ПЕРСПЕКТИВЫ РАЗВИТИЯ АППАРАТНЫХ ТЕХНОЛОГИЙ И ИХ ПРИМЕНЕНИЕ В СУПЕРКОМПЬЮТЕРАХ ЭКЗАФЛОПСНОГО УРОВНЯ

23 октября 2013 г.

Андрей Слепухин
Главный системный архитектор
andrey.slepuhin@t-platforms.ru

Требования к Exascale-технологиям

	2012 (BG/Q)	2018-2020 (Exascale)	Относительно 2012
System peak	20 Pflops	1 Eflops	$O(10^2)$
Power	8.6 MW	~20 MW	
System memory	1.6 PB	32-64 PB	O(10)
Node performance	205 Gflops	1.2 or 15 Tflops	$O(10)$-$O(10^2)$
Node memory BW	42.6 GB/s	2-4 TB/s	$O(10^3)$
Node concurrency	64 threads	$O(10^3)$ or $O(10^4)$	$O(10^2)$-$O(10^3)$
Node interconnect BW	20 GB/s	200-400 GB/s	O(10)
System size (nodes)	98304	$O(10^5)$ or $O(10^6)$	$O(10)$-$O(10^2)$
Total concurrency	5.97 M	$O(10^9)$	$O(10^3)$
MTTI	4 days	O(<1 day)	-O(10)

Ключевые технологии для Exascale

- Процессоры и ускорители
- Память
- Интерконнект и оптические коммуникации

Прогноз развития микроэлектроники

- Технологический процесс производства микросхем (оптимистичный)

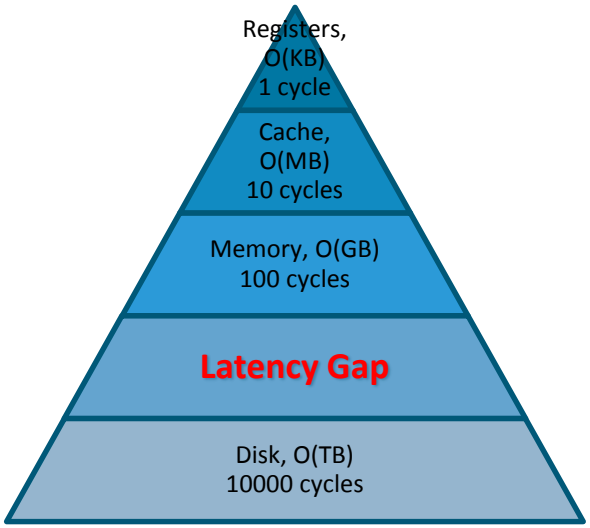
	2012	2013	2014	2015	2016	2017	2018	2019	2020
Intel	22nm FinFET	14nm DP		10nm QP		7nm EUV			5nm
Другие фабрики	28nm	20nm DP	14/16nm FinFET		10nm QP		7nm EUV		

- Технологический процесс производства микросхем (пессимистичный)

	2012	2013	2014	2015	2016	2017	2018	2019	2020
Intel	22nm FinFET		14nm DP		10nm QP			7nm EUV	
Другие фабрики	28nm		20nm DP	14/16nm FinFET		10nm QP			7nm EUV

Развитие технологий памяти

- Новые виды памяти
  - SCM (Storage-Class Memory)
    - PCRAM
    - STT-MRAM
    - Redox RAM
  - Увеличение пропускной способности и интеграция
    - HMC (Hybrid Memory Cube)
    - Wide-IO memory



