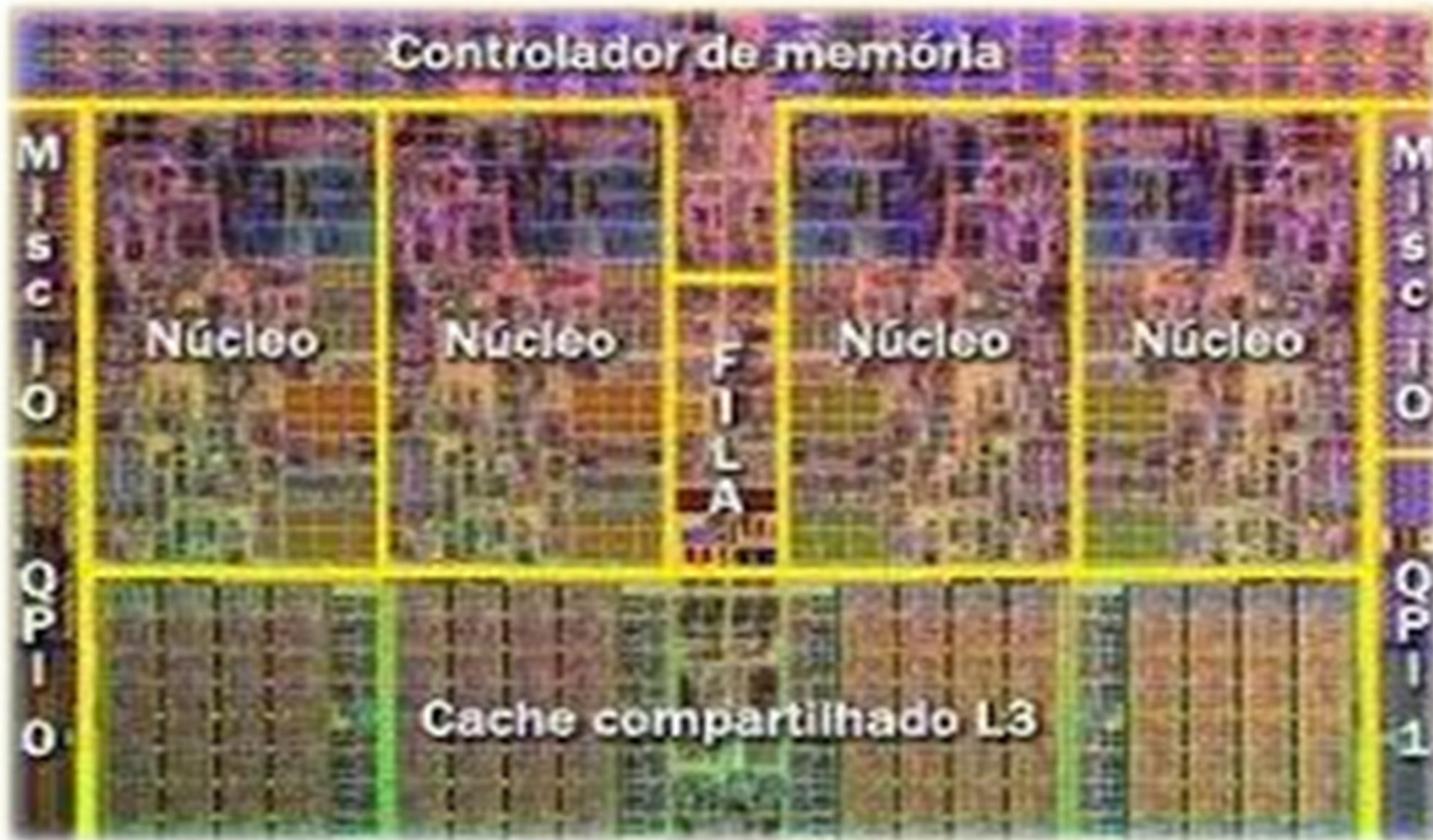


MEMÓRIA- SRAM-CACHE



MEMÓRIA SRAM

A memória SRAM foi desenvolvida para auxiliar a CPU (processador) devido a sua alta velocidade em fornecer os dados para o processamento.

Ela surgiu devido o atraso (delay) que temos dentro da DRAM (DDRx). A DRAM tem uma limitação na tecnologia que não permitir ter um tempo de acesso pequeno, por ex. pulso de refresh .

$$1 \text{ GHz} \rightarrow 1.000.000.000 \text{ Hz}$$

$$d = \frac{1 \text{ SEG}}{1.000.000.000} =$$

$$\frac{1 \text{ SEG}}{1.000.000.000} = 1 \text{ nano segundo } \underline{\text{ns}}$$

$$\frac{1 \text{ SEG}}{1000} = 1 \text{ milésimo} = \text{ms}$$

$$\frac{1 \text{ SEG}}{1.000.000} = 1 \text{ micronésimo de segundo} - \underline{\mu\text{s}}$$

MEMÓRIA SRAM

Perceba que uma DDR4 tem um atraso interno cerca de 14ns

SPEED VS. LATENCY AS MEMORY TECHNOLOGY HAS MATURED (INDUSTRY STANDARDS)				
TECHNOLOGY	MODULE SPEED (MT/s)	CLOCK CYCLE TIME (ns)	CAS LATENCY (CL)	TRUE LATENCY (ns)
SDR	100	8.00	3	24.00
SDR	133	7.50	3	22.50
DDR	335	6.00	2.5	15.00
DDR	400	5.00	3	15.00
DDR2	667	3.00	5	15.00
DDR2	800	2.50	6	15.00
DDR3	1333	1.50	9	13.50
DDR3	1608	1.25	11	13.75
DDR4	1866	1.07	13	13.93
DDR4	2133	0.94	15	14.06
DDR4	2400	0.83	17	14.17
DDR4	2666	0.75	18	13.50

MEMÓRIA SRAM

A média de um tempo de acesso dentro da DDR está em torno de 14 ns ou seja 14 nanosegundo:

Supondo um clock de 1 GHz (1 bilhão de Hz), equivale um pulso de clock igual a:

$$1 \text{ ns} = \frac{1 \text{ seg}}{1.000.000.000}$$

Um computador com um clock de 3 GHz, equivale um pulso de clock de :

$$0,33 \text{ ns} = \frac{1 \text{ seg}}{3.000.000.000}$$

MEMÓRIA SRAM

Isto significa que este processador é 42 vezes mais rápido que uma DDR4 ($3_{\text{ (clock)}} \times 14_{\text{ (tempo de acesso)}}$), ou seja esta DDR4 é 42 vezes mais lenta que o processador.

