

Примеры архитектур: Intel x86 и RISC-V

Луцив Дмитрий Вадимович

Кафедра системного программирования СПбГУ



- 1 Архитектура и система команд x86
 - Регистры и адресация
 - Система команд
- 2 Архитектура и система команд RISC-V
 - Расширения и профили
 - Регистровый файл
 - Соглашение о вызовах
 - Команды и формат машинного кода
 - Формат машинного кода
 - Кросс-компиляция

Архитектура и система команд x86

- Регистры и адресация
- Система команд

Регистры (для 32-битной ЭВМ)

- EAX (общий, аккумулятор), EDX (умножение и деление вместе с EAX), EBX (указатели), ECX (счетчик)
- EDI (dest index), ESI (source index)
- EBP, ESP, EIP
- CS, SS, DS, ES, FS, GS — сегментные
- EFLAGS

Регистры (для 32-битной ЭВМ)

- EAX (общий, аккумулятор), EDX (умножение и деление вместе с EAX), EBX (указатели), ECX (счетчик)
- EDI (dest index), ESI (source index)
- EBP, ESP, EIP
- CS, SS, DS, ES, FS, GS — сегментные
- EFLAGS

Фрагменты регистров

- _H, _L — 8-разрядные
- _X, _S — 16-разрядные
- E_ — 32-разрядные
- R_ — 64-разрядные

Например, для аккумулятора

(AH (8), AL (8)) → AX (16) → EAX (32) → RAX (64)

Регистр флагов EFLAGS. Весь регистр 32-битный (начиная с 80386).

Основные флаги (с 8086):

- OF — флаг переполнения
- DF — флаг направления
- IF — флаг прерывания
- TF — флаг трассировки
- SF — флаг знака
- ZF — флаг нуля
- AF — флаг дополнительного переноса (для упакованных двоично-десятичных операций)
- PF — флаг четности
- CF — флаг переноса

- Непосредственная (аргументы в коде)
- Регистровая (номер регистра в коде)
- Память[E_X + смещение], Память[EBP + смещение], + возможно префиксы сегментов

- MOV память обменивается только с арифметическими регистрами, ESI, EDI
- XCHG reg, mem/reg
- LAHF, SAHF — флаги \leftrightarrow AH

Логические

AND, OR, XOR, NOT

Арифметические

- ADD, SUB, ADC, SBB, INC, DEC, NEG
- MUL (reg/mem), DIV (reg/mem), IMUL, IDIV,
- CWD (EAX → EDX:EAX)

Сдвига

- ROR, ROL
- RCL, RCR — с переносом
- SHL, SHR — без переноса
- SAL, SAR — со знаковыми битами

ASCII и BCD — для быстрого преобразования двоично-десятичных чисел

- PUSH, POP
- PUSH, POPA
- Косвенно — CALL, RET, INT, IRET

Переходы Безусловные

- JMP FAR, NEAR, JMP M[xx], JMP REG

Команды Сравнения

- CMP — как SUB
- TEST — как AND
- CMPS — CMPSB, CMPSW, CMPSD

Переходы Безусловные

- JMP FAR, NEAR, JMP M[xx], JMP REG

Команды Сравнения

- CMP — как SUB
- TEST — как AND
- CMPS — CMPSB, CMPSW, CMPSD

Compare-exchange

- CMPXCHG dest, src — Сравнивает аккумулятор (8-64 bits) с dest. Если равны, то в dest грузят src, иначе в аккумулятор загружают dest

Команды сравнения и передачи управления

Переходы Безусловные

- JMP FAR, NEAR, JMP M[xx], JMP REG

Команды Сравнения

- CMP — как SUB
- TEST — как AND
- CMPS — CMPSB, CMPSW, CMPSD

Compare-exchange

- CMPXCHG dest, src — Сравнивает аккумулятор (8-64 bits) с dest. Если равны, то в dest грузят src, иначе в аккумулятор загружают dest

Ужас. Кошмар. Для чего она?.. [↗](#)

По результату R или итогам сравнения $A ? B$, в зависимости от получившихся значений флагов.

Беззнаковые

- JA/JNBE — если $A > B$;
- JAE/JNB/JNC — если $A \geq B$;
- JB/JNAE/JC — если $A < B$;
- JBE/JNA — если $A \leq B$.

Знаковые

- JG/JNLE — если $A > B$;
- JGE/JNL — если $A \geq B$;
- JL/JNGE — если $A < B$;
- JLE/JNG — если $A \leq B$;
- JNS — если $R \geq 0$;
- JS — если $R < 0$.

По результату R или итогам сравнения $A ? B$, в зависимости от получившихся значений флагов.

- JE/JZ – если $A = B \vee R = 0$;
- JNE/JNZ – если $A \neq B \vee R \neq 0$;
- JNO – $\neg OF$;
- JO – OF ;
- JCXZ – $CX = 0$ – для организации циклов
do ... while(--CX) ;;
- JNP/JPO – $\neg PF$;
- JP/JPE – PF .

