма, стоимость SVD разложения 2^k траекторных матриц с $n \approx m \approx \frac{N_1}{2}$ примерно

равна: $2^k * \left\{ \frac{N_1^2}{3} + O(\frac{N_1^2}{4}) \right\}$ операций. Таким

образом, стоимость вычисления SVD разложения при использовании предлагаемого алгоритма уменьшается в 2^{2k} раза.

Заключение

Сингулярный Спектральный Анализ в сочетании с вейвлет – пакетным разложением может служить реальным инструментом корректного выделения тренда и сглаживания сильно зашумленных больших временных рядов. Однако в изложенном здесь алгоритме применено пороговое отбрасывание всех

сингулярных троек сингулярные числа которых меньше некорого значения. При этом могут быть отброшены скрытые периодические и квазипериодические слабые сигналы, сингулярные числа в разложении которых также меньше пороговых. При этом мы лишены возможности подвергнуть соответствующие сингулярные вектора дополнительному анализу ,т.к. каждый из наших рядов вейвлет коэффициентов отвечает всему временному ряду в своем масштабе. При этом возникают трудности выделения полезного сигнала при отношении сигнал/шум меньше единицы. Эти трудности могут быть обойдены если из исходного сигнала вычесть найденный при помощи этого алгоритма тренд и анализировать оставшуюся часть. Второй алгоритм предлагает более оптимальный путь решения задачи.

Литература

1 R. Vautard, P. Yiou, M. Ghil. Physica D., 1992, v.58, p. 95 – 126.

2 D.S. Broomhead, G.P. King. Physica D., 1986, v.20, p 217 – 236

3 Дж. Деммель, Вычислительная линейная алгебра, М., Мир, 2001г.

4 Э. Столниц, Т. ДеРоуз, Д. Салезин, Вейвлеты в компьютерной графике, Ижевск, НИЦ «Регулярная и хаотическая динамика», 2002г.

Two algorithm denoising, smoothing and trend identify for large noisy time series: Algorithm1.

M.L. Bakhmutsky

Abstract: The singular-spectrum analysis main specific feature for large nonstationary time series are provided. It is shown that the trends large time series can be determined by wavelet-packet transformation of such series and use singular-spectrum analysis for wavelet coefficient (algorithm1.). It is shown that specific features for large nonstationary noisy time series can be determined by the wavelet – packet transformation of its large trajectory matrix (algorithm2).

LLVM как инфраструктура разработки компиляторов для встроенных систем

Н.И. Вьюкова, В.А. Галатенко¹, С.В. Самборский

1 – доктор физико-математических наук

В статье представлен обзор инфраструктуры LLVM (Low-Level Virtual Machine) и ее оценка с учетом особенностей разработки компиляторов для микропроцессоров, используемых во встроенных системах. Рассматривается общая структура LLVM, промежуточное представление программы, средства генерации кода, средства описания целевой архитектуры.

1. Введение

К встроенным системам обычно предъявляются определенные требования по производительности при значительных ограничениях на энергопотребление и размеры; к ним также могут предъявляться повышенные требования сбоеустойчивости. Разработка аппаратных средств для встроенных систем зачастую предполагает расширение системы команд инструкциями, специфическими для типичных задач данной предметной области, при этом, по результатам экспериментальных исследований, система команд может неоднократно меняться с целью достижения указанных выше характеристик.

Как правило, в микропроцессорах для встроенных систем не используются средства, требующие сложных решений, такие как аппаратное аппаратных планирование потока команд, ротация регистрового файла, предикатное исполнение. В то же время они имеют высокий потенциал параллелизма на уровне команд, который на практике бывает довольно сложно эффективно использовать. В связи с этим целесообразным представляется применение точных методов оптимизации кода (см. [1,2,3]), использующих методы математического программирования. Все это предъявляет определенные требования к компилятору к инфраструктуре, в рамках которой ОН И разрабатывается. Она должна включать:

- удобные средства описания целевой архитектуры для генерации кода, его ассемблирования и дизассемблирования;

- внугреннее представление, достаточное для описания семантики программ и удовлетворяющее потребностям анализа, оптимизации, генерации кода;

- простые возможности для добавления новых или альтернативных проходов анализа и оптимизации кода;

- возможности подключения внешних программных средств, таких как системы решения задач математического программирования.

Данная работа посвящена анализу инфраструктуры построения компиляторов LLVM с учетом перечисленных выше требований. Рассматриваются основные аспекты LLVM, существенные с точки зрения проблемы генерации кода для встроенных систем: внутреннее представление программы LLVM, структура генератора кода, средства описания целевых платформ, а также существующие в настоящее время средства, обеспечивающие интерфейс к внутренним структурам LLVM на языке Python.

2. Общий обзор LLVM

Работа над проектом LLVM была начата в 2000 г. в университете Иллинойса. Целью проекта было создание гибкой инфраструктуры, позволяющей разрабатывать различные, как экспериментальные, так и промышленные средства анализа, оптимизации, трансформации и компиляции программ [4, 5]. Внутренняя структура GCC (Gnu Compiler Collection), начало которого относится к 80-м годам прошлого века, к тому времени стала слишком громоздкой, неудобной, трудно поддающейся для введения новых или альтернативных проходов анализа и оптимизации кода. Довольно сложными и преимущественно процедурными являются средства описания целевых микропроцессорных архитектур в GCC. Хотя разработчики постоянно предпринимают усилия по совершенствованию внутренней организации GCC, многие изначально заложенные подходы затрудняют решение указанных проблем. К числу причин такого положения можно отнести использование языка Си, «монолитность» кода, применение большого числа макросов и глобальных переменных, не всегда удачные структуры данных.

В [6] приводятся примеры, которые дают представление о степени компактности и простоты понимания программного кода LLVM в сравнении с GCC. В этих примерах сравниваются фрагменты кода этих двух инфраструктур, реализующие одну и ту же функциональность.

1) Генерация одной команды внутреннего представления:

LLVM: Value *B = IRB.CreateLoad(IRB.CreateIntToPtr(A, Ty));

GCC: t = build1 (INDIRECT_REF, shadow_type, build1 (VIEW_CONVERT_EXPR, shadow_ptr_type, t)); t = force_gimple_operand (t, &stmts, false, NULL_TREE); gimple_seq_add_seq (&seq, stmts); shadow_value = make_rename_temp(shadow_type,""); g = gimple_build_assign (shadow_value, t);
2) Установка признака **noreturn** для вызова функции, которая не возвращает управление:

LLVM: Call->setDoesNotReturn();

GCC: TREE_THIS_VOLATILE (call) = 1;

В 2010 г. начались подготовительные работы по переходу к С++ как к языку разработки GCC. В 2012 vжe официально объявлено. что лля сборки последующих версий GCC будет использоваться g++, разработка компилятор новой и функциональности также будет вестись на С++; постепенно также будет осуществляться перенос существующего кода на С++.

В проекте LLVM, с учетом опыта GCC, а также методов современного компиляторостроения, изначально были заложены решения, обеспечивающие хорошо структурированную, гибкую внутреннюю организацию.

1) Понятие абстрактной низкоуровневой виртуальной машины (Low-Level Virtual Machine -LLVM) и ее язык, составляющие идейную основу, вокруг которой выстроен весь набор инструментов LLVM. Внутреннее представление входной программы это запись ее семантики на языке виртуальной машины. Отличительной чертой внутреннего языка LLVM является низкоуровневый набор инструкций в сочетании с высокоуровневой системой типов данных, что позволяет в компактной форме отражать семантику программ на различных языках, эффективно проводить анализ и оптимизирующие преобразования программ, генерировать машинный код и отладочную сравнения, информацию. промежуточное Для представление GIMPLE, используемое в GCC, не является самодостаточным, поскольку при генерации DWARF информации отладочной компилятор обращается к представлению операндов в форме AST (в виде синтаксическое дерева).

2) Трехфазная структура компилятора, обеспечивающая поддержку множества входных языков и множества целевых платформ.

LLVM в настоящее время используется для множества языков, в частности, языков, поддерживаемых GCC, а также .NET, Python, Ruby, Scheme, Haskell, D и др. В качестве целевых архитектур поддерживаются X86, X86_64, Sparc, ARM, PowerPC, MIPS, Cell, PTX, XCore, MBlaze.

Хотя трехфазная структура считается классической и описана едва ли не в любом учебнике по построению компиляторов, на практике она далеко не всегда бывает реализована в полной мере. Так, разработчики, пытавшиеся использовать frontend'ы GCC для целей анализа, рефакторинга и т.п., оказывались перед необходимостью включать в приложение практически весь код GCC.

3) Использование языка С++ и разумных стандартов кодирования, что обеспечивает высокий уровень абстракции, компактность и читаемость кода. Достаточно подробная документация, как внешняя, написанная вручную, так и внутренняя, сгенерированная при помощи Doxygen.

4) Развитые, преимущественно декларативные (в отличие от GCC), средства описания целевых платформ, которые позволяют задавать информацию как для генерации кода, так и для его ассемблирования и дизассемблирования.

5) Представление всех видов анализа и преобразования кода в виде проходов компилятора.

6) Модульная инфраструктура оптимизатора, возможность произвольно комбинировать проходы анализа и преобразований.

7) Набор средств для отладки и тестирования, включающий графическую визуализацию внутреннего представления, инструмент генерации для минимального тестового кода, на котором воспроизводится ошибка (bugpoint), инструмент для генерации случайного кода BO внутреннем представлении (llvm-stress), инструмент выборки файла заданной функции ИЗ с внутренним представлением и др.

8) Организация LLVM в виде набора библиотек, которые могут использоваться при построении как базового, стандартного набора инструментов, так и специфических инструментов, создаваемых сторонними группами разработчиков. Ниже перечислены основные инструменты, входящие в состав комплекта Clang-LLVM.

clang – компилятор с языков С, С++, Objective С или Objective C++.

орt – оптимизатор внутреннего представления LLVM.

llc – генератор машинного кода.

11і – интерпретатор внутреннего представления.



llvm-as – утилита, преобразующая внутреннее представление из текстового формата в бинарный.

llvm-dis – утилита, преобразующая внутреннее представление из бинарного формата в текстовый.

llvm-ld – утилита, позволяющая скомпоновать файлы внутреннего представления (с подключением библиотек) в один файл внутреннего представления или в выполняемый файл.

llvm-prof – печать данных профилирования.

llvm-link – утилита, позволяющая скомпоновать несколько файлов внутреннего представления в один.

llvm-ar – архиватор файлов внутреннего представления.

llvm-diff – структурное сравнение модулей во внутреннем представлении.

llvm-stress – генератор случайных файлов внутреннего представления (для тестирования).

bugpoint – утилита, позволяющая автоматически сократить тест, на котором происходит ошибка.

llvm-extract – утилита для выборки заданной функции из файла внутреннего представления.

LLVM приобрел признание и популярность как среда для разработки экспериментальных и промышленно используемых инновационных методов анализа и оптимизации, а также инструментальных средств разработки для встроенных систем (см. [7]).

3. Внутреннее представление LLVM

3.1. Общая характеристика

Внутреннее представление программы LLVM представляет собой последовательность меток и низкоуровневых команд над значениями, принадлежащими к высокоуровневому набору типов, включающих элементарные типы (целочисленные, вещественные, метки) и производные типы, такие как указатели, массивы, векторы и структуры, а также функции. Поддерживается RISC-подобный набор команд, в основном, трехадресных, имеющих два аргумента и один результат.

Низкоуровневость набора команд внутреннего представление дает такие преимущества, как компактность возможность кода. производить широкий Применение спектр трансформаций. высокоуровневой информации о типах обеспечивает возможности для глубокого межпроцедурного анализа и интенсивных оптимизаций, в том числе, на стадии компоновки, выполнения, или реоптимизации в периоды простоев системы.

Внутреннее представление может храниться в трех различных формах: в виде данных в оперативной памяти, в виде биткода на диске и в текстовом виде, пригодном для чтения и анализа программистом; все три формы семантически эквивалентны друг другу.

3.2. Структура модуля

Внутреннее представление программы в LLVM состоит из модулей. Модуль представляет одну единицу трансляции программы на исходном языке. Модуль состоит из определений функций,

определений глобальных переменных и деклараций. Рассмотрим примеры определений и деклараций в приведенном ниже внутреннем представлении программы

```
#include <stdio>
int main () {
   printf ("hello world!");
}
```

Соответствующий модуль состоит декларации строковой переменной @str, определения функции @main и декларации функции @puts:

```
@str = private unnamed_addr constant
    [13 x i8] c"hello world!\00"
define i32 @main() nounwind {
    entry:
        %puts = tail call i32 @puts(i8*
        getelementptr inbounds ([13 x i8]*
        @str, i32 0, i32 0))
    ret i32 0
}
```


Спецификатор private в описании переменной @str означает, что она видна только в пределах соответствует данного модуля; constant спецификатору const языка С. Спецификатор unnamed addr означает, что адрес переменной несуществен, так что определения константных переменных с одинаковыми значениями могут быть слиты. Далее следует тип переменной: [13 x i8] (массив из 13 элементов типа і8 – целое размера 8) и начальное значение, заданное строкой символов в кавычках. Помимо этих элементов, декларация переменной может содержать имя секции, в которую должна быть помещена переменная, выравнивание, а также номер адресного пространства, например:

@G = addrspace(5) constant float 1.0, section "foo", align 4

Определение функции @main в данном примере содержит тип (іЗ2 – целое размера 32), список аргументов в скобках (в данном случае пустой), атрибут nounwind, который означает, что функция не может завершиться особым образом в результате исключительной ситуации. Другие примеры возможных атрибутов функций: noinline (He выполнять inline-подстановку данной функции), noreturn (данная функция не возвращает управление).

Тело функции состоит из двух инструкций – вызова (tail call – хвостовой вызов) функции puts и инструкции возврата ret. Аргумент функции задается как адрес строки @str, получаемый при помощи команды внутреннего языка LLVM getelementptr (описание этой команды приведено далее).

Программный код, составляющий тело функции, имеет представление SSA (Single Static Assignment): последовательность команд внутреннего «машинного» языка LLVM, где каждой внутренней переменной (виртуальному регистру) значение присваивается только один раз. Для слияния переменных, значения которым могли быть присвоены в разных базовых блоках, используется инструкция **phi**.

Последняя строка модуля содержит декларацию стандартной функции **puts** с одним аргументом типа «указатель на **і8**». Аргумент имеет атрибут **nocapture**, который означает, что функция не создает копий указателя, которые могли бы быть использованы где-либо, помимо самой функции puts. Другие примеры атрибутов, поддерживаемых для аргументов функций: noalias – близок по смыслу спецификатору restrict языка C, он означает, что указуемый объект не перекрывается с объектами, на которые могут указывать другие аргументы указательного типа; sret - означает, что данный аргумент является указателем на область памяти, в функция должна которой записать результат, имеющий агрегатный тип.

3.3.Типы данных

Система типов является одним из важнейших свойств внутреннего представления LLVM. Благодаря тому, что с операндами инструкций связаны типы, многие виды трансформаций могут выполняться непосредственно, без предварительных проходов анализа. Типы данных подразделяются на простые и производные.

3.3.1. Простые типы

Целочисленные типы различных размеров, задаваемых в битах: **i1**, **i2**, **i3**, ..., **i64**, ... Максимальный размер 2^23-1.

Типы данных с плавающей точкой. К ним относятся типы

half – 16-битные значения,

float – 32-битные значения,

double – 64-битные значения,

fp128 – 128-битные значения со 112-битной мантиссой,

х86 fp80 – 80-битные значения (Х87),

ррс fp128 – 128-битные значения (PowerPC).

Х86ттх – представляет значения, хранимые на регистрах ММХ процессора x86.

Tun **void** не представляет какое бы то ни было значение, не имеет размера.

Тип **label** представляет метки в коде программы.

Tun **metadata** представляет встроенные метаданные программы. Этот тип не может быть использован для создания производных типов.

3.3.2. Производные типы

Производный тип может быть построен на основе простых или других производных типов. Поддерживаются следующие производные типы: массивы, функции, структуры, указатели, векторы.

Массив – это набор однотипных значений, хранящихся в памяти последовательно. Тип элементов – любой тип, имеющий размер. Примеры:

- [40 x i32] одномерный массив из 40 32-битных целых.
- [2 x [3 x [4 x i16]]] многомерный массив из 2x3x4 16-битных целых.

