

Processador ARM Cortex-A9

Erick Nogueira do Nascimento
Instituto de Computação
Unicamp
ra032483@students.ic.unicamp.br

Franz Pietz
Instituto de Computação
Unicamp
ra076673@students.ic.unicamp.br

Lucas Watanabe
Instituto de Computação
Unicamp
ra134068@students.ic.unicamp.br

Categorias e Descritores de Assunto

D.3.3 [Arquitetura de Computadores]: Microprocessadores – *pipelining, microarquitetura, datapath, desempenho, eficiência energética.*

Palavras-chave

Microprocessadores, pipelining, microarquitetura, datapath, desempenho, eficiência energética

1. INTRODUÇÃO

O ARM Cortex-A9 é um processador da ARM, conceituada por processadores para sistemas embarcados, que tenta suprir as necessidades atuais de alto desempenho e com baixo consumo de energia. Os processadores da linha Cortex-A9 estão sendo utilizados em celulares, dispositivos de rede e equipamentos eletrônicos portáteis (câmeras, vídeo games, etc). O processador implementa a arquitetura ARMv7-A, é *superescalar multi-issue*, possui pipeline de 8 estágios com especulação, execução *out-of-order*, além de caches L1 com tamanho configurável.

Os processadores baseados na arquitetura ARM são desenvolvidos, em sua maioria, para produtos onde existem limitações de dissipação térmica e consumo de energia, onde o custo deve ser baixo e o desempenho deve ser alto para as aplicações que necessitem, tais como em Smartphones e Tablets.

Existem três famílias de processadores ARM. A família Cortex-M foi desenvolvida como processadores embarcados, otimizados para aplicações onde o custo é muito importante, microcontroladores e para FPGA. A família Cortex-R foi desenhada para sistemas embarcados de tempo real e, a família Cortex-A, da qual falaremos neste trabalho, são processadores para aplicações, suportando sistemas operacionais complexos e várias aplicações dos usuários.

Basicamente desenvolvido para aplicações móveis, os processadores ARM são desenvolvidos pensando em reduzir o custo final dos equipamentos e em sua eficiência. Os processadores ARM Cortex-A9 são baseados na arquitetura ARMv7, implementando suas características que já são suportadas por vários dispositivos. Estes processadores estão entre os mais eficientes de sua categoria.

Existem duas versões: a *single-core* e a *multicore*. A versão *single-core* oferece maior desempenho para dispositivos que utilizam ARM11, aumentando a funcionalidade e diminuindo o consumo da bateria de dispositivos móveis. Já a versão *multicore*, o Cortex-A9 MPCore, apresenta um controle avançado de consumo de energia, próprio para atender diferentes necessidades de aplicações que demandam alto desempenho e baixo consumo de energia. O desempenho dos processadores pode ser incrementado utilizando o Cortex-A9 NEON, uma engine SIMD de 64 ou 128 bits da ARM para aceleração de algoritmos multimídia e de processamento de sinais, e uma unidade de ponto flutuante.

Nesse trabalho, apresentaremos informações sobre a arquitetura do processador, a microarquitetura avançada, suas funcionalidades, vantagens e comparações com modelos similares.

2. MICROARQUITETURA

O ARM Cortex-A9 utiliza técnicas de ILP, como execução *superescalar* com escalonamento dinâmico. É um processador *multi-issue*, *out-of-order* e com um pipeline de 8 estágios com especulação. A cache L1 é quatro-associativa e pode ter 16, 32 ou 64KB de tamanho. A Figura 1 mostra a arquitetura geral do ARM Cortex-A9 *single-core*.

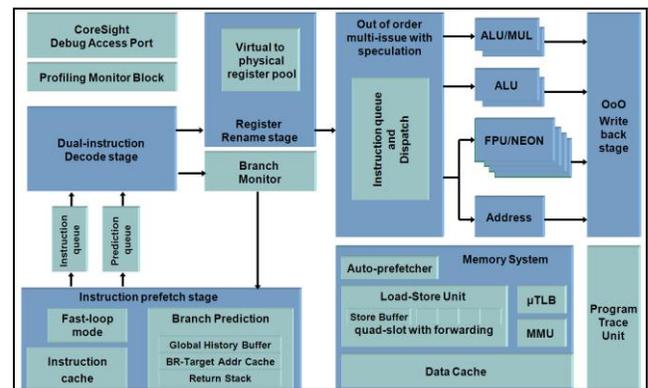


Figura 1. Microarquitetura do ARM Cortex-A9 *single-core*.