- Вызовы
 - CALL адрес
- Прерывания
 - Управление STI, CLI
 - Ожидание (HALT)

IN (mem/DX), OUT (mem/DX) — с AL

- REP, REPE, REPZ, REPNE, REPNZ
- LODS (загружает в аккумулятор),
STOS (пишет из аккумулятора),
MOVS (B-W-D — пересылка память-память),
CMPS(сравнение память-память),
SCAS (вычитает из аккумулятора)

- REP, REPE, REPZ, REPNE, REPNZ
- LODS (загружает в аккумулятор),
STOS (пишет из аккумулятора),
MOVS (B-W-D — пересылка память-память),
CMPS(сравнение память-память),
SCAS (вычитает из аккумулятора)

Команды учитывают DF — флаг направления. Выставив его «неправильно» можно быстро размножить участок памяти

Стековые

- FLD, FSTP — загрузить из памяти / выгрузить в память, формат
- FILD, FLD, ... — загрузить из регистра целое / выгрузить в регистр целое
- FMUL, ... — операции и функции
- FWAIT — синхронизация для старых процессоров

Регистровые

MULSD, MULSF — работают с векторными регистрами (появились в Pentium MMX, позволяют производить по несколько операций с числами разной длины)

Загрузка таблиц страниц и дескрипторов, изменение режимов работы процессора

Архитектура и система команд RISC-V

- Расширения и профили
- Регистровый файл
- Соглашение о вызовах
- Команды и формат машинного кода
- Формат машинного кода
- Кросс-компиляция

Расширения и профили

Базовые наборы и расширения

- Есть минимальные базовые наборы инструкций, регистров, прочих свойств процессора
- Есть стандартные расширения, пополняющие базовые наборы

Википедия [↗](#)

Профили и реализации

- Наборы расширений образуют профили, профили объединяются в семейства профилей
- Реализации могут включать разные расширения и реализовывать разные профили

Например, [профили для микроконтроллеров](#) [↗](#) или [профиль для запуска полновесных ОС](#) [↗](#)

Да давайте сразу [в английскую Википедию](#)  !

Обращаем внимание:

- Регистров много (характерно для RISC), поскольку для работы с ОЗУ команды отдельные
- У них есть разные названия (просто номер и смысловое) — для более дружественного кода на языке ассемблера
- Есть *интересные* регистры, точнее их предназначения:
 - `x0 zero` — тождественный ноль, наподобие `/dev/zero`
 - `x1 ra` — адрес возврата в регистре!
 - `x10-x17 a0-a7` — специальные регистры для аргументов и возвращаемых функциями значений
 - Оберегаемые/сохраняемые (`x18-27 s2-11`, вызываемая функция должна их восстанавливать) и не оберегаемые/временные (`x28-31 t3-6`, вызываемая функция не должна их восстанавливать) регистры

Что за адрес возврата в регистре

- Листовые функции — которые сами никого не вызывают, адрес возврата в вызывающую функцию при вызове сохраняется в регистре `ra`
- Не листовые функции — которые вызывают другие функции, и перед вызовом должны сохранять значение `ra` (на стеке, во временном регистре — это их дело), а потом восстановить его

Это позволяет быстро и часто вызывать «мелкие» функции. Подробнее:

- В английской Википедии => [↗](#)
- *Сара Л. Харрис, Дэвид Харрис. Цифровая схемотехника и архитектура компьютера: RISC-V / пер. с англ. В. С. Яценкова, А. Ю. Романова; под ред. А. Ю. Романова. – М.: ДМК Пресс, 2021. – 810 с.: ил.*

Переходы

Безусловные
содержимое...

Условные
содержимое...

Безусловные

содержимое...

Условные

содержимое...

Конвейер распознаёт конфликты

Потому что реализации стремятся быть (и являются) достаточно умными, и поддерживать:

- Суперскалярность
- Предсказание переходов
- В пределе — внеочередное исполнение

Т.е. у конвейера нет шансов остаться таким же простым, как и у MIPS

- При помощи **BuildRoot** [↗](#)
- С отладкой — при помощи **Ripes** [↗](#)

Вопросы

- 1 Приведите примеры арифметико-логических команд x86
- 2 Что такое микроциклы?
- 3 Приведите примеры и опишите работу нескольких команд условного перехода
- 4 Что такое лексема?
- 5 Что такое синтаксическое дерево?
- 6 Назовите обязательные стадии трансляции
- 7 Назовите опциональные стадии трансляции

Упражнения

- 1 Попробуйте воспользоваться претрансляцией для любого языка (Scheme, Julia, Nemerle, ...)
- 2 Скомпилируйте программу из примера для любой незнакомой архитектуры; пользуясь справочниками, объясните действия всех машинных команд

Вопросы



[EDU.DLUCIV.NAME](https://edu.dluciv.name) 