



Estrutura do tema ISC

1. Representação de informação num computador
2. Organização e estrutura interna dum computador
3. Execução de programas num computador
4. Análise das instruções de um processador
5. **Evolução da tecnologia e da eficiência**



Componentes (físicos) a analisar:

- a unidade de processamento / o processador:
 - o nível ISA (*Instruction Set Architecture*):
tipos/formatos de instruções, acesso a operandos, CISC/RISC...
 - CISC *versus* RISC
 - **melhoria de eficiência** no processador: com paralelismo
 - **melhoramentos** fora do processador (ou *core*)
 - **evolução** da arquitetura x86 da Intel
- a hierarquia de memória:
cache, memória virtual, ...
- periféricos...

CISC versus RISC

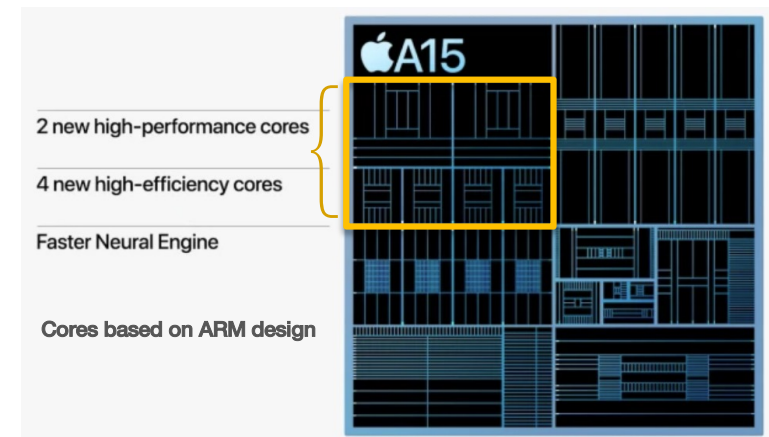
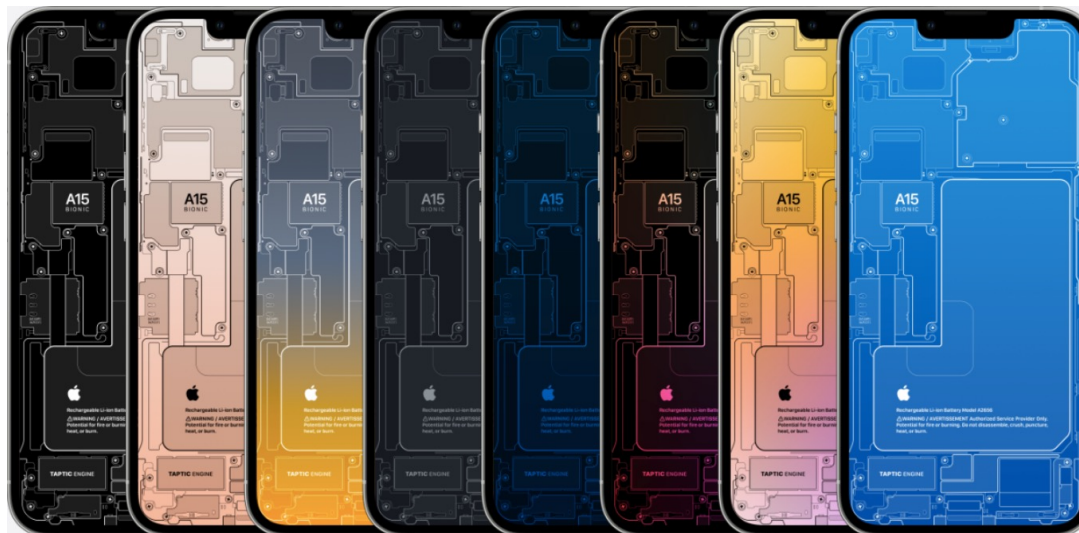


A “revolução” dos Reduced Instruction Set Computers

Caracterização das arquiteturas RISC

- conjunto reduzido e simples de instruções
- formatos simples de instruções
- uma operação elementar por ciclo máquina
- operandos sempre em registos
- modos simples de endereçamento à memória

Arquiteturas RISC: **em todos os smartphones!**



Eficiência nos Sistemas de Computação: oportunidades para melhorar



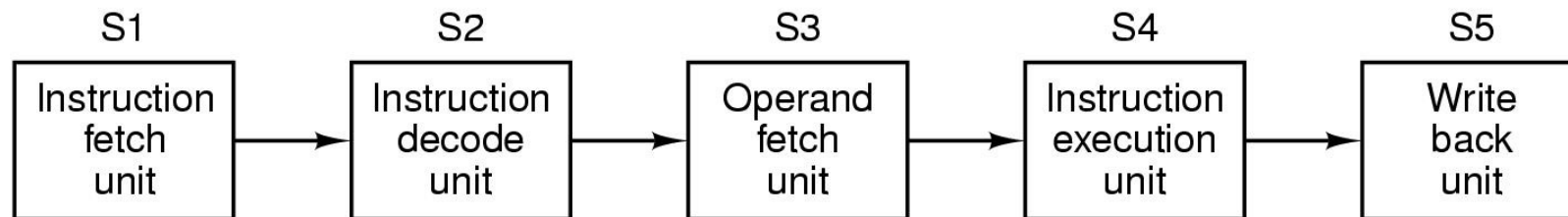
Oportunidades para melhorar o desempenho / eficiência

- com introdução de **paralelismo**
 - na execução de instruções em binário (*Instruction Level Parallelism*)
 - paralelismo desfasado ou execução encadeada (*pipeline*)
 - paralelismo nos dados (processamento vetorial, MMX/SSE/AVX...)
 - paralelismo nas operações (VLIW, superescalar)
 - com execução de instruções fora de ordem (*out-of-order execution*)
 - no acesso à memória
 - paralelismo "real" (maior largura do *bus*)
 - paralelismo desfasado (*interleaving*)
 - ao nível da aplicação (sistemas concorrentes/paralelos/distribuídos)
 - com fios de execução (*multithreading => multicore/multichip c/ mem partilhada*)
 - com processos (com memória distribuída)
- com **hierarquia de memória** (para diminuir latência)
 - *cache ...*