Os passos de despacho, execução, e finalização (completion) são *out-of-order*. O suporte a *out-of-order* é implementado sem o uso de *reorder buffers* custosos (em recursos e energia). Suporta o despacho de 4 instruções e a finalização de 7 instruções por ciclo. Provê compatibilidade com ótimo desempenho para binários compilados para processadores ARM anteriores.

Conforme já dissemos anteriormente, o processador ARM Cortex-A9 *single-core* é um processador que se utiliza de um pipeline de 8 estágios. Além disso, dependendo da versão do processador, junto com a unidade de ponto flutuante, há uma unidade chamada NEON, que é uma unidade que acelera o processamento de multimídia. A FPU foi revista nesta versão, dobrando o desempenho se comparado com a versão anterior da arquitetura.

2.1.1 Pipeline

O pipeline do Cortex-A9 tem como características principais um avançado processamento do fetch das instruções e predição de branches, utilizando instruções de stall na memória, pode realizar prefetching de até quatro instruções. Entre duas e quatro instruções podem ser decodificadas por ciclo, há também o modo *fast-loop*, o qual reduz o consumo de energia na execução de pequenos loops.

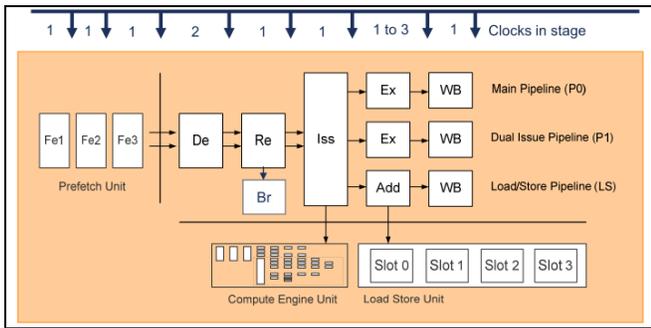


Figura 2. Estrutura do pipeline.

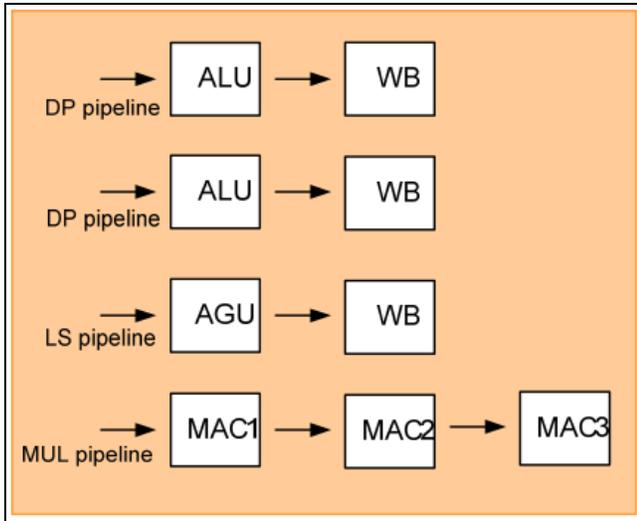


Figura 3. Detalhe sobre o pipeline.

O processador especula a execução de instruções, com renomeação dos registradores, renomeação virtual de registradores. Possui um pipeline *out-of-order*, podendo escolher quaisquer das quatro próximas instruções para iniciar o pipeline, pode executar até duas instruções aritméticas completas e duas de load-store, e é capaz de resolver um branch a cada ciclo. Em média, instruções comuns levam 9 ciclos desde o fetch até a finalização, enquanto instruções complexas levam 11 ciclos (Figura 2). A Figura 3 mostra com mais detalhes os pipelines no estágio de execução (clocks “1 a 3” na Figura 2).

