

Разработка таблицы поколений LDGT в архитектуре микропроцессоров «Эльбрус»

Выпускная квалификационная работа
на соискание степени бакалавра
Выполнил: Григорьев Алексей 816 гр. МФТИ (ГУ)
Научный руководитель: Слесарев М. В.

Модели обращения в память

В многопоточных, многопроцессорных программах:

- Intel “processor ordering model” – аппаратная гарантия:
 - store не обгоняет load
 - load не обгоняет load
 - store не обгоняют store
- В «Эльбрус» “weak consistency model” - ослабленный порядок обращения в память. Модель Intel не поддерживается

Возможное решение

- Использовать собственную модель обращения в память
- Для поддержки модели памяти Intel использовать аппаратно-программные механизмы слежения, добавив дополнительные буферы, отслеживающие очередность.
- Механизм описан далее

Цель работы

- спецификация буфера
- rtl описание
- предварительный синтез устройства
- выявление критических путей и их устранение
- верификация устройства

Базовые понятия

Введение 2 режимов работы:

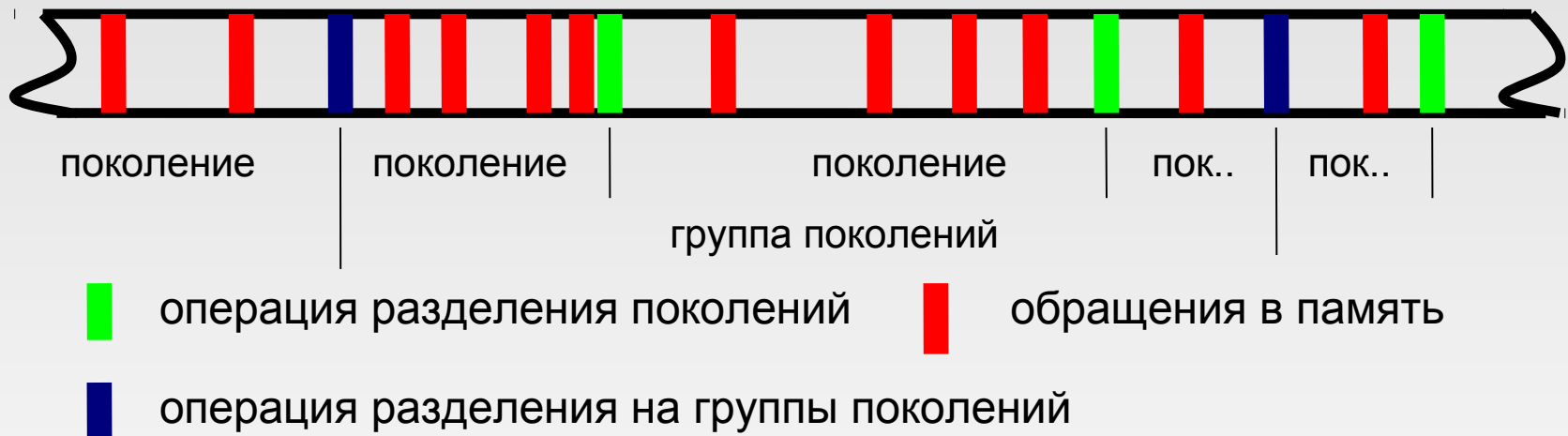
- **NRM** (Non-Recoverable Mode) — не оптимизированный режим, в котором команды идут в программном порядке.
- **ORM** (Optimized Recoverable Mode) — оптимизированный режим со слежением очередности и возможностью восстановления (откат).

В ORM происходит деление на поколения:

- ◆ Ограниченность размеров буферов
- ◆ Откат на большие дистанции
- ◆ Возможность забросить операции «вверх по коду»

Реализация деления на поколения

- Операции записи в память объединяются в поколения (generation).
- Несколько поколений объединяются в группы поколений – поколения чтения