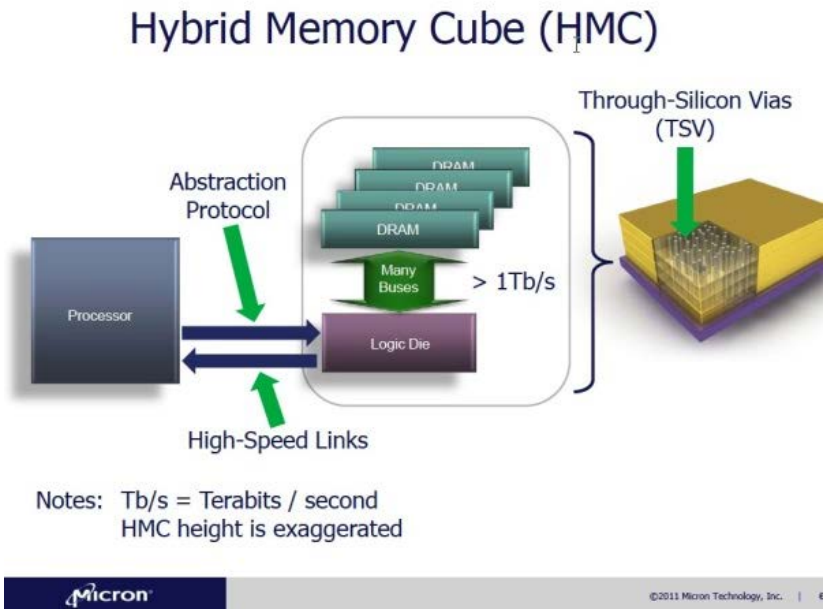
Существующие технологии памяти

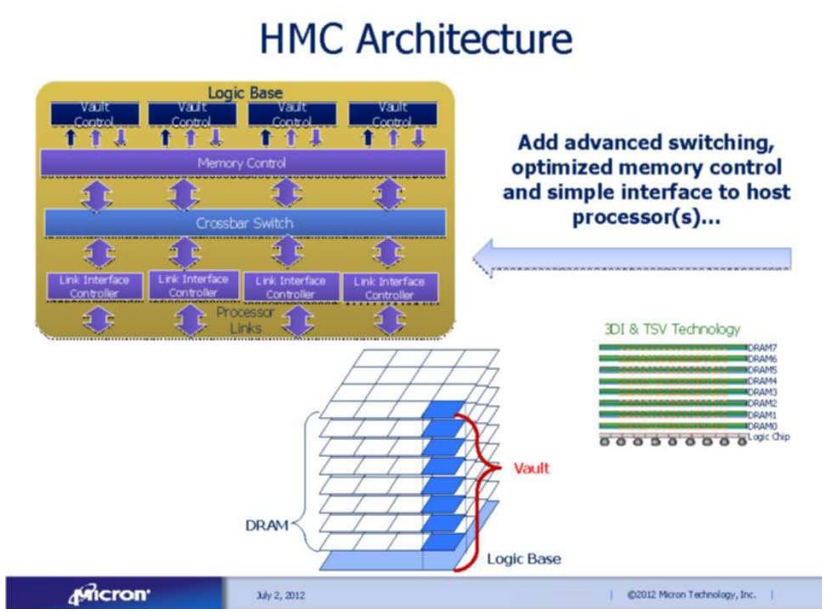
		Baseline Technologies					Prototypical technologies [A]		
		DRAM		SRAM [C]	Flash		FeRAM	STT-MRAM	PCM
		Stand-alone [A]	Embedded [C]		NOR Embedded [C]	NAND Stand-alone [A]			
Storage Mechanism		Charge on a capacitor		Inter-locked state of logic gates	Charge trapped in floating gate or in gate insulator		Remnant polarization on a ferroelectric capacitor	Magnetization of ferromagnetic layer	Reversibly changing amorphous and crystalline phases
Cell Elements		1T1C		6T	1T		1T1C	1(2)T1R	1T(D)1R
Feature size F, nm	**2011**	36	65	45	90	22	180	65	45
	2024	9	20	10	25	8	65	16	8
Cell Area	**2011**	6F2	(12-30)F2	140 F2	10 F2	4 F2	22F2	20F2	4F2
	2024	4F2	(12-50)F2	140 F2	10 F2	4 F2	12F2	8F2	4F2
Read Time	**2011**	<10 ns	2 ns	0.2 ns	15 ns	0.1ms	40 ns [G]	35 ns [J]	12 ns [K]
	2024	<10 ns	1 ns	70 ps	8 ns	0.1ms	<20 ns [H]	<10 ns	< 10 ns
W/E Time	**2011**	<10 ns	2 ns	0.2 ns	1ms/10ms	1/0.1 ms	65 ns [G]	35 ns {J]	100 ns [K]
	2024	<10 ns	1 ns	70 ps	1ms/10ms	1/0.1 ms	<10 ns[H]	<1 ns	<50 ns
Retention Time	**2011**	64 ms	4 ms	[D]	10 y	10 y	10 y	>10 y	>10 y
	2024	64 ms	1 ms	[D]	10 y	10 y	10 y	>10 y	>10 y
Write Cycles	**2011**	>1E16	>1E16	>1E16	1E5	1E4	1E14	>1E12	1E9
	2024	>1E16	>1E16	>1E16	1E5	5E3	>1E15	>1E15	1E9
Write Operating Voltage (V)	**2011**	2.5	2.5	1	10	15	1.3-3.3	1.8	3 [K]
	2024	1.5	1.5	0.7	9	15	0.7–1.5	<1	<3
Read Operating Voltage (V)	**2011**	1.8	1.7	1	1.8	1.8	1.3–3.3	1.8	1.2
	2024	1.5	1.5	0.7	1	1	0.7–1.5	<1	<1
Write Energy (J/bit)	**2011**	4E-15 [B]	5.00E-15	5.00E-16	1E-10 [E]	>2E-16 [F]	3E-14 [I]	2.5E-12 [A]	6E-12 [L]
	2024	2E-15 [B]	2.00E-15	3.00E-17	1E-11 [E]	>2E-17 [F]	7E-15 [I]	1.5E-13 [A]	~1E-15 [M]