Ainda sobre o pipeline, caso as instruções de load-store sejam dependentes, o processador pode realizar o *forward* do valor carregado para a instrução store que necessita deste valor. O início do processamento da instrução (após o fetch) é *out-of-order*, mas a finalização é *in-order*, garantindo os dados na ordem correta.

2.1.2 Renomeação de Registradores

O processador Cortex-A9 utiliza a técnica de renomeação de registradores explícita, através de um mapeamento dos 32 registradores arquiteturais (visíveis ao programador) para os 56 registradores físicos (não visíveis ao programador). A Figura 4 mostra um trecho de código assembly original (“ARM source”), com os registradores arquiteturais alocados pelo compilador, e o código de fato executado pelo processador (“Renamed to”), com os registradores físicos alocados.

As principais vantagens do uso da renomeação de registradores são a remoção dos *hazards* WAW (*write-after-write*) e WAR (*write-after-read*) no pipeline. Uma vantagem adicional é o desdobramento completo de loops pequenos por hardware.

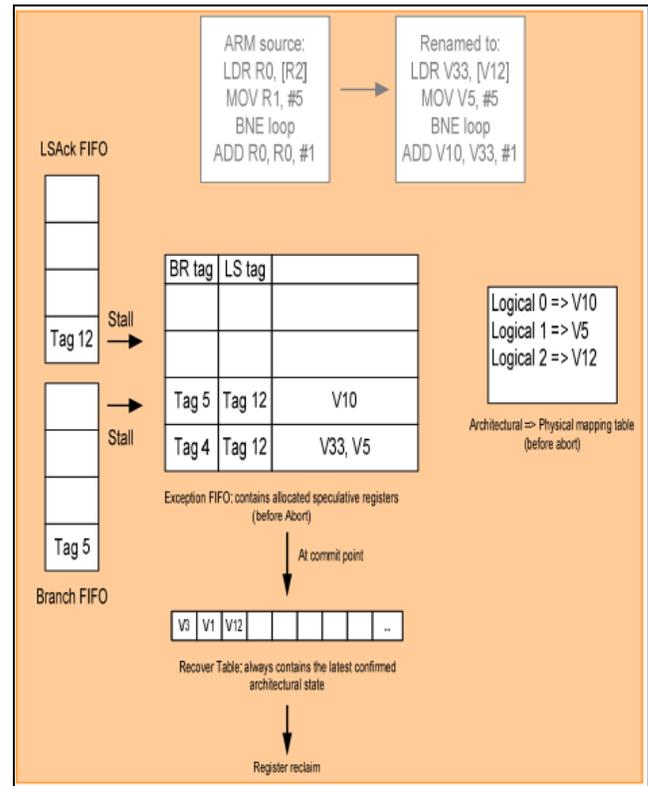


Figura 4. Renomeação de registradores.

2.1.3 Hierarquia de Memória

A cache L1 é dividida em I e D (Instruções e Dados), com 16KB, 32KB ou 64KB de tamanho cada uma.

A TLB tem dois níveis, no nível 1 (Micro TLB) há duas TLBs, I e D, a primeira tem 32 entradas, e a segunda tem 32 ou 64 entradas. A Micro TLB é completamente associativa, e consulta sua tabela de endereços virtuais para endereços físicos em um único ciclo de clock.

Se ocorre um *hit* na Micro TLB, o endereço físico associado é enviado diretamente a cache correspondente (I ou D) que prosseguirá na obtenção da página solicitada. Caso contrário, se ocorre um *miss*, a busca é realizada pela Main TLB, a qual tem uma estrutura física e de associatividade diferente da Micro TLB. A Main TLB é composta por um *array* de 4 elementos completamente associativo, e por uma estrutura 2-associativa com 2x32, 2x64, 2x128 e 2x256 entradas.

O gerenciamento da cache L2, utiliza-se de um controlador externo (PrimeCell PL310) e foi desenhado para otimizar o controle e aumentar a performance. Ele suporta uma razão síncrona de 1/2 clock, reduzindo a latência, principalmente em processadores com o clock mais elevado. Ainda pode utilizar até 8 MB, sendo entre 4-associativa e 16-associativa. Suporta paridade e memórias RAMs com correção de erros (ECC).

