

Photonic and Electronic Convergence Technologies for 1-chip Data Center

Jan. 27th 2015

**Photonics Electronics Technology Research Association(PETRA)
Jun Inasaka**

Agenda

- **Hardware Issues of Large Scale System**
- **Optical Interface inside the sever**
- **Photonics Electronics Convergence Technology for Power-Reducing Jisso System” project**
- **Photonics with Electronics**
- **Conclusions**

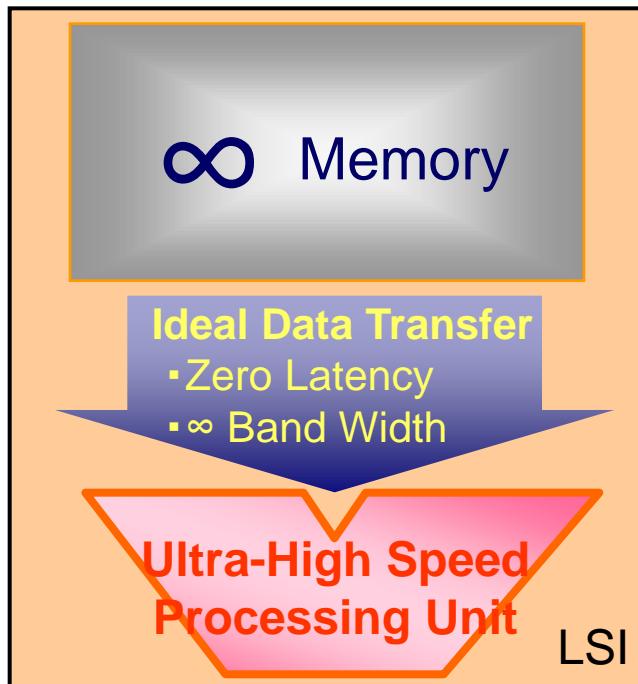