Перспективные технологии памяти

		A. Emerging Ferroelectric memory	B . Nanomechanical Memory	C. Redox Memory	D. Mott Memory	E. Macromolecular Memory	F. Molecular Memories
Storage Mechanism		Remnant polarization on a ferroelectric dielectric	Electrostatically-controlled mechanical switch	Ion transport and redox reactions	Multiple mechanisms	Multiple mechanisms	Multiple mechanisms
Cell Elements		1T or 1T1R or 1D1R	1T1R or 1D1R	1T1R or 1D1R	1T1R or 1D1R	1T1R or 1D1R	1T1R or 1D1R
Device Types		1) FET with FE gate insulator 2) FE barrier effects	NEMS	1) cation migration 2) anion migration	Mott transition	M-I-M (nc)-I-M	Bi-stable switch
Feature size F	Min. required	<65 nm	<65 nm	<65 nm	<65 nm	<65 nm	<65 nm
	Best projected	22 nm [A1]	>50 nm [B1, B2]	5 nm [C1]	5-10 nm	5-10 nm	5 nm [F1]
	Demonstrated	0.6 mm [A2]	500 nm [B3, B4]	30 nm [C2], 9nm [C7]	10 mm [D1]	130 nm [E1]	30 nm [F2]
Cell Area	Min. required	8F2	8F2	8F2	8F2	8F2	8F2
	Best projected	4F2	4F2	4F2	4F2	4F2	4F2
	Demonstrated	Data not available	Data not available	4F2 [C2], 8F2 [C3]	Data not available	4F2 [E1]	Data not available
Read Time	Min. required	<15 ns	<15 ns	<15 ns	< 15 ns	<15 ns	<15 ns
	Best projected	2.5 ns	<10 ns	<10 ns	< 10 ns	<10 ns	<10 ns [F1]
	Demonstrated	20 ns [A3]	Data not available	<50 ns [C3]	Data not available	10 ns [E1]	Data not available
W/E time	Min. required	Application dependent	Application dependent	Application dependent	Application dependent	Application dependent	Application dependent
	Best projected	2.5 ns [A1]	<1 ns [B1, B2]	<1 ns [C4]	<1 ns [D2]	<10 ns	<40 ns [F1]
	Demonstrated	20 ns [A4]	~5 ns [B3, B4]	0.3ns [C5]	< 20 ns [D3]	15 ns [E2]	10s [F6],0.2 s [F3]
Retention Time	Min. required	>10 y	>10 y	>10 y	>10 y	>10 y	>10 y
	Best projected	>10 y [A4]	>10 y	>10 y	Not known	Not known	Not known
	Demonstrated	~3.5 month [A6]	~days	>10 y [C2]	Not known	~year [E3]	1 hour [F6], 2 months [F4]
Write Cycles	Min. required	>1E5	>1E5	>1E5	>1E5	>1E5	>1E5
	Best projected	>1E16	>1E16	>1E16	>1E16	>1E16	>1E16
	Demonstrated	2E11 [A5]	~1E3 [B4]	1E12 [C2]	~1E2 [D4]	~1E5 [E4]	~2E3 [F2]
Write operating voltage (V)	Min. required	Application dependent	Application dependent	Application dependent	Application dependent	Application dependent	Application dependent
	Best projected	<0.9 V [A1]	>1 V [B1, B2]	<0.5 V [C6]	Not known	<1 V [E5]	80 mV[F5]
	Demonstrated	±4[A4]	5 V [B3, B4]	0.6/-0.2 [C3]	1.25/0.75 V [D1]	~±2 V [E3]	4V[F6], ~±1.5 V [F2]
Read operating voltage (V)	Min. required	2.5	2.5	2.5	2.5	2.5	2.5
	Best projected	0.7	0.7	<0.2 V [C6]	Not known	0.7	0.3 [F1]
	Demonstrated	2.5 [A3]	1 [B3]	0.15 [C3]	0.2 [D1]	0.5 V [E3]	0.5V [F6],0.5 V [F2]
Write energy (J/bit)	Min. required	Application dependent	Application dependent	Application dependent	Application dependent	Application dependent	Application dependent
	Best projected	2E-15 [A7]	1E-17 [B5]	1E-17 [C4]	Not known	Not known	2E-19 [F6]
	Demonstrated	Data not available	Data not available	1E-13 [C7]	5E-13 [D5]	5E-11 [E6]	Data not available