2.2 ARM Cortex-A9 MPCore

Sobre os processadores ARM Cortex-A9 *multicore* ou, na denominação da ARM, os Cortex-A9 MPCore, cada processador pode ter uma quantidade de cache L1 específica, FPU ou qualquer outro componente diferente do outro núcleo. A Figura 5 mostra uma visão geral da arquitetura de um Cortex-A9 MPCore com 4 núcleos, apontando os principais componentes responsáveis pela comunicação e coerência entre os núcleos.

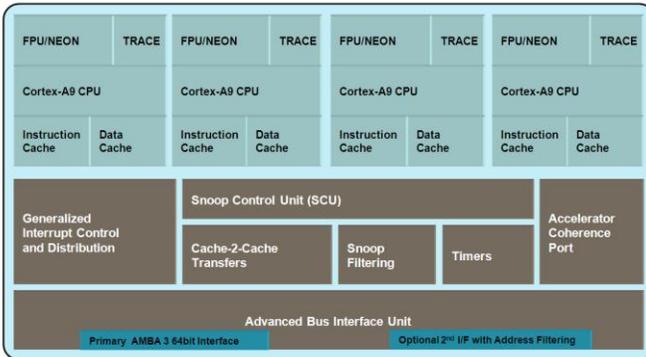


Figura 5. Microarquitetura do ARM Cortex-A9 MPCore.

2.2.1 Coerência de Cache

Há, ainda, uma unidade chamada *Snoop Control Unit* (Figura 6), responsável pela interconexão, como árbitro, comunicação, transferência entre os caches e a memória e todas as funções necessárias para garantir o funcionamento coerente e correto.

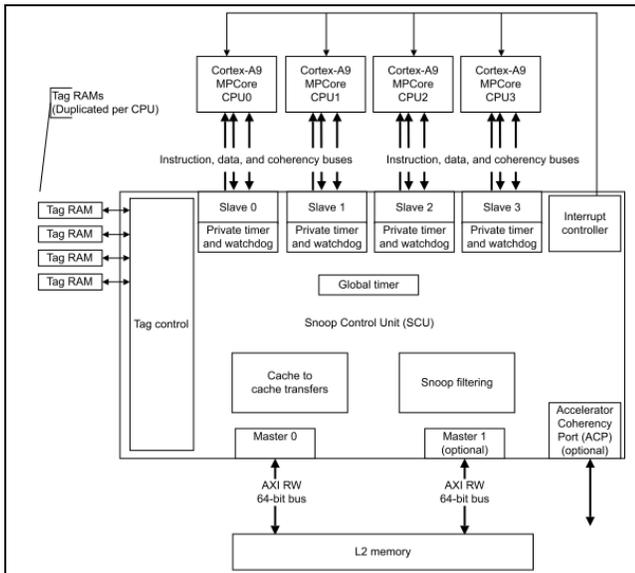


Figura 6. Snoop Control Unit.

Em sistemas que utilizam aceleradores ou periféricos externos a CPU, a CPU expõe a *Accelerator Coherence Port* (ACP) para estes componentes, a qual provê acesso à hierarquia de cache da CPU (L1 e L2) com coerência, proporcionando um bom desempenho e baixo consumo de energia do que outras arquiteturas para resolver este problema.

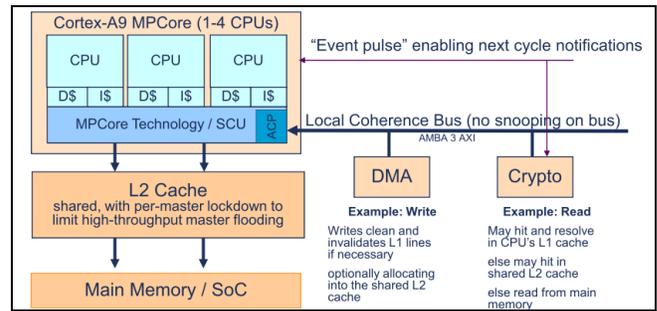


Figura 7. Cortex-A9 MPCore - Accelerator Coherence Port.