Функция. Тип функции представляет сигнатуру функции, он содержит тип результата и типы параметров. Тип результата может быть простейшим (кроме **х86mmx**) или производным (за исключением функции). Тип функции не имеет размера.

Примеры:

- float (i16, i32 *) функция с результатом типа float, имеющая два параметра: i16 и указатель на i32.
- i32 (i8*, ...) функция с результатом типа i32, имеющая переменное число аргументов. Первый аргумент обязателен и имеет тип «указатель на i8».

Структура. Тип структуры представляет набор элементов различных типов, расположенных в памяти последовательно. Типы элементов структур должны иметь размер.

Структуры могут быть упакованными И неупакованными. В упакованных структурах элементы выравниванием байт. В располагаются с 1 неупакованных структурах элементы имеют выравнивание, определяемое ABI целевой платформы. Примеры:

- { float, i32 (i32) * } структура из двух элементов; первый элемент имеет тип float, второй – указатель на функцию типа i32 (i32).
- <{ i8, i32 }> упакованная структура из двух элементов, размера 5 байт.

Скрытая (opaque) структура соответствует декларации именованной структуры без описания элементов. Пример описания скрытой структуры: **%X** = type opaque.

Указатель. Тип указателя служит для описания позиций в памяти. Он используется для доступа к объектам в памяти.

С типом указателя может быть связан номер адресного пространства, в котором находятся указуемые объекты. По умолчанию подразумевается нулевое адресное пространство. Семантика ненулевых адресных пространств зависит от целевой платформы. Примеры:

- [4 x i32]* указатель на массив из четырех элементов i32.
- **i32** addrspace (5) * указатель на значение **i32**, находящееся в адресном пространстве 5.

В LLVM не допускаются указатели **void*** и указатели на метки (label*); вместо этого следует использовать i8*.

Вектор. Этот тип представляет векторы элементов простейших типов, используемые как операнды SIMDинструкций. Его определение включает число и тип элементов. Тип элементов может быть целочисленным типом, типом с плавающей точкой или указателем на эти типы. Примеры:

<4 х і32> – вектор из четырех 32-битных целых.

- **x** float> вектор из восьми 32-битных значений с плавающей точкой.
- <4 x i64*> вектор из четырех указателей на 64битные целые.

3.4. Инструкции

Множество инструкций внутреннего представления LLVM включает следующие группы: терминирующие инструкции, бинарные арифметические операции, бинарные побитовые операции, векторные операции, агрегатные операции, операции адресации и доступа к объектам в памяти, операции преобразования типов и прочие операции.

3.4.1. Терминирующие инструкции

Терминирующие инструкции – это инструкции, которыми заканчиваются базовые блоки кода. К ним относятся следующие инструкции:

ret – возврат из функции. Аргументом может быть возвращаемое значение функции или **void** (возврат из **void**-функции).

br – условный или безусловный переход на другой базовый блок данной функции. Инструкция условного перехода имеет 3 аргумента: условие и две метки; в случае безусловного перехода задается один аргумент – метка.

Пример использования инструкций br и ret:

```
Test:

%cond = icmp eq i32 %a, %b

br i1 %cond, label %IfEqual,

label %IfUnequal

IfEqual:

ret i32 1

IfUnequal:

ret i32 0
```

switch – переключатель, позволяет разветвить поток управления по более чем двум направлениям. Инструкция имеет три аргумента: целочисленное значение (переменная), метка перехода по умолчанию и массив пар «константа, метка», задающий константы, с которыми сравнивается переменная, и метки переходов. Пример:

```
switch i32 %val, label %otherwise
[ i32 0, label %onzero
```

i32 1, label %onone i32 2, label %ontwo]

indirectbr – инструкция косвенного перехода.

invoke – инструкция вызова функции с возможностью возврата по одной из двух меток – для случая нормального возврата и случая возникновения исключительной ситуации.

resume – эта инструкция является частью системы обработки исключительных ситуаций LLVM. Она служит для возобновления распространения исключительной ситуации, развертка которого была прервана выполнением инструкции **landingpad**.

unreachable – указывает оптимизатору, что данный участок кода является недостижимым.

3.4.2. Бинарные арифметические операции над целыми значениями

Все эти операции имеют два аргумента и один результат. Аргументы имеют одинаковые целочисленные типы – скалярные или векторные. Результат имеет тот же тип, что и аргументы.

- add сложение целых значений
- **sub** вычитание целых значений.

mul – умножение целых значений.

Результат трех перечисленных выше инструкций корректен как для знаковых, так и для беззнаковых типов, поскольку для отрицательных значений используется дополнительное представление, и результат имеет тот же размер, что и операнды.

Если при вычислении результата произошло беззнаковое переполнение, то берется результат по модулю **2^n**, где **n** – число бит результата.

Для этих операций поддерживаются модификаторы **nuw**, **nsw** (или оба), которые обозначают, что при беззнаковом и/или знаковом переполнении результат некорректен (poisoned). Примеры:

```
<result> = add i32 4, %var
```

```
Результат = 4 + %var, тип результата i32
<result> = sub i32 0, %val
Результат = -%val, тип результата i32
```

udiv – беззнаковое деление целых значений.

sdiv – знаковое деление целых значений. Остаток имеет тот же знак, что и делитель.

Для операций деления поддерживается модификатор **exact**, который означает, что результат некорректен, если остаток от деления ненулевой.

urem – беззнаковая инструкция вычисления остатка от деления целых значений.

srem – знаковая инструкция вычисления остатка от деления целых значений.

3.4.3. Бинарные арифметические операции над значениями с плавающей точкой

Аргументы этих операций имеют одинаковые типы с плавающей точкой – скалярные или векторные. Результат имеет тот же тип, что и аргументы.

- **fadd** сложение.
- **fsub** вычитание.
- **fmul** умножение.
- **fdiv** деление.

frem – взятие остатка от деления.

3.4.4. Бинарные побитовые операции

Операции этой группы реализуют различные манипуляции над битами значений. Они могут появляться как результат оптимизаций понижения мощности операций (например, деление или умножение целого значения на степень числа 2 реализуются как сдвиги). Оба операнда должны иметь один и тот же целый скалярный или векторный тип.

shl – сдвиг влево. Результат равен **opl * 2^op2**. Второй операнд трактуется как беззнаковое значение. Модификатор **nuw** означает, что результат некорректен, если был удален хотя бы один ненулевой бит; модификатор **nsw** означает, что результат некорректен, если в результате знаковый бит изменился. То есть, данные модификаторы здесь работают так же, как в соответствующей операции умножения на **2^op2**.

lshr – логический сдвиг вправо. Старшие биты результата заполняются нулями. Модификатор **exact** означает, что результат некорректен, если в результате сдвига удален хотя бы один ненулевой бит.

ashr – арифметический сдвиг вправо. Старшие биты результата заполняются знаковым битом первого операнда. Модификатор **exact** означает, что результат некорректен, если был удален хотя бы один ненулевой бит.

and – побитовая логическая операция «и».

ог – побитовая логическая операция «или».

хог – побитовая логическая операция «исключающее или».

3.4.5. Векторные операции

extractelement – извлечение компонента вектора по номеру.

insertelement – запись компонента вектора по номеру.

shufflevector – создание вектора из элементов двух входных операндов-векторов. Эта операция имеет три операнда: два вектора **v1** и **v2** одинакового типа и маску – массив из **m** элементов типа **i32**.

Элементы маски задают номера элементов вектора, являющегося конкатенацией v1 и v2, из которых должен быть составлен результат. Результирующий вектор имеет m элементов, тип которых совпадает с типом элементов входных векторов.

3.4.6. Операции над агрегатными значениями

extractelement – позволяет выбрать элемент из структуры или массива. Первый операнд – значение агрегатного типа – структура или массив. Последующие операнды – индексы, относящиеся к последовательным уровням иерархии агрегатного типа. Например, для выборки элемента массива **B**[5][13] нужно задать индексы 5, 13. Аналогично используются индексы для доступа к элементам вложенных структур.

insertelement – производит запись элемента в структуру или массив. Первый аргумент задает агрегатное значение, второй – записываемое значение, последующие аргументы задают индексы для доступа к агрегатному значению, так же как в операции extractelement.

3.4.7. Операции адресации и доступа к объектам в памяти

В эту группу включены операции выделения памяти, адресации, чтения, записи, а также операции атомарного доступа к памяти и барьеры памяти, близкие по своей семантике к средствам, вошедшим в последние стандарты языков С [8] и С++ [9].

В командах атомарного доступа к памяти и барьеров памяти задается операнд, задающий тип синхронизации операций с памятью, который может принимать одно из следующих значений: unordered, monotonic (то же, что relaxed в C/C++), acquire, release, acl_rel, seq_cst. Для этих команд также поддерживается необязательный модификатор singlethread (синхронизация только с обработчиком исключений, исполняемым в рамках потока).

alloca – инструкция, выделяющая область памяти в стеке. В качестве аргументов задается тип значения, под которое нужно выделить память, возможно – число элементов, если память выделяется под массив, а также, возможно, выравнивание. В качестве результата возвращается указатель на выделенную память.

load – инструкция чтения значения из памяти. Аргументом служит указатель на читаемый элемент. Необязательный аргумент – выравнивание. Может присутствовать модификатор volatile, а также аргументы, используемые оптимизатором. Они могут указывать, что читаемые данные не будут переиспользованы в кэше или что они находятся в памяти, содержимое которой не меняется в ходе выполнения программы.

Поддерживается также формат команды load atomic, которая реализует атомарное чтение объекта. В этом случае задается тип синхронизации и возможен модификатор singlethread.

store – запись объекта в память. Аргументы этой команды аналогичны аргументам **load**. Поддерживается также операция **store atomic**.

fence – установить барьер между операциями доступа к памяти. Аргументы – необязательный модификатор **singlethread** и тип синхронизации.

стрясhg – атомарная условная модификация значения в памяти. Значение загружается из памяти, сравнивается с заданным значением и, в случае равенства, в память записывается новое значение. Аргументы команды: адрес, значение для сравнения, новое значение, необязательный модификатор singlethread и тип синхронизации.

atomicrmw – атомарная модификация значения в памяти. Аргументы команды: адрес в памяти **ptr**, операция, операнд **val**, необязательный модификатор **singlethread** и тип синхронизации. Из памяти по указанному адресу считывается значение ***ptr**, к нему и заданному операнду применяется операция, и результат записывается в память по адресу **ptr**; все эти действия выполняются атомарно. Допустимые операции и их семантика:

xchg: *ptr = val add: *ptr = *ptr + val sub: *ptr = *ptr - val and: *ptr = *ptr & val nand: *ptr = ~(*ptr & val) or: *ptr = *ptr | val xor: *ptr = *ptr ^ val max: *ptr = *ptr > val ? *ptr : val (3HaKOBOE CPABHEHUE) min: *ptr = *ptr < val ? *ptr : val (3HaKOBOE CPABHEHUE) umax: *ptr = *ptr > val ? *ptr : val (6e33HaKOBOE CPABHEHUE) umin: *ptr = *ptr < val ? *ptr : val (6e33HaKOBOE CPABHEHUE)

getelementptr – операция вычисления адреса элемента в агрегатной структуре данных. В качестве аргумента задается указатель на значение, имеющее тип структуры, массива или вектора. Далее, как в команде extractelement, задаются индексы. Первый индекс относится к самому указателю (т.е. считается, что по указателю может находиться массив объектов). Последующие индексы относятся к соответствующим уровням вложенности в иерархии типов (структур, массивов, векторов).

В качестве первого аргумента может также быть задан вектор указателей; тогда результатом также будет вектор указателей на элементы данных, в соответствии с индексами. Модификатор **inbounds** указывает, что все индексы должны быть в пределах границ, определяемых типами данных, в противном случае результат будет считаться некорректным.

3.4.8. Операции преобразования типов

Операции этой группы имеют в качестве аргументов значение и тип, к которому оно должно быть преобразовано. Результатом является преобразованное значение.

trunc – преобразует целое (или вектор целых) к целому (или вектору целых) меньшей ширины.

zext, **sext** – эти команды преобразуют целое (или вектор целых) к целому (или вектору целых) большей ширины. **zext** заполняет старшие биты нулями, **sext** – размножает знаковый бит.

fptrunc – преобразует значение с плавающей точкой к заданному типу с плавающей точкой меньшего размера (например, **double** к **float**). Если значение не может быть представлено в рамках указанного типа, то результат неопределен.

fpext – преобразует значение с плавающей точкой к заданному типу с плавающей точкой большего размера (например, **float** к **double**).

fptoui, **fptosi** – преобразуют значение (или вектор значений) с плавающей точкой к ближайшему целому (или вектору целых) заданного типа, с округлением в сторону нуля. **fptoui** осуществляет преобразование к беззнаковому целому, **fptosi** – к знаковому. Если преобразованное значение Не представим в рамках указанного целого типа, то результат неопределен.

uitopf, **sitofp** – преобразуют целое значение (или вектор целых значений) к значению (или вектору значений) с плавающей точкой. **uitopf** трактует входное значение как беззнаковое, **sitofp** – как знаковое. Если преобразованное значение не представимо в рамках указанного целого типа, то результат неопределен.

ptrtoint – преобразует указатель (или вектор указателей) к целому (или вектору целых) заданного типа.

inttoptr – преобразует целое (или вектор целых) к указателю (или вектору указателей) заданного типа.

bitcast – преобразует входное значение к заданному типу, не меняя битовое представление значения. Тип результата должен иметь тот же размер, что и входное значение. Примеры преобразований, которые могут осуществляться при помощи bitcast: float к i32, указатель к указателю другого типа, вектор целых к вектору целых другого типа, имеющему тот же размер, например <2 x i32> к <8 x i8>).

3.4.9. Прочие операции

істр – возвращает битовое значение (или вектор битов), соответствующий результату сравнения целых, указателей (или векторов целых или указателей). Поддерживаемые операции сравнения: «равно», «не равно», а также знаковые и беззнаковые «больше», «меньше», «больше или равно», «меньше или равно».

fcmp – возвращает битовое значение (или вектор битов), соответствующий результату сравнения значений (или векторов значений) с плавающей точкой. Отметим, что генератор кода в настоящее время не поддерживает **fcmp** с векторными операндами.

Поддерживаются операции сравнения «истина» (всегда 1), «ложь» (всегда 0), а также упорядоченные и неупорядоченные «равно», «не равно», «больше», «меньше», «больше или равно», «меньше или равно». Неупорядоченные операции отличаются тем, что возвращают 1, если хотя бы один операнд является QNAN. Имеются также операции, проверяющие, что оба операнда не являются QNAN и что хотя бы один операнд – QNAN. **phi** – реализует узел phi в SSA-представлении функции. В качестве аргументов задается последовательность пар (массивов) вида [*значение*, *метка*], где *метка* определяет базовый блок. Пример:

```
Loop: ; бесконечный цикл...
%indvar = phi i32 [0, %LoopHeader],
[%nextindvar, %Loop]
%nextindvar = add i32 %indvar, 1
br label %Loop
```

Здесь команда **phi** сообщает оптимизатору, что значение переменной **%indvar** равно либо 0, либо **%nextindvar**, а также метки базовых блоков, в которых переменная могла получить значение.

select – условный выбор одного из двух входных значений. Пример:

%X = select i1 %cond, i8 17, i8 42

Переменная **%X** получает значение **17**, если **%cond == 1**, и **42** в противном случае.

call – обычный вызов функции. Операнды – имя функции и ее аргументы. Могут также присутствовать атрибуты функции и ее результата, а также модификатор, указывающий, что допустима оптимизация данного вызова как хвостового.

va_arg – служит для доступа к аргументам, переданным функции с переменным числом аргументов.

landingpad – является частью системы обработки исключений LLVM.

3.5. Внутренние функции

поддерживает внутренние LLVM (intrinsic) функции, аналогичные встроенным (builtin) функциям GCC. Вызовы внутренних функций преобразуются в последовательность команд, реализующую семантику функции. Определены, в частности, внутренние функции, соответствующие стандартным функциям языка С: (тетсру, тетточе и др.). Предусмотрен механизм определения внутренних функций, специфических для целевой архитектуры (например, через внутренние функции реализована поддержка библиотеки Altivec для архитектуры PowerPC).

3.6. Выводы

Внутренний язык LLVM достаточен лля представления семантики программ на современных языках программирования, включая такие механизмы как обработка исключений и поддержка атомарных операций. Наличие высокоуровневой информации о типах, развитая система команд внутреннего языка позволяют эффективно проводить различные оптимизации и обеспечивают потребности генерации Механизм внутренних функций кода. лает возможность программистам использовать специфические команды процессора в программах на языках высокого уровня.