Hybrid Memory Cube



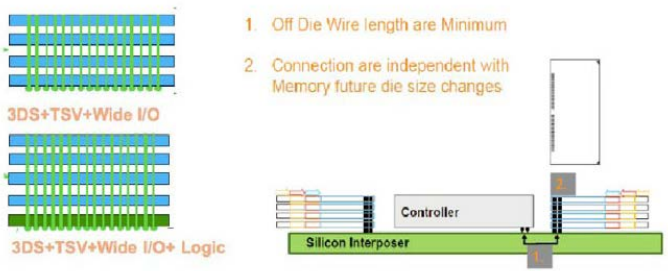


Highest Performance and Most Energy Efficient DRAM Memory in the Industry

Technology	VDD	IDD	BW GB/s	Power (W)	mW/GB/s	pj/bit
SDRAM PC133 1GB ECC Module	3.3	2.33	1.06	7.69	7226.50	903.31
DDR-333 1GB ECC Module	2.5	3.00	2.66	7.50	2815.32	351.91
DDRII-667 2GB ECC Module	1.8	2.89	5.34	5.20	974.89	121.86
DDR3-1333 4GB ECC Module	1.5	3.07	10.66	4.61	431.83	53.98
DDR4-2667 8GB ECC Module	1.2	2.83	21.34	3.40	159.17	19.90
GDDR5 Die	1.35	2.00	20.00	2.70	135.00	16.88
LPDDR2-1066 X32 Die	1.2	0.356	4.26	0.43	100.28	12.54
HMC Gen1 512MB Cube	1.2	6.64	128.00	7.97	62.23	7.78

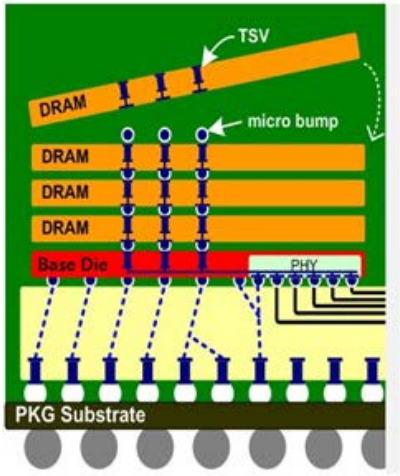
JEDEC HIGH-BANDWIDTH MEMORY (HBM)

JEDEC HBM:



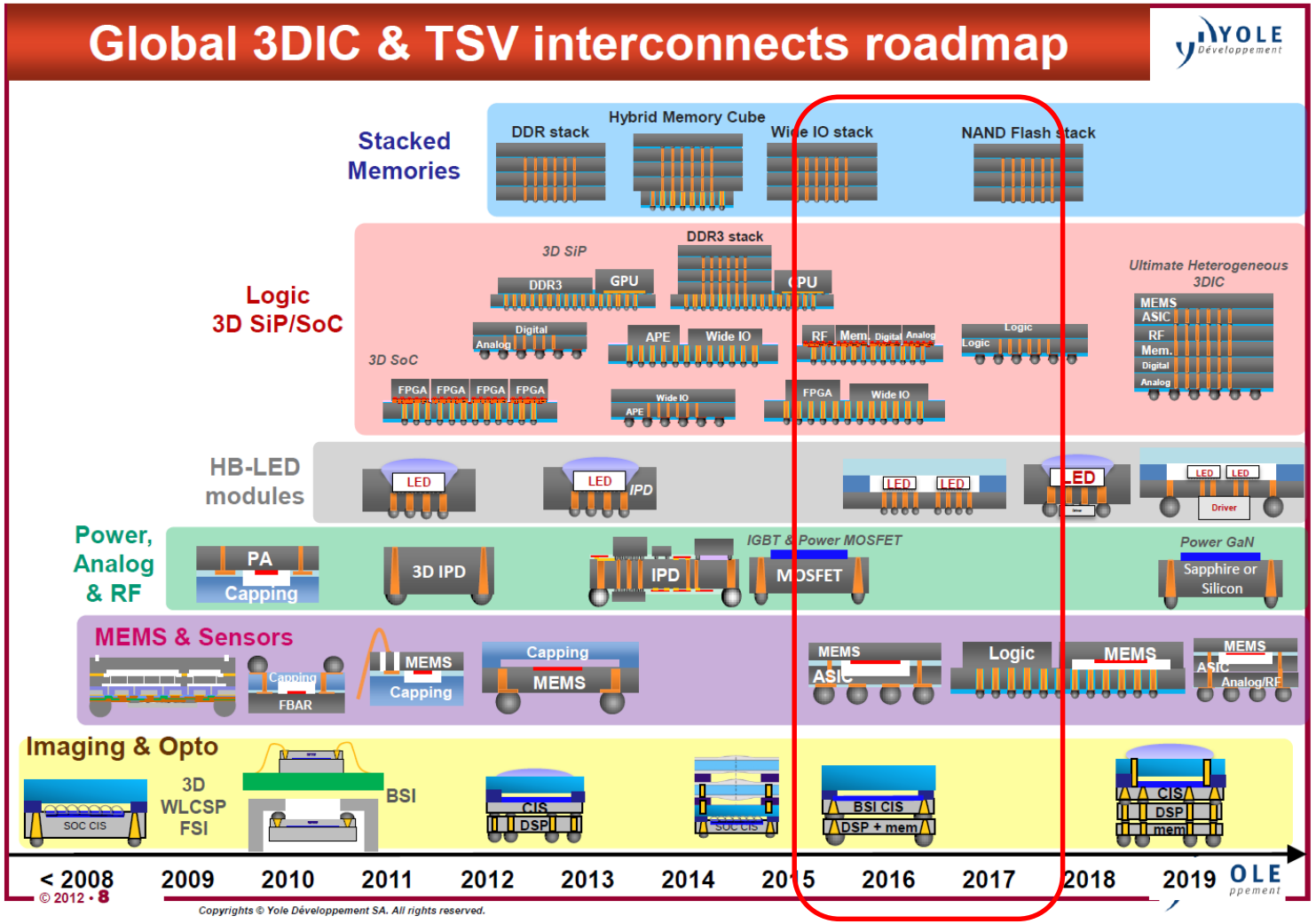
- Phase 1: focus on device architecture with wide I/O
- Phase 2: 3DS+TSV stacking
- Phase 3: 3DS cube +Logic.

JEDEC — Global Standards for the Microelectronics Industry



ITEM	TARGET
Burst Length	2, 4
Stack Density	1GByte per stack (2Gbit per slice)
Channel / Slice	2
Banks / Channel	8
IO / Channel	128
Prefetch / Channel	32B (128x2bit)
Channels / Stack	8
Total TSV Data IO Width	1024
Clock Speed	500MHz
Peak Read BW / Stack	128 GB/s
Page Size	2KB
Data Parity	1 bit / 32 bit
DRAM Core Voltage	1.2V
Logic Buffer IO Voltage	1.2V