3. CARACTERÍSTICAS AVANÇADAS

3.1 Gerenciamento de Energia

O processador Cortex-A9 provê mecanismos para controlar dissipação de energia estática e dinâmica. Dentre as características do processador que contribuem para o baixo consumo de energia estão:

1. Previsão de desvio e de retorno de função precisos, reduzindo o gasto energético incorrido nas penalidades de miss;
2. Uso de caches fisicamente endereçáveis, reduzindo o número de *flushes* e *refills* de cache;
3. Uso de micro TLBs (um dos dois níveis de TLB) reduz o consumo de energia na tradução de endereço e verificação de bits proteção;
4. Loops de instrução que são menores do que 64B podem ser executados sem acessos adicionais a cache (além do *misses* compulsórios).

O processador pode estar em um dos modos de energia abaixo, os quais podem ter submodos (não apresentados):

1. *Run Mode*: modo normal de operação, no qual todas as funcionalidades do dispositivo estão disponíveis;
2. *Standby*: neste modo a maioria dos sinais de clock estão desabilitados. O subsistema de lógica, apenas a lógica de *wake-up* fica habilitada. Os bancos de memória RAM e caches permanecem ativos. Este modo reduz a dissipação de energia estática;
3. *Dormant*: Apenas os bancos de memória RAM e caches permanecem ativos;
4. *Shutdown*: todos os subsistemas são desligados.

Em processadores Cortex-A9 MPCore, cada núcleo pode estar em um modo de energia diferente dos outros núcleos. Deste modo, quando aplicações que demandam muito desempenho da CPU estão ativas, o sistema operacional (SO) pode deixar todos os núcleos em *Run Mode* e, quando não há aplicações exigentes ativas, o SO pode deixar apenas um dos núcleos em *Run Mode*.

Como um exemplo de implementação do gerenciamento avançado de energia, os chips A5 utilizados pela Apple em seus produtos consistem em dois processadores ARM Cortex-A9, com clock variável, para que possa existir um menor consumo de energia quando há a possibilidade. Em outros equipamentos, como os que utilizam o processador OMAP4400 da Texas Instruments, que é

um processador dual-core, há a possibilidade de um dos *cores* ser desligado para economia de energia, caso não exista necessidade de alto desempenho pelas aplicações em execução.

3.2 Advanced Bus Interface Unit

O Cortex-A9 MPCore possui uma interface de barramento avançada, chamada Advanced Bus Interface Unit, com as seguintes características principais:

1. Até 16 transações de barramento pendentes por processador (núcleo);
2. Transferências de cache L1 para cache L1 diretas, reduzindo consumo energético;
3. Suporte a razão de barramento de meio-clock, síncrona e de baixa latência, provendo maior desempenho ao subsistema de memória do que um barramento que usa um clock completo;

4. COMPARAÇÃO COM O CORTEX-A8

O processador ARM Cortex-A8 é o antecessor do Cortex-A9 dentro da família Cortex. Está presente no tablete iPad 1 e em smartphones.

É vendido como IP core (FOOTNOTE: núcleo do processador, o qual é feito para ser incorporado a outros componentes de lógica), disponível em:

- hard IP cores: alto desempenho (tempo ou energia), baixa flexibilidade de configuração;
- soft IP cores: desempenho em geral inferior ao hard, maior flexibilidade.

O processador é capaz de emitir duas instruções por ciclo de clock com um clock rate máximo de 1GHz.

4.1 Hierarquia de Memória

A cache L1 é composta por um par de caches, I e D (Instrução e Dados). Cada uma pode ter 16KB ou 32KB, com blocos de 64B. A cache L1 é 4-associativa, possui *way-prediction*, a política de substituição é *random replacement*, é virtualmente indexada e fisicamente *tagged*. A latência é de 1 ciclo para *hazards* do tipo *load-use*.