Se o processador enviar uma solicitação de escrita/leitura para esta DDR4, a CPU ficará ociosa cerca de 42 pulsos de clock até receber da DDR4 a confirmação desta solicitação.

MEMÓRIA SRAM

Porém a CPU não pode ficar sem atividade é preciso que a memória entregue as informações o mais rápido possível, mas como vimos a DRAM tem sua limitação de tecnologia.

Então foi preciso desenvolver uma outra memória muito mais rápida que a DRAM (DDR) para abastecer a CPU de instruções, para que ela não fique ociosa.

MEMÓRIA SRAM

Esta Memória é a **SRAM** (Static RAM ou RAM estática) também conhecida como memória **CACHE**.

A **CACHE** é intermediária entre a CPU e a DDR. O Sistema operacional e os aplicativos que estão no HD são transferidos para a **DRAM**, esta para a **CACHE** e esta para a CPU processar as instruções.

MEMÓRIA SRAM

Sistema sem SRAM (CACHE)



14 ns

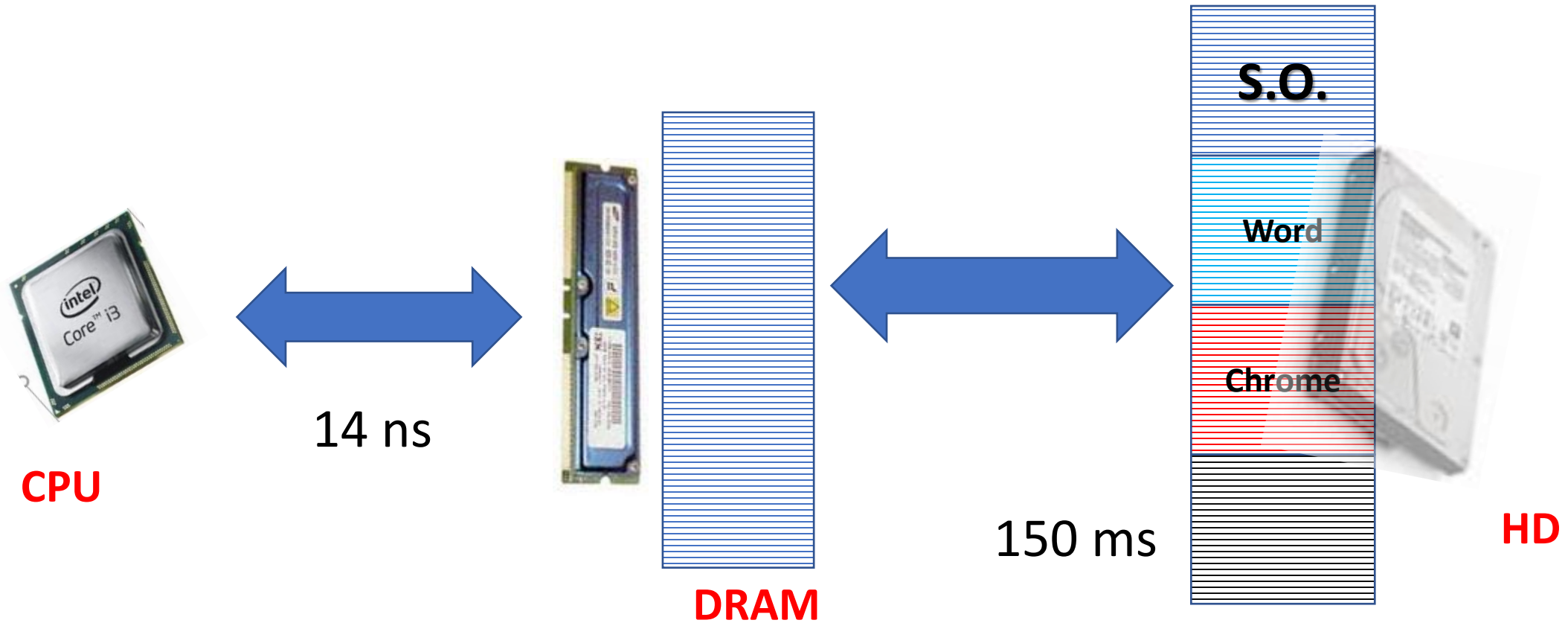


150 ms



MEMÓRIA SRAM

Sistema sem SRAM (CACHE)



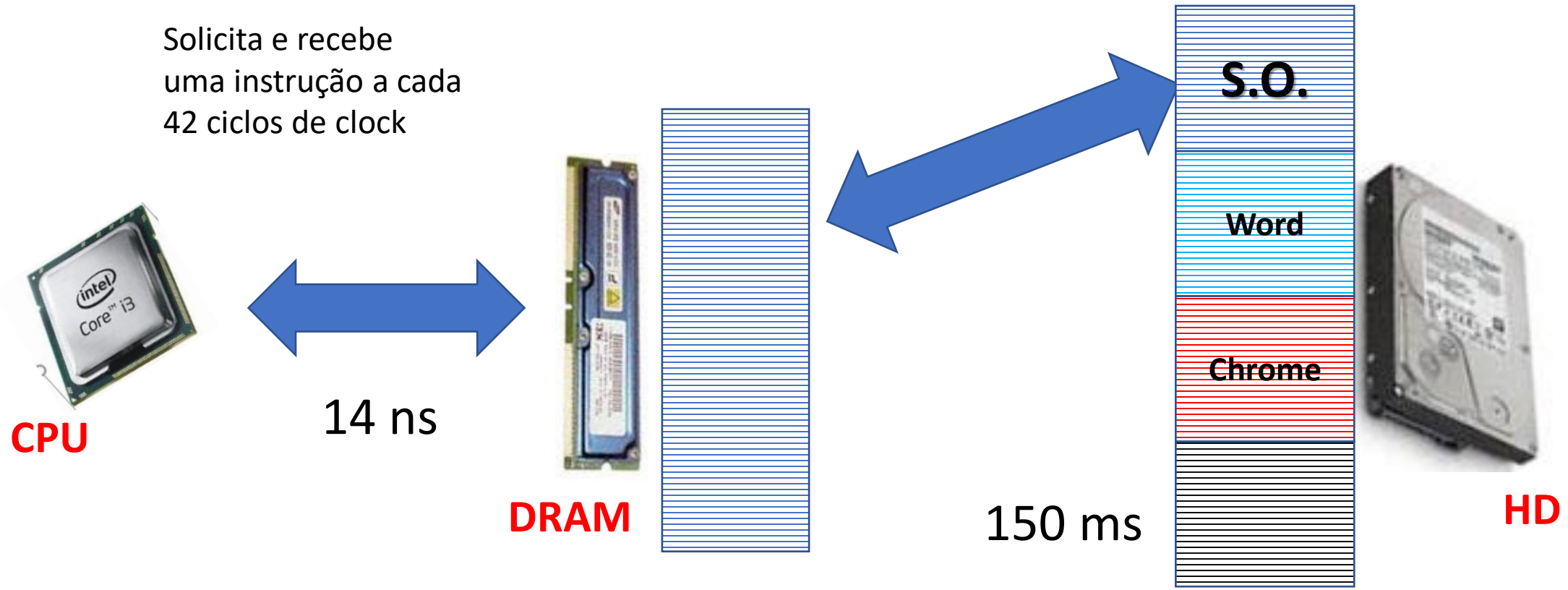
MEMÓRIA SRAM

A CPU irá solicitar cada **instrução** da DRAM para ser processada uma de cada vez, logo para cada instrução ela ficará ociosa, no nosso exemplo cerca de 42 ns (uma eternidade para um processador).