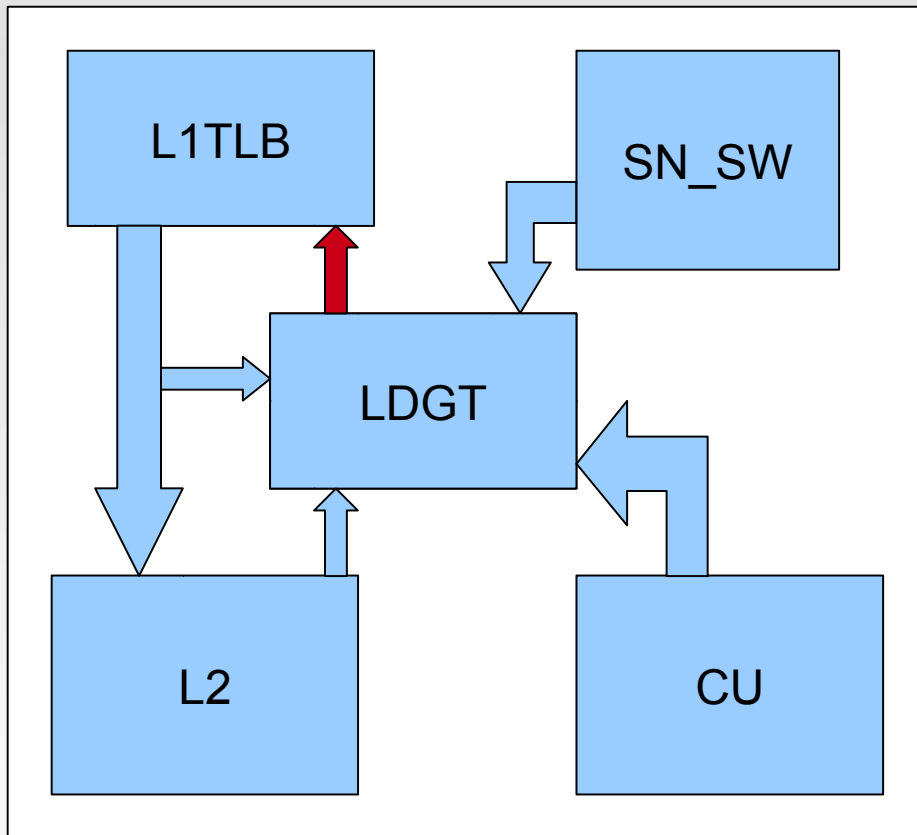


- Информации об операциях чтения сохраняется в LDGT.
- Любое кэширующее обращение в память вызывает snoop запрос, который обрабатывается в LDGT .
- LDGT просматривает свои таблицы и сверяет со snoop запросом.
- В случае конфликта вызывает прерывание.

Требования к устройству

- Обеспечение работы устройства на частоте 1ГГц при изготовлении по 65нм нормам
- Разделение по каналам
- Оптимальный размер буфера
- Введение «LifeTime» - времени жизни операции в терминах поколений
- Когерентность LDGT и кэшей всех уровней (L1, L2)