A cache L2 é opcional, pode conter 128KB a 1MB, com blocos de 64B. Pode ser organizada em 1 a 4 bancos, permitindo transferências concorrentes. É fisicamente indexada e fisicamente *tagged*. O barramento entre cache L2 e memória pode ser de 64 ou 128bits.

A TLB tem um nível, e é composta por duas TLBs, I e D. Ambas são completamente associativas, com 32 entradas. A política de substituição de entradas da TLB é o *round-robin*. Os tamanhos de página disponíveis para o sistema de memória virtual são 4KB, 16KB, 64KB, 1MB ou 16MB.

O desempenho da hierarquia de memória foi avaliado via simulação, e os resultados estão na Figura 8 (taxa de miss) e Figura 9 (penalidade de miss), ambos foram obtidos com o *benchmark* Minnespec *integer*. A largura de banda da memória é 12GB/s (pico).

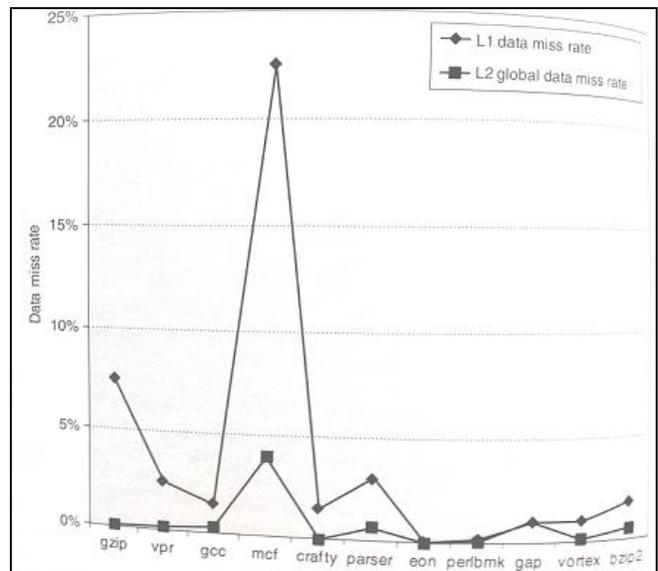


Figura 8. Taxa de miss para Cortex-A8 com cache L1 de 32KB e taxa global de miss para uma cache L2 de 1MB.

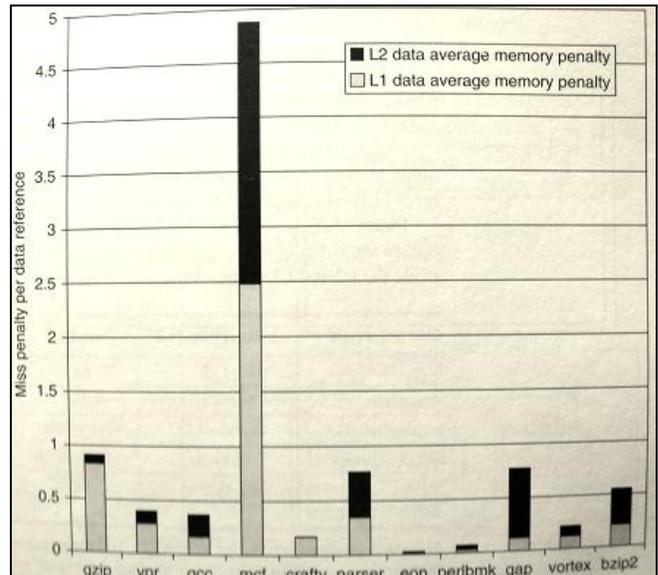


Figura 9. Cortex-A8 - Penalidade de miss média para as caches L1 e L2.

4.2 Pipeline

O Cortex-A8 suporta emissão de até 2 instruções por ciclo, é superscalar e tem escalonamento estático. O processador apresenta um CPI médio de 2, onde os *pipeline stall cycles* são o fator preponderante (Figura 10).