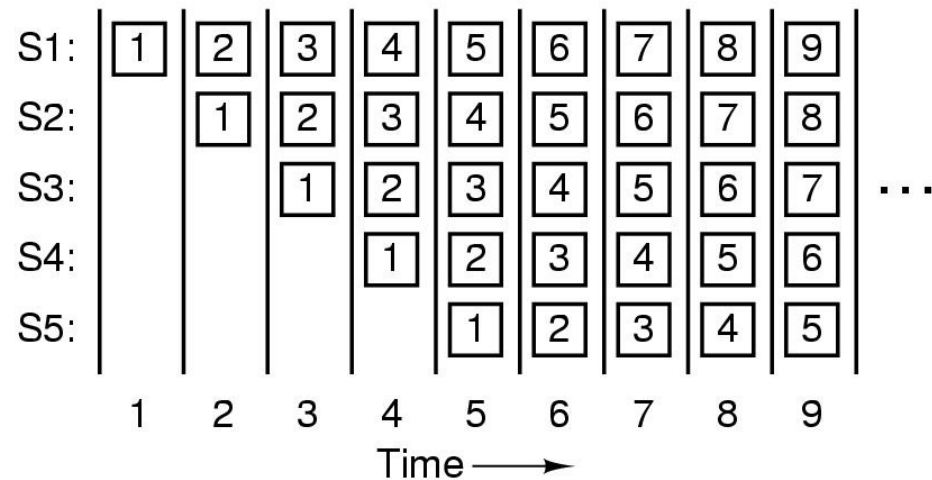
Paralelismo no processador: exemplo



Exemplo de *pipeline*

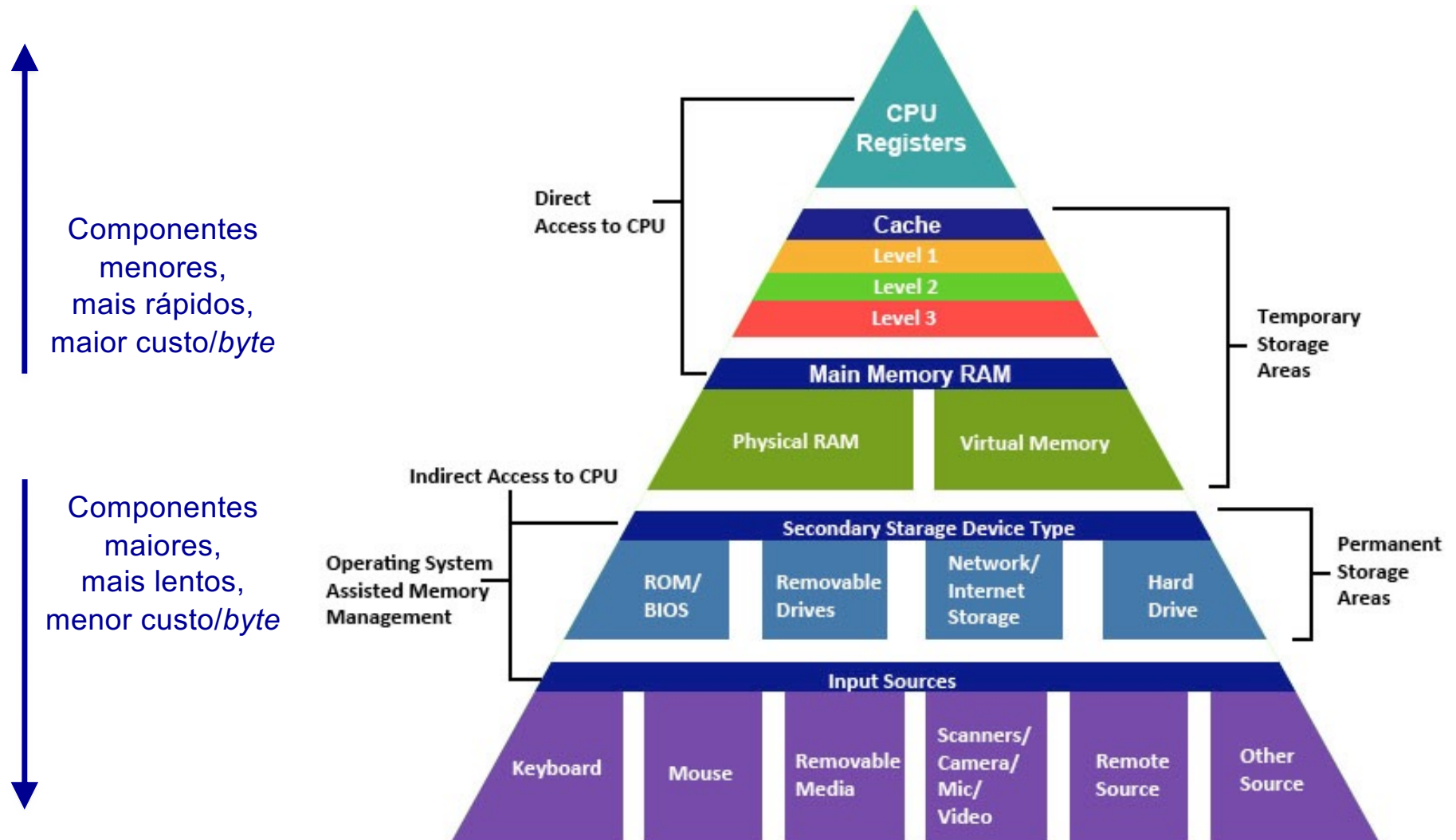


(a)



(b)

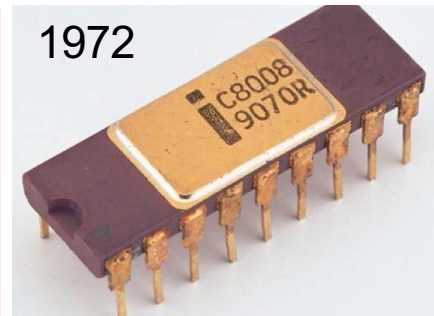
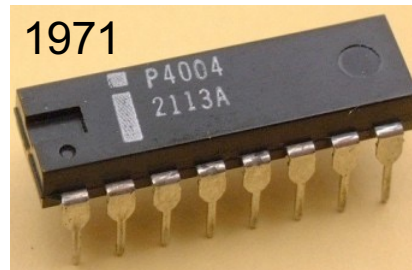
Hierarquia de memória



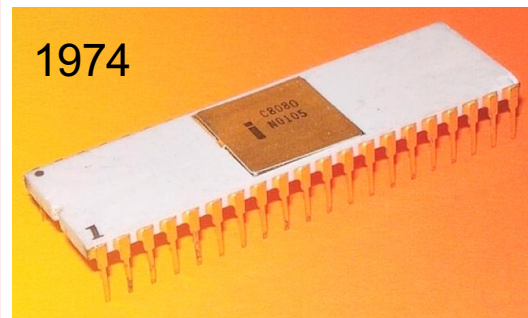
Evolução dos processadores da Intel até à família Intel x86