Estatisticamente, cerca de 90% a 95% de um programa /aplicativo tem as suas instruções de forma sequencial, ou seja as instruções são chamadas um após a outra.

MEMÓRIA SRAM

Sistema sem SRAM (CACHE)



MEMÓRIA SRAM

Esta sequência só será interrompida quando houver um desvio (uma condição), mas a princípio a maioria das instruções está sequencial .

MEMÓRIA SRAM

```
#include<stdio.h>

/* Programa que utiliza os operadores aritméticos*/

void main()
{
    int x,y,soma,sub;
    float modulo,div,mult;
    x=69;
    y=24;

    //utilizando os operadores aritmeticos:
    soma=x+y;
    sub=x-y;
    mult=(float) x*y;
    div=(float) x/y; /*aqui é utilizado além do operador aritmetico de divisão
                     o operador Cast.*/
    modulo=(x % y);

    printf("%d + %d=%d\n",x,y,soma); //imprime a soma dos dois números
    printf("%d - %d=%d\n",x,y,sub);  //imprime a subtração
    printf("%d x %d=%f\n",x,y,mult); //imprime a multiplicação
    printf("%d / %d=%f\n",x,y,div);  //imprime a divisão
    printf("%d mod %d=%f\n",x,y,modulo); //imprime o resto da divisão
}
```

MEMÓRIA SRAM

Uma vez que as instruções estão, grande parte dela, sequenciais, foi desenvolvida uma memória muito rápida entre a CPU e a DRAM, que é a MEMÓRIA CACHE (SRAM).

Logo quando a CPU solicita uma instrução, UM BLOCO DE INSTRUÇÕES é transferida da DRAM para a CACHE

MEMÓRIA SRAM

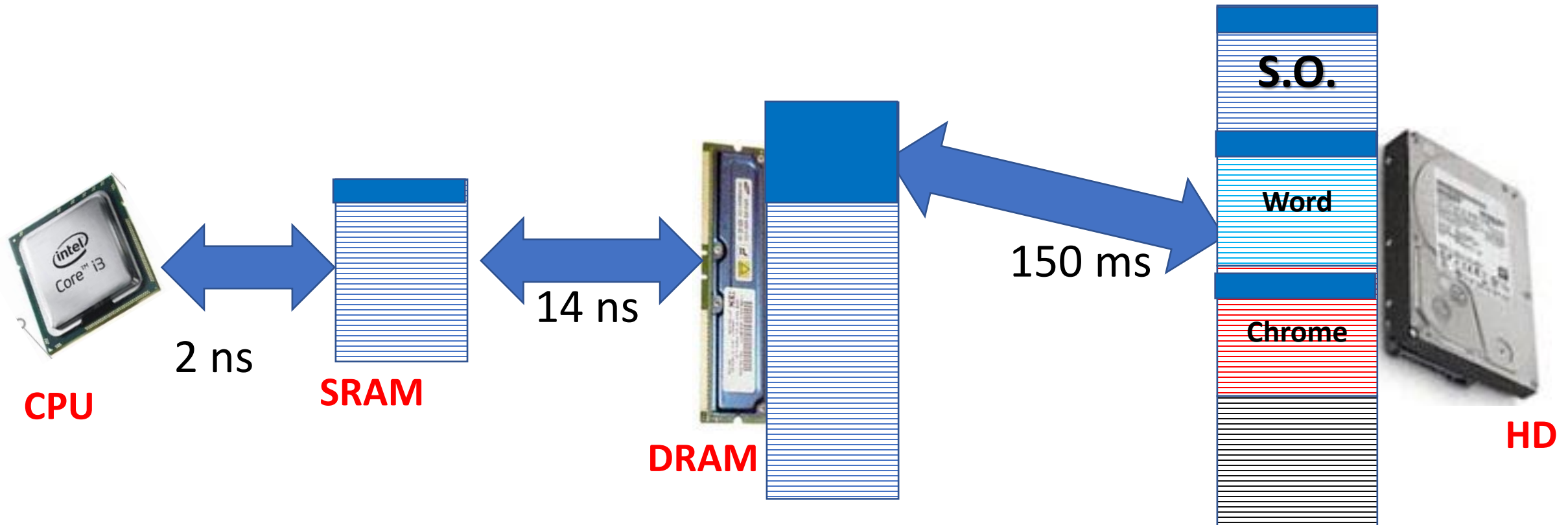
Neste instante a CPU irá solicitar as instruções diretamente da CACHE, com um tempo de acesso muito mais baixo, por exemplo, de 2ns.

Agora a CPU ficará ociosa somente 6 pulsos de clock ($3_{\text{(clock)}} \times 2 \text{ ns}$).

As primeiras CACHE (SRAM) ficavam instaladas na placa-mãe.