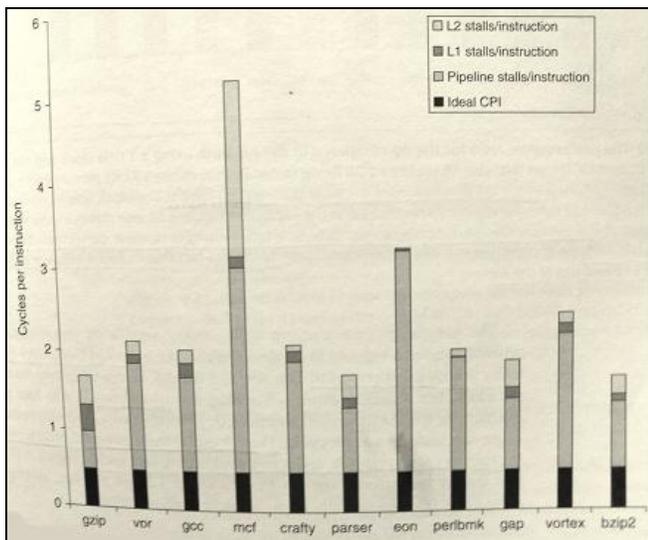


Figura 10. Composição do CPI para o Cortex-A8. Estimado a partir da taxa de miss e da penalidade de miss.

A previsão de desvio tem dois níveis, contendo: branch target buffer com 512 entradas, histórico global de 4K, *return stack* com 8 entradas.

4.3 Comparação com o ARM Cortex-A9

Cortex-A8 tem escalonamento estático enquanto Cortex-A9 possui escalonamento dinâmico. Cortex-A9 é mais rápido por um fator de 1.28, em média (Figura 11).

As medidas (simulações) de desempenho de ambos os processadores foram realizadas com a versão *single-core* do Cortex-A9 com a mesma taxa de clock e o mesmo tamanho de cache. A cache L1 do A8 é de 32KB e a L2 tem 1MB, sendo 8-associativa no A8 e 16-associativa no A9.

As principais razões para o desempenho melhor do Cortex-A9 são o uso de escalonamento dinâmico, um pipeline de multiplicação mais sofisticado, além de um preditor de desvio, uma cache L1 não bloqueante e um prefetch de instrução mais poderosos.

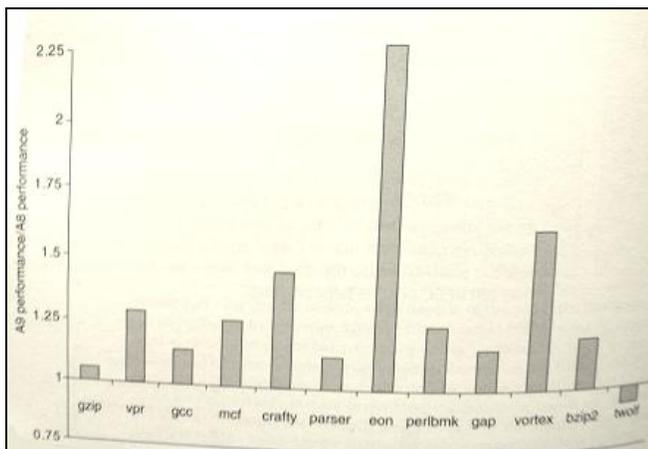


Figura 11. Razão de Desempenho entre o Cortex-A9 e o Cortex-A8. Ambos rodando a 1GHz e com caches L1 e L2 de mesmo tamanho. O Cortex-A9 é 1.28 vezes mais rápido do que o A8, em média.

4.4 Comparação de desempenho e características com outros processadores ARM

A Figura 12 mostra a evolução dos processadores ARM, comparando suas características fundamentais, como número de instruções despachadas, ordem de execução, e operação do barramento, além do desempenho alcançado por cada um no benchmark Dhrystone.