 **Memory Wall**

 **Power Wall**

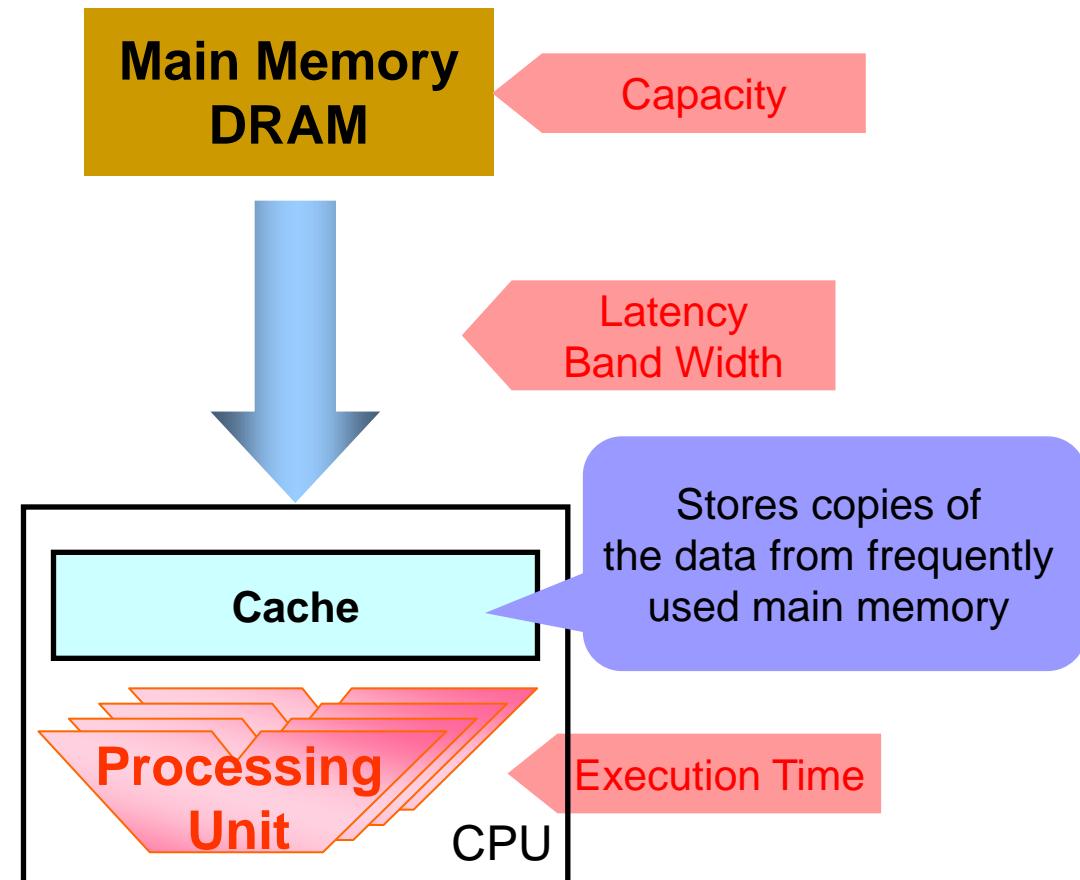
 **Electrical Transmission Wall**

 **Large Scale System**

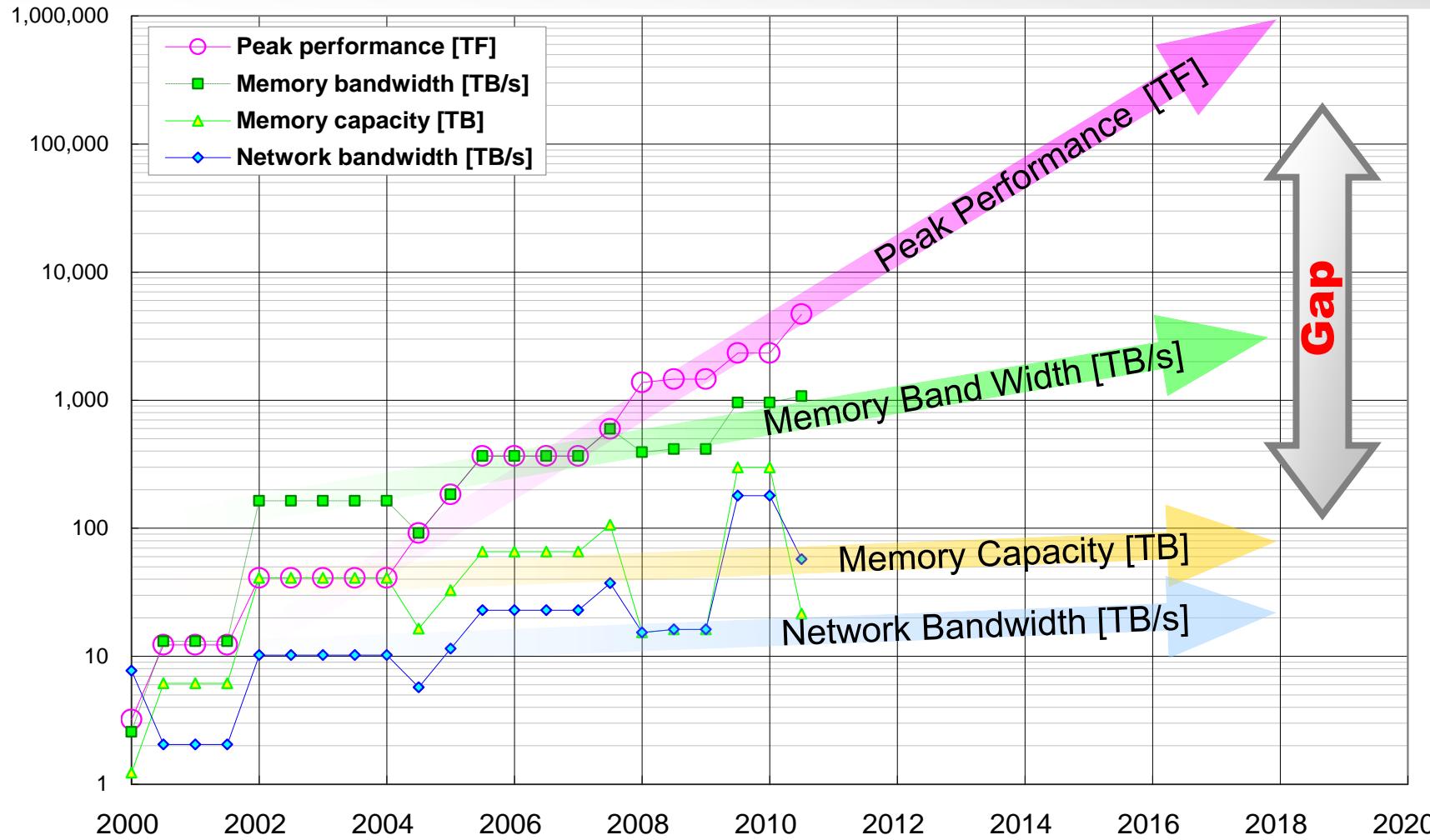
Ideal



Real . . .

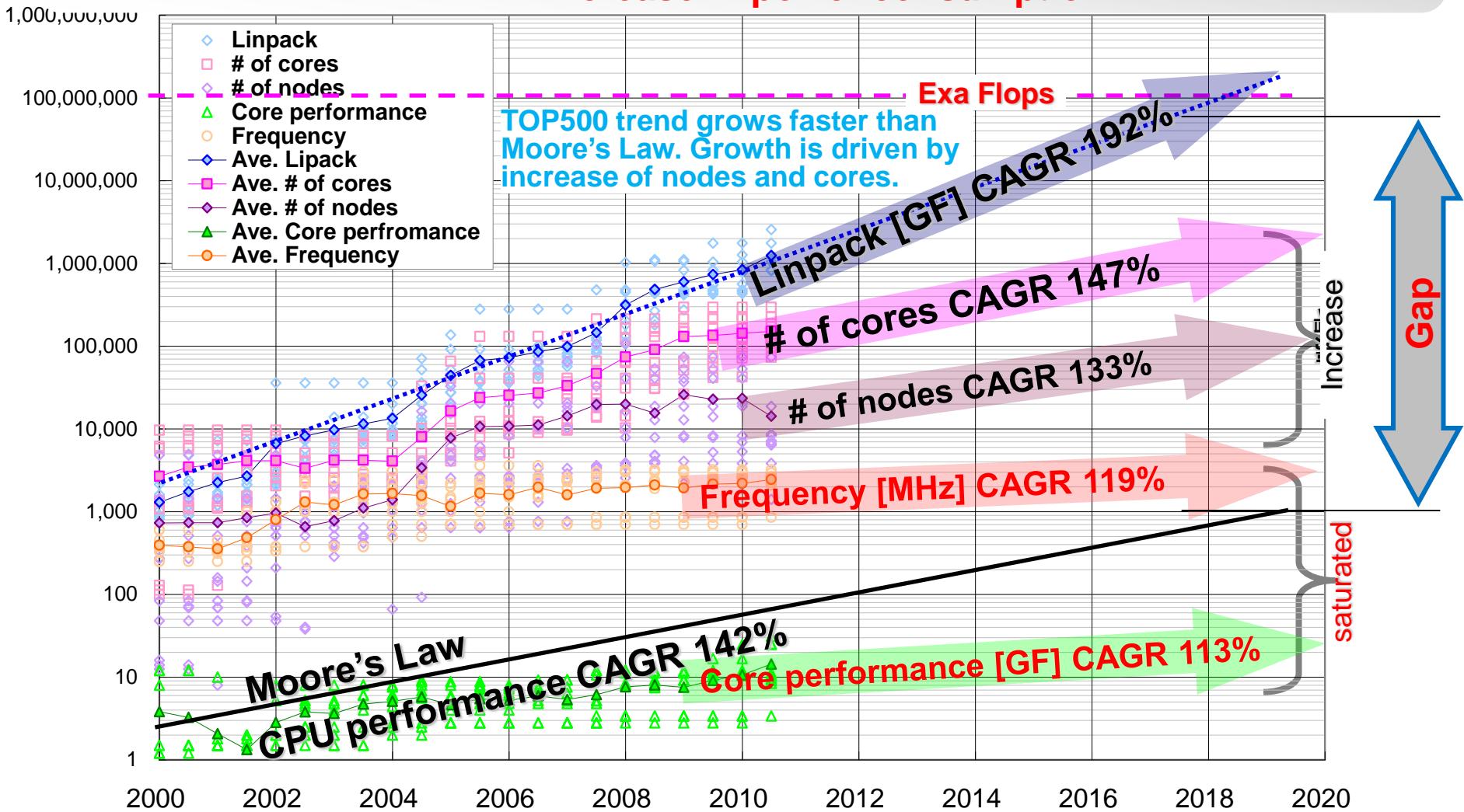


- The Gap between Peak (Linpack) and Memory Bandwidth become larger
 - Unbalanced Processing Performance, Memory Bandwidth & Capacity, and NW Bandwidth
- ⇒ Reducing Sustained Performance Efficiency ratio of Application