Рабочее окружение



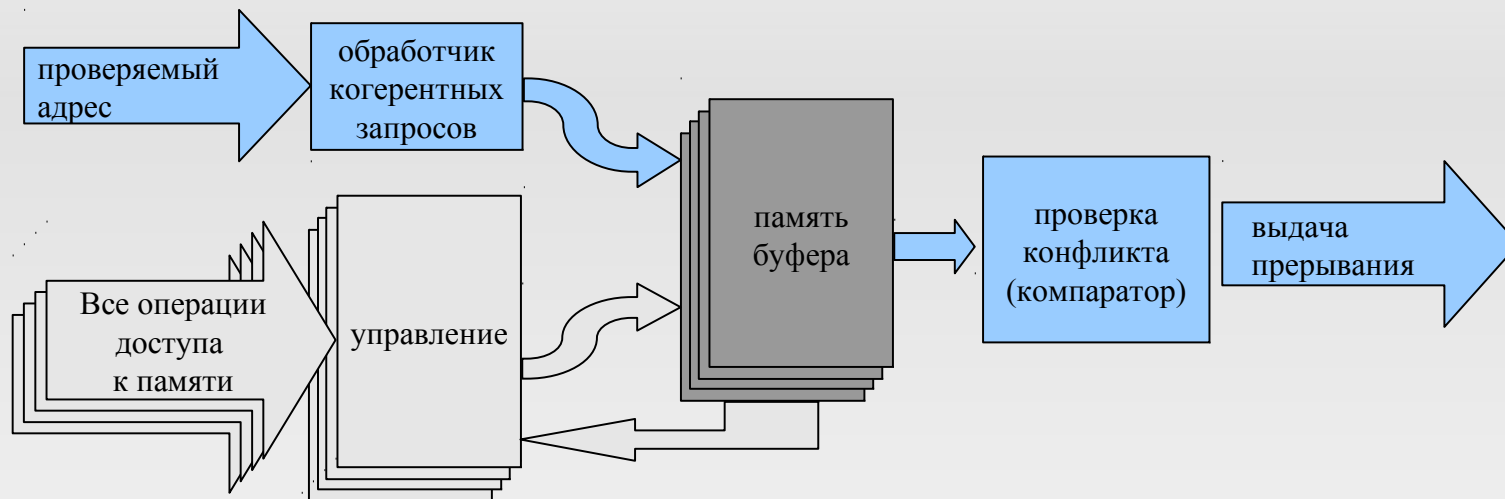
L1TLB — кэш 1-го уровня и TLB

L2 — кэш 2-го уровня

SN_SW — когерентный коммутатор

CU — устройство управления

Реализация буфера



Проверяемый адрес = адрес снуп запроса

НЕ ВСЕ операции доступа к памяти сохраняются в буфере

Организация буфера

На основе требований выбрана конфигурация буфера:

- Разделение по каналам (0,2,3,5) исполнительных устройств
- Тип буфера - кэш
- Число строк 16
- Число столбцов 8 (ассоциативность)

val	LifeTime	par	TagAddr[39:10]
-----	----------	-----	----------------

Формат строки:

val — значимость [1 бит]

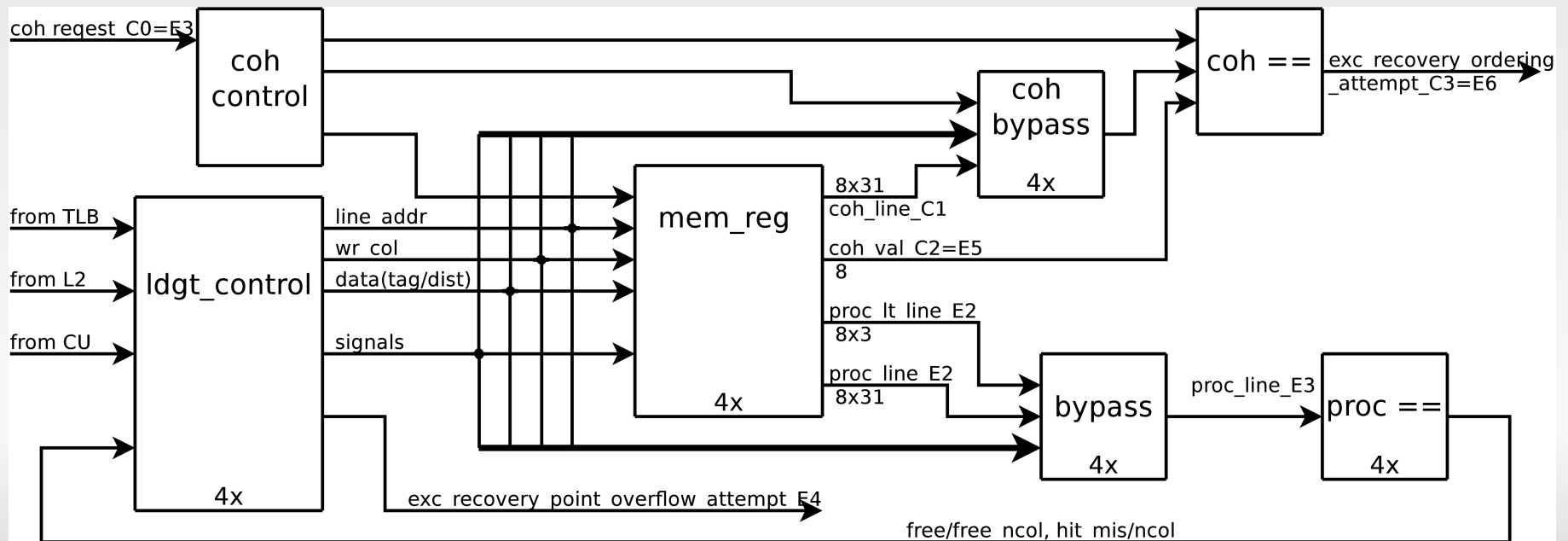
LifeTime — время жизни строки в терминах поколения [2 бита]

Par — проверочный бит [1 бит]