Недостатком представляется отсутствие комплексных типов и, соответственно, операций над

данными этих типов. При генерации кода лля процессоров, поддерживающих операции нал комплексными значениями оказывается невозможным использовать эти средства автоматически. Для того чтобы использовать команды, работающие С комплексными данными, программист должен, вместо стандартных средств языка высокого уровня применять вызовы внутренних функций, что приводит к снижению мобильности программного обеспечения.

Хотя LLVM в принципе предусматривает расширение системы типов и набора команд внутреннего языка, использование этой возможности трудоемко, поскольку необходимо реализовать поддержку вновь вводимых команд и типов во многих компонентах компилятора, включая трансляцию с входного языка на внутренний язык LLVM, проходы оптимизации и анализа, генераторы кода для различных архитектур.

4. Средства генерации кода

Подсистема генерации кода LLVM состоит из машинно-независимого генератора кода и описаний целевых архитектур. В настоящее время поддерживаются архитектуры Alpha, ARM, CellSPU, IA64, MIPS, PowerPC, Sparc, X86, X86_64, XCore, PIC-16. Реализована также генерация выходного кода на языках С и MSIL.

4.1. Машинно-независимый генератор кода

Процесс генерации кода LLVM состоит из следующих основных стадий:

- выбор инструкций,
- планирование команд,

- оптимизации машинного кода в SSA-представлении,

- распределение регистров,
- генерация пролога и эпилога,
- финальные оптимизации,
- вывод машинного кода.

На стадии выбора инструкций программный код из промежуточного представления LLVM преобразуется в код, выраженный в терминах машинных инструкций целевого процессора. При этом сохраняется SSAпредставление, использующее виртуальные регистры (а также, возможно, физические регистры там, где этого требуют соглашения о связях или аппаратные ограничения).

В генераторе кода LLVM используется метод, основанный на применении графа выбора (Selection DAG – Directed Acyclic Graph, направленный ациклический граф). Узлы графа представляют собой объекты класса **SDNode**, основными элементами которого являются код операции и операнды. Поддерживаются узлы операций, порождающих одно или несколько значений; несколько значений может порождать, например, команда, вычисляющая частное и остаток от деления, или команда, вычисляющая сумму и разность двух значений.

Каждому операнду соответствует дуга, указывающая на узел, в котором вычисляется значение операнда. Различаются два типа дуг – дуги,

описывающие зависимости по данным и дуги, описывающие зависимости по управлению.

Важное понятие, связанное с графом выбора, – понятие легального и нелегального графа. Легальный граф для данного целевого процессора содержит только операции, поддерживаемые процессором и операнды, имеющие типы, которые поддерживаются процессором. За легализацию графа в процессе выбора инструкций отвечают фазы легализации типов и легализации операций.

Рассмотрим процесс генерации кода более детально. Он состоит из следующих фаз.

1. Построение исходного, нелегального, графа выбора по внутреннему представлению программы.

2. Оптимизация графа. Выполняются некоторые упрощения, а также выявление идиом, таких как вычисление частного и остатка деления для процессоров, поддерживающих соответствующую команду.

Важный класс упрощений – удаление избыточных операций **zext** и **sext** (приведение типов с дополнением нулями или знаковым типом).

3. Легализация типов. Граф выбора преобразуется таким образом, чтобы исключить типы операндов, не поддерживаемых целевым процессором. Например, если процессор не поддерживает SIMD-команды, то операции над векторами преобразуются в подходящие последовательности операций над скалярными операндами.

4. Оптимизация графа. Проводится устранение избыточности, которая могла возникнуть в результате предыдущей фазы.

5. Легализация операций. Операции, которые не поддерживаются целевым процессором, заменяются последовательностями команд, реализующих соответствующую функциональность. Например, операции атомарного доступа к памяти для RISC-процессора заменяются последовательностями, включающими пары команд LL (Load Linked) и (SC – Store Conditional).

6. Оптимизация графа. Проводится устранение избыточности, которая могла возникнуть в результате предыдущей фазы.

7. Выбор инструкций из графа. Здесь применяются шаблоны подстановки, представленные в описании команд целевой архитектуры. Например, если процессор поддерживает команды умножения с накоплением **madd**, то фрагменты графа вида **add** (x, mul(y,z)) будут заменены узлами, содержащими команду **madd**.

8. Планирование. Команды, составляющим преобразованный, легальный граф выбора, присваивается легальный порядок. В LLVM применяется стандартный алгоритм планирования по списку, с поддержкой различных модификаций, включая планирование сверху вниз и снизу вверх.

После выполнения этих фаз граф выбора уничтожается, и далее используется линейное представление программы.

На стадии планирования инструкции, сформированные в результате выбора, упорядочиваются с учетом зависимостей так, чтобы оптимизировать задержки, связанные с латентностями команд и конкуренцией по использованию вычислительных ресурсов.

9. Стадия оптимизаций кода в SSA-представлении является необязательной. На этой стадии могут выполняться такие оптимизации, как планирование по модулю (конвейеризация циклов).

10. Распределение регистров. На этой стадии код, включающий неограниченный набор виртуальных регистров, преобразуется в код, использующий физические регистры процессора. При этом, в случае нехватки регистров, в программу может быть добавлен спилл-код, обеспечивающий временное выталкивание в память некоторых значений и их последующее восстановление.

Поддерживается несколько методов распределения регистров:

Быстрый. Выполняется локальное распределение регистров на уровне базовых блоков; при этом делаются попытки по возможности переиспользовать регистры. Этот метод используется по умолчанию при отладочных сборках.

Базовый. Инкрементальный метод распределения, при котором регистры назначаются для интервалов жизни значений в порядке, управляемом эвристикой. Хотя этот метод нельзя считать промышленным, он может быть основой для реализации более интересных подходов, а также как отправная точка для сравнения производительности.

<u>Жадный метод</u>, используется по умолчанию. Это усовершенствованная версия базового алгоритма, включающая разбиение глобальных диапазонов жизни значений. Этот метод направлен на минимизацию накладных расходов, связанных с выталкиванием регистров в стек (register spilling).

<u>PBQP</u> – метод, при котором задача распределения регистров сводится к задаче PBQL (Partitioned Boolean Quadratic Programming). Назначение регистров выполняется в соответствии с полученным решением задачи PBQP. Этот метод считается наиболее подходящим для архитектур с нерегулярным регистровым файлом.

Генератор кода позволяет явно задать метод распределения регистров при помощи опции.

11. Генерация пролога и эпилога. К этому моменту становится известен размер кадра стека, необходимый с учетом команд alloca, спилл-кода и количества сохраняемых регистров, поэтому возможно формирование кода для пролога и эпилога. На этой стадии выполняются также такие оптимизации, как исключение указателя кадра стека (frame pointer), а также упаковка кадра стека с учетом размеров и выравнивания хранящихся в нем объектов. В результате абстрактные позиции стека, которые использовались на предыдущих стадиях, заменяются на фактические адреса в памяти.

12. Стадия финальных оптимизаций может включать планирование спилл-кода и другие специфические машинно-зависимые оптимизации, например, заполнение гнезд задержки (для MIPS).

13. Вывод кода. Сгенерированный код выводится в виде ассемблерного или объектного файла.

Важная составная часть машинно-независимого генератора кода – базовые классы для описания целевых процессоров, например:

LLVMTargetMachine – класс, предоставляет виртуальные методы для доступа к разным составным частям описания архитектуры – регистры, набор инструкций, структура кадра стека и др.

TargetRegisterInfo – класс, предоставляющий информацию о регистровом файле.

TargetInstrInfo – класс, содержащий информацию о наборе команд.

TargetSubtarget – класс, представляющий конкретный процессор из семейства процессоров данной архитектуры.

4.2. Средства описания целевых процессоров

4.2.1. Табличные средства описания

Инфраструктура LLVM предоставляет язык для описания многих аспектов целевой архитектуры в табличной форме, а также инструмент TableGen, который служит для трансляции табличных описаний в фрагменты кода на языке С++.

Табличный язык, называемый далее языком TableGen, позволяет описывать сущности архитектуры, такие как регистровый файл, классы регистров, инструкции, и т.д. в более компактной и наглядной форме по сравнению с соответствующими описаниями на C++. Файл на табличном языке состоит из записей двух типов: определения и классы.

Определение начинается словом **def**; оно описывает конкретный объект, например, можно определить объект «регистр **zero**»:

```
def reg_zero {
  string Namespace = "Mips";
    // perистр Mips
  string AsmName = "zero";
    // ассемблерное имя
  list<string> AltNames = ["0"];
    // список альтернативных имен
  int DwarfNum = 0;
    // номер для DWARF
}
```

Определение содержит имя объекта, за которым в фигурных скобках следуют определения полей. Определение поля включает тип, имя поля и его значения. В данном примере использованы типы

string – строка символов,

int – 32-битное целое,

list<*mun*> – список. В угловых скобках задается тип элементов списка.

Другие типы, поддерживаемые TableGen: bit – булево значение, 0 или 1, bits<n> – битовое поле ширины n, класс – имя класса, задает объект данного класса, dag – направленный ациклический граф объектов, **code** – произвольный текст, например, фрагмент кода на С++.

Язык TableGen позволяет также описывать классы. Рассмотрим, например, класс **Register** ¹, представляющий регистры целевой архитектуры:

```
class Register
```

```
{
  string Namespace;
    // архитектура
  string AsmName;
    // ассемблерное имя
  list<string> AltNames = [];
    // альтернативные имена
  int DwarfNum;
    // номер для DWARF
}
```

Описав класс, можно определять объекты этого класса. Ниже приведены примеры описания регистров **zero** и **at**:

```
def ZERO : Register {
   let Namespace = "Mips";
   let AsmName = "zero";
   let AltNames = ["0"];
   let DwarfNum = 0;
}
def AT : Register {
   let Namespace = "Mips";
   let AsmName = "at";
   let AltNames = ["1"];
   let DwarfNum = 1;
}
```

В описаниях этих объектов можно «вынести за скобки» общее поле **Namespace**, используя конструкцию **let**:

```
let Namespace = "Mips" in {
  def ZERO : Register {
    let AsmName = "zero";
    let AltNames = ["0"];
    let DwarfNum = 0;
  }
  def AT : Register {
    let AsmName = "at";
    let AltNames = ["1"];
    let DwarfNum = 1;
  }
}
```

Для того чтобы сделать описания объектов более компактными, можно ввести параметризованный класс (шаблон класса), например:

class Register<string n, list<string> altNames = [],

¹ Здесь приведено упрощенное описание класса по сравнению с тем, которое фактически используется в LLVM.

118

}

```
int dwarfNum >
{
 string Namespace = "";
 string AsmName = n;
 list<string> AltNames = altNames;
 int DwarfNum = dwarfNum;
}
let Namespace = "Mips" in {
 def ZERO : Register<"zero", ["0"],0>;
 def AT : Register<"at", ["1"], 1>;
...
```

При определении классов можно использовать наследование, т.е. задавать базовый класс. Например, можно определить класс **MipsRegister**, используя класс **Register** как базовый:

Здесь для поля **Namespace** из базового класса при помощи конструкции **let** задается предопределенное значение. Вводится также дополнительное поле **Code**, задающее кодировку регистра в машинных командах. Определения регистров на основе класса **MipsRegister**:

LLVM предоставляет набор базовых классов TableGen для всех основных сущностей, используемых в описании целевых архитектур: регистры, классы регистров, инструкции, описатели таблиц резервирования ресурсов и латентностей инструкций, описатели целевых процессоров данной архитектуры, описатели соглашений о связях и др.

Еще один способ сделать описания более компактными – мультиклассы. Мультикласс задает шаблоны сразу для нескольких определений. Например, следующий мультикласс в описании набора инструкций MIPS вводит шаблоны для трех инструкций преобразования плавающих в целые (на основе базового класса **FFR1**):

```
multiclass FFR1 W M<bits<6> funct,
```

```
"w.d", FGR32, FGR64>,
Requires<[IsFP64bit]> {
let DecoderNamespace = "Mips64";
}
}
```

Далее при помощи мультиопределения

defm ROUND_W : FFR1_W_M<0xc, "round">;

вводятся описания инструкций округления плавающих значений: инструкция **round.w.s** и два варианта инструкции **round.w.d**, соответствующие случаям 32-битных и 64-битных регистров плавающей арифметики. Варианты инструкции **round.w.d** различаются классом регистров, используемых для входного аргумента типа **double**: **AFGR64** – 16 пар соседних 32-битных физических регистров, **AFGR64** – 16 пар соседних 32-битных физических регистров, **FGR64** – 32 64-битных регистра. Тот или иной вариант выбирается в зависимости от истинности предикатов **NotFP64bit** и **IsFP64bit**, которые определяются в зависимости от опций командной строки.

4.2.2. Элементы описания целевой платформы

Ниже перечислены основные элементы описания целевой платформы, которые необходимо реализовать для поддержки генерации кода [10].

- Описание машины и опций генерации кода (например, для различных процессоров данной архитектуры, различных АВІ и т.п.).

- Описание регистров и классов регистров архитектуры.

- Описание соглашений о связях, включая способы передачи аргументов и возврата значений функций, множество зарезервированных регистров и др.

- Описание набора инструкций целевой архитектуры.

- Описание шаблонов и реализация процедур для трансформации исходного графа выбора в легальный для заданного целевого процессора граф.

- Реализация программного кода для преобразования внутреннего представления в ассемблерный код.

 Необязательный элемент – поддержка JIТкомпиляции, когда машинный код формируется непосредственно в оперативной памяти.

4.2.3. Целевая архитектура, процессоры, опции кодогенерации

Целевая архитектура задается на языке C++ как определение класса XXXTargetMachine (здесь и далее буквы XXX обозначают имя архитектуры, например: MIPSTargetMachine, SparcTargetMachine и т. п.), базирующегося на предопределенном классе LLVMTargetMachine и предоставляющего такие методы как getInstrInfo (информация о наборе инструкций), getRegisterInfo (информация о регистрах), getFrameInfo (информация о структуре кадра стека) и др. Описания этого класса для различных архитектур достаточно однотипны, отличаются в основном именем целевой машины.

Опции кодогенерации задаются в виде определений TableGen на базе предопределенного в LLVM класса **SubtargetFeature**, например:

```
def FeatureVFPU : SubtargetFeature<
    "vfpu",
    "HasVFPU",
    "true",
    "Enable vector FPU instructions.">;
```

Аргументы описания – имя опции командной строки компилятора, имя поля в объекте класса **xxxSubtarget**, значение этого поля и строка описания опции.

Также необходимо создать описание класса **XXXSubtarget**, включив в него поля, соответствующие опциям кодогенерации.

Описания процессоров задаются в табличном формате на основе класса **Processor**, пример:

Аргументы описания – имя процессора (может быть задано в командной строке как аргумент ключа – **march**), описатель таблиц резервирования, используемый планировщиком, и список опций кодогенерации, соответствующих по умолчанию данному процессору.

4.2.4. Регистры

Описания регистров задаются в табличной форме на основе предопределенного класса **Register** (или, чаще, на основе производных классов). Класс **Register** включает поля, описывающие имя регистра, его альтернативные имена, список регистров, с которыми он перекрывается, список субрегистров, номер, используемый при выводе отладочной информации и др. Примеры описания регистров приведены выше, в разделе 0.

Каждый регистр относится к некоторому классу регистров. Классы регистров также описываются в табличной форме, например, так выглядит описание общих 32-битных регистров MIPS:

```
def CPURegs : RegisterClass<"Mips",
        [i32], 32, (add
    // Return Values and Arguments
    V0, V1, A0, A1, A2, A3,
    // Not preserved across procedure
    // calls
    T0, T1, T2, T3, T4,
    T5, T6, T7, T8, T9,
    // Callee saved
    S0, S1, S2, S3, S4, S5, S6, S7,
    // Reserved
    ZERO, AT, K0, K1, GP, SP, FP, RA)
>;
```

В нем указан список типов значений, которые могут храниться в регистрах этого класса [i32], выравнивание 32 и dag, содержащий регистры, относящиеся к данному классу (регистры должны быть предварительно описаны). Заметим, что здесь используется не список регистров, а вырожденный граф (dag) с узлом add.

Следует отметить простоту описания регистров и классов регистров в LLVM по сравнению с GCC.