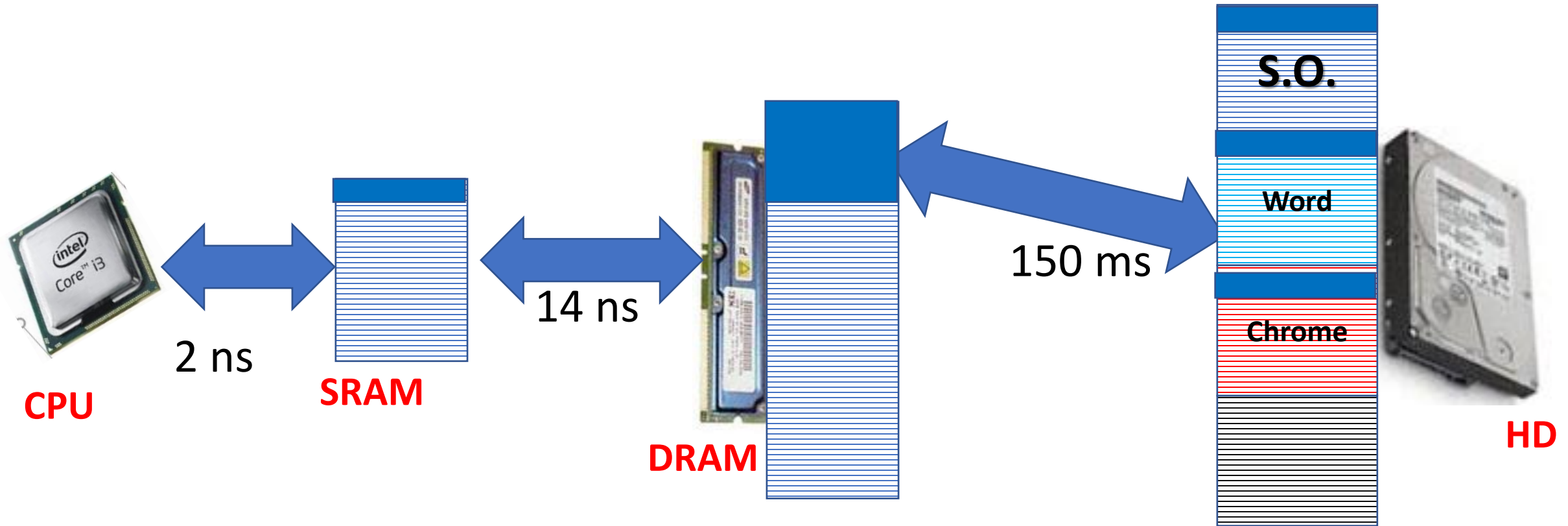
MEMÓRIA SRAM

Sistema com SRAM (CACHE)



MEMÓRIA SRAM

Sistema com SRAM (CACHE)



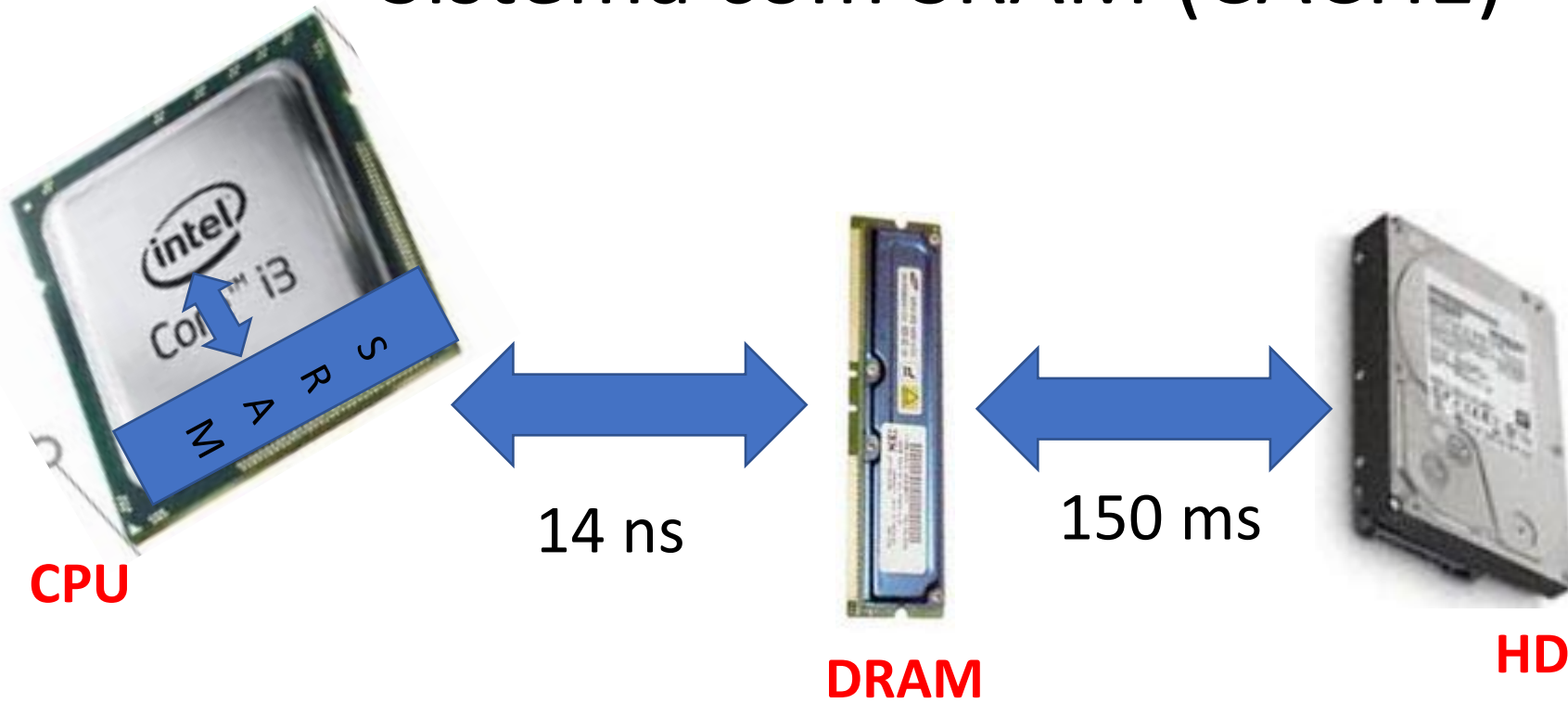
MEMÓRIA SRAM

Para conseguir que a CPU ficasse ainda com menos tempo ociosa, foi necessário inserir a CACHE dentro do processador.

Agora a CPU solicita as instruções da CACHE com um tempo de acesso MUITO próxima do processador.

MEMÓRIA SRAM

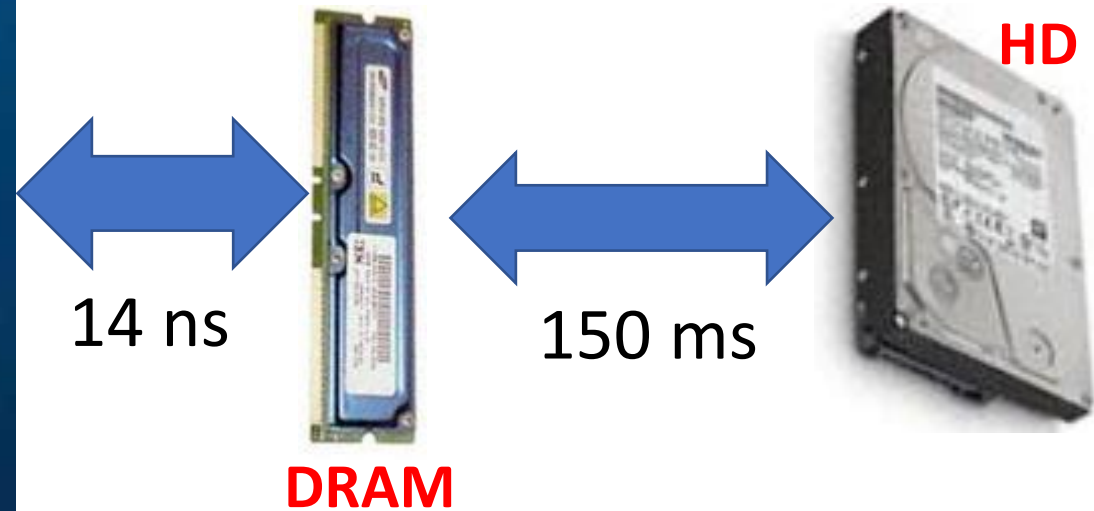
Sistema com SRAM (CACHE)