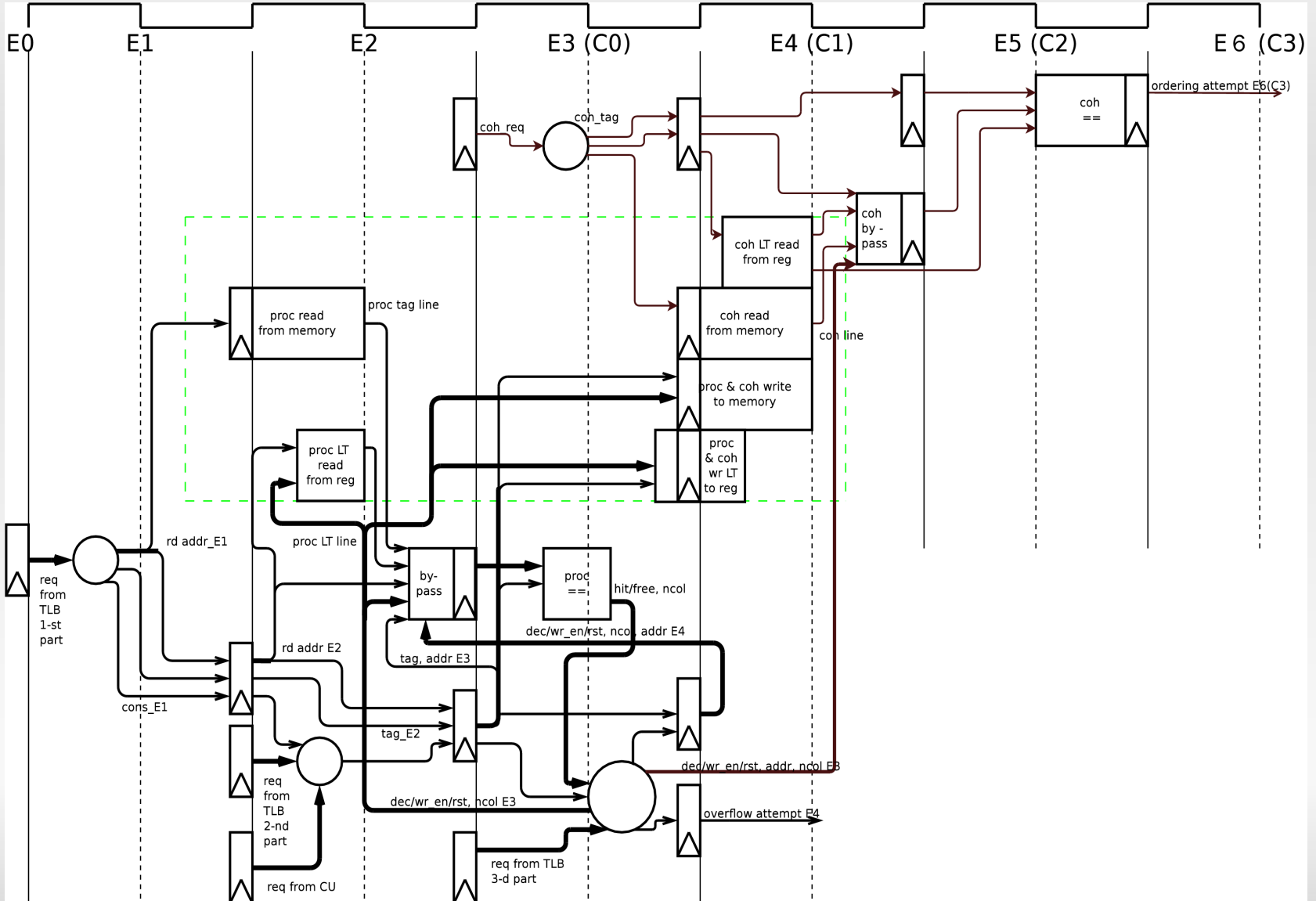
TagAddr — тэг адреса чтения [30 бит]

Принципиальная схема LDGT

- управляющий модуль (ldgt_control)
- обработчик когерентных запросов (coh_control)
- байпас когерентных обращений (coh_bypass)
- байпас процессорных обращений (proc_bypass)
- компаратор когерентных обращений (coh ==)
- компараторы процессорных обращений (proc ==)
- модуль памяти (mem_reg)