Описание регистров включает также реализацию класса **XXXRegisterInfo**, который базируется на предопределенном классе **RegisterInfo**. Методы этого класса предоставляют информацию об использовании регистров, и выполняют некоторые действия по реорганизации кода. Ниже приведены примеры методов.

getCalleeSavedRegs — возвращает список регистров, значения которых сохраняются вызываемой функцией. Сами списки могут быть описаны табличным способом (см. далее «Соглашения о связях»); метод должен лишь выбрать список, соответствующий текущему ABI.

getReservedRegs — возвращает битовую маску зарезервированных (недоступных для обычных вычислений в программе) регистров.

eliminateCallFramePseudoInstr – удаляет псевдоинструкции, которые могут использоваться при создании кадра стека для вызываемой функции.

Методы базового класса возвращают нулевые или пустые значения; для тех из них, которые должны возвращать содержательные значения, необходимо предоставить реализации.

4.2.5. Соглашения о связях

Этот раздел описания содержит следующую информацию о передаче аргументов и возврате результатов функций, а также о структуре кадра стека, например:

- в каком порядке передаются аргументы функций;

- где размещаются аргументы и результаты функций (на регистрах, в стеке);

- какие регистры используются для передачи аргументов и возврата результата функций;

Основные интерфейсы, которые для этого используются – **ССІfТуре** (проверяет предикат) и **ССАssignToReg** (назначает первый свободный регистр из списка). Следующий пример описывает регистры, на которых должны возвращаться результаты функций, для архитектуры MIPS при заданном ABI=o32:

```
def RetCC_MipsO32 : CallingConv<[
    // i32 are returned in registers
    // V0, V1, A0, A1
    CCIfType<[i32],
        CCAssignToReg<[V0, V1, A0, A1]>>,
    // f32 are returned in registers
    // ray = ray
```

```
// F0, F2
CCIfType<[f32],
CCAssignToReg<[F0, F2]>>,
```

```
// f64 are returned in registers
// D0, D1
CCIfType<[f64],
     CCIfSubtarget<"isNotSingleFloat()",
          CCAssignToReg<[D0, D1]>>>
]>;
```

Здесь строка CCIfType<[f32], CCAssignToReg<[F0, F2]>>, означает, что результат типа float (f32) возвращается на регистре F0 или F1. (Два регистра могут использоваться, например, для возврата комплексного результата.) CCIfSubtarget в последней строке служит для проверки опции генерации кода isNotSingleFloat, которая означает, что все вычисления должны выполняться с одинарной точностью.

Примеры описания сохраняемых регистров для различных ABI архитектуры MIPS:

```
def CSR_032 : CalleeSavedRegs<(add
  (sequence "D%u", 15, 10), RA, FP,
  (sequence "S%u", 7, 0))>;
```

```
def CSR_N32 : CalleeSavedRegs<(add
D31_64, D29_64, D27_64,
D25_64, D24_64, D23_64,
D22_64, D21_64, RA_64,
FP_64, GP_64,
(sequence "S%u_64", 7, 0))>;
```

```
def CSR_N64 : CalleeSavedRegs<(add
  (sequence "D%u_64", 31, 24),
  RA_64, FP_64, GP_64,
  (sequence "S%u 64", 7, 0))>;
```

Здесь оператор **sequence** задает последовательность регистров: первый аргумент определяет формат печати, второй и третий – диапазон чисел. Например, (**sequence** "S%u", 7, 0) задает регистры **S7**, **S6**, **S5**, **S4**, **S3**, **S2**, **S1**, **S0** (в порядке, в котором они будут выделяться при распределении).

Еще один употребительный интерфейс – **CCDelegateTo** – используется для выбора подходящего соглашения из множества возможных, в зависимости от заданных параметров генерации кода. Например:

```
def RetCC_Mips : CallingConv<[
    CCIfSubtarget<"isABI_EABI()",
    CCDelegateTo<RetCC_MipsEABI>>,
    CCIfSubtarget<"isABI_N32()",
    CCDelegateTo<RetCC_MipsN>>,
    CCIfSubtarget<"isABI_N64()",
    CCDelegateTo<RetCC_MipsN>>,
    CCDelegateTo<RetCC_MipsO32>
]>;
```

Эта спецификация определяет соглашение о возврате значений функций для MIPS в зависимости от заданного в командной строке ABI. Здесь объект **RetCC_MIPS** определяется как выбор одного из соглашений в соответствии с текущим ABI: **EABI**, **N32**, **N64**, **O32** (по умолчанию).

4.2.6. Набор инструкций целевой архитектуры

Наиболее объемная часть этого раздела описания архитектуры – табличные описания машинных инструкций.

TableGen позволяет описывать одновременно низкоуровневую и высокоуровневую информацию: кодировку инструкций, соответствие между инструкцией и подграфом DAG, входные и выходные операнды различных типов.

LLVM предоставляет шаблон класса **Instruction** для описания инструкций. Как правило, в описаниях инструкций используют иерархию производных классов, базирующихся на этом классе, где каждый «нижележащий» класс содержит уточнения, соответствующие различным форматам машинных инструкций.

Рассмотрим некоторые классы, используемые в описаниях инструкций архитектуры MIPS. Класс **MipsInst**, производный от **Instruction**, описывает общий формат инструкций архитектуры MIPS.

```
// Обший формат инструкций MIPS
class MipsInst<dag outs, dag ins,
    string asmstr,
    list<dag> pattern,
    InstrItinClass itin,
    Format f>:
    Instruction
```

// 32-битное командное слово, //соответствующее инструкции field bits<32> Inst; // Формат, или тип инструкции Format Form = f;

let Namespace = "Mips";

{

// Размер командного слова (4 байта) let Size = 4;

```
// Код операции, 6 бит
bits<6> Opcode = 0;
```

```
// Код операции заносится в
// старшие биты командного слова
let Inst{31-26} = Opcode;
```

```
// Выходные операнды
let OutOperandList = outs;
// Входные операнды
let InOperandList = ins;
```

```
// Формат ассемблерной команды
let AsmString = asmstr;
// Фрагмент графа выбора, с
// которым сопоставляется команда
let Pattern = pattern;
// Таблица резервирования
```

```
// вычислительных устройств
//(для планирования)
let Itinerary = itin;
// Другие поля
//...
}
```

MipsInst имеет аргументы **outs** – выходные операнды инструкции, **ins** – входные операнды, **asmstr** – форматная строка для печати ассемблерной команды, **pattern** – шаблон фрагмента графа выбора, с которым сопоставляется инструкция, **itin** – таблица резервирования для планировщика, **f** – идентификатор формата командного слова инструкции. Идентификаторы форматов для архитектуры MIPS вводятся как объекты простого класса с одним числовым аргументом.

На базе **MipsInst** определяются другие классы, описывающие группы инструкций, имеющие общие характеристики формата и одинаковые наборы операндов. Рассмотрим класс **CP2BIN**, описывающий свойства арифметических команд сопроцессора комплексной арифметики процессора К64-РИО архитектуры MIPS (в данном случае эти команды используются для реализации вычислений над данными типа **float**).

```
// Арифметические команды
// сопроцессора 2 процессора K64rio
class CP2BIN<bits<2> a,
             bits<5>op,
             dag outs,
             dag ins,
             string asmstr,
             list<dag> pattern> :
 MipsInst<outs, ins, asmstr,
          pattern, CP2BINitin,
           FrmOther>
{
  // Операнды:
 bits<6> fd; // выходной регистр
 bits<6> fs; // входной регистр
 bits<6> ft; // входной регистр
 bits<6> fr; // дополнительный
               // входной регистр
               // команды MADDPS
  // Команды этого класса имеют
  // код операции 0b010010
  let Opcode = 0b010010;
  // Бит 25 командного слова равен 1
  let Inst{25-25} = 0b1;
  // Заполнение полей командного слова,
  // соответствующих операндам
 let Inst{24-19} = ft;
  let Inst{18-13} = fs;
  let Inst{12-7} = fd;
  // Биты 5-6 командного слова
  // заполнить значением аргумента а
  let Inst{6-5} = a;
  // Биты 0-4 командного слова
  // заполнить значением аргумента ор
```

let Inst{4-0} = op;

}

Шаблон класса СР2ВІМ имеет 4 аргумента, совпадающие с первыми аргументами MipsInst. Последние два аргумента MipsInst в СР2ВІМ фиксированными определяются значениями CP2BINitin (таблица резервирования) и FrmOther - идентификатор формата командного слова. На базе CP2BIN описываются конкретные машинные инструкции: сложение, умножение, вычитание и умножение с накоплением. (в других случаях классы, описывающие группы инструкций, могут образовывать более глубокую иерархию И наследования.)

```
// Следующая строка указывает,
// что данные инструкции могут
// использоваться, соответствующий
// только если задана опция hasCP2
let Predicates = [HasCP2] in {
// Значение поля isCommutable=1 класса
// Insntruction означает,
// операция, реализуемая инструкцией,
// коммутативна
let isCommutable = 1 in {
// Инструкция сложения
def CADD : CP2BIN<0b01, 0b00000,
    (outs CP2R32:$fd),
    (ins CP2R32:$ft, CP2R32:$fs),
    "cadd\t$fd, $fs, $ft",
    [(set CP2R32:$fd,
      (fadd CP2R32:$fs, CP2R32:$ft))]>;
// Инструкция умножения
def MULPS : CP2BIN<0b00, 0b01010,
    (outs CP2R32:$fd),
    (ins CP2R32:$ft, CP2R32:$fs),
    "mulps\t$fd, $fs, $ft",
    [(set CP2R32:$fd,
      (fmul CP2R32:$fs, CP2R32:$ft))]>;
} // конец коммутативных инструкций
// Инструкция вычитания
def CSUB : CP2BIN<0b10, 0b00000,
    (outs CP2R32:$fd),
    (ins CP2R32:$ft, CP2R32:$fs),
    "csub\t$fd, $fs, $ft",
    [(set CP2R32:$fd,
      (fsub CP2R32:$fs, CP2R32:$ft))]>;
// Инструкция умножения с накоплением
def MADDPS : CP2BIN<0b01, 0b01010,
    (outs CP2R32:$fd),
    (ins CP2R32:$ft, CP2R32:$fs,
         CP2R32:$fr),
    "maddps\t$fr, $fs, $ft",
    [(set CP2R32:$fd,
      (fadd
        (fmul CP2R32:$fs, CP2R32:$ft),
        CP2R32:$fr))]>
  {let Constraints = "$fr = $fd";}
}
```

Рассмотрим наиболее сложное описание инструкции **MADDPS**. Операнды **0b01**, **0b01010** соответствуют аргументам **a** и **ор** класса **CP2BIN**. Они задают значения двух полей в командном слове.

Аргумент (outs CP2R32:\$fd) описывает список выходных операндов инструкции – в данном случае один регистр класса CP2R32 (этот класс включает регистры сопроцессора комплексной арифметики).

Аргумент (ins CP2R32:\$ft, CP2R32:\$fs, CP2R32:\$fr) задает список входных операндов – три регистра класса CP2R32.

Далее в списке аргументов следует форматная строка для вывода ассемблерного текста инструкции "maddps\t\$fr, \$fs, \$ft".

Последний аргумент задает шаблон для сопоставления с графом выбора поддерево (set CP2R32:\$fs, CP2R32:\$fd, (fadd (fmul CP2R32:\$ft), **CP2R32:\$fr))**: «Присвоить регистру \$fd результат вычисления (\$fs*\$ft)+\$fr».

Наконец, строка {let Constraints = "\$fr = \$fd";} задает значение поля Constraints базового класса Instruction. Данное поле описывает ограничения кодогонерации; здесь "\$fr = \$fd" означает, что для выходного операнда \$fd следует использовать тот же регистр, что и для входного операнда \$fr.

Помимо описаний собственно инструкций, обычно требуется еще задать шаблоны, описывающие сопоставление фрагментов графа выбора с командами целевого процессора. Например, следующий шаблон описывает, что сопроцессор комплексной арифметики процессора К64-РИО позволяет реализовать деление при помощи команды умножения (MULPS) и команды взятия обратной величины (C2RECIPnl):

Следующие примеры описывают способы получения 16-битного и 32-битного непосредственного значения для архитектуры MIPS. В этих примерах используются определения анонимных (неименованных) объектов, представляющих шаблоны.

```
// Small immediates
def : Pat<(i32 immSExt16:$in),
    (ADDiu ZERO, imm:$in)>;
// Arbitrary immediates
def : Pat<(i32 imm:$imm),
    (ORi (LUi (HI16 imm:$imm)),
        (LO16 imm:$imm))>;
```

В первом случае используется команда **ADDiu** (сложение **0** с 16-битной константой), во втором значение вычисляется из двух 16-битных значений при

помощи команд **LUi** (загрузить 16-битную константу в старшую половину слова) и **ORi** (побитовое **or** с 16-битной константой).

К описанию системы команд также относится реализация класса, производного от **TargetInstrInfo**, который задает общую информацию о системе команд и некоторый методы, используемые генератором кода, такие как:

isLoadFromStackSlot – проверяет, является ли заданная машинная инструкция загрузкой значения из стека. Если да, то возвращает целевой регистр и индекс кадра стека.

isStoreToStackSlot – проверяет, является ли заданная машинная инструкция записью значения в стек. Если да, то возвращает записываемый регистр и индекс позиции в стеке.

соруРhysReg – реализует копирование значения одного физического регистра в другой. Этот метод должен быть реализован для всех сочетаний классов регистров целевой архитектуры.

storeRegToStackSlot – реализует запись регистра в стек.

loadRegFromStackSlot – реализует считывание значения из стека в регистр.

storeRegToAddr – реализует запись регистра в память по указанному адресу.

loadRegFromAddr – реализует считывание значения из памяти в регистр.

4.2.7. Таблицы резервирования

Для описания таблиц резервирования необходимо определить функциональные устройства (объекты класса **FuncUnit**). Далее для каждого типа процессора целевой архитектуры описывается объект класса **ProcessorItineraries**, который содержит список таблиц резервирования (объекты класса **InstrItinData**) для различных групп команд. Каждая таблица, в свою очередь, представляет собой список описателей стадий конвейера (объектов класса **InstrStage**), а также латентности и описатели **bypass** по каждому операнду.

Описатель стадии содержит число тактов, которое она занимает, и список функциональных устройств, которые могут быть использованы. Описатель стадии также может содержать число тактов до начала следующей стадии (по умолчанию – 1 такт).

Язык описания таблиц резервирования GCC предоставляет аналогичные средства с некоторыми дополнительными возможностями. Например, он позволяет указывать, что некоторые функциональные устройства являются взаимно исключающими, т.е. не могут работать одновременно. В GCC также можно точно задавать латентности для **bypass**, в то время как в LLVM предполагается что **bypass** всегда сокращает обычную латентность на 1.

Также полезной представляется имеющаяся в GCC возможность указывать, что на некоторой стадии работают сразу несколько функциональных устройств.

4.2.8. Выбор инструкций

Основная часть информации для реализации выбора инструкций содержится непосредственно в описаниях инструкций, где она задается в виде выбора, фрагментов графа сопоставляемых с Дополнительная машинными инструкциями. информация предоставляется в виде отдельных шаблонов графа выбора, которые могут быть реализованы последовательностями машинных команд.

Как правило, не все узлы исходного графа выбора могут быть сопоставлены с машинными командами при помощи шаблонов. Кроме того, генератору кода требуется сообщить дополнительную информацию об аппаратных возможностях архитектуры и реализовать нестандартные способы генерации кода. Для этого служат методы класса **XXXTargetLowering**, производного от **TargetLowering**.

Во-первых, необходимо зарегистрировать определенные ранее классы регистров, указав таким образом генератору кода, какие типы данных поддерживаются непосредственно и какие классы регистров им соответствуют. Например, для MIPS :

```
// Set up the register classes
addRegisterClass(MVT::i32,
    Mips::CPURegsRegisterClass);
```

```
if (HasMips64)
addRegisterClass(MVT:::i64,
Mips::CPU64RegsRegisterClass);
```

```
if (HasCP2)
```

```
{
    addRegisterClass(MVT::f32,
        Mips::CP2R32RegisterClass);
    addRegisterClass(MVT::v2f32,
        Mips::CP2RVRegisterClass);
}
```

```
if (!TM.Options.UseSoftFloat) {
   addRegisterClass(MVT::f32,
        Mips::FGR32RegisterClass);
// ...
```

Для узлов, не имеющих прямой аппаратной реализации, необходимо задать способы их обработки. Для этого класс **TargetLowering** предоставляет несколько callback-методов, например:

```
setOperationAction (для общих операций),
setLoadExtAction (для чтения с расширением),
setIndexedLoadAction и другие.
```

В качестве аргументов им задается тип узла графа выбора и действие: **Promote**, **Expand**, **Custom** или **Legal**.