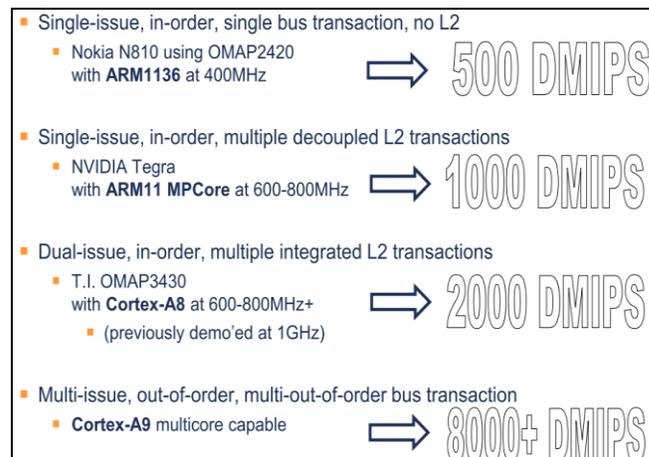


Figura 12. Evolução dos processadores ARM, desde o ARM1136 até o Cortex-A9. A medida de desempenho utilizada é o DMIPS (Dhrystone MIPS).

5. REFERÊNCIAS

- [1] ARM Holdings. 2007. *Details of a New Cortex Processor Revealed - Cortex-A9*. (October, 2-4, 2007, Santa Clara Convention Center CA, USA). Disponível em <http://www.arm.com/products/processors/cortex-a/presentations.php> (Acessado em 10/06/2012).
- [2] ARM Holdings. 2007. *Cortex-A9 Processor Microarchitecture*. (October, 2-4, 2007, Santa Clara Convention Center, CA, USA). Disponível em <http://www.arm.com/products/processors/cortex-a/presentations.php> (Acessado em 10/06/2012).
- [3] ARM Holdings. 2008. *ARM MPCore Architecture Performance Enhancement*. (July 16-17, 2008 Aoyama Diamond Hall, Omotesando, Tokyo, Japan). Disponível em <http://www.arm.com/products/processors/cortex-a/presentations.php> (Acessado em 10/06/2012).
- [4] ARM Holdings. 2009. *Circuit Design: High performance AND low power, the ARM way*. (September 22-23, 2009 Sophia Antipolis, France). Disponível em <http://www.arm.com/products/processors/cortex-a/presentations.php> (Acessado em 10/06/2012).
- [5] ARM Holdings. 2009. *2GHZ Capable Cortex-A9 Dual core Processor Implementation*. (September 16th, 2009, Internal presentation). Disponível em <http://www.arm.com/products/processors/cortex-a/presentations.php> (Acessado em 10/06/2012).
- [6] ARM Holdings. 2012. *Cortex-A9 Technical Reference Manual - Revision: r4p0*. Disponível em

- http://infocenter.arm.com/help/topic/com.arm.doc.ddi0388h/DDI0388H_cortex_a9_r4p0_trm.pdf (Acessado em 10/06/2012).
- [7] ARM Holdings. 2012. *Cortex-A9 MPCore Technical Reference Manual - Revision: r4p0*. Disponível em http://infocenter.arm.com/help/topic/com.arm.doc.ddi0407h/DDI0407H_cortex_a9_mpcore_r4p0_trm.pdf (Acessado em 10/06/2012).
- [8] Hennessy, J. L. and Patterson, D. A. 2011. *Computer Architecture: A Quantitative Approach, Fifth Edition*. Morgan Kaufmann Press, MA, USA.
- [9] Reis, D., Lino, D., Taborga, A., Saboia, V. Facó, P., Reis, F., Araújo, D. 2012. *Mobile Embedded System Platform – MESP*. Anais da Info Brasil - TI e Telecom 2012. Disponível em: http://www.infobrasil.inf.br/userfiles/16-S4-4-97247-Mobile%20Embedded%20System_.pdf (Acessado em 10/06/2012).
- [10] Rodrigues, G. R. 2009. *Smartphones e suas tecnologias*. Trabalho de Conclusão de Curso. Universidade de São Paulo. Disponível em: www.tcc.sc.usp.br/tce/disponiveis/18/180450/tce-23042010-094332/publico/Rodrigues_Guilherme_Rodrigues_e.pdf (Acessado em 10/06/2012).
- [11] Volpato, D. G. 2011. *Exploração de diferentes níveis de paralelismo visando a redução da área de processadores embarcados*. Trabalho de Conclusão de Curso. Universidade Federal do Rio Grande do Sul.