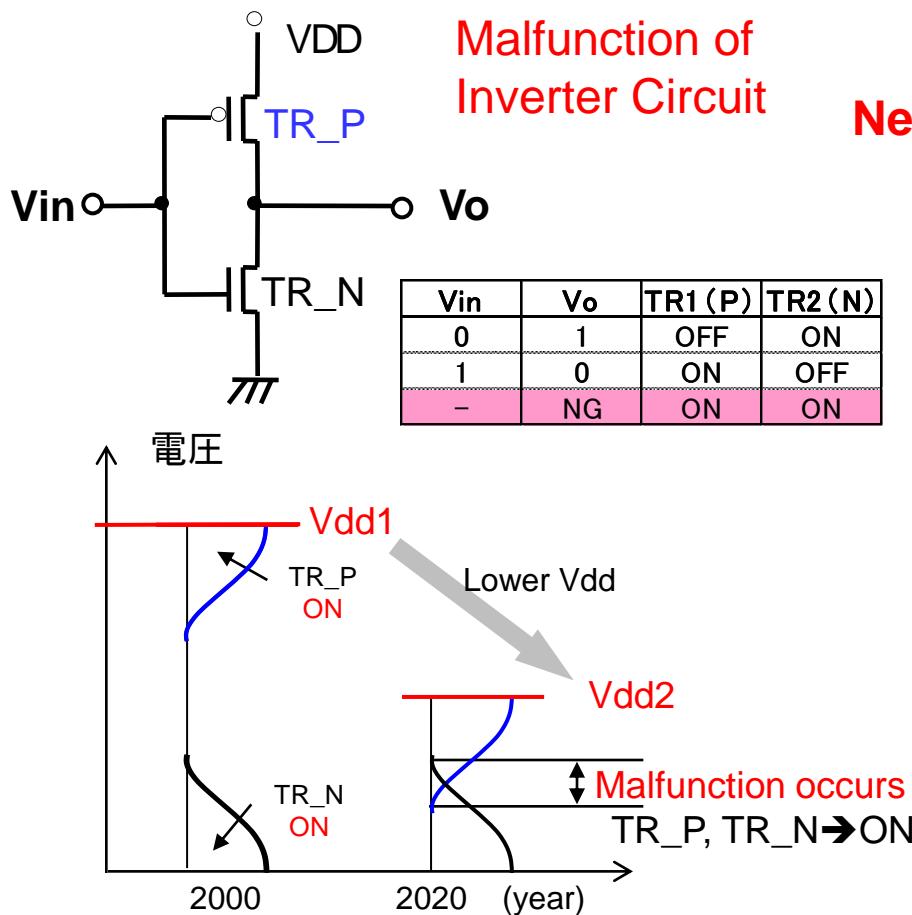
Frequency and core performance is nearly constant.

TOP500 performance now depends on system scale which requires more cores and more nodes. ⇒ increase in power consumption



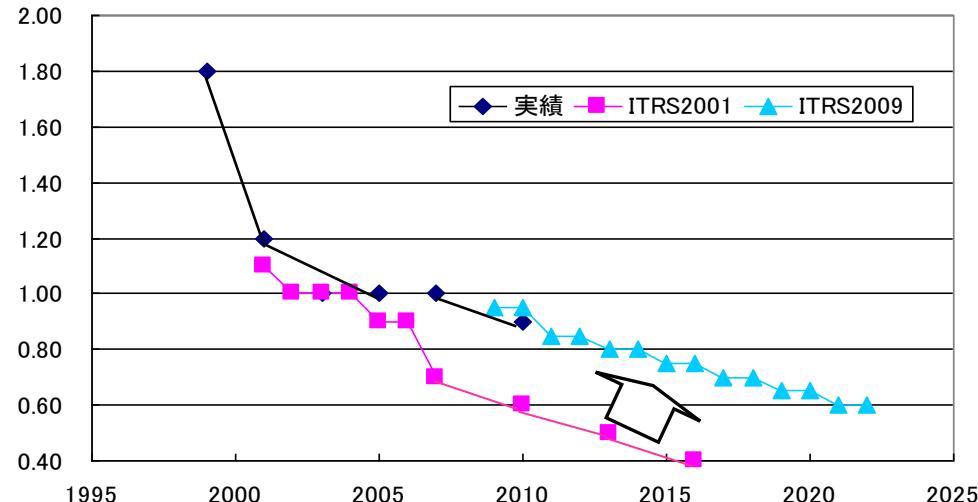
Lower Operating Voltage issue

Page 7



$$P = f c V^2 = f \times 2c \times 0.7^2$$

Next Gen. process : density x2, Voltagex0.7
→ Same Power /die of Previous Process



ITRS-Trend has been changed

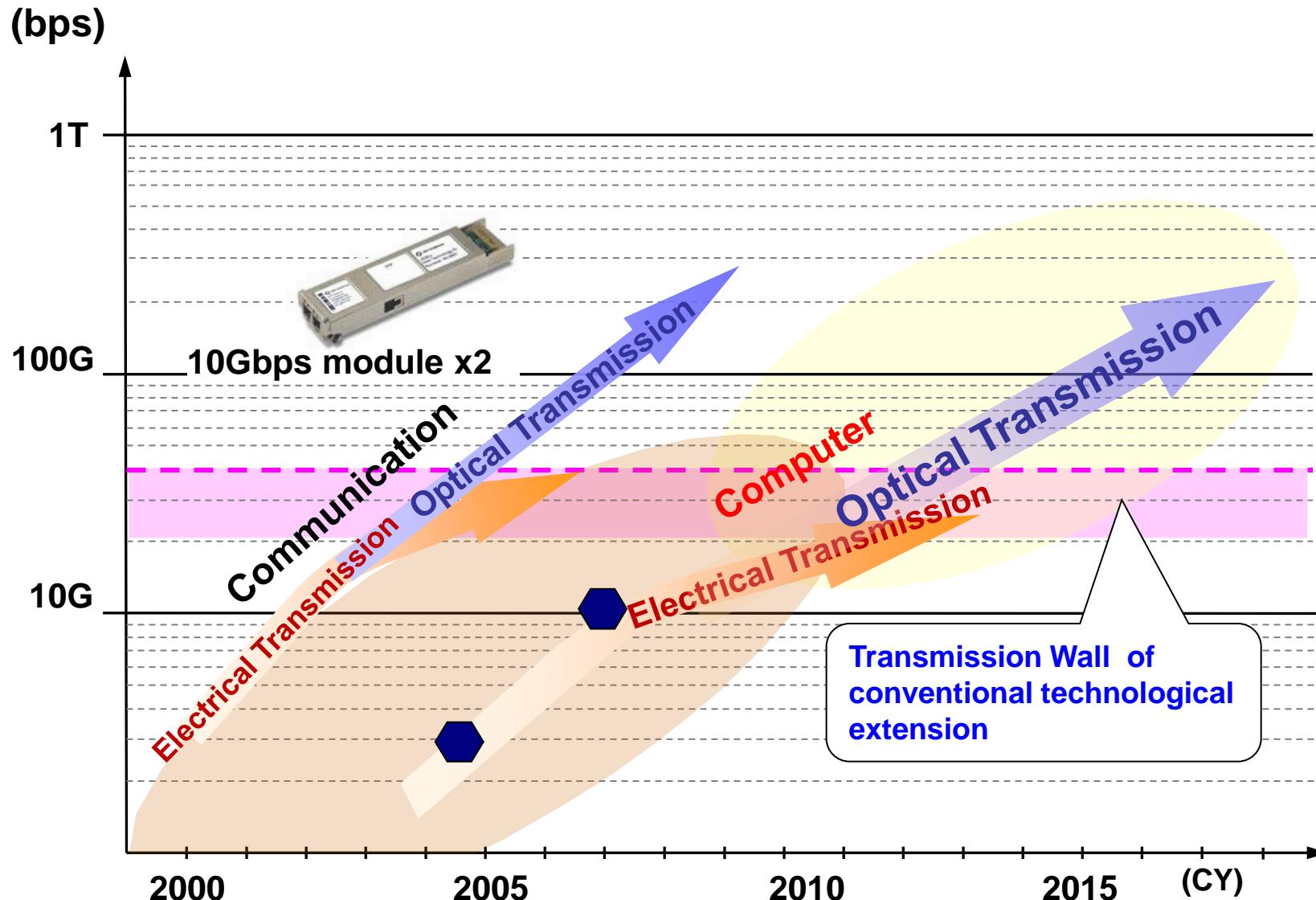
LSI Operated Voltage will be down to
“only” 0.6~0.8V @2020