Promote означает, что эту операцию нужно реализовать через тип большего размера, например, на MIPS загрузка с расширением нулевым битом для типа **i1** реализуется через другие целочисленные типы:

Expand означает, что нужно попытаться реализовать данную операцию через другие операции или через библиотечный вызов.

Custom означает, что для узла требуется специальная обработка. В таком случае должна быть реализована соответствующая функция, а в переключатель, который содержится в методе **LowerOperation**, должен быть добавлен оператор **case** с именем узла, включающий вызов этой функции.

Legal означает, что узел обрабатывается стандартным методом. Это действие может быть полезно, если аппаратная поддержка для некоторого узла имеется не для всех процессоров.

Следует отметить, что данная часть описания системы команд может быть довольно трудоемкой, если специальная (Custom) обработка требуется для значительного числа типов узлов.

4.2.9. Вывод ассемблерного кода

За вывод ассемблерного кода отвечает класс XXXAsmPrinter, производный от предопределенного класса AsmPrinter. Необходимо реализовать метод для вывода функции, описывающий вывод начальных и завершающих директив для каждой функции. Могут потребоваться также вспомогательные функции для вывода различных директив или нестандартных инструкций. Вывод ассемблерных команл составляющих тело функции, выполняется по форматам, содержащимся в описаниях машинных инструкций.

4.3. Выводы

Хотя LLVM имеет развитые средства для представления основных свойств целевых процессоров в табличной форме, значительная часть описания должна быть представлена в программной форме. Фактически, размер (число строк) программных частей описаний архитектур в 2-3 раза превышает размер табличных описаний.

Существенным недостатком представляется также отсутствие такой важной оптимизации, как циклов 1 программная конвейеризация Хотя программной конвейеризации реализация в значительной мере машинно-зависима, многие ее элементы (выявление циклов, построение графа зависимостей с дистанциями, алгоритм планирования) могли бы быть реализованы в машинно-независимом генераторе.

Средства описания таблиц резервирования не позволяют специфицировать некоторые детали работы конвейера, в частности, число тактов, на которое

¹ В ранних версиях LLVM присутствовала реализация алгоритма планирования по модулю (только для процессора Sparc), но в дальнейшем, вероятно, из-за сложности сопровождения она была исключена.

сокращается латентность при передаче операндов по **bypass**, взаимоисключение функциональных устройств. Подобные детали могут быть несущественными при использовании эвристических методов планирования команд, но применение точных методов оптимизации должно быть основано на точном описании архитектуры.

Отметим также, что генерация кода для некоторых архитектур не позволяет пока использовать все их аппаратные возможности. Например, для архитектуры MIPS в настоящее время реализована поддержка нескольких систем команд (ISA), таких как mips32, mips64, mips64r1, но нет поддержки различных процессоров семейства MIPS (на уровне возможных различий в системе команд и в свойствах конвейера). Не реализована также поддержка векторных команд (формата paired single).

5. Опыт настройки LLVM для генерации кода с использованием сопроцессора комплексной арифметики K64RIO CP2

Авторами была выполнена настройка генератора кода LLVM для использования сопроцессора комплексной арифметики (СР2) процессора К64-РИО архитектуры MIPS.

Сопроцессор СР2 предназначен для работы с векторами из двух 32-битных значений с плавающей точкой. Он поддерживает обычные векторные операции, комплексные операции (над операндами, рассматриваемыми как комплексные значения), а также раз специализированных операций. По сравнению с сопроцессором плавающей арифметики (FPU), CP2 обеспечивает лучшие характеристики по латентностям и степени конвейеризованности команд. эксперимента являлось исследование Целью возможности применения данного сопроцессора для выполнения вычислений над данными с плавающей точкой одинарной точности (float), используя в вычислениях только один компонент вектора.

Для этого было дополнено описание архитектуры MIPS, имеющееся в LLVM. В частности, было введено описание процессора k64rio и опции компиляции HasCP2. Были описаны регистры CP2 и соответствующий класс регистров, а также арифметические команды, команды чтения и записи, команды пересылок. Были дополнены методы генерации кода, реализующие пересылки между регистрами СР2 и другими классами регистров и сохранение/восстановление этих регистров в стеке. были Также составлены описания таблин резервирования для k64rio.

Эксперименты по генерации кода с использованием СР2 показали, что увеличения производительности в среднем, на векторных и матричных вычислениях, не наблюдается. Прирост производительности возможен только на некоторых рекуррентных вычислениях, где работа с памятью во внутреннем цикле отсутствует. Это объясняется тем, что k64rio не содержит команд, позволяющих загрузить 32-битное значение в старшую или младшую часть регистра CP2 и записать такое значение в память. Поддерживается только загрузка и сохранение 64битных значений по адресам, выровненным на 8 байт. Поэтому загрузка и запись значений типа **float** в/из регистра CP2 возможна только через промежуточный регистр (FPU или общий), что препятствует повышению производительности. Дополнительные пересылки могут возникать также из-за того, что некоторые команды CP2 используют один и тот же регистр для входного и выходного операнда.

Пересылки могли бы выполняться параллельно с вычислительными командами, но для этого необходима конвейеризация или развертка циклов. Хотя развертка циклов в LLVM реализована, ее использование контролируется самой средой оптимизации, и опции для управления ею со стороны пользователя отсутствуют, поэтому получить удовлетворительный результат не удалось.

Хотя реализованный метод генерации кода не привел к повышению производительности, полученный опыт показал, что в рамках LLVM можно достаточно легко опробовать различные экспериментальные подходы к генерации кода.

6. Python-интерфейс к LLVM

Разработка точных методов оптимизации кода является сложной задачей, работа над которой обычно требует длительного периода экспериментирования. При использовании методов математического программирования, например, ЦЛП, приходится опробовать различные варианты формулировок задачи ЦЛП, системы решения таких задач и их настройки, варьировать подходы к генерации кода. Весьма сложным также может быть С/С++-интерфейс с внешними системами решения задач ЦЛП.

Опыт показывает, что существенно ускорить и упростить этот процесс позволяет использование языков высокого уровня абстракции, таких как Python [11]. На базе Python peaлизована, в частности, система PuLP [12], позволяющая в удобном и естественном виде задавать формулировки задач линейного и целочисленного линейного программирования и применять для их решения свободно-доступные (GLPK [13], Coin-or [14]) и коммерческие (Cplex [15], Gurobi [16]) системы решения.

Python-интерфейса Наличие представляется важным условием для разработки точных методов кода с использованием метолов оптимизации математического программирования. В 2010 г. был реализован проект llvm-ру [17], работавший с Python 2.4 и llvm 2.7 – 2.8. В дальнейшем он некоторое время не поддерживался, и только в октябре 2012 г. вышла новая версия, работающая с llvm 3.1. Важно отметить, что llvm-ру предоставляет интерфейс к базовому внутреннему представлению, а для реализации методов оптимизации на стадии генерации кода необходимо либо реализовать интерфейс к пониженному (машинно-зависимому) внутреннему представлению, либо полностью реализовать генерацию кода на языке Python.

д в условиях дефици

Также в рамках самой инфраструктуры LLVM начата разработка Python-привязки к LLVM, которая должна обеспечивать интерфейс к основным классам внутреннего представления, но эта работа пока не завершена, поддерживается только интерфейс для работы с объектными файлами и дизассемблер.

7. Заключение

LLVM представляет собой развитую, хорошо структурированную инфраструктуру для создания компиляторов и других средств разработки программ. Он обладает рядом свойств, которые представляются важными и полезными при разработке и реализации промышленных и экспериментальных средств оптимизации и генерации кода:

- внутреннее представление, пригодное для отображения семантики современных языков программирования и позволяющее эффективно производить анализ и трансформации программы;

- гибкий механизм сборки приложений, простота добавления новых компонентов, библиотек;

- простые средства встраивания дополнительных или альтернативных проходов оптимизации;

- внутреннее представление на стадии кодогенерации (DAG), позволяющее в принципе применять различные методы выбора команд;

- развитые средства описания целевых процессоров и внутреннего представления этих описаний;

- разнообразные инструменты отладки и тестирования, включая дампы промежуточного представления на каждом проходе, отладочные дампы, визуализация графа выбора, генерация случайных тестов.

Все это делает LLVM привлекательным инструментом для разработки экспериментальных и промышленно используемых инновационных методов анализа и оптимизации, а также инструментальных средств разработки для встроенных систем.

В то же время, для реализации точных методов оптимизации кода на базе LLVM требуются значительные доработки, включая расширение Pythonинтерфейса, уточненное описание таблиц резервирования, инфраструктуру для программной конвейеризации циклов.

Литература

1. Н. И. Вьюкова, В. А. Галатенко, С. В.Самборский. Совместное решение задач выбора и

планирования команд в условиях дефицита регистров. – М. Программная инженерия, 2012, N2, с. 35-40. 2. С. В. Самборский. Формулировка задачи

 С. В. Самоорскии. Формулировка задачи планирования линейных и циклических участков кода.
 Программные продукты и системы. – 2007. – Т.79, # 3. – С. 12-16.

3. Н. И. Вьюкова, С. В. Самборский. Многомерная конвейеризация циклов. Программные продукты и системы. – 2008. – # 4. – С. 8-13.

4. C. Lattner. LLVM: An Infrastructure for Multi-Stage Optimization. – Computer Science Dept., University of Illinois at Urbana-Champaign. – Masters Thesis. – 2002.

5. Chris Lattner and Vikram Adve. LLVM: A Compilation Framework for Lifelong Program Analysis & Transformation. Proc. of the 2004 International Symposium on Code Generation and Optimization (CGO'04), Palo Alto, California, Mar. 2004.

6. Konstantin Serebryany, Dmitry Vyukov. Finding races and memory errors with compiler instrumentation. AddressSanitizer. ThreadSanitizer. GNU Tools Cauldron 10 Julv 2012 http://gcc.gnu.org/wiki/cauldron2012?action=AttachFile& do=get&target=kcc.pdf

7. Giuliano de Souza Vilela Cid. A LLVM based development environment for embedded systems software targeting the RISCO processor. Universidade Federal do Rio Grande do Norte. 2010.

8. ISO/IEC 9899:2011 – Information technology – Programming Languages – C.

9. ISO/IEC 14882:2011 – Information technology – Programming Languages – C++.

10. LOPES, B. C. Understanding and Writing an LLVM compiler backend. In: ELC'09: Embedded Linux Conference, 2009.

11. Язык программирования Python. http://www.python.ru/

12. Stuart Mitchell. PuLP: A Linear Programming Toolkit for Python. http://www.optimizationonline.org/DB_FILE/2011/09/3178.pdf

13. Gnu linear programming kit. http://www.gnu.org/software/glpk/glpk.html

14. R. Lougee-Heimer, The common optimization interface for operations research – IBM Journal of Research and Development, vol. 47, no. 1, pp. 57-66, January 2003.

15. IBM ILOG CPLEX Optimizer. http://www.ilog.com/products/cplex/

16. Gurobi Optimization. http://www.gurobi.com/

17. LLVMPY. <u>http://www.llvmpy.org/</u>

LLVM as infrastructure for compiler development to embedded systems

N. I. Vyukova, V.A. Galatenko, S.V. Samborskij

Abstract. The paper presents an overview of the infrastructure LLVM (Low-Level Virtual Machine) and its evaluation with respect to peculiarities of compilers for microprocessors used in embedded systems. We consider the general structure of the LLVM, an intermediate representation of the program, code generation tools, tools for describing the target architecture.

Анализ трудоемкости процессов переноса информации между базами данных, существующими в средах с разными операционными системами

Г.Л.Левченкова, Г.А.Прилипко, А.Г.Прилипко

Настоящая статья является логическим завершением рассмотренного примера автоматизации организации переноса информации из системы «1С:Бухгалтерия 7.7», функционирующей в ОС Windows, в складскую систему учета, работающую в среде типа ОС Unix [1]. Помимо технологии переноса данных важно понимать степень удобства процедур заполнения форм, а также необходимо оценить трудоемкость всех процессов.

1. Введение

Задача по переносу данных сводится к извлечению информации в электронном виде из системы «1С:Бухгалтерия 7.7» и последующему заполнению форм в системе поступлений в «складскую систему учета» и в системе хранения в «складской системе учета». При этом выполняются определенные процедуры переноса данных. Работы по конвертации можно разбить на несколько основных технологических этапов, трудозатраты по которым зависят от объема и сложности переносимой информации: Данные мигрируют из базы-источника в базуприёмник. База-источник жестко структурирована в части хранящейся информации: фиксированные формы представления данных, ограничения по количеству символов, строк, наличие справочников для хранения наименований и других учетных данных; и имеет налаженное сопровождение, а база-приемник «гибкая» в программном смысле и проста в использовании.

Из базы данных «1С:Бухгалтерия 7.7» (из базыисточника) извлекаются сведения для заполнения формы «складской системы учета» (база-приёмник) следующего вида (см. рис.1):

N	Регистр. пор, номер поступления	Наименова- ние *	Кол-во	Ед. измерения	Цена (без НДС)	Сумма (без НДС)	Ячейка до складу	Основание передачи	Примечание *	Признак обработки	ндс	Сумма с НДС	Номер накладной	Дата накладной	Но менклатур- нъй номер
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	521 2011	Кеп-67	3	ШT.	1.20	3.60	25633		PM	s 08 2011		3.60	874	02.08.11	51
	_	тумблер													
2	521_2011	Стул 5В	1	ШТ.	22221.10	22221.10			OC	1		22221.10	874	02.08.11	03

Рис. 1. Форма «файла поступления» в «складскую систему учета» (имя файла: 521 2011)

1) анализ состава и структуры переносимой информации;

2) выявление типовых ошибок ввода данных и настройка автоматической фильтрации некорректной информации;

3) отработка зависимостей и выбор варианта более полного переноса информации;

4) тестовые выгрузки-загрузки и отбор записей, не удовлетворяющих условиям.

Для корректного переноса данных (до непосредственной записи этих сведений в базуприемник) необходим промежуточный этап визуализации собранной информации с целью анализа корректности переносимых сведений. Технологически промежуточный этап состоит в заполнении первичными данными таблиц значений, допускающих анализ, правку и сохранение в виде временных текстовых файлов.

* эти поля могут иметь продолжение, которое выглядит как символы, расположенные в переделах разделителей соответствующих полей в последующих строках, не имеющих «Регистрационного порядкового номера поступления». Так, слово «тумблер» поля 3 на рис.1 является комментарием к «Кеп-67» - значащему значению поля 3 первой записи документа.

Экспорт данных из «1С:Бухгалтерия 7.7» условно можно разделить на два этапа:

Шаг 1. Запуск системы с указанием информационной базы, из которой требуется загрузить данные.

Шаг 2. Указание параметров в обработке (период выгрузки, типы переносимых документов и справочников, каталог – место выгрузки) и запуск самой обработки.

Остальное программа сделает автоматически.

Однако далеко не всех «программистов Unix» устраивает результат «типовой» выгрузки из

«1С:Бухгалтерия 7.7». В результате «стандартной» обработки по выгрузке, например, счетов-фактур, мы, конечно, получим в файле формата XML все необходимые данные, однако файл получится слишком громоздким последующая обработка И его программными способами может занять длительное время, да и на исследование такого вывола потребуются определенные «силы». Также возможны случаи, когда требования к переносу данных из «1С:Бухгалтерия 7.7» более специфичны, чем функционально заложенные в наборе типовых обработок, входящих в комплект поставки типовых конфигураций. В этих случаях необходимо обратиться к программистам 1С, для написания дополнительных обработок по выгрузке данных.

При переносе данных ИЗ системы «1С:Бухгалтерия 7.7» в систему учета под ОС Unix следует обратить особое внимание на корректное заполнение «полей по длине». Дело в том, что «построчность» обработки файлов в ОС Unix имеет елинственное неудобство: «табличные попя» организуются посредством использования в текстах символов специальных (например, псевдографических). Строка файла не должна быть длиннее определенного количества символов, чтобы её можно было распечатать или увидеть на экране в приемлемом виде. Таким образом, длинное содержимое поля, которое будет импортироваться из «1С:Бухгалтерия 7.7», следует «разрезать» на несколько «строк», первая из которых булет фигурировать в значащей строке «таблицы учета», а остальные будут зрительно располагаться ниже, в строках «комментариях», оформленные в виле кусочков текста, окруженных пробелами И псевдографическими символами-разделителями таблицы Unix-файла. Так, например, три поля таблицы «1С:Бухгалтерия 7.7», которые выглядели как (см. рис. 2), после выгрузки в файл-Unix будут выглядеть, как две строки вида (см. рис. 3):

трудоемкость процессов создания программных средств и занесения с помощью них сведений.