Конвейерное представление



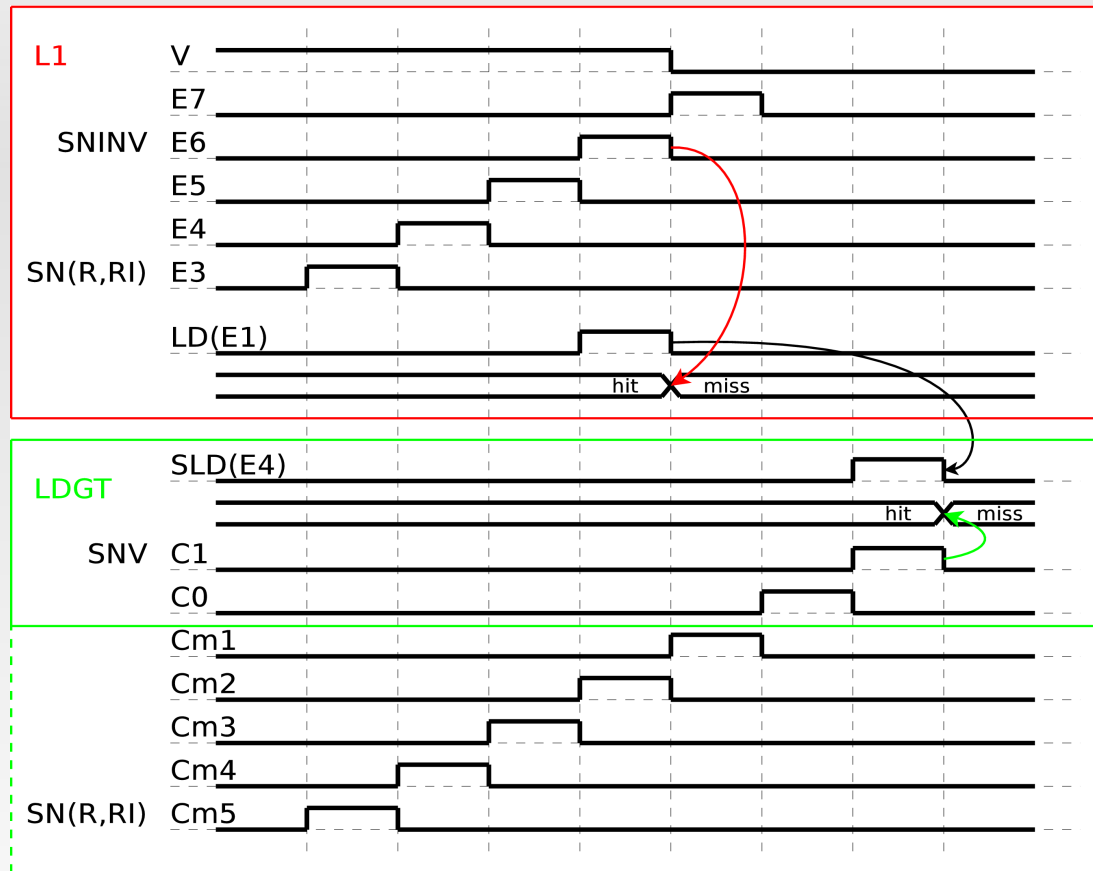
Стадии конвейера

- прием запроса, подготовка адреса для просмотра буфера (E1);
- просмотр буфера входным запросом (E2);
- прием подтверждения, формирование адреса для записи в буфер (E3);
- запись в память буфера (E4).

Когерентные запросы также проходят стадии конвейера:

- прием когерентного запроса и задержка для согласования ($Cm5 \rightarrow Cm1$);
- формирование адреса для просмотра буфера (C0);
- просмотр буфера когерентным запросом (C1);
- сравнение поступившего запроса с записями в буфере (C2);
- выдача попытки прерывания (C3).

Когерентность LDGT и кэшей всех уровней (L1, L2)



SN(R,RI) — снуп запрос

LD(E1) — чтение пришедшее на фазе E1 в L1 кэш

SLD(E4) — фиксирование того-же самого чтения в LDGT

Результаты работы

- Написана спецификация на устройство
- RTL — описание
- Предварительное физическое проектирование в DC по нормам 65нм:
 - Возможность работать на частоте 1ГГц
 - Занимаемая площадь – 0.3~0.35 мм²
- Выполнена верификация и отладка модулей и устройства в целом.

Спасибо за внимание