LSI, System Power will increase

- ✓ Due to variation in 2B Transistors on LSI
- ✓ The further Low Operating Voltage(Vdd2)
cause Malfunctions

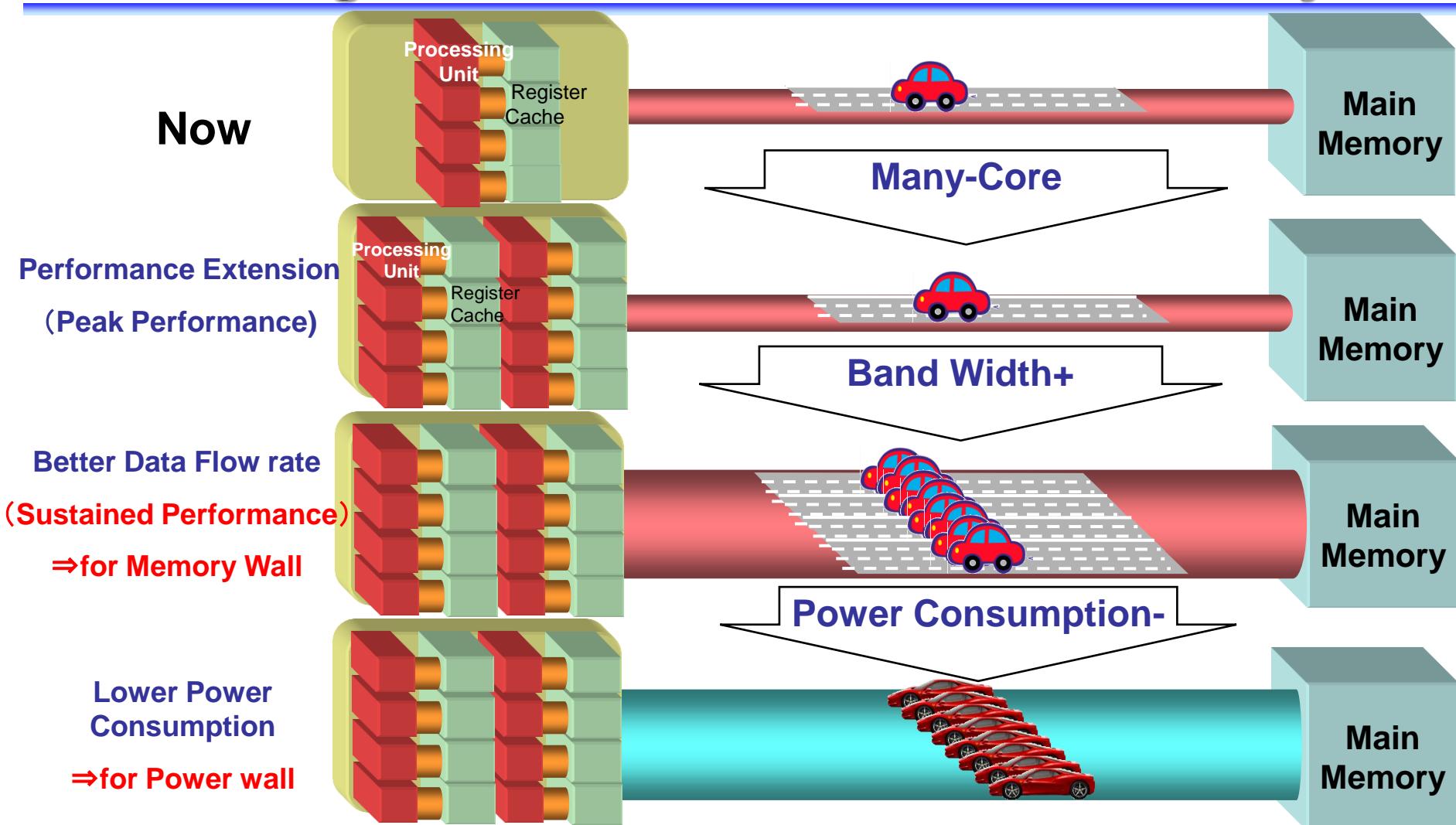
Electrical Transmission Wall

Page 8



Maximizing Performance and Power Efficiency

Page 9



✓ Optical Interface inside with Si photonics
⇒1st:Between Nodes(CPUs)----->last:CPU-Memory



Optical Interface inside the sever

✓ Bandwidth

⇒ Product Performance Extension

- Server/File System/Network
- Game/ Entertainment Visual/Sensor Medical



✓ Electro Magnetic Immunity

⇒ Cost Reduction for Countermeasure

- No EMI effect
- Less Pin Count due to No return path
- Reduce PWB Layers
- Robot, Industrial Machine



✓ Light weight

⇒ Reducing Operational Cost

- Aviation, Automobile, Railroad

Data Rate: 40Gbps

⇒ 4K × 2K Uncompressed Video

※From 『Nikkei Elec.』 May 14th 2014

✓ Design

⇒ Easier Design for High speed Transmission Line

No Reflection, No Attenuation and No X-talk on the transmission line

Shorter Development Time / Common PF

Key Factor : Cost, Reliability

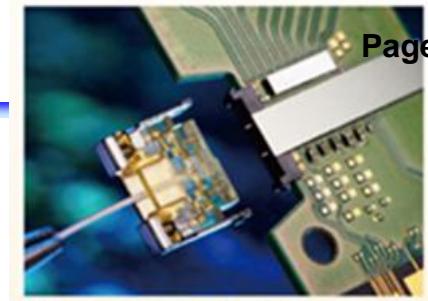
Si Photonics (Intel)

Page 12

✓ Developing Rack Scale Architecture with Facebook

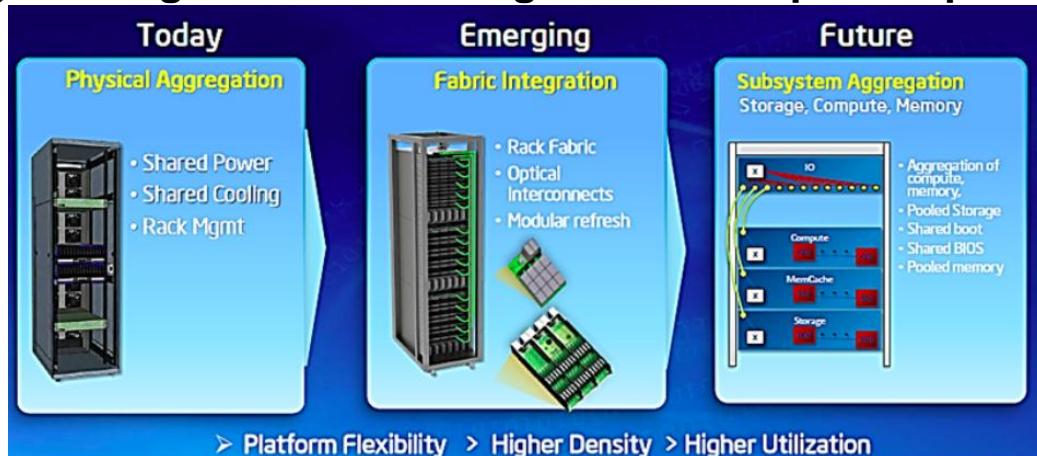
Now : Processor, Memory Storage in Server, Rack and DC

→ Connected with “Expensive Infiniband”



Near Future : Mesh Optical Network in the Rack w/ “Reasonable” Silicon Photonics

Future : Disaggregate Rack : Dividing an integrated whole thing into its component parts



✓ Xeon Phi(Knights Landing) with Omni Scale Fabric

⇒ For I/O bottle neck

⇒ Integration to Xeon

Unveiling Details of Knights Landing (Next Generation Intel® Xeon Phi™ Products)

Platform Memory: DDR4 Bandwidth and Capacity Comparable to Intel® Xeon® Processors

Compute: Energy-efficient IA cores²
▪ Microarchitecture enhanced for HPC³
▪ 3X Single Thread Performance vs Knights Corner⁴
▪ Intel Xeon Processor Binary Compatible⁵

On-Package Memory:

- up to 16GB at launch
- 5X Bandwidth vs DDR4⁷
- 1/3X the Space⁶
- 5X Power Efficiency⁶

Jointly Developed with Micron Technology

2nd half '15
1st commercial systems
3+ TFLOPS¹
In One Package
Parallel Performance & Density

Announcing Intel® Omni Scale—The Next-Generation Fabric

INTEGRATION

- Designed for Next Generation HPC
- Host and Fabric Optimized for HPC
- Supports Entry to Extreme Scale
- End-to-End Solution

Coming in '15

- ✓ PCIe Adapters
- ✓ Edge Switch
- ✓ Director Systems
- ✓ Intel Silicon Photonics
- ✓ Open Software Tools*

Intel® True Scale Fabric Upgrade Program Helps Your Transition

*OpenFabrics Alliance
Co-brands and names are the property of their respective owners

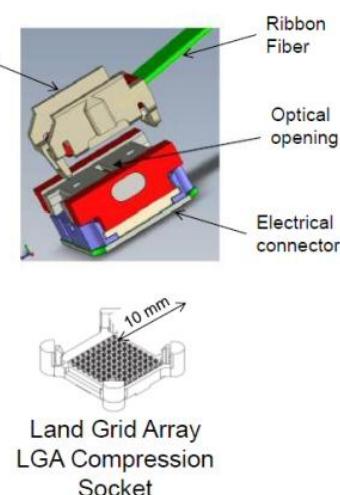
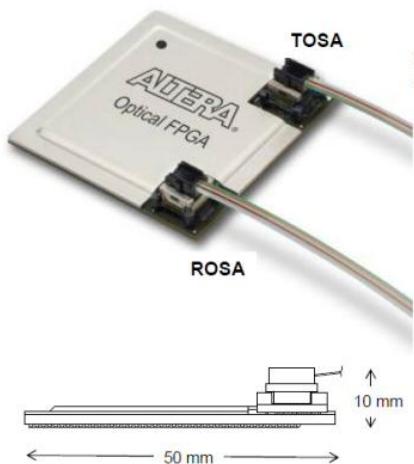
Intel Facebook, Open Compute

Knights Landing
Future 14nm generation

✓ Integrate FPGA and Optical Transceivers / Receivers into MCM@1st Quarter 2012

- 40nm FPGA “Stratics IV” + Avago “Mico-Pod” → Optical FPGA

Optical FPGA R&D Demo



Driving 100GE with Optical FPGAs

■ Set up

- Optical FPGA Board
- 100m MMF in loopback

■ FPGA Design

- 100GE MAC
- Random packets
 - Generator
 - Checker
- 10 x 10.3125 Gbps
 - 2 channels are unused

■ Status

- Ran 72-hours error free



※From Altera

http://www.altera.com/corporate/about_us/history/optical/abt-optical-interconnects.html



Photonics Electronics Convergence Technology for Power-Reducing Jisso System” project

① Research on Photonics-Electronics Convergent System Technology

(FIRST Program by the Council for Science and Technology Policy in the Cabinet Office)

② Photonics Electronics Convergence Technology for Power-Reducing Jisso System project (“Future Pioneering Projects” entrusted by the Minister of Economy, Trade and Industry (METI))

⇒ Photonics Electronics Technology Research Association (PETRA)

Project Leader : Yasuhiko Arakawa, Professor of the University of Tokyo

Member companies: AIST, Fujitsu, Furukawa Electric, NEC, NEL, NTT, OITDA, OKI, Toshiba

Joint research organization: Univ. of Tokyo, Kyoto Univ., Tokyo Institute of Technology,
Yokohama National Univ., WASEDA Univ.

FY	total	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Billon Yen	29.1	6.0	3.0	3.0	2.8	2.8	2.5	2.5	2.5	2.0	2.0

Target

Photonics Electronics

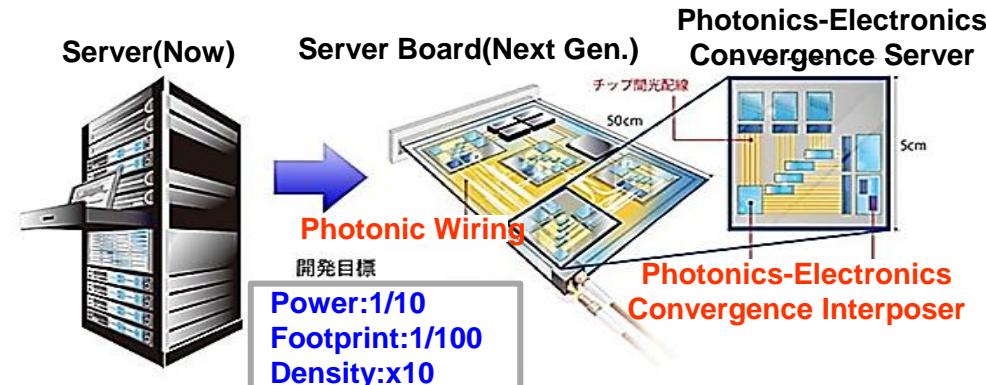
Convergence Device Technology

→ Si Photonics Engine (Optical IO Core)

Systemizing Technology

International Standardization

→ Power-Reducing New Optical
Interface System



- Develop a systemizing technology to apply IT systems by combining the Photonics Electronics Convergence Jisso Key Technology
⇒ Developing the Turn Key Solution of Si photonics for Practical use

Photonics-Electronics System Development

Page 16

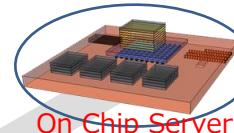
Photonics-Electronics
Convergence "JISSO" System
@ Member Company

Server

Board, Rack

- : Photonics-Electronics Convergence Interposer
⇒ Photonics Electronics Convergence Server
- : SSD w/ Optical SATA Interface
- : AOC ⇒ LSI w/ Optical Interface

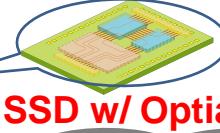
Photonics Electronics
Convergence Server



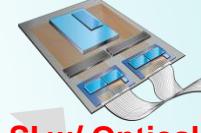
On Chip Server



Server



SSD w/ Optial I/O



LSI w/ Optical I/O



Inter-Board



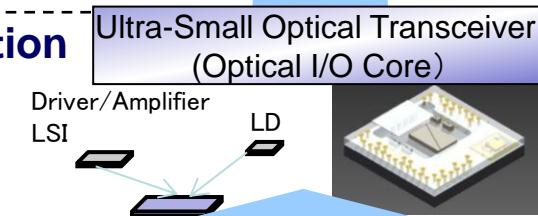
Inter-Rack

Broadband (FTTH) delivering Service to anywhere

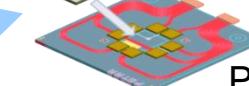
Photonics-Electronics Integration

@PETRA Tsukuba

Core Technology



Photonics Electronics Hybrid Package



Photonics-Electronics Design

Photonic Integration @PETRA Tsukuba

PECST

Parallel / Multiplex
Photonic Integration Device

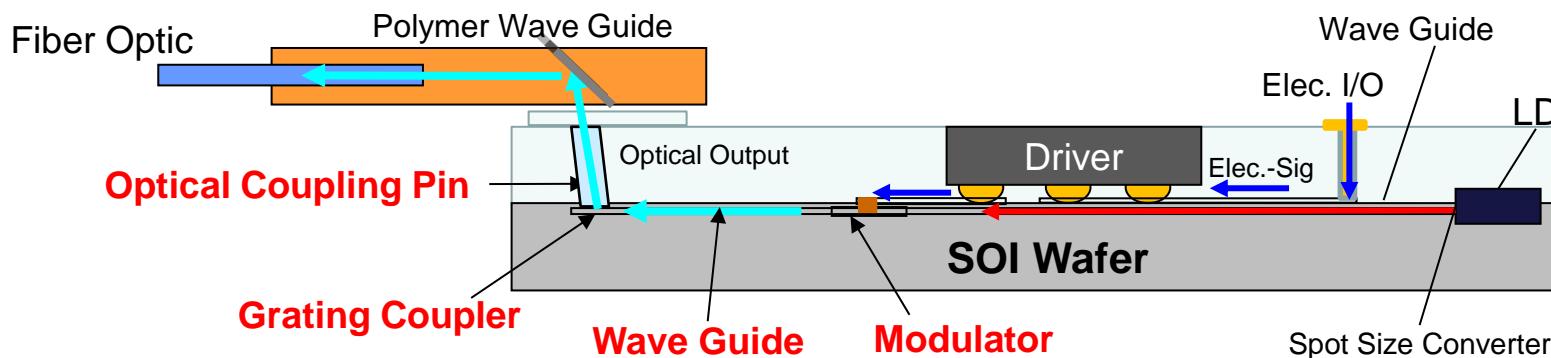
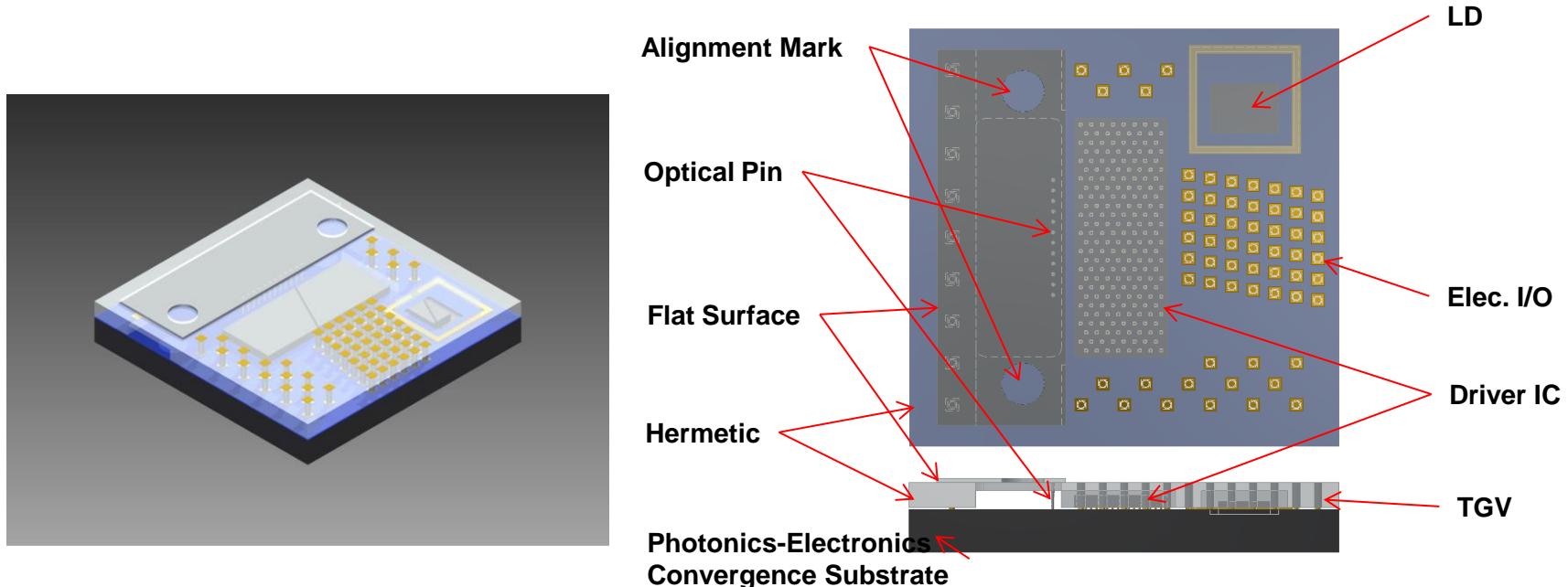
Modulators / Photodetectors Optical waveguide
Process Integration
Low Power • High Density Device

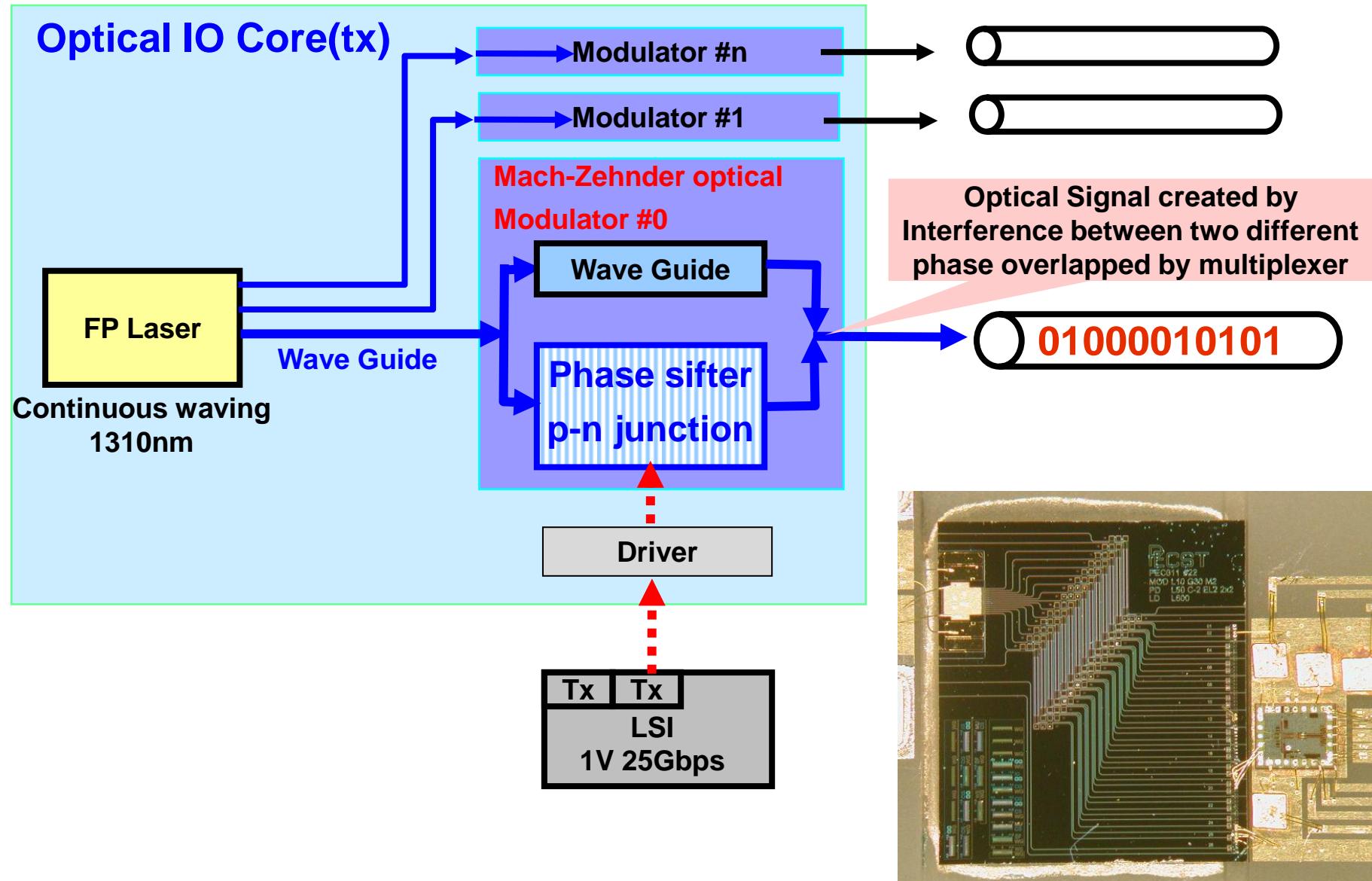
Advanced Device

Advanced Device @Univ.

Optical I/O Core External View / Structure(Tx)

Page 17





✓ **25Gbps Optic / Elec. Conversion Device w/ Si photonics Tech.**

➤ **1310nm Laser CW + External Modulation**

⇒ High Reliability

➤ **Low Power**

↔ 1/3 of VCSEL based Optical Module
(15mW/bps ↔ 5mW/bps @ 25Gbps)

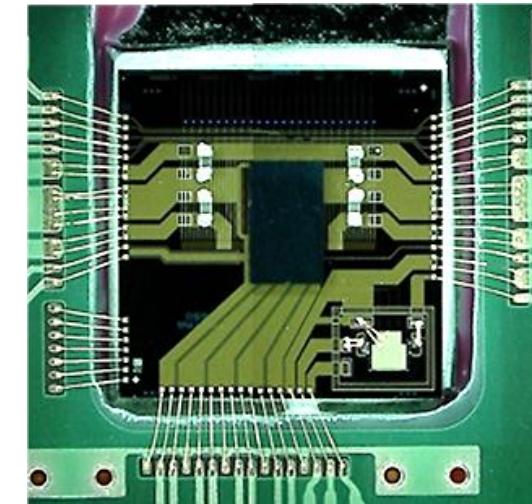
➤ **Small Form Factor (25Gbpsx12ch @ 5mm sq.)**

⇒ More I/O Channel beside LSI

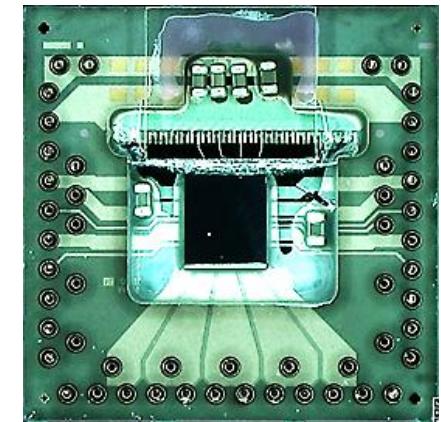
➤ **Mounted as Typical SMT Device**

➤ **CMOS Process based device**

⇒ Good Price/Performance (\$/Gbps)



Wire Bond ver.-Tx

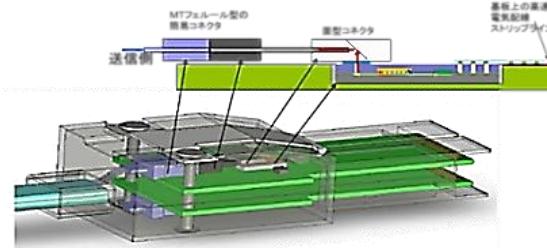


Flipped Chip ver.-Rx

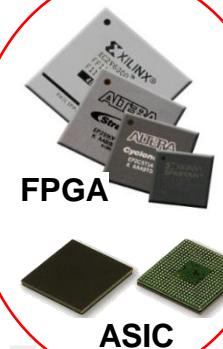
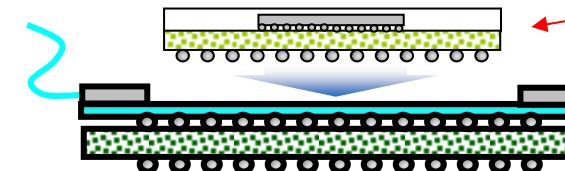
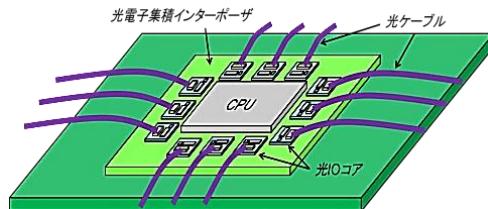
Project Target Milestone

Page 20

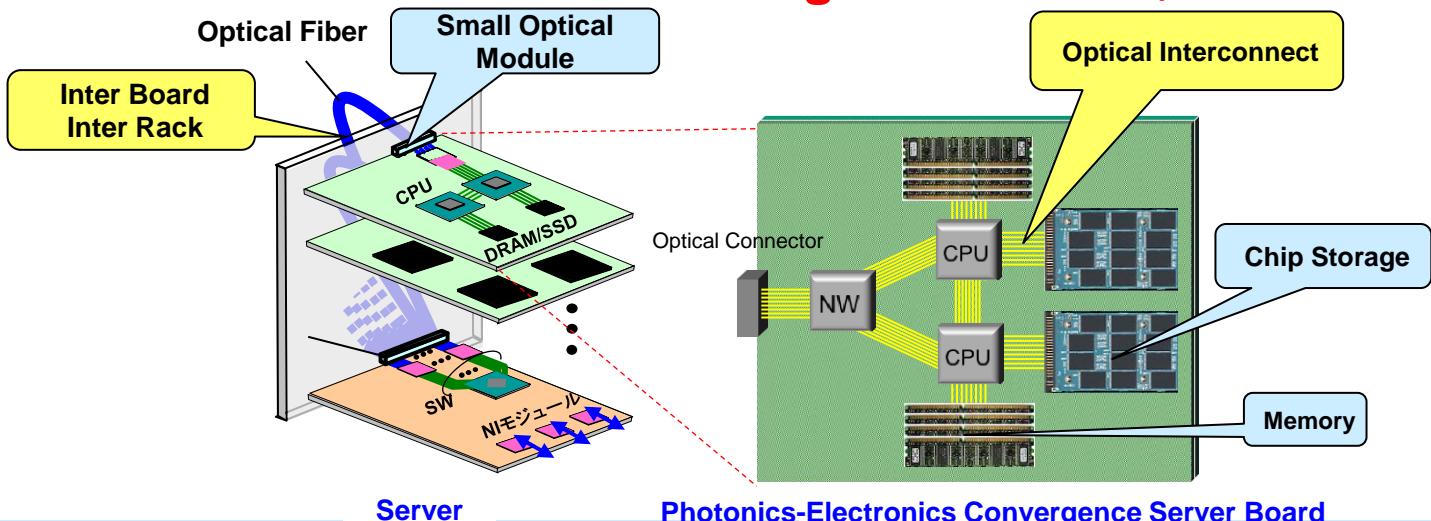
- ✓ ~Mar. 2015 Optical Engine(Optical I/O Core), Active Optical Cable



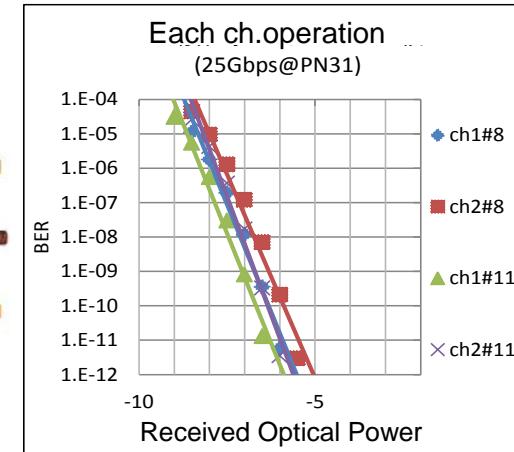
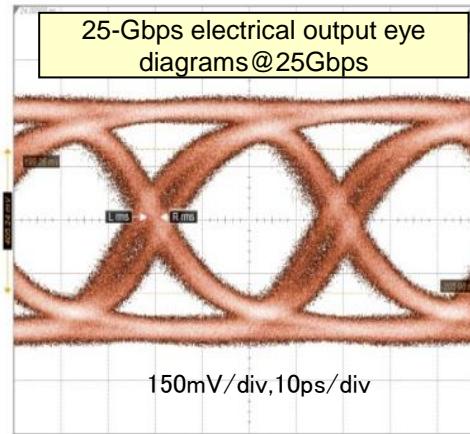
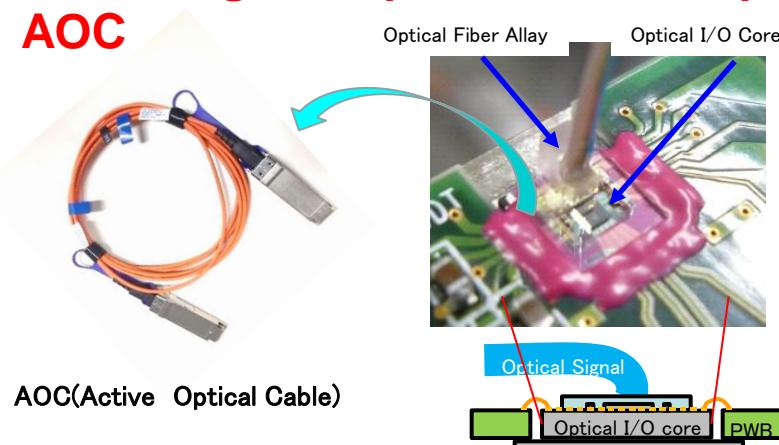
- ✓ ~Mar. 2017 Photonics-Electronics Convergence Interposer



- ✓ ~Mar. 2021 Photonics-Electronics Convergence Server (←Fundamental Tech.)



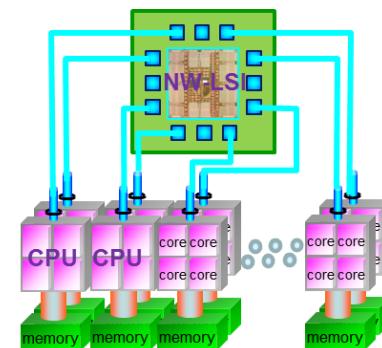
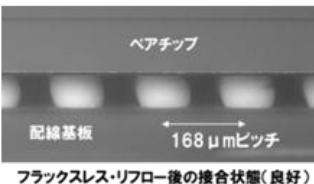