2. Трудоемкость создания программ переноса данных

Когда возникает задача переноса данных из одной системы учета в другую базу данных, большое значение имеет корректная постановка задачи. К процедуре миграции нужно тщательно подготовиться: разработать регламент выполняемых действий. провести перенос на тестовой копии системы, убедиться в корректности настроек процедур, предусмотреть и разработать откат изменений в базе в случае неудачной миграции. Подготовительная работа состоит в согласовании вопросов: какую информацию и в какую структуру предстоит вносить; если необходимо, доработка структуры базы данных; перевод нужных сведений в базу данных.

Основной отличительной чертой ОС Unix, которая с одной стороны облегчает, а с другой, наоборот, затрудняет передачу данных между базами данных Windows и Unix, является построчная организация файлов и строковая обработка содержимого файлов программами и процессами, функционирующими в Unix-среде. Таким образом, чтобы организовать процесс переноса информации между базами данных, существующими в этих ОС, можно использовать комбинацию процедур «выгрузки» сведений из базы, конвертации выходной информации, результатом которой будет построчная организация данных, и, затем, преобразование полученных сведений в строчный формат, который «понимается» интерфейсными программами Unix.

Анализ того, что «нужно иметь на входе в складскую систему», займет не очень много времени, поскольку потребуется лишь проанализировать печатную форму накладной на передачу МЦ на склад.

АСМИТ 3.487-Л 10КПН преобразователь напряжения	2	ШТ	

Рис. 2. Три поля таблицы «1С:Бухгалтерия 7.7»

Ι	АСМИТ 3.487-Л 10КПН	I	2	I	ШТ	Ι
Ι	преобразователь напряжения	Ι		Ι		Ι

Рис. 3. Строки из файла складской системы (Unix)

На рисунке 3 символ «І» является разделителем полей Unix-таблицы, первая строка - значащая, а все пустоты являются на самом деле «пробелами», иначе «таблица» будет нечитаемой:

ІАСМИТ З..487-Л 10КПНІ2ІштІ

Преобразователь напряжения III

Минимизация ручного ввода позволит избежать большого количества неточностей, а наличие соответствующих программных процедур позволит автоматизировать процесс переноса и занесения новой информации, что значительно увеличит скорость обработки документов и файлов данных. Оценим Именно из накладной будут извлекаться сведения, которые следует автоматически перенести в складскую систему из «1С:Бухгалтерия 7.7».

Для выгрузки данных из «1С:Бухгалтерия 7.7» потребуется периодический запуск специального модуля или обработки – либо раз за период, либо после каждого факта появления нового документа. Перед «озадачиванием» специалиста по 1С в части разработки специализированной обработки по выгрузке необходимых данных нужно исследовать установленную систему «1С:Бухгалтерия 7.7». Может оказаться, что некоторые поставляемые с типовой

конфигурацией обработки удовлетворяют необходимым требованиям. Анализ действующей конфигурации в части уже существующих обработок займет не более одного дня, поскольку всего лишь требуется получить результат выгрузки документов и «высмотреть» в файле-результате нужные для переноса значения полей форм. В случае, если ни одна «типовая» обработка не содержит одновременно всех необходимых данных, то написание специального модуля выгрузки (нетипового переноса) данных у специалиста по 1С не займет более 6-ти часов.

Поскольку рассматриваемая в нашем примере «складская система учета» уже имеет сложившуюся структуру данных, то определить требования к порядку заполнения полей форм достаточно просто. Для создания процедур загрузки данных важно также определить заранее ширину полей таблиц, поскольку, как уже отмечалось, в ОС Unix и в рассматриваемой базе данные имеют построчную организацию, следовательно, для соблюдения форматов вывода нужно будет предусмотреть заполнение «пробелами» полей форм с целью выравнивания границ форм. Подробный анализ структуры складской системы не займет более четырех часов, поскольку количество и вид полей, импортируемых из «1С:Бухгалтерия 7.7», уже определены, и их заведомо конечное число, причем зрительно умещающееся в пределах печатной формы накладной на передачу МЦ (на листе формата А4 максимум).

Самая трудоемкая часть работы по организации импорта информации делится на 4 стадии:

1) определение совпадающих (одинаковых изначально и не изменяющихся в процессе занесения) и наследуемых (перезаписывающихся или обновляемых в зависимости от изменения «главенствующего» документа) данных полей форм

складской системы учета. В зависимости от этого будет определяться набор процедур (скриптов), которые следует создать для возможно более гибкой обработки поступающих данных;

 создание программ (процедур), обеспечивающих корректную подготовку и занесение информации в файлы базы данных склада;

3) отладка работы созданных скриптов;

4) создание процедур-комбинаций из уже созданных программ, которые позволят сократить количество запускаемых процедур по обработке ввода данных.

Поскольку пользоваться написанными скриптами надлежит работникам склада, то при написании программ следует заложить разветвленную систему оповещения о «накладках» и ошибках, чтобы сотрудники самостоятельно смогли выявить причину неудачи работы процедур. Скрипты должны быть снабжены подробным описанием параметров, указываемых при вызове процедур, а также сведениями о результатах работы каждой из программ. В этом случае сотрудники смогут самостоятельно комбинировать имеющиеся в их распоряжении программные средства, для более тонкой настройки под свои нужды по вводу информации. Еще раз отметим необходимость создания программ-проверок корректности ввода данных, например, касающихся попыток повторного импорта данных в базу-приёмник за один и тот же период - загруженные ранее документы должны обновляться или корректироваться в соответствии с изменениями в базе-источнике, либо должны быть предусмотрены возможности частичной корректировки данных.

Трудоемкость всех четырех стадий можно оценить в пять «трудодней» максимум, при условии работы над составлением программ 2-х специалистов, хорошо



Рис.4. Трудоемкость переноса сведений из «1С:Бухгалтерия 7.7» в «складскую систему учета»

знающих среду ОС Unix.

Таким образом, через максимум 2 недели (с учетом 3-4 дней, отведенных на «непредвиденные» ситуации) после постановки задачи экспорта данных из «1С:Бухгалтерия 7.7» в «складскую систему учета» мы будем иметь механизм импорта сведений, позволяющий оперативно и с наименьшим искажением информации перенести сведения из одной базы данных в другую.

На рисунке 4 (выше) схематично представлена трудоемкость переноса сведений из «1С:Бухгалтерия 7.7» в «складскую систему учета».

3. Трудоемкость переноса информации

Благодаря существованию возможности написания обработок в системе «1С:Бухгалтерия 7.7», а также использованию типовых обработок, поставляемых в программой, экспорт комплекте с данных производится довольно просто. Чтобы выгрузить сведения из системы «1С:Бухгалтерия 7.7», требуется всего лишь запустить соответствующую обработку (нажатием кнопки). Минимум времени, которое нужно затратить на такую операцию, - несколько секунд (если экспортировать один документ), а максимум – 10-15 минут, если выгрузке подлежит кусок базы данных за определенный период.

Пересылка файлов обмена информацией может быть организована через почтовую службу, либо копированием файлов в папку общего доступа двух операционных систем, а перекодировка файлов нами включена в процедуры импорта. Значит, передача сведений в виде файлов занимает очень мало времени.

Вся трудоемкость, таким образом, определяется

процессом переработки экспортированной информации и её вводом в базу данных склада.

По графикам, приведенным в рисунке 5, можно оценить время работы сотрудников, вводящих информацию базу данных склада В «автоматизированным» И «ручным» способами. (Процесс автоматизированного переноса. изображенный на графике ниже, учитывает время выгрузки информации из «1С:Бухгалтерия 7.7»).

ОС Unix является типичной интерактивной операционной системой, или системой с разделением времени. Это означает, что каждый пользователь системы взаимодействует с системой со своего собственного терминала в интерактивном режиме, задавая системе команды и получая на экран своего терминала ответные сообщения системы. Другими словами – сразу несколько работников склада смогут одновременно обрабатывать каждый свой приходный документ, параллельно, что сокращает время работы по вводу сведений в базу.

Кроме того, любая выполняемая программа может быть запущена в интерактивном режиме как команда shell или выполнена в отдельном процессе, образуемом уже запущенной программой, или может быть включена в скрипт-процедуру. Используя сочетания созданных программистами скриптов, стандартных команд Unix, опираясь на описания shell-процедур, анализируя сведения «об ошибках», возникающих в процессе работы программ, применяя возможности переназначения ввода-вывода, механизм каналов, а также задавая шаблоны имен файлов, требующих обработки, сотрудники способны самостоятельно создать для себя удобный набор команд для ввода и корректировки информации, заносимой в «складскую систему учета».



Рис.5. Скорость ввода данных «автоматизированным» и «ручным» способами

Поскольку складская система (под OC Unix) работает в текстовом режиме, то есть отображает только текст, то работает очень быстро, оперативно загружая информацию ИЗ сети. позволяет просматривать тексты в любой кодировке и применять любые фильтры ввода/вывода. Применяя процедуры загрузки и корректировки данных сотрудники экономят время на вводе информации. Единственное, что следует учесть, оценивая трудоемкость процесса время обучения персонала. Поскольку программисты, создавая shell-процедуры, учитывают пожелания работников, то и скрипты, которые предлагаются для работы, максимально соответствуют требованиям загрузки данных в базу, а, значит, сотрудники быстро осваивают способы применения той или иной программы.

Предположим, что сотрудники уже составили для себя список наиболее часто применяемых в процессе работы команд по вводу данных. Тогда действия по импорту или корректировке будут сводиться к запуску той или иной программы с последующими дополнениями или уточнениями, вносимыми вручную в уже созданные строки файлов «складской системы учета». Так, например, ввод в базу данных накладных с 300 по 499 номер сотрудником «Горлов П.П.» займет несколько минут (а, может, и секунд), если командная строка будет иметь вид (см. рис.6):

производится вручную, в связи с чем наименования материалов, фигурирующие в системах учета склада, нередко отличаются от наименований ТМЦ (товарноматериальных ценностей), числящихся в бухгалтерских базах данных.

Современные системы учета и базы данных имеют возможность экспорта содержащихся в них сведений, а также почти всегда сопровождаются прикладными осуществляющими процедурами, импорт информации. Вполне естественно, что в целях уменьшения ручного ввода и, как следствие, уменьшения ошибок, опечаток и неточностей, для реализации ведения корректного учета И сопровождения перемещения материальных внутри предприятия целесообразно ценностей автоматизировать занесение данных путем переноса из одной системы в другую посредством программных процессов.

В настоящее время неизбежно мирное сосуществование двух операционных систем Unix и Windows; ряд критически важных задач по-прежнему остаются в сфере влияния Unix, поэтому естественно обеспечить их совместную работу и облегчить процессы переноса информации между базами данных, существующими в средах с этими ОС. Однако обе операционные системы имеют существенные различия как на уровне приложений, оборудования, так и

procedura export[34][0-9][0-9]_1c; imena -i"Горлов П.П." [34][0-9][0-9]_prih -n521_2011; prihod [34][0-9][0-9]_hran -f"s_08_2011"

Рис. 6. Пример командной строки для загрузки сведений в базу данных склада

После выполнения такой команды останется только лишь:

- проставить в файле-оглавлении у соответствующих строк признак 3 "занесение всех данных из накладной в систему хранения в «складской системе учета»". Это тоже может быть сделано автоматически, путем выполнения, например, команды:

priznak [34][0-9][0-9]_prih -ini"3"

- внести необходимые комментарии и уточнения (которые, кстати, не требуют немедленного занесения в базу данных, а могут заполняться позднее).

Таким образом, длительность обработки информации будет определяться скоростью работы прикладных программ, а «трудозатраты сотрудника» будут выражаться всего в запуске двух процедур.

4. Заключение

Несмотря на то, что уже существуют системы, позволяющие используя одну базу данных вести бухгалтерский и складской учеты, многие предприятия в силу своей специфики, не имеют возможности сконфигурировать «складскую» составляющую «покупной» системы таким образом, чтобы она удовлетворяла всем необходимым требованиям. Поэтому зачастую организация предпочитает иметь базу данных, созданную «своими силами» и заточенную под конкретные нужды предприятия. Занесение информации в такие складские системы менталитета работающего с ними персонала. Нередко ситуация разрешается посредством создания промежуточного ПО, ориентированного, в частности, на организацию обмена информацией и данными между системами, функционирующими в различных OC.

Благодаря тому, что платформа «1С» предоставляет разнообразные инструменты для обмена данными и интеграции прикладных решений, экспортимпорт данных может быть осуществлен посредством комбинаций использования обработок, созданных специально либо типовых, поставляемых в комплекте «1С:Бухгалтерия 7.7», и прикладных программ, функционирующих как в среде ОС Windows, так и в среде OC Unix.

ОС Unix позволяет пользователю достаточно быстро создавать простые программные комплексы, выполняющие повторяющиеся действия как по команде пользователя, так и в автономном режиме. Этому способствуют такие основные особенности системы, как: универсальная форма представления информации - текстовый файл, понимаемый как последовательность символов, некоторые из которых имеют специальный смысл (разделяют строки (записи), поля и слова); возможность переназначения ввода-вывода и использования регулярных выражения формальный, достаточно гибкий и более-менее интуитивный способ указать программе, что следует искать в потоке данных, стандартные средства текстовой обработки опираются именно на этот механизм.

В процессе сопровождения и модернизации сложных программных комплексов внедрение и поддержка соответствующих правил обработки информации требует незначительных трудозатрат и упрощает процесс переноса данных между базами данных.

Литература

1. Г.Л.Левченкова, А.Г.Прилипко. Перенос информации между базами данных, существующими в средах с разными операционными системами // Труды НИИСИ РАН. - Том 1 №1. Математическое и компьютерное моделирование систем. Теоретические и прикладные аспекты. ISSN 2225-7349, Москва, 2011, С. 77-85

2. Г.Л.Левченкова. Система "ЗАКУПКА И РАСПРЕДЕЛЕНИЕ". Обеспечение потребностей предприятия в комплектующих, материалах и готовых изделиях // под редакцией академика РАН В.Б.Бетелина, Москва, НИИСИ РАН, 2006, С. 76-234

3. Введение в конфигурирование 1С:Бухгалтерии 7.7. - М.: Фирма «1С», 2000

4. 1С:Бухгалтерия 7.7 Руководство по ведению учета (Описание типовой конфигурации редакции 4.5). Введение в конфигурирование 1С:Бухгалтерии 7.7. - М.: ЗАО «1С», 2004

5. Д. Пик, Т О'Райли, М. Лукидис, Т. О'Райли. Unix: инструментальные средства. 2-е издание. : BHV-Киев, 1999

6. Виктор Вагнер. True Unix GUI. // LinuxRSP, 1999-2011

The labor input analysis of information transfer processes between the databases existing in environments with different operating systems

G.L.Levchenkova, G.A.Prilipko, A.G.Prilipko

Abstract. This paper summarizes the previously described experience [1] of automation of data translation from the "1C Buhgalteria 7.7" system running under Microsoft Windows to a warehouse accounting system running under a Unix-like operating system. Besides the technology of data translation, it is important to evaluate the convenience of data form filling procedures as well as the complexity of the entire process.

У истоков вычислительной математики и программирования: Лазарь Аронович Люстерник

В.Б.Демидович¹, В.М.Тихомиров²

1 – кандидат физико-математических наук, 2 – доктор физико-математических наук



Статья посвящена творчеству замечательного советского учёного, специалиста в области фундаментальной и прикладной математики, члену-корреспонденту АН СССР Лазарю Ароновичу Люстернику (31.12.1899-23.07.1981).

Введение

В своём письме (от 4 августа 1915 года) выдающийся математик Николай Николаевич Лузин (1883-1950) сетовал другу (богослову-философу с математическим образованием) Павлу Александровичу Флоренскому (1882-1937), что "у нас в России не было живой науки, было преподавание, была таинственность, тьма, стена, пробить которую предоставлялось каждому, кто как хочет и может. Не было *школы*, как живой творческой атмосферы". И он создал *Московскую математическую школу*. Ярким представителем этой школы является Лазарь Аронович Люстерник, стоявший, в частности, у истоков исследований по вычислительной математике и программированию в СССР.