MEMÓRIA SRAM

Sistema com SRAM (CACHE) interna ao
processador

CPU



MEMÓRIA SRAM

Hoje temos 4 níveis (Levels) de CACHE : L4, L3, L2 e L1

L4 - está mais próxima da DRAM (DDR) e

L1 – mais próxima do processador (CPU)

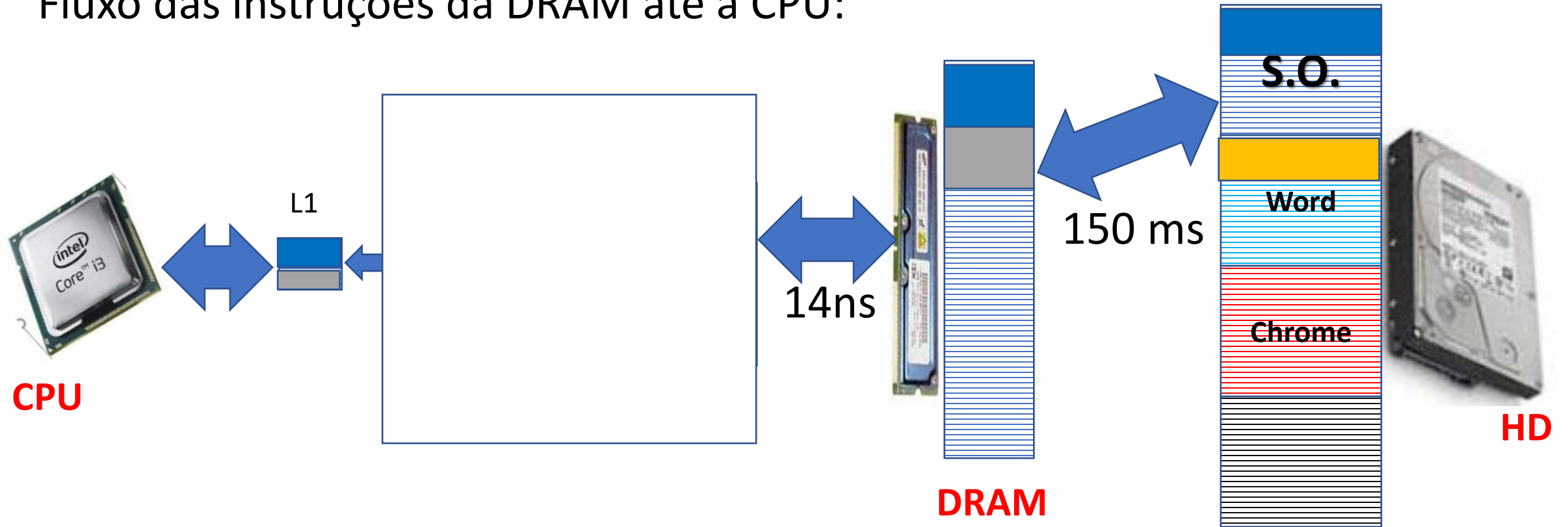
As caches podem estar instaladas na placa-mãe e/ou interna a CPU, não necessariamente os 4 níveis, p.ex somente 2 níveis (L1 e L2) na placa mãe, ou na placa mãe (L3) e na CPU (L1 e L2) ou somente na CPU (L1 e L2)

Processador ou CPU é único componente que utiliza 100% do clock.

Outros componentes utilizam porções do clock

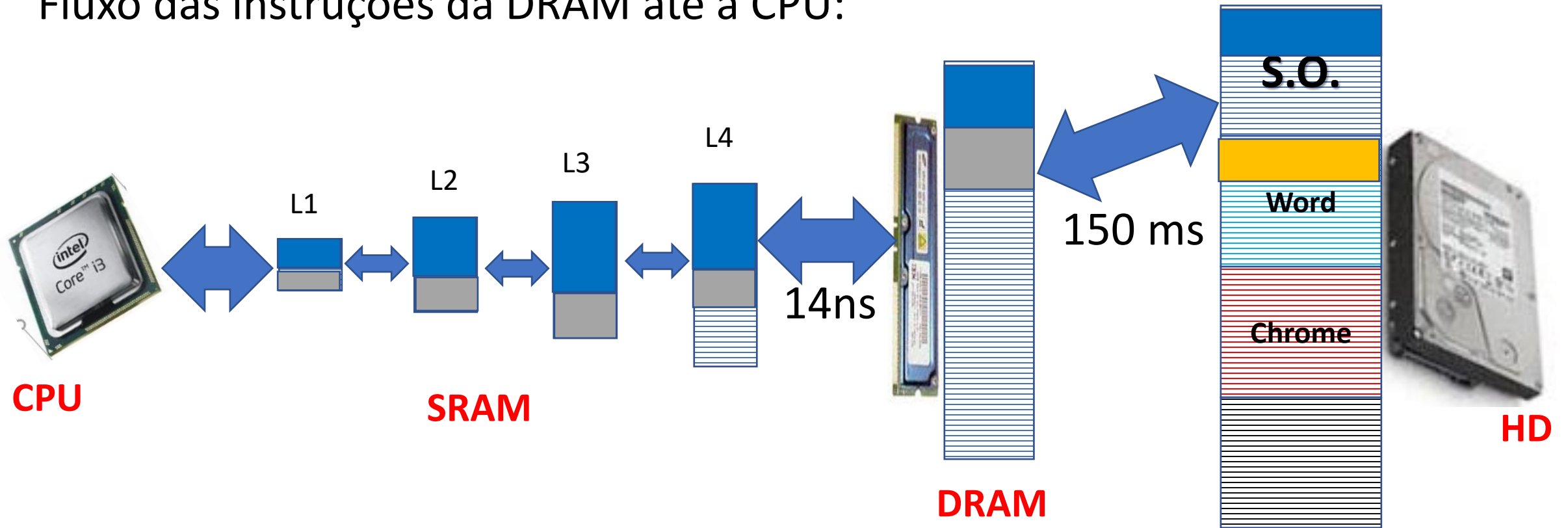
MEMÓRIA SRAM

Fluxo das instruções da DRAM até a CPU:



MEMÓRIA SRAM

Fluxo das instruções da DRAM até a CPU:



MEMÓRIA SRAM

A cache L4 tem maior capacidade de armazenamento, geralmente está nas propagandas dos anúncios da memória, por exemplo:



- Intel® Pentium® Gold 5405U (2.3 GHz, **cache de 2MB**, dual-core, 8ª geração)
- Windows 10 Home Single Language, de 64 bits - em Português (Brasil)
- Placa de vídeo integrada Intel® UHD Graphics 610
- Memória de 4GB (1x4GB), DDR4, 2400MHz; Expansível até 16GB (1 slot soDIMM, sem slot livre)
- HD de 500GB (5400 RPM) SATA 2.5"

Preço

R\$ 2.176,00

[Detalhes da oferta](#)

MEMÓRIA SRAM

Devido o alto preço da SRAM, temos processadores com diferentes preços dentro da mesma família.

Vide Aula Passada

MEMÓRIA SRAM

Lembram da aula passada sobre a memória DRAM

DRAM – Dynamic RAM

Suas características:

- Armazena grande quantidade de Bytes (4GB, 8GB, 16GB ...)
- Desenvolvida com baixa tecnologia
- Mais barata em relação a cada Byte armazenado
- Vendidas nas lojas para computadores
- É expansível (depende da placa mãe)
- Normalmente não esquenta (exceção a memória para jogos)
- Para armazenar um único bit utiliza-se 1 transistor e 1 capacitor
- Necessita de um pulso de REFRESH

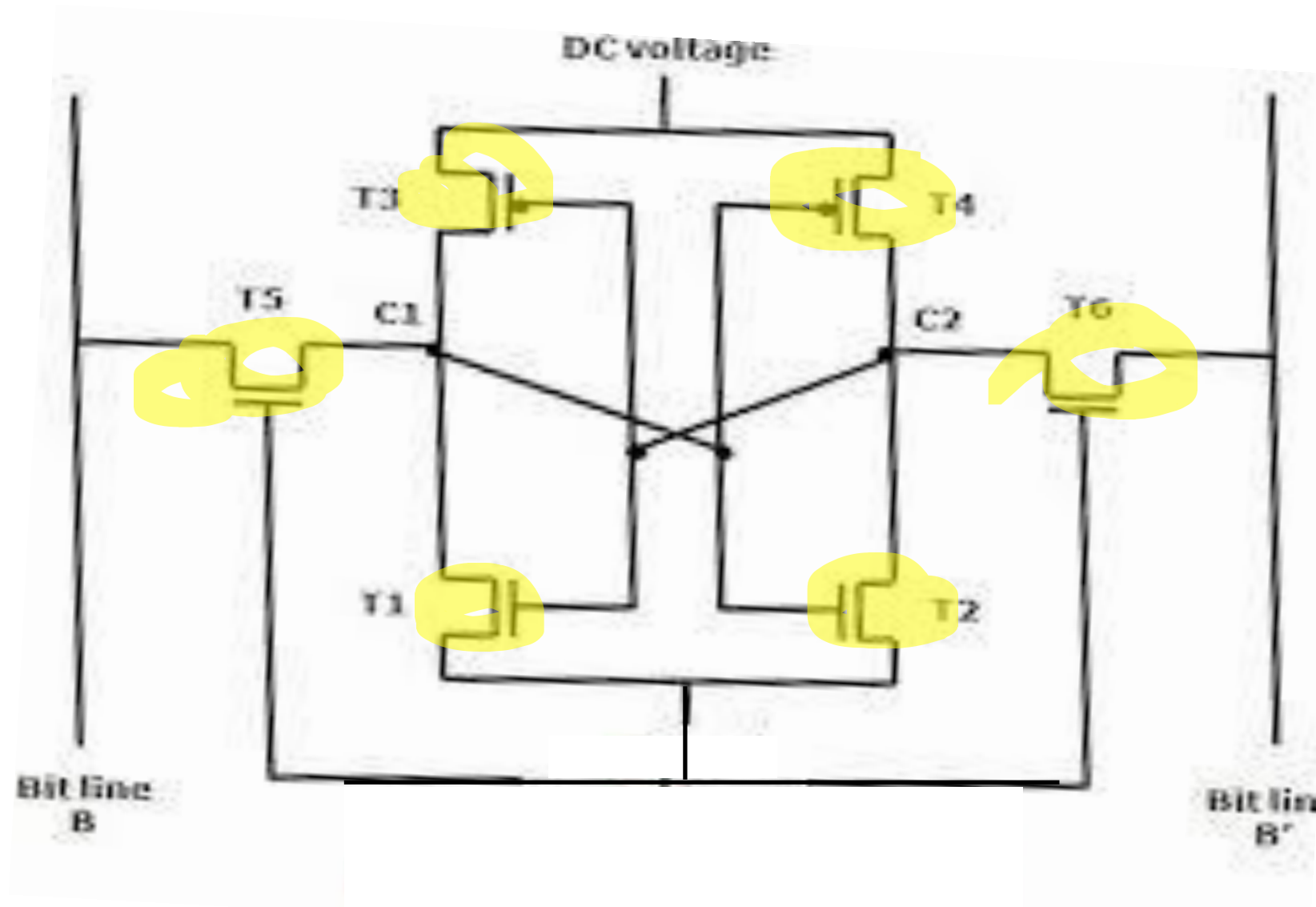
MEMÓRIA SRAM

SRAM – STATIC RAM

Suas características:

- Armazena pequenas quantidade de Bytes (16MB, 8MB, 256KB ...)
- Desenvolvida com alta tecnologia (mesma da CPU)
- Muito mais cara em relação a cada Byte armazenado
- Não são vendidas nas lojas para computadores
- Não é expansível
- Esquenta junto com o processador
- Para armazenar um único bit utiliza-se de 6 a 8 transistores
- Não necessita de um pulso de REFRESH

MEMÓRIA SRAM



MEMÓRIA SRAM

Geralmente as memórias CACHE, têm estas grandezas de capacidades:

L4 – dezenas de MB (32MB)

L3 – unidades de MB (12MB)

L2 – centenas de KB (512 – 4 MB)

L1 – dezenas/centenas de KB (128 KB – 256 KB)

MEMÓRIA SRAM

Para agilizar ainda mais o processamento temos:

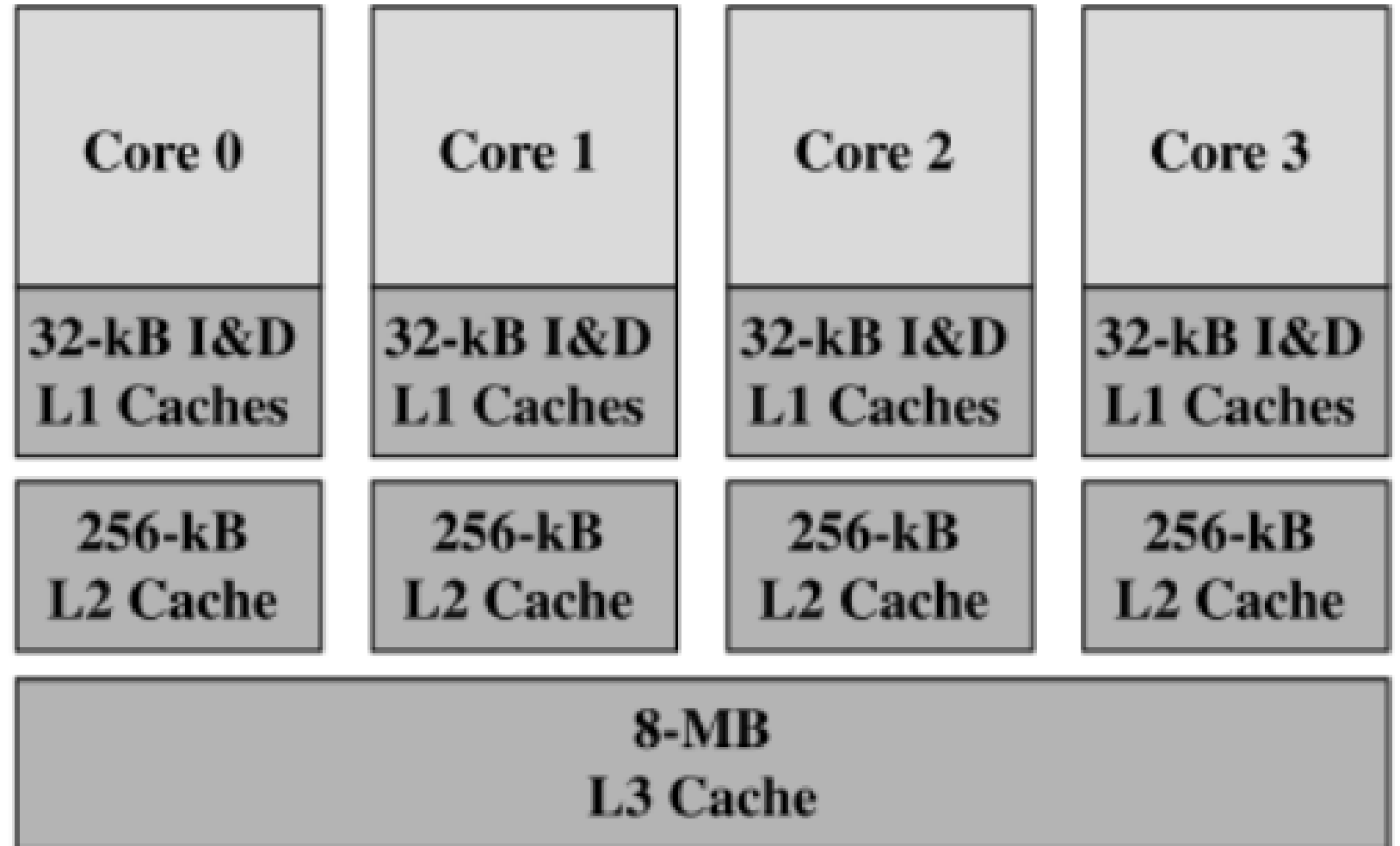
CACHE dedicada : onde totalmente a cache serve um núcleo do processador

CACHE compartilhada : onde uma cache serve todos os núcleos do processador

MEMÓRIA SRAM

L1 e L2 –dedicada a cada núcleo (core)

L3 - compartilhada a cada núcleo (core)



MEMÓRIA SRAM

Outra forma de agilizar ainda mais o processamento temos:

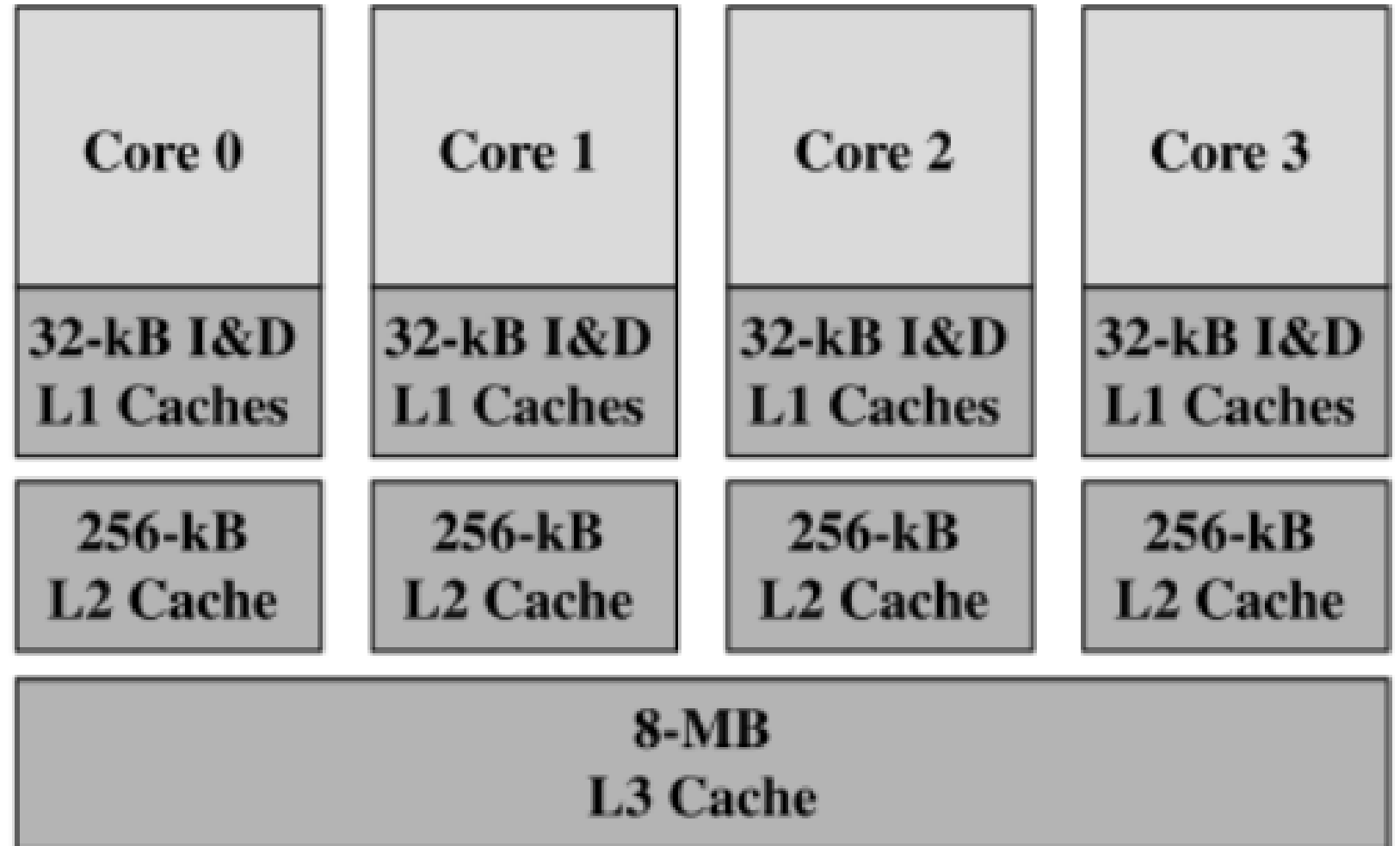
CACHE somente de instrução : onde temos as instruções do programa

CACHE somente para dados : onde temos os dados das instruções do programa.

MEMÓRIA SRAM

L1 e L2 –dedicada a cada núcleo (core)

L3 - compartilhada a cada núcleo (core)



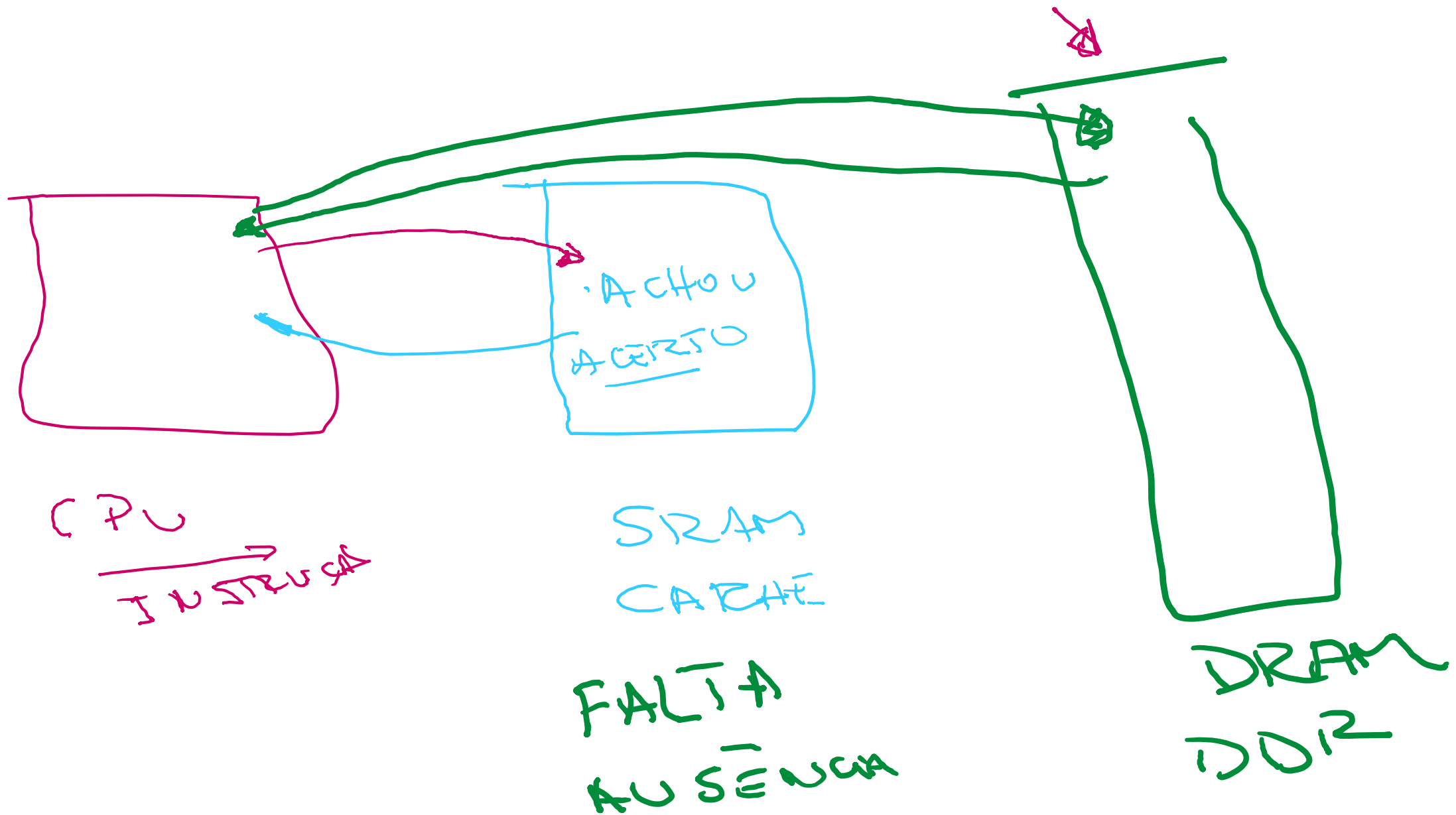
MEMÓRIA SRAM

A CPU quando for solicitar uma instrução, PRIMEIRAMENTE ela irá verificar se esta instrução já está na CACHE, se estiver, a instrução será processada.

Neste caso informamos que houve HIT (acerto).

Porém se a instrução não estiver na CACHE a CPU irá solicitar à DRAM (DDR).

Neste caso informamos que houve um MISS (falta-ausência)



MEMÓRIA SRAM

Durante um processamento se o contador de **HIT (acerto)** está alto e o contador de **MISS (falta-ausência)** está baixo, significa que a cache do seu equipamento está **bem dimensionada**, pois as instruções a serem executadas já estão dentro da cache.

Logo o processamento terá um bom desempenho.



MEMÓRIA SRAM



No caso inverso, se durante um processamento se o contador de **MISS (falta-ausência)** está alto e o contador de **HIT (acerto)** está baixo, significa que a cache do seu equipamento está **subdimensionada**, pois as instruções a serem executadas ainda estão na DRAM(DDR).

Logo o processamento terá um mal desempenho, logo preciso fazer um upgrade de processador e ou placa mãe.

MEMÓRIA SRAM

OBRIGADO!

DÚVIDAS.