Развитие технологий 3D-интеграции



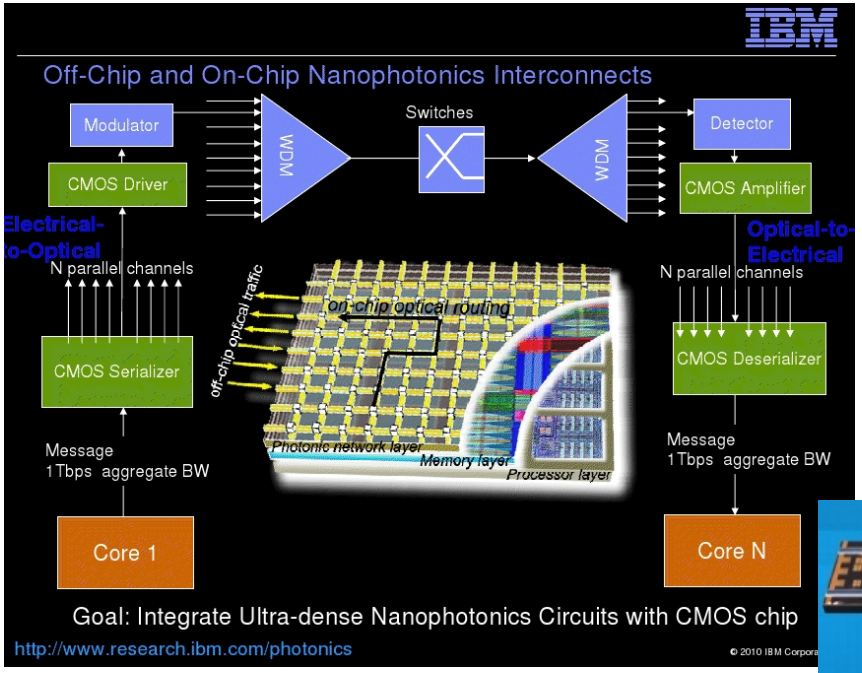
Интеграция компонентов

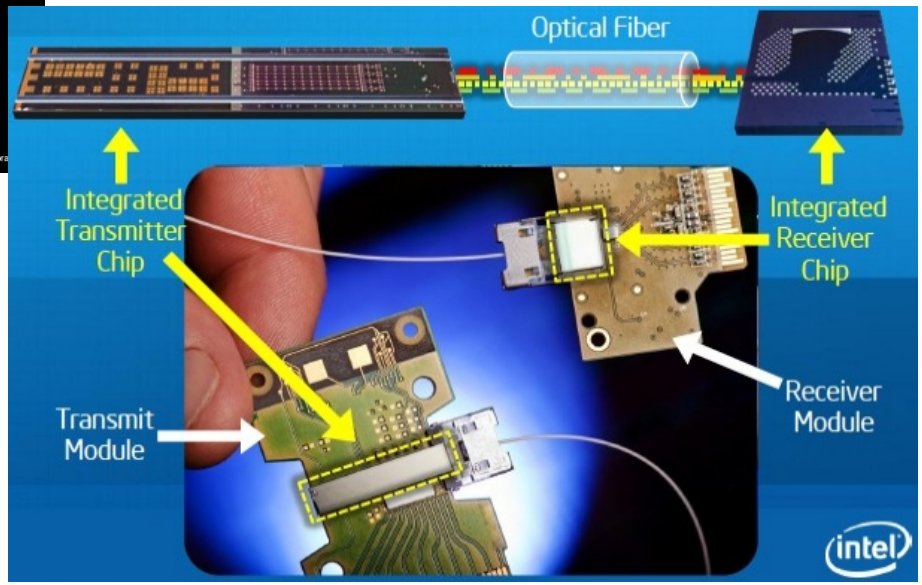
- Интеграция процессора и интерконнекта
- Интеграция процессора и памяти
- К 2020 году 1 узел ≈ 1 чип (возможно с дополнительной памятью большого объема)
- Перспективы развития:
  - 2015-2016:
    - Интеграция интерконнекта в процессор
    - Интеграция процессора и памяти в одном корпусе (system-in-a-package)
  - 2017-2018
    - 3D-stacking процессора и памяти

Развитие оптических технологий

- Возможности высокоскоростной передачи данных по медным проводникам ограничены
  - 40Gbit/sec => несколько сантиметров по плате, ≈1 м по кабелю
- Оптическая передача данных обеспечивает намного лучшую плотность и энергоэффективность
- Перспективы развития:
  - Интеграция оптических трансиверов в одном модуле с логическими устройствами (опытные образцы есть уже сейчас): 2015-2016
  - Интеграция оптических волноводов на печатную плату: 2017-2018
  - Полный переход к кремниевой фотонике, оптические коммуникации внутри микросхемы: после 2018

Развитие оптических технологий

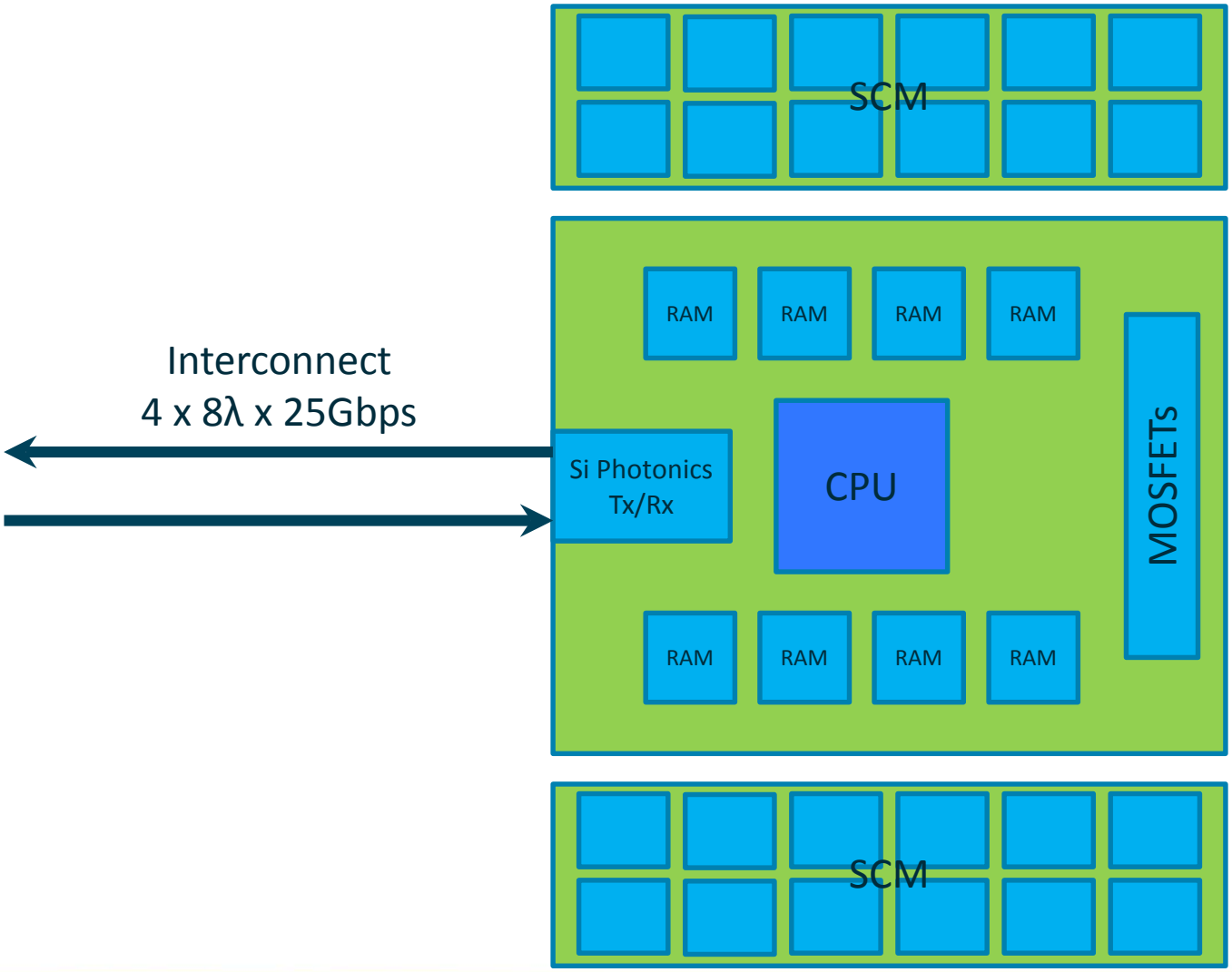




Вычислительный узел для Exascale

- Технологический процесс: 10 nm
- Производительность: ~4 Tflops
- Объем интегрированной памяти: 64-128GB
- Пропускная способность памяти: ~1TB/sec
- Пропускная способность интерконнекта: ~100GB/sec
- Объем SCM: 512GB-1TB
- Пропускная способность SCM: ~100GB/sec
- Энергопотребление: ~150W

Вычислительный узел для Exascale



СПАСИБО!

Андрей Слепухин
Главный системный архитектор
andrey.slepuhin@t-platforms.ru