1. Краткая биография

Лазарь Аронович Люстерник родился 31 декабря 1899 года в небольшом городке Царства Польского Здуньска Воля (Zducska Wola), близ Лодзи. Его отец, Арон Ерукимович Люстерник, был директором и совладельцем текстильной фабрики, а мать, Мария Фёдоровна (урожд. Русинова), не работала, воспитывая семерых детей.

В 1908 году Лазарь Аронович поступил в Лодзинскую гимназию. Но в 1915 году, с развитием немецкого наступления на Российском фронте в ходе І Мировой войны, семья эвакуировалась в Смоленск, и Л.А. стал учеником Смоленской общественной гимназии (кстати, той же, которую в 1913 году окончил Павел Сергеевич Александров (1896-1982)). В 1918 году, по окончании Смоленской средней школы (так стала именоваться эта гимназия), он уехал в Москву для дальнейшего своего образования, а его семья возвратилась в провозгласившую независимость Польшу. В том же 1918 году Л.А. поступил на Физикоматематический факультет «1-го Московского университета» (так тогда стап называться «Императорский Московский университет»: «Московским Государственным университетом», /коротко, «МГУ»/ его назвали в 1930 году, «МГУ им. М.Н.Покровского» - в 1932 году, снова просто в 1937 году, -«МГУ «МГУ» а им М.В.Ломоносова» - в 1940 году).

По разному сложилась судьба возвратившихся в Польшу членов семьи Лазаря Ароновича. Его отец скончался ещё до II Мировой войны. Мать, два брата и одна сестра Л.А. погибли в Польше во время её немецкой оккупации. Другая сестра Л.А. перед оккупацией Польши успела эмигрировать во Францию. Ещё две сестры Л.А., оставаясь в Польше, оккупацию пережили и умерли уже в послевоенные годы (одна из них, став актрисой под артистическим псевдонимом «Иза Фаленьска», в 1946 году также переехала во Францию).

Обучаясь в Университете, Лазарь Аронович вошёл в круг учеников Николая Николаевича Лузина. Его студенческие годы совпали с той порой нашей страны, о которой П.С.Александров писал так: «То были годы необычайного подъёма И увлечения внезапно открывшимися новыми творческими возможностями, годы подлинного цветения для многих молодых людей, впервые вкусивших радость творческого соприкосновения с наукой. Мало найдётся в истории математики периодов столь горячего энтузиазма, как начало двадцатых годов в Московском университете, когда в столь короткий срок, буквально в несколько лет, возникла большая научная школа, в значительной степени определившая развитие математики в нашей стране и сразу выдвинувшая целый ряд выдающихся ученых".

Лазарь Аронович принадлежал ко второму поколению «Лузитании» - так назывался коллектив

молодых математиков, группировавшихся вокруг их общего учителя Н.Н.Лузина. Молодости «Лузитании» Л.А. посвятил, впоследствии, ряд замечательных статей и поэтических строк. Среди них такая: «Дни незабвенной Лузитании, дни вдохновений и исканий». Но вот как Лазарь Аронович описывал (в стихотворной форме) атмосферу в Университете к началу 1930-ых годов («Божество», с которого начинается нижеследующий отрывок - это сам Николай Николаевич Лузин):

«...Но Божество уж окружало Созвездие полубогов: Иван Иванович Привалов, Димитр Евгеньевич Меньшов, И Александров остро взвинчен, И милый Павлик Урысон, И философствующий Хинчин, И множество других персон ...»

А в 1936 году, «по указанию сверху», началась всенародная травля Н.Н.Лузина. К сожалению, в ней приняли участие и многие члены «Лузитании», в том числе (пусть в мягкой форме) и Лазарь Аронович. Что же тут поделать - из песни слов не выкинешь ...

По окончании «1-го Московского университета» по специальности «математика» (1921) Лазарь Аронович был оставлен в качестве «научного сотрудника 2 -го разряда» в только что созданном при Университете Научно-исследовательском институте математики и механики (НИИММ 1-го МУ). А с организацией при этом Институте «аспирантуры» (1922) он становится её аспирантом.

В 1924 г., будучи ещё аспирантом, Лазарь Аронович впервые выступает в Московском математическом обществе с докладом «О кривизне пространств n измерений, вложенных в (n+1)-мерное пространство», а затем публикует по этой тематике (1925) свою первую работу. Его заключительная аспирантская научную работа «О прямом методе в вариационном исчислении и дифференциальных уравнениях» (1926),представлению НИИММ 1-го МУ, награждается премией Главного управления научными учреждениями РСФСР. В том же году он командируется в Данциг (ныне «Гданьск») для участия в Международной математической конференции.

Лазарь Аронович обладал замечательным чувством юмора. После своего первого доклада на Московском математическом обществе, он был принят в члены этого Общества. И тогда же ему был вручён ключ от специального туалета, куда допускались лишь члены Общества. «Это был первый закрытый распределитель в моей жизни» - говорил Л.А. об этом впоследствии.

В 1927 году Лазарь Аронович стал приват-доцентом 1-го МУ и начал читать на его физико-математическом факультете свои спецкурсы. В 1928 году он участвовал в работе VIII Международного математического конгресса в Болонье, где выступил с докладом «Топологические методы в дифференциальной геометрии».

В том же 1928-ом году Лазарь Аронович был избран

профессора Нижегородского на должность Госуниверситета. Преподавал он там на физикоматематическом отделении Педагогического факультета (в 1930 году этот факультет был реорганизован в Нижегородский педагогический институт). Среди его нижегородских учеников был Николай Николаевич Баутин (1908-1993). В статье Е.А.Андроновой, Б.Н.Скрябина «Николай Николаевич Баутин (к 100летию со дня рождения)» (Нижний Новгород, «Математика в высшем образовании», № 6 (2008), с. 111можно прочесть такое занимательное мнение 122) Н.Н.Баутина о преподавательском искусстве молодого Л.А.:

«В те времена (до 1932 года) в ВУЗах практиковался лабораторно-бригадный метод обучения ... Его особенность состояла в том, что традиционная форма обучения – лекционное изложение материала – отвергалось, а на смену предлагались занятия в предметных лабораториях во главе с преподавателем-консультантом, которому мелом, тряпкой и доской пользоваться не рекомендовалось....

Лазарь Аронович ... из сложившейся ситуации с бригадным методом выходил следующим образом. Во время занятий он садился на стул лицом к слушателям и, если возникала необходимость, писал формулы мелом на доске, не вставая и почти не поворачивая головы. Когда свободное место на доске, находящееся в пределах его досягаемости, исчерпывалось, он, подпрыгивая вместе со стулом, перемещался дальше, и таким образом получал новое место для написания формул ...»

В 1931-ом году Лазарь Аронович был избран на должность профессора МГУ. Звание профессора, а одновременно (без защиты диссертации) и степень доктора физико-математических наук, ему присвоили в 1935 году, с образованием в нашей стране ВАК.

Широта научных интересов Лазаря Ароновича была необыкновенной: дифференциальные уравнения, топология, вариационное исчисление, функциональный анализ, геометрия, вычислительная математика, специальные функции и многое другое. Он был и геометром и аналитиком. Для его математического стиля характерно движение «от простого к сложному» - в основе далёких обобщений у него всегда лежала простая модель.

О некоторых его работах, в частности, об его исследованиях по методу сеток и вариационному исчислению, мы поговорим ещё более подробно. Здесь же мы лишь приведём его шутливые поэтические строчки по этому поводу:

«...Я стал работать в направленьях Тогда в Москве совсем не модных -Вариационном исчисленьи, Задачах в частных производных... ОЯ метод сеток развивал ...»

В 1920-ые - 1930-ые годы Лазарь Аронович (совместно со своим другом Львом Генриховичем Шнирельманом (1905-1938)) создаёт совершенно новую математическую область - топологические методы нелинейного анализа. Л.А. вводит новый гомотопический инвариант - категорию, и в работе, совместной с Л.Г.Шнирельманом, успешно применяет это понятие к оценке числа критических точек гладкой функции на гладком многообразии. Итогом их совместной деятельности явилось решение знаменитой проблемы Пуанкаре о трёх геодезических.

Предвоенные исследования Лазаря Ароновича были посвящены теории обыкновенных дифференциальных уравнений (им были получены замечательные результаты по качественному поведению собственных функций нелинейных задач типа Штурма-Лиувилля), функциональному аннализу (доказанная им теорема касательном пространстве лежит в самой основе 0 современной теории экстремальных задач, о чём также пойдёт речь далее), геометрии, (общеизвестным стало его обобщение неравенства Брунна-Минковского об объёме векторной суммы выпуклых тел на случай произвольных множеств).

В 1934 году Лазарь Аронович Люстерник, оставаясь механико-математическом «совместителем» на факультете «МГУ им. М.Н.Покровского» (факультет, далее сокращённо называемый «Мехмат МГУ», был создан в Университете год назад - весной 1933 года, а 1-ым деканом его был Владимир Васильевич Голубев (1884-1954)),перешёл на «основную работу» в Математический Институт им. В.А.Стеклова AH CCCP (институт, коротко, называемый тогла «МИ АН СССР», а ныне «МИ РАН», был создан в Москве как раз в 1934 году, и его 1-ым директором стал Иван Матвеевич Виноградов (1891-1983)).

В 1942-1948 годы Лазарь Аронович возглавлял в МИ АН СССР Отдел приближённых вычислений, где выполнялись «спецзадания» оборонного значения. В частности, в 1942 году им были созданы методы эффективного расчёта аэро- (и морской) навигации по радиопеленгу, позволившие разработать «штурманскую таблицу» быстрого определения курсового угла и расстояния - необходимых пилоту сведений для прокладывания нужного курса. До Войны эти сведения вычислялись достаточно сложно. В разработанной же Л.А. таблице, по координатам начального и конечного пунктов, сразу же определялись путевые углы и длина *пути* при полёте самолёта «по геодезической линии». В теоретическом плане им была решена задача о том, как функцию от трёх переменных оптимально представить суперпозицией функций двух переменных – одну заданной точно, а другую - с достаточной степенью точности.

Таблицы эти были сразу внедрены в военную практику. Вот как это произошло, согласно воспоминаниям сотрудника Отдела приближённых вычислений МИ АН СССР Израиля Яковлевича Акушского (1911-1992) об этой деятельности Лазаря Ароновича.

Шла война, время было суровое. Неожиданно Лазарь Аронович был вызван на Лубянку. В течение нескольких часов, ни жив ни мёртв, он ждал своей участи у кабинета

какого-то высокого чина НКВД. Тот принял его только глубокой ночью. Размахивая перед Л.А. какой-то небольшой книжицей типа карманного блокнота, хозяин кабинета принялся осыпать его нецензурной бранью за то, что тот, якобы, зря ест государственный хлеб, не помогает Родине в трудный час, и грозился «стереть его в порошок». Насладившись произведенным эффектом, начальник перешёл к делу. А суть дела состояла в следующем. У сбитого немецкого лётчика нашли эту самую книжицу, и когда спросили пленного, для чего ему она, он объяснил (крайне удивившись, что этого не знают русские), что по ней в течение минуты он может проложить курс следования самолёта в любую заданную точку. А у нас на это тратились часы. Вот Л.А. и было приказано в кратчайший срок создать такую же для советских лётчиков. Л.А. просил на её книжиич разработку месяц, ему дали две недели, а создал он её в течение нескольких дней. Ещё через неделю его «штурманская книжка» была направлена по фронтам Красной Армии ...

Для реализации таких работ требовались вычисления большого объёма. Это привело к пониманию необходимости создания в МИ АН СССР специальных вычислительных средств.

В связи с этим в 1943 году, по инициативе бывшего тогда заместителя директора МИ АН СССР Анисима Фёдоровича Берманта (1904-1959), при Отделе Лазаря Ароновича организуется «первая в СССР экспериментальная счётная станция, оборудованная рядом счётных приборов, в том числе *счётноаналитических машин* (сокращённо, *САМ*), т.е. счётных автоматов перфорационного действия» (см. УМН, т. 1 (1946), вып. 3-4, с. 202-205).

Вообще уже в те годы Лазарь Аронович проявил глубокое перспектив понимание развития математических методов прикладных расчётов. Некоторые его ученики и сотрудники уверяют, что сами «вычислительная математика» термины И «вычислительная техника» были впервые употреблены именно Л.А.! Потому не удивительно, что в 1945-1948 годы Л.А. был назначен заместителем директора МИ АН СССР, курирующим в Институте деятельность по развитию вычислительной математики и созланию счётно-аналитических машин.

В 1946 году Лазарь Аронович был избран членомкорреспондентом АН СССР. В том же году он был удостоен и звания лауреата Сталинской премии высшего премиального отличия в те времена.

В 1948-1954 годы Лазарь Аронович работал в Институте точной механики и вычислительной техники АН СССР (кратко, ИТМиВТ АН СССР). Этот Институт был организован в Москве как раз в 1948 году на базе соответствующего отдела МИ АН СССР. Л.А. был одним из инициаторов открытия в СССР такого Института, также как и создания в МГУ и в некоторых других ВУЗах нашей страны кафедр вычислительной математики.

Первым директором ИТМиВТ АН СССР был генерал-лейтенант авиации, академик АН СССР Николай

Григорьевич Бруевич (1896-1987). При нём в ИТМиВТ АН СССР в 1952 году появилась советская ЭВМ -БЭСМ-1, одним из разработчиков которой был Владимир Андреевич Мельников (1928-1993). А самая первая в СССР ЭВМ - МЭСМ - была создана в Киеве, под руководством Сергея Алексеевича Лебедева (1902-1974), лишь двумя годами раньше. В 1953 году С.В.Лебедев переехал в Москву и возглавил ИТМиВТ (ныне Институт носит его имя).

В этих обстоятельствах в ИТМиВТ АН СССР Лазарь Аронович стал активно заниматься кругом проблем, связанных с программированием. Благодаря ему в Советском Союзе появилась первая научная группа по работе на ЭВМ (вместо ЭВМ тогда ещё применялась аббревиатура АЦМ – автоматическая цифровая машина), а затем и первая советская книга (для служебного пользования) по программмированию (Л.А.Люстерник, А.А.Абрамов, В.И.Шестаков. М.Р.Шура-Бура «Решение математических задач на автоматических цифровых машинах». М., "Изд-во АН СССР", 1952). Л.А. явился и первым в СССР лектором по курсу программирования.

В 1950-ые - 1960-ые годы у Лазаря Ароновича происходит новый творческий взлёт - и появляется блестящий цикл его работ (с Марком Иосифовичем Вишиком (1921-2012)) по асимптотическим разложениям решений уравнений с малым параметром, по задачам с барьерами и с быстро меняющимися граничными функциями, по возмущению несимметрических матриц и операторов.

Огромную роль сыграл Лазарь Аронович в истории механико-математического факультета МГУ. Он кто прочёл на Мехмате МГУ курс был первым, лекций по функциональному анализу и открыл по нему (совместно с переехавшим из Германии в Советский Союз в 1932 году Абрамом Иезекииловичем Плеснером (1900-1961)) научно-исследовательский семинар. Вместе с Михаилом Александровичем Лаврентьевым (1900-1980) он коренным образом модернизировал мехматский курс вариационного исчисления. Одним из первых он объявил факультетский спецсеминар по вычислительной математике - ещё до организации в 1949 году самой кафедры вычислительной математики Мехмата МГУ, профессором которой он впоследствии работал (до перехода в 1970 году на кафедру ОПУ – общих проблем управления). Он же организовал затем спецсеминар по математическим вопросам управления производством. Добавим, что он был организатором и заведующим кафедрой функционального анализа Мехмата МГУ, одним из инициаторов проведения первой Московской школьной математической олимпиады, первым ответственным редактором журнала «Успехи математических наук». Среди учеников Лазаря Ароновича 20 докторов наук и около 50 кандидатов наук.

С 1969 года под редакцией Лазаря Ароновича на Мехмате МГУ стали выходить сборники трудов из серии «Математические вопросы управления производством».

Серия возникла в связи с работой одноименного семинара под его руководством, но в дальнейшем включила в свой круг интересов и другие исследования, проводившиеся на кафедре ОПУ. Всего было издано девять сборников этой серии. Первый из соавторов данной статьи, являвшийся ответственным секретарём редколлегии указанной серии, с благодарностью вспоминает свои «исторические беседы о Мехмате МГУ» с Лазарем Ароновичем.