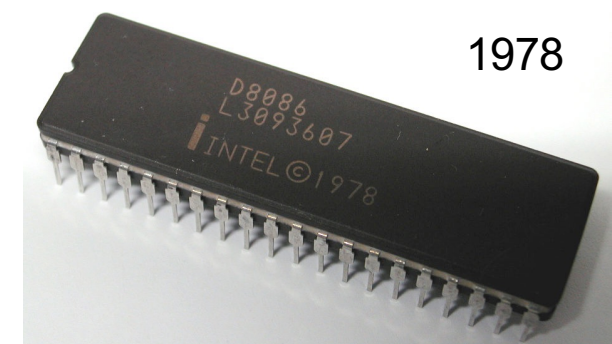
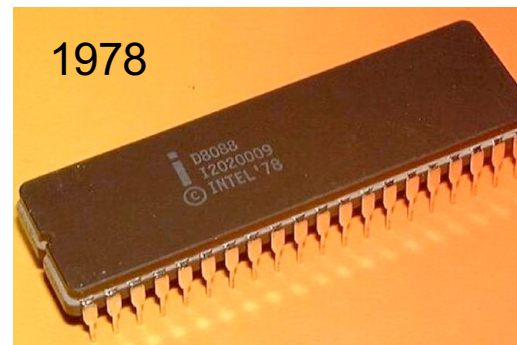
4004: 1º processador
num único *chip*
(microprocessador)



8008 e 8080:
1ºs microprocessadores de 8 bits



8088 e 8086:
1ºs microprocessadores
de 16 bits
(selecionados para o IBM PC)



Evolução do Intel x86 : pré-Pentium

(visão do programador) (1)



<i>Nome</i>	<i>Data</i>	<i>Nº transístores</i>	
8086	1978	29K	
<i>– processador 16-bits (registos + ALU); base do IBM PC & DOS</i>			
<i>– espaço de endereçamento limitado a 1MiB (DOS apenas vê 640Ki)</i>			
80286	1982	134K	
<i>– endereço 24-bits e protected-mode; base do IBM PC-AT & Windows</i>			
386	1985	275K	→ <u>primeiro IA-32 !!</u>
<i>– estendido para 32-bits: registos + op. inteiros + endereçamento</i>			
<i>– memória segmentada+paginada, capaz de correr Unix</i>			
486	1989	1.9M	
<i>– integração num único chip: 386, co-proc 387, até 16KiB cache L1</i>			
<i>– poucas alterações na arquitetura interna do processador</i>			

Evolução do IA-32: família Pentium (visão do programador) (2)



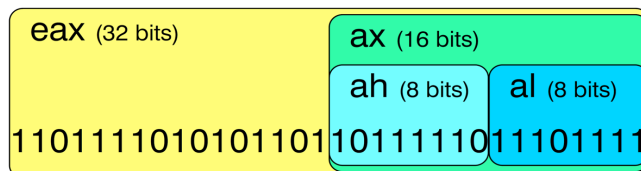
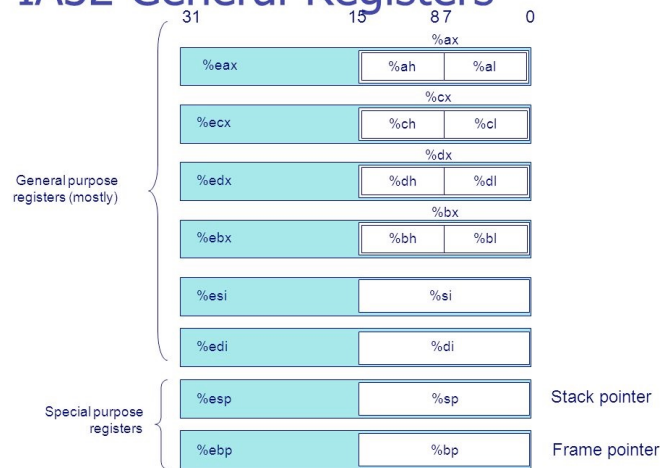
Pentium	1993	3.1M	(= P5 , aka i586)
– arquitetura superescalar, com 2 pipelines de inteiros (de 5 níveis)			
PentiumPro	1995	5.5M	(= P6 , aka i686)
– out-of-order execution, 14 níveis de pipeline, 3-issue superscalar			
– endereço 36-bits, cache L2 on-package			
Pentium/MMX	1997	4.5M	
– SIMD: opera com vetores de 64-bits, tipo <i>int</i> de 1, 2, ou 4 bytes			
Pentium II	1997	7.5M	(= <i>Pro</i> + <i>MMX</i>)
Pentium III	1999	8.2M	(+Celeron, +Xeon)
– “Streaming SIMD Ext”, SSE: vetores 128-bits, <i>int</i> / <i>fp</i> 1/2/4 bytes			
Pentium 4	2000	42M	(= NetBurst , aka i786)
– trace cache, pipeline muito longo (20 ou 31 níveis), suporta multi-threading			
– SSE2: mais instruções e com dados <i>fp</i> de 8-bytes			
Pentium M	2003	77M	(= P-M)
– arquitetura mais próxima do Pentium III (eficiência energética)			

Evolução do IA-32 para Intel 64 (visão do programador) (3)



- IA-32 ou x86 open architecture cresce para 64-bits
 - HP e Intel propõem arquitetura incompatível c/ IA-32: IA-64 (Itanium PU)
 - AMD anuncia em 1999 extensão do x86: x86-64
 - Intel segue AMD: IA-32e (Fev-04), EM64T (Mar-04), ou **Intel 64** (2006)
 - AMD64 e Intel 64 diferentes, mas compiladores usam sub-set comum

IA32 General Registers



x86-64 Integer Registers

%rax	%eax	%r8	%r8d
%rbx	%ebx	%r9	%r9d
%rcx	%ecx	%r10	%r10d
%rdx	%edx	%r11	%r11d
%rsi	%esi	%r12	%r12d
%rdi	%edi	%r13	%r13d
%rsp	%esp	%r14	%r14d
%rbp	%ebp	%r15	%r15d

- Twice the number of registers
- Accessible as 8, 16, 32, 64 bits

Intel 64 ≠ IA-64 (Itanium): nos registros

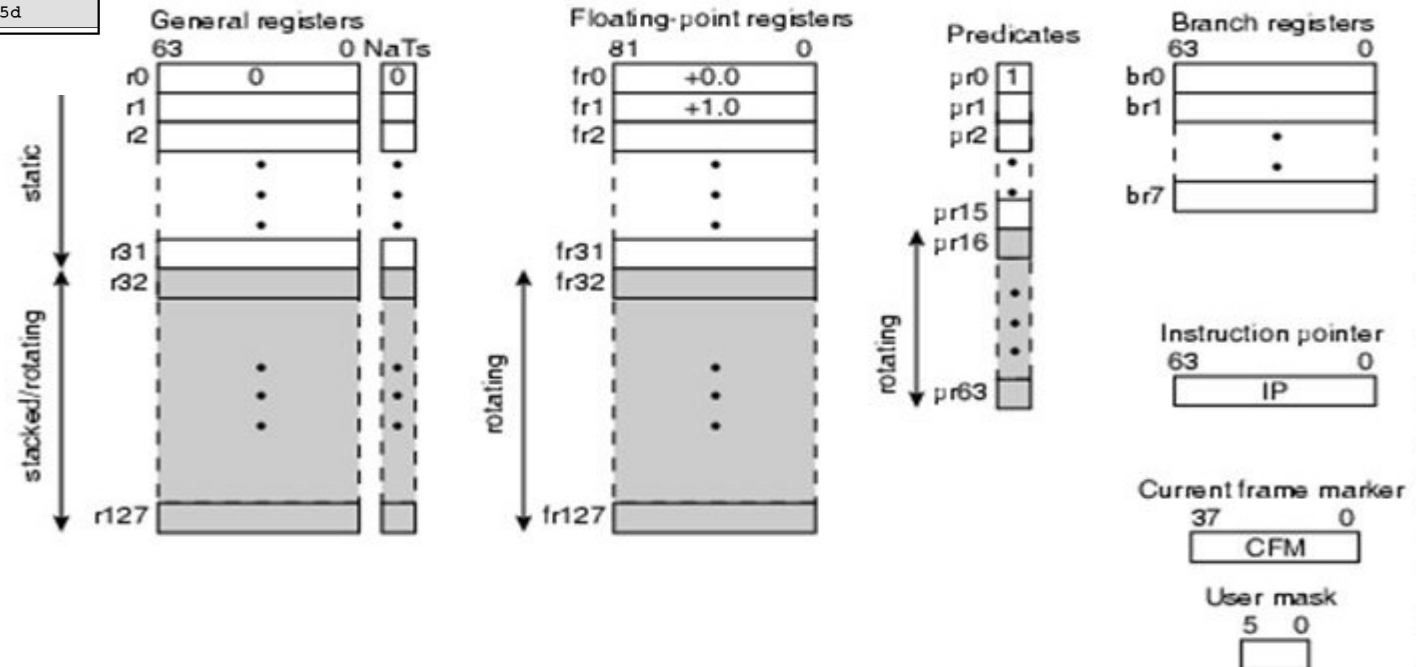


x86-64 Integer Registers

%rax	%eax	%r8	%r8d
%rbx	%ebx	%r9	%r9d
%rcx	%ecx	%r10	%r10d
%rdx	%edx	%r11	%r11d
%rsi	%esi	%r12	%r12d
%rdi	%edi	%r13	%r13d
%rsp	%esp	%r14	%r14d
%rbp	%ebp	%r15	%r15d

- Twice the number of registers
- Accessible as 8, 16, 32, 64 bits

IA-64 Register Set



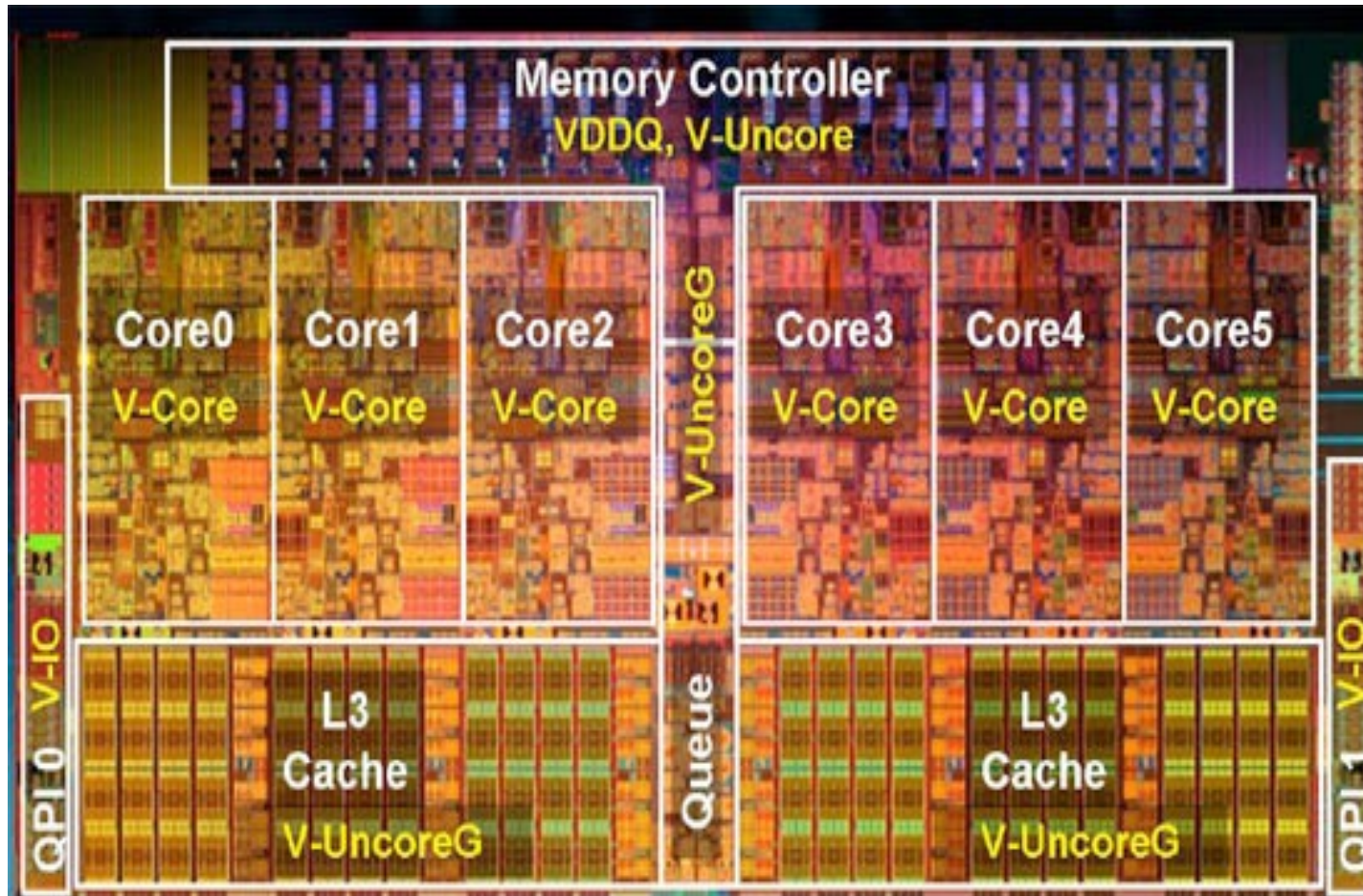
Arquiteturas Intel 64 com maior integração (visão do programador) (1)



- arquitetura **Core** surge em 2006 (*151M transístores*)
 - desenvolvida pela mesma equipa que o P-M (Israel)
 - 14 níveis de pipeline (como P6), mas 4-issue superscalar
 - 2 níveis de cache on-chip
 - multi-core on-chip e virtualização por h/w
 - suporta fusão de instruções RISC (μ -ops na terminologia Intel)
 - arquitetura Core 2 é integralmente 64-bit (Intel 64)
- arquitetura **Nehalem** anunciada em 2008 (*731M transístores*)
 - inspirada no NetBurst (com multi-threading e maiores frequências de clock)
 - 2 a 8 cores por chip, com cache L3 on-chip
 - com conexão ponto-a-ponto inter-CPU_chips
 - integra controlador de memória, rumo a servidores com arquitetura NUMA



Intel Hex-Core Nehalem (1.17 mil milhões de transístores)



Arquiteturas Intel 64 com maior integração

(visão do programador) (2)

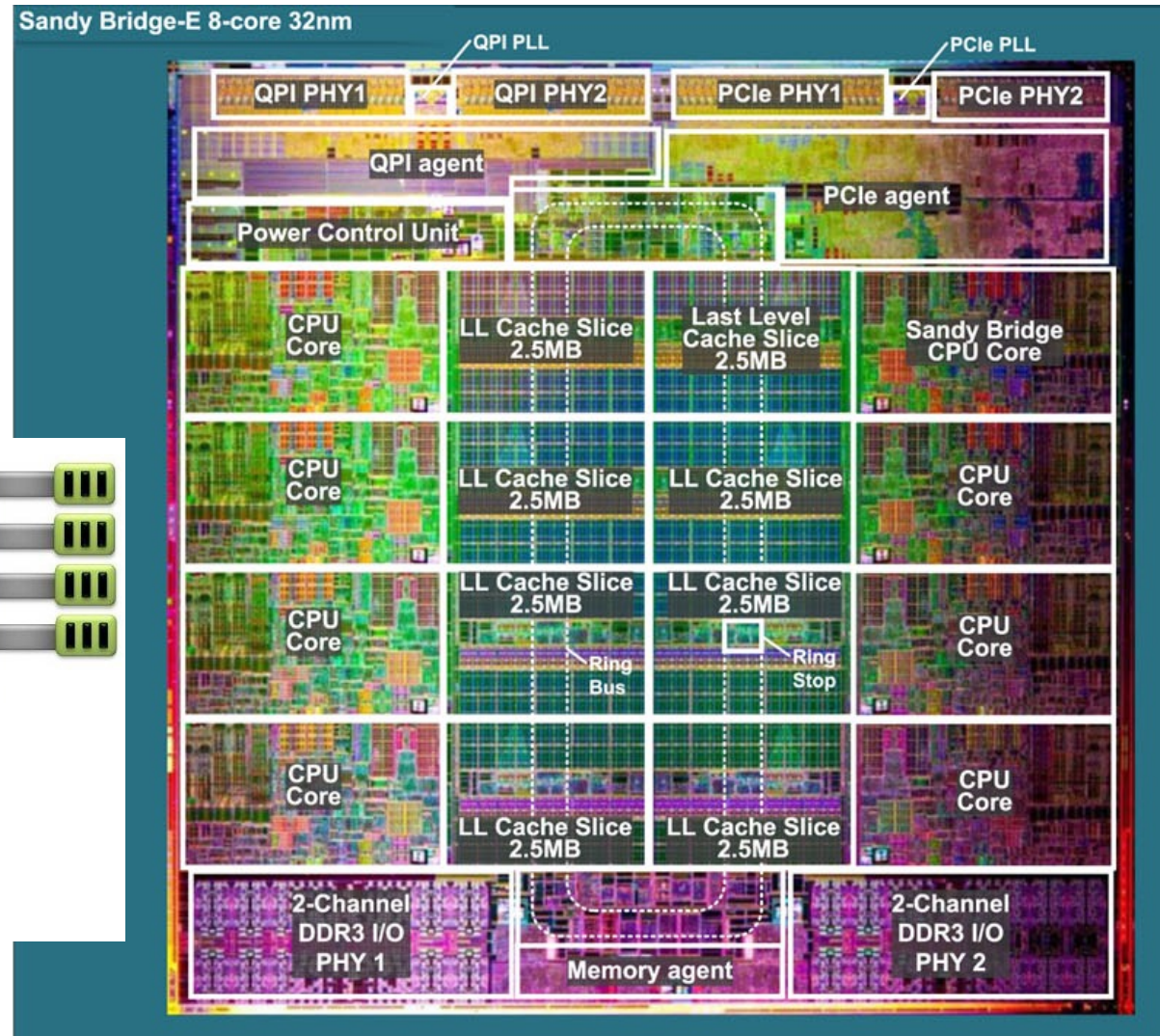
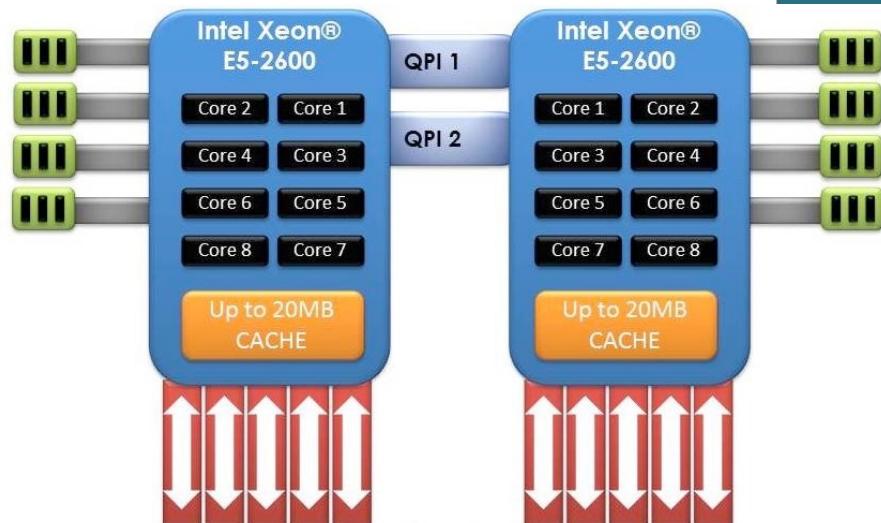


- arquitetura **Sandy Bridge** anunciada em 2010 (>1,000M transístores)
 - ensaio de arquitetura híbrida multicore, integrando o processador gráfico
 - interface com o processador gráfico através da cache L3
 - processamento vetorial de fp estendido para 256-bits (AVX)
 - integração no chip do interface PCIe 16x
- arquitetura **Haswell** anunciada em 2013 (>5,500M transístores)
 - nível adicional de cache para μ -ops (formato RISC)
 - processamento vetorial integral com 256-bits (AVX2)
 - 2 unidades vetoriais para operações com inteiros
 - até 22 cores e 55 MiB de cache L3 (Xeon)
 - introdução de cache L4 de 128 MiB , eDRAM (off-chip. on-package)
- arquitetura **Skylake** anunciada em 2015 (>8,000M transístores)
 - mais 1 unidade vetorial para operações com inteiros (total: 3)
 - AVX-512 para topo de gama Xeon (Skylake-SP)

Xeon Sandy Bridge



Arquitetura NUMA



Arquiteturas Intel 64 com maior integração (visão do programador) (3)

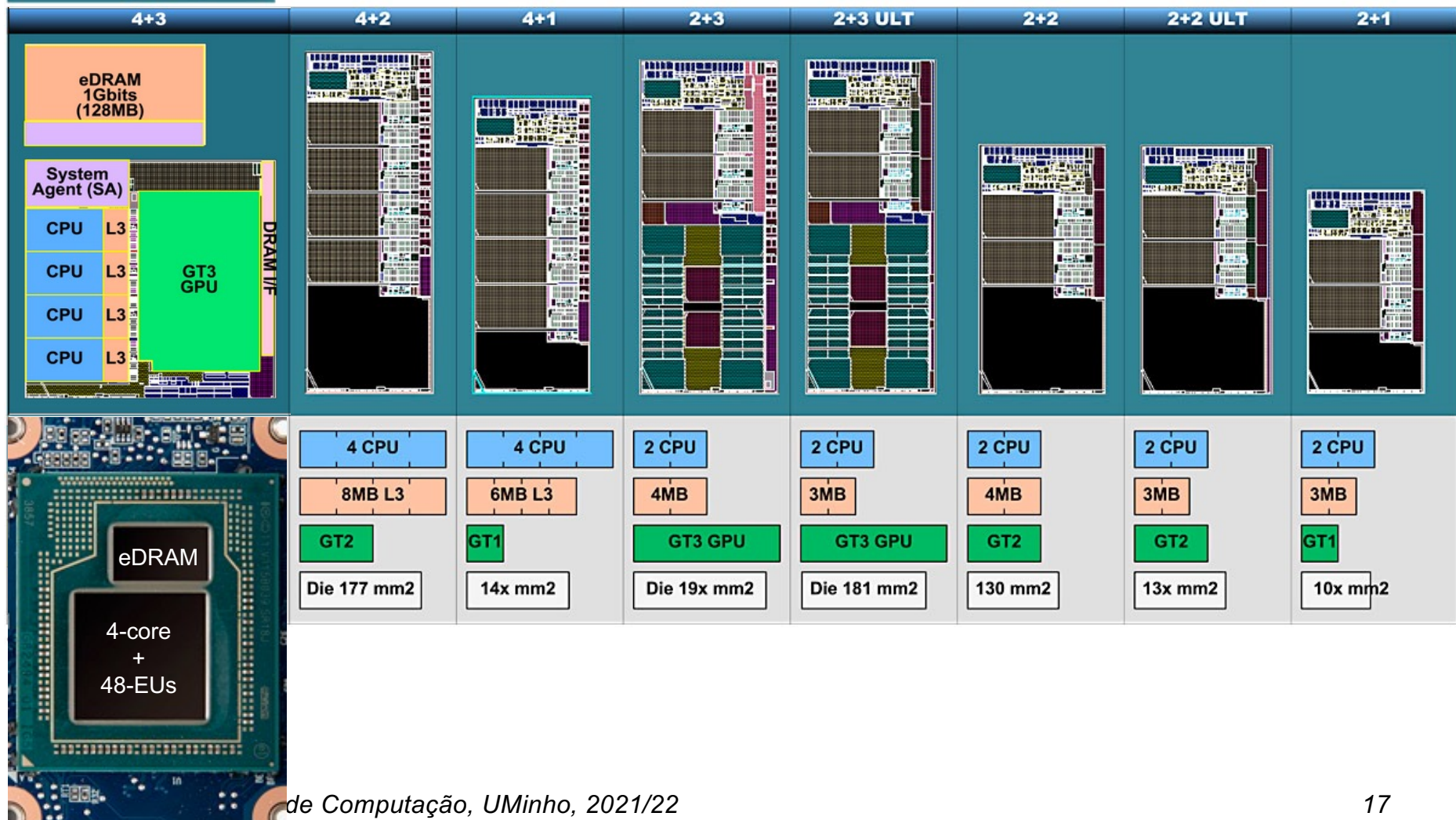


- arquitetura **Sandy Bridge** anunciada em 2010 (>1,000M transístores)
 - ensaio de arquitetura híbrida multicore, integrando o processador gráfico
 - interface com o processador gráfico através da cache L3
 - processamento vetorial de fp estendido para 256-bits (AVX)
 - integração no chip do interface PCIe 16x
- arquitetura **Haswell** anunciada em 2013 (>5,500M transístores)
 - nível adicional de cache para μ -ops (formato RISC)
 - processamento vetorial integral com 256-bits (AVX2)
 - 2 unidades vetoriais para operações com inteiros
 - até 22 cores e 55 MiB de cache L3 (Xeon)
 - introdução de cache L4 de 128 MiB , eDRAM (off-chip, on-package)
- arquitetura **Skylake** anunciada em 2015 (>8,000M transístores)
 - mais 1 unidade vetorial para operações com inteiros (total: 3)
 - AVX-512 para topo de gama Xeon (Skylake-SP)

Lançamento da Intel em 2013: 8 configurações para i5 & i7 Haswell



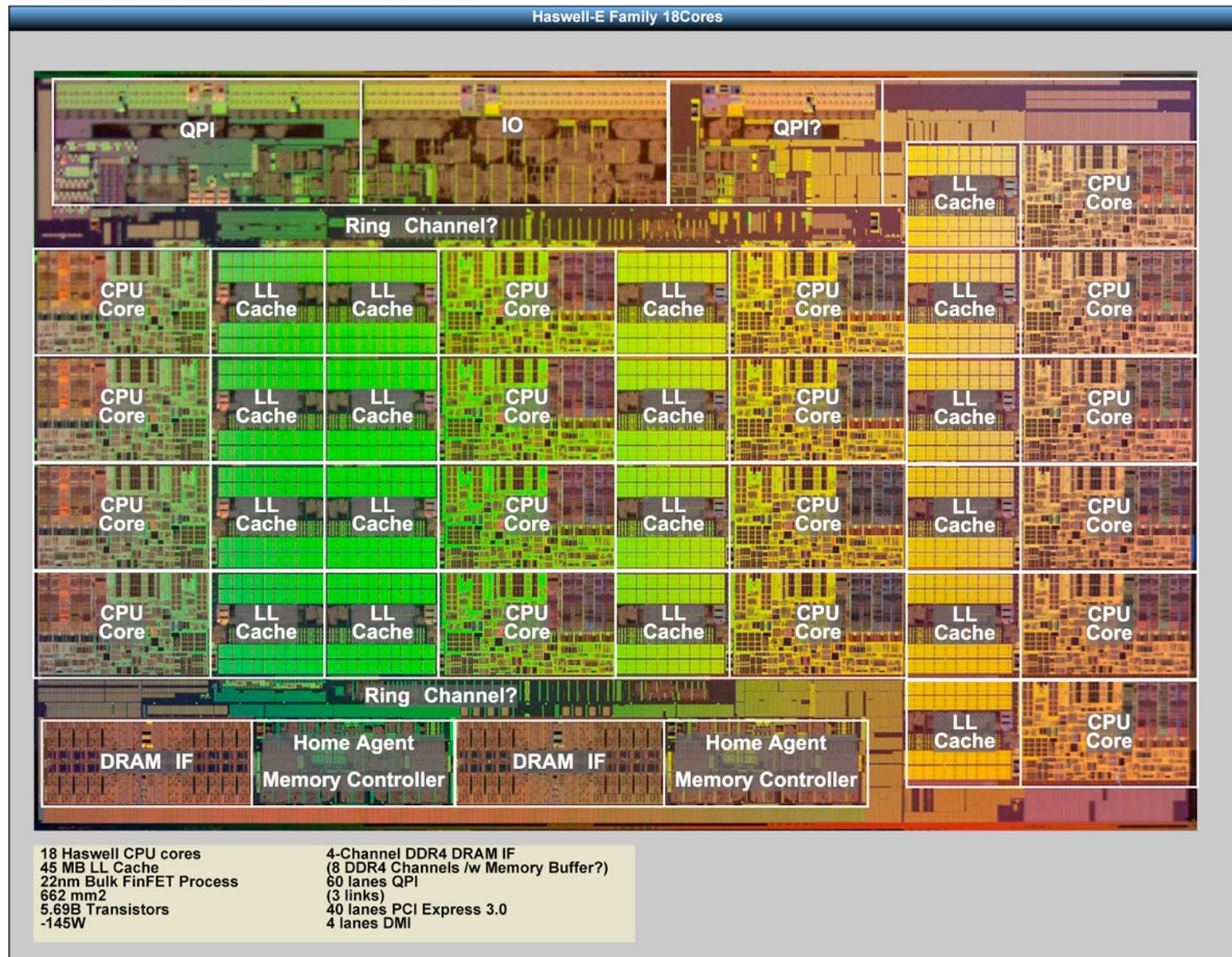
Haswell Family Die Layout



Lançamento da Intel em 2016: 18-core Xeon Haswell



Haswell-E Family (EP/EX) 18 Cores Die Layout





Arquiteturas Intel 64 com maior integração

(visão do programador) (4)



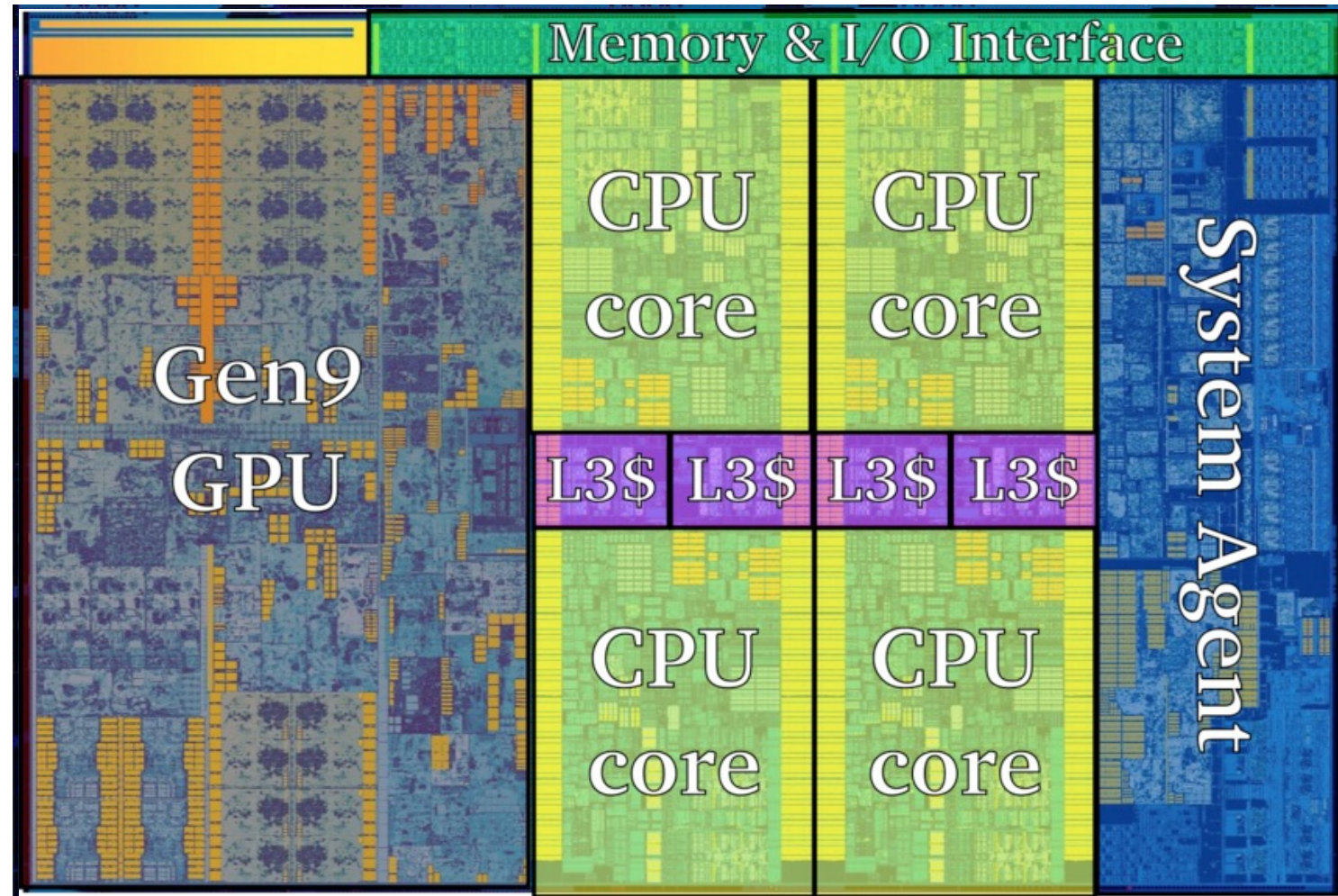
- arquitetura **Sandy Bridge** anunciada em 2010 (>1,000M transístores)
 - ensaio de arquitetura híbrida multicore, integrando o processador gráfico
 - interface com o processador gráfico através da cache L3
 - processamento vetorial de fp estendido para 256-bits (AVX)
 - integração no chip do interface PCIe 16x
- arquitetura **Haswell** anunciada em 2013 (>5,500M transístores)
 - nível adicional de cache para μ -ops (formato RISC)
 - processamento vetorial integral com 256-bits (AVX2)
 - 2 unidades vetoriais para operações com inteiros
 - até 22 cores e 55 MiB de cache L3 (Xeon)
 - introdução de cache L4 de 128 MiB , eDRAM (off-chip, on-package)
- arquitetura **Skylake** anunciada em 2015 (>8,000M transístores)
 - mais 1 unidade vetorial para operações com inteiros (total: 3)
 - AVX-512 para topo de gama Xeon (Skylake-SP)
- arquitetura Xeon a seguir: **Sunny Cove**, anunciada a 6-abr-21
 - Xeon Ice Lake, redução de 14nm para 10nm, com 2 anos de atraso...
 - caches de maior dimensão

Intel 4-core Skylake

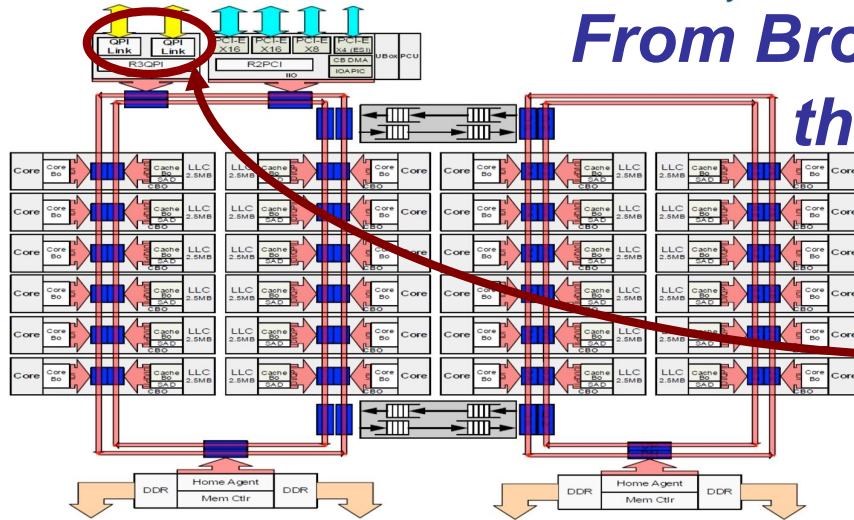


GPU integrado
com 24 EU's

Chip layout
anotado



From Broadwell to Skylake (server): the move from ring to mesh



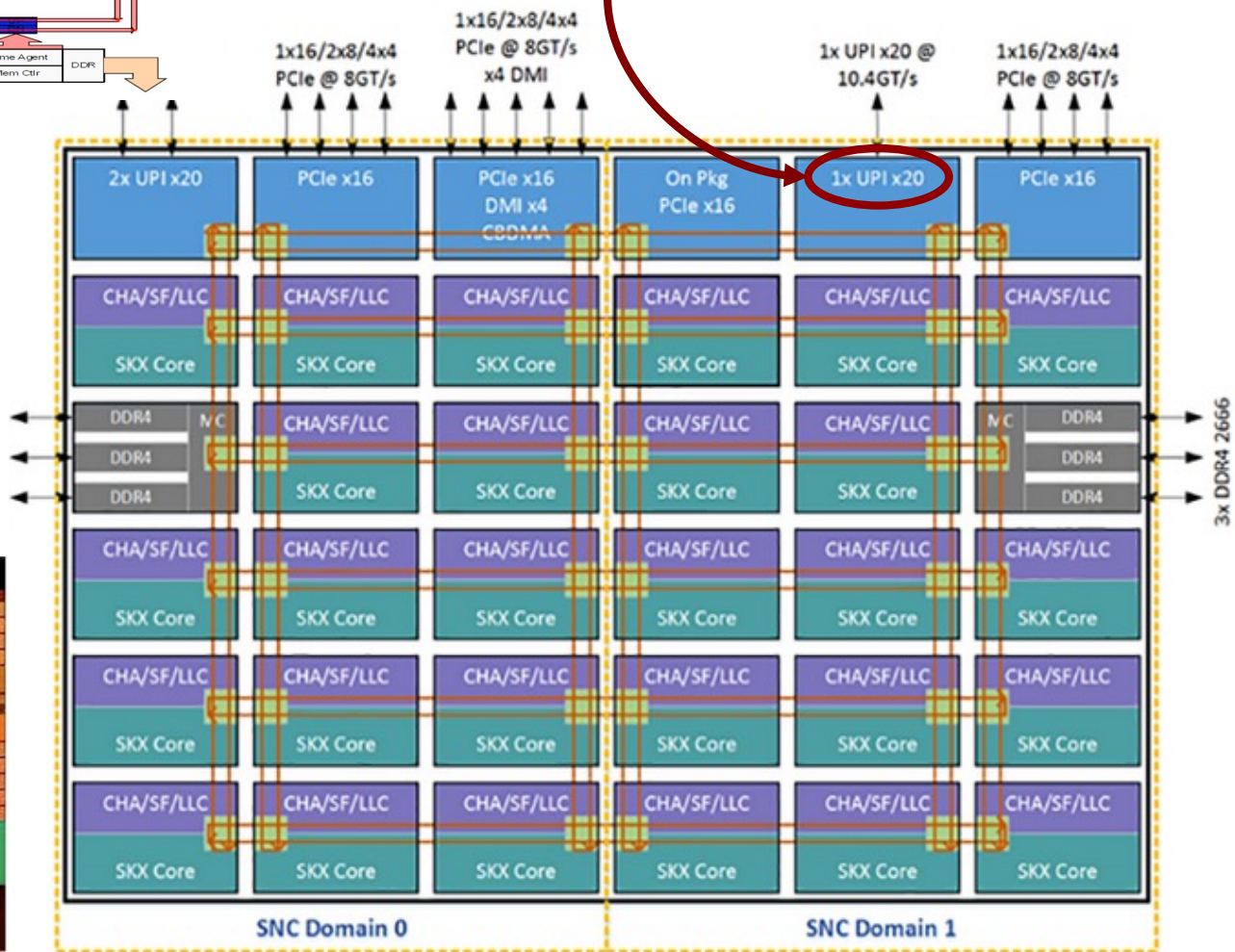
UPI required for dual-socket
(Ultra Path Interconnect)

Broadwell

ring interconnection
does not scale for
large #cores

Skylake (server)

(mesh follows KNL)



Intel CPU core roadmaps