Как уже отмечалось, последние годы своей жизни Лазарь Аронович провёл на кафедре общих проблем управления Мехмата МГУ. Дело в том, что С в 1970 году в МГУ факультета организацией Вычислительной математики и кибернетики (кратко, ВМиК) во главе с Андреем Николаевичем Тихоновым кафедра вычислительной математики (1906-1993). Мехмата МГУ перешла туда. А Лазарь Аронович не пожелал расставаться с Мехматом МГУ, и из всех мехматских кафедр выбрал для своей дальнейшей работы кафедру ОПУ. Мы всегда воспринимали этот выбор его как «подарок судьбы».

Лазарь Аронович, как и многие его великие сверстники, был легендарной фигурой. Но про него можно было услышать и много смешных рассказов. Часто люди боятся показаться смешными, и, как правило, это свидетельствует либо об ограниченности их личности, либо о том, что они преисполнены не добра, а недоброжелательства. Однако светлая личность высвечивается лучами юмора. Вот два эпизода из жизни Лазаря Ароновича, подтверждающие сказанное.

Один эпизод, рассказанный Николаем Николаевичем Константиновым, был связан с приёмом экзамена Лазарем Ароновичем у Евгения Михайловича Ландиса (1921-1997), только что демобилизованного, ходившего на факультет ещё в гимнастёрке. Е.М.Ландис, будучи тогда студентом, должен был сдать Л.А. спецкурс. Они встретились в назначенное время. Л.А задал ему вопрос и куда-то вышел. Вскоре он вернулся, но не начал экзамена, а стал разговаривать с Е.М.Ландисом о математике. Это длилось около часа. Наконец, Л.А. нетерпеливо посмотрел на часы и сказал: «Ну куда же он пропал?» «Кто-то должен подойти?» - спросил Е.М.Ландис. «Ну конечно, где же тот студент, которого мы с Вами должны были экзаменовать?»

Другой эпизод из жизни Лазаря Ароновича был связан с его женой Ираидой Фоминичной. Это была строгая женщина. Сначала все как-то весело восприняли новость о женитьбе Лазаря Ароновича, но познакомившись с женой, шутить перестали. Даже Нина Карловна Барии (1901-1961) - ох, какая остренькая на язычок, не упускавшая случая съязвить, и та сочинила в стихах не сатиру, а оду:

Был наш Лазарь неизменно И нечёсан, и небрит. Но жена его мгновенно Привела в нормальный вид.

Ираида Фоминична не требовала, не просила, не

умоляла – она ПОВЕЛЕВАЛА. Иногда словом, чаще взглядом. И ослушаться было невозможно. Мужа она в разговоре называла ОН. При этом, как правило, объяснялось, что ОН делал что-нибудь не то или не так.

Один раз она повелела ему сводить себя в Большой театр. Сказано - сделано. Открылся занавес, и Ираида Фоминична стала наслаждаться искусством. А Лазарь Аронович продолжал думать о некоторых проблемах нелинейного анализа. В антракте Ираида Фоминична, отдав Л.А. свой номерок, поручила ему спуститься в гардероб и принести ей платок, который остался в кармане её пальто. Л. А. пошёл выполнять поручение. Он усвоил первую цель - надо идти в гардероб. Но пока он шёл туда, продолжая думать о проблемах нелинейного анализа, вторая цель - платок - как-то растворилась в его сознании. Он дошел до гардероба и отдал гардеробщице свой номерок. Она принесла ему его пальто. Он машинально его принял, оделся, вышел на улицу, взял такси и оказался дома.

Дома жены не было. Это его не удивило: ведь она собиралась в театр ... Он поразмышлял еще над своими проблемами. Дело не пошло, и он прилёг на диван.

Ираида Фоминична весь второй акт сидела, как на иголках. Где ОН? Что случилось? Случился приступ? Отвезли в больницу? Едва дождавшись антракта, бедная женщина побежала в гардероб, бурной атакой добыла пальто без номерка (в ярости она сметала все препятствия), взяла такси и примчалась домой к телефону, чтобы звонить в больницы, в милицию, в морг ... Но тут она обратила внимание на свет в гостиной. Рванулась туда ...

На диванчике, в своём парадном костюме, не сняв ботинок, сном праведника спал ОН ...

В 1975 году Лазарь Аронович вышел на пенсию. Незадолго до этого он определил следующим образом тот момент, когда профессор МГУ должен выходить на пенсию. А именно, многие знают, что для того, чтобы открыть парадную дверь главного входа в МГУ требуются немалые усилия. Так вот Л.А. высказал мысль, что профессор МГУ может работать лишь до того момента, пока он сам, без посторонней помощи, способен открывать эту дверь.

Лазарь Аронович скончался 23 июля 1981 года. В память о нём в сентябре 1999 года кафедра ОПУ Мехмата МГУ, совместно с МИ РАН и Международным Банаховым центром, провели в Варшаве и Бендлево (Bedlewo - местечко близ Познани) мини-симпозиум, посвященный 100-летию со дня рождения Лазаря Ароновича Люстерника

2. О математическом творчестве

А теперь коснёмся подробнее *лишь трёх* математических тем в творчестве Лазаря Ароновича Люстерника.

1. Лазарь Аронович стоял у истоков прямых методов и методов сеток в численном анализе.

Ещё в 1924 году он первым применил метод конечных разностей для доказательства существования решений уравнений с частными производными (уравнений Лапласа) в криволинейных областях. «Мне показалось естественным, - пишет он по этому поводу в своих воспоминаниях, - рассматривать вариационные задачи как пределы соответствующих задач лля в конечном числе точек. Тогда, функций, заданных например, условие Якоби положительности второй вариации возникает как предельное для условия Сильвестра. Поскольку в аспирантскую программу входило доказательство существования решения задачи Дирихле для плоского уравнения Лапласа, я попробовал доказать это предельным переходом от соответственной «сеточной задачи», и был даже удивлен что такой «грубый» метод доказывал теорему в «тонких» случаях - даже когда граница не содержит континуальных компонент». От этой работы потянулся след, ощутимый и в наши дни. Тогда же состоялись плодотворные научные контакты Лазаря Ароновича с Иваном Георгиевичем Петровским (1901-1973), приведшие к первой публикации И.Г.Петровского.

2. Важными результатами Лазаря Ароновича служат его «абстрактные обобщения» классического метода Ньютона и теоремы об обратной функции.

Свои основные открытия в области математического анализа Исаак Ньютон совершил в начале шестидесятых годов XVII столетия, когда ему было около двадцати лет. В частности, он научился решать (конечно, *приближённо*) уравнения f(x) = y, где f - функция одного переменного. Свой метод он проиллюстрировал на примере решения уравнения $x^3 - 2x = 5$ (см. И.Ньютон «Математические работы». М-Л, Издательство технико-теоретической литературы, 1937, с. 9). Поясним, как он это делал.

Первой аппроксимацией, для нахождения корня x данного уравнения, у Ньютона служил выбор числа 2 в качестве его начального приближённого значения, то есть он положил $x_1 = 2$.

Далее, представляя x в виде суммы x = 2 + p и, подставляя это представление x в исходное уравнение $x^3 - 2x = 5$, он приходил к уравнению $p^3 + 6p^2 + 10p - 1 = 0$,

«у которого, - пишет Ньютон, - следует определить корень p, чтобы прибавить его к первому результату. Отсюда (пренебрегая $p^3 + 6p^2$ по малости) имеем приблизительно 10p-1=0 или p=0.1». Тем самым, второй аппроксимацией, для нахождения корня x данного уравнения, у Ньютона служило число $x_2 = 2.1$.

Далее, представляя уже *р* в виде суммы

p = 0.1 + q и, подставляя это представление p в уравнение $p^3 + 6p^2 + 10p - 1 = 0$, Ньютон получал уравнение вила $q^{3} + 6.3q^{2} + 11.23q + 0.061 = 0$. Снова, «пренебрегая по малости» величиной $q^3 + 6.3q^2$, он равенство 11.23q + 0.061 = 0, откуда получал заключал, что $q \approx -0.0055$. Тем самым, получалось, $p \approx 0.1 - 0.0055 = 0.0945$ что И $x = 2 + p \approx 2.0945$. Поэтому третьей аппроксимацией, для нахождения корня х исходного уравнения, у Ньютона служило число $x_3 = 2.0945$.

После следующей итерации Ньютон получил, что $x \approx 2.09455147$, то есть он вычислил корень исходного уравнения с точностью до восьмого знака после запятой, так как «точным ответом» здесь служит «равенство» x = 2.094551481...

В нашем описании этого процесса по сути дела изложен *метод Ньютона* для решения уравнения f(x) = y, состоящий в том, что, после выбора начального приближения x_1 , далее применяется итеративная процедура:

 $x_{k+1} = x_k + (f'(x_k))^{-1}(y - f(x_k)) \ (k = 1, 2, ...).$

Роль, которую суждено было сыграть методу Ньютона в истории математики, совершенно исключительная. Одно из важнейших приложений его доказательство теоремы об обратной (и неявной) функции.

Сформулируем сначала теорему об обратной функции в самом простейшем (опять-таки одномерном) случае.

Теорема 1 (об обратной функции в одномерном случае).

Пусть f - функция одного переменного, определённая в окрестности нуля, равная нулю в нуле, непрерывно дифференцируемая в окрестности нуля, причём $f'(0) \neq 0$.

Тогда найдётся такое число $\delta > 0$, что для любого числа у такого, что $|y| < \delta$,

существует единственное решение x(y) уравнения f(x) = y.

Смысл этой теоремы заключается в том, что, подбирая «достаточно точно» соответствующие начальные приближения, обратная функция x(y) уравнения f(x) = y может быть «достаточно точно» построена с помощью *модифицированного* метода Ньютона:

$$x_1 = 0, \quad x_{k+1} = x_k + (f'(0))^{-1}(y - f(x_k)), \quad k \ge 1.$$

Этот результат естественно приписать самому Ньютону. Двумерные обобщения такой теоремы появились лишь в XIX веке. В конце XIX века и в первом десятилетии XX столетия теорема получила многомерное развитие. А в 1934 году Лазарь Аронович дал бесконечномерное обобщение этой теоремы. Оно было опубликовано в журнале «Математический сборник» (Л.А.Люстерник «Об условных экстремумах функционалов». Матем. сб., т .41 (1934), N 3 с.390-401).

Именно, сначала Лазарем Ароновичем было получено обобщение Теоремы 1 на бесконечномерную ситуацию в виде:

Теорема 2 (об обратной функции в бесконечномерном случае).

Пусть X и Y - банаховы пространства, U окрестность нуля в X, $F: U \to Y$ - отображение из X в Y, непрерывно дифференцируемое в окрестности нуля. Положим, что F(0) = 0, причём F'(0) отображает X на всё Y.

Тогда найдётся такое $\delta > 0$, что для любого $y \in Y$ такого, что $||y||_{Y} < \delta$, существует решение x(y) уравнения F(x) = y такое, что $||x(y)||_{X} \le K ||y||_{Y}$ (для некоторой константы K > 0).

Упомянутое здесь понятие банахова пространства обобщает понятие конечномерного евклидова пространства. Ныне студенты университета знакомятся с этим понятием на первых курсах - оно (как и начала теории банаховых пространств) стали теперь абсолютно общепринятыми.

А в том далёком 1934 году теории банаховых пространств исполнилось всего лишь два гола. Автором теории был замечательный польский математик Стефан Банах /Stefan Banach/ (1892-1945), опубликовавший в 1932 году свой знаменитый мемуар «Theorie des operations lineaires» (Теория линейных операций), в котором он изложил основы своей теории, ставшей одной из существенных составных частей функционального анализа. Знания по теории банаховых пространств в те годы в Москве черпались из двух экземпляров этого мемуара Банаха: один принадлежал А.И.Плеснеру (и потому, вероятно, был доступен Лазарю Ароновичу), другой - Андрею Николаевичу Колмогорову (1903-1987). Поразительно, что Лазарь Аронович так быстро сумел извлечь из новой теории столь фундаментальный результат.

Читатель, не владеющий понятием банахова пространства, может считать, что X - это, например, трёхмерное, а Y - двумерное пространство с евклидовой нормой. При доказательстве Теоремы 2

Лазарь Аронович использовал только самые обычные свойства нормы (в основном, *неравенство треугольника*) и один из важнейших принципов линейного анализа (т.е. теории банаховых пространств), а именно, *теорему* Банаха об обратном операторе. Сформулируем её здесь в ослабленном варианте, как теорему о правом обратном операторе, достаточном для доказательства Теоремы 2:

Если X и Y - банаховы пространства и $\Lambda: X \to Y$ - непрерывный линейный оператор, отображающий X на всё Y, то существуют оператор $R: Y \to X$ и константа C > 0 такие, что $R\Lambda x = x$ ($\forall x \in X$) и || R(y) || $\leq C || y ||$.

Лазарь Аронович использует подобное утверждение при доказательстве Теоремы 2, но, почему-то, при этом ссылается не на Банаха, а на немецкого математика Феликса Хаусдорфа /Felix Hausdorf/ (1868-1942), не указывая даты публикации его работы. Было бы полезно разобраться, что это за работа Хаусдофа и как она связана с исследованиями Банаха.

3. Для общей теории экстремума фундаментальным результатом оказалась теорема Лазаря Ароновича о касательном подпространстве в банаховом пространстве.

Пусть X - банахово пространство и M некоторое его подмножество. Элемент x называется *касательным* к M в точке $\hat{x} \in M$, если существует отображение $r:[-1,1] \to X$ такое, что $\hat{x}+tx+r(t) \in M$ ($\forall t \in [-1,1]$) и ||r(t)||=o(t) при $t \to 0$.

Множество касательных элементов к M в точке \hat{x} обозначается $T_{\hat{x}}M$. Если это множество является подпространством X, то оно называется касательным подпространством к множеству M в точке \hat{x} .

Теорема 3 (Люстерника о касательном подпространстве).

Пусть X и Y - банаховы пространства, U окрестность точки \hat{x} из X и $F: U \rightarrow Y$ отображение, непрерывно-дифференцируемое в U, причём $F'(\hat{x})$ отображает X на всё Y.

Тогда если $M = \{x \in X \mid F(x) = F(\hat{x})\},$ то $T_{\hat{x}}M$ является касательным подпространством к множеству M в точке \hat{x} , причём $T_{\hat{x}}M = \text{Ker}F'(\hat{x}),$ где под $\text{Ker}F'(\hat{x})$ понимается ядро отображения $F'(\hat{x}).$

Происхождение этой теоремы, по-видимому, связано с работой Лазаря Ароновича (совместно с М.А.Лаврентьевым) над учебником по вариационному

исчислению. Он был издан в следующем году (М.А.Лаврентьев, Л.А.Люстерник «Основы вариационного исчисления». М-Л., ОНТИ, 1935, тт. 1 и 2.). В упомянутой ранее статье Лазаря Ароновича 1934 года из «Математического сборника» явно указывается, что одной из целей статьи является предстоящее включение вариационного исчисления в общую схему теории экстремума в функциональном анализе. Но почему-то в вышеуказанном учебнике теория условий экстремума была построена традиционным путём, без применения методов функционального анализа.

Спустя сорок пять лет после выхода вышеуказанного двухтомника, второй из соавторов настоящей статьи написал (совместно с Владимиром Михайловичем Алексеевым (1932-1980)) учебник по оптимальному управлению (В.М.Алексеев, В.М.Тихомиров «Оптимальное управление». М.: Наука, 1979). Книгу эту он подарил Лазарю Ароновичу. А через некоторое время (когда В.М.Алексеева уже не было в живых) Лазарь Аронович вдруг позвонил ему, поблагодарил его за подаренную книгу и сказал с некоторым смущением, что «даже и не подозревал о том, что его теорема может лечь в основание общей теории экстремума».

Скорее всего, Лазарь Аронович немного лукавил. Конечно, он нечто подобное «подозревал», но что-то, наверное, отвлекло его тогда от осуществления широкой программы модернизации теории экстремальных задач. Потом же он просто позабыл о своих замыслах.

А теорема Люстерника ныне - одна из самых цитируемых его теорем ...

At the roots of Computational Mathematics and Programming: Lazar Aronovich Lusternik

V.B. Demidovich, V.M. Tikhomirov

Abstract: This article is devoted to the work of a remarkable Soviet scientist, specialist in the field of Fundamental and Applied Mathematics, corresponding member of the Academy of Sciences of the USSR Lazar Aronovich Lusternik (31.12.1899-23.07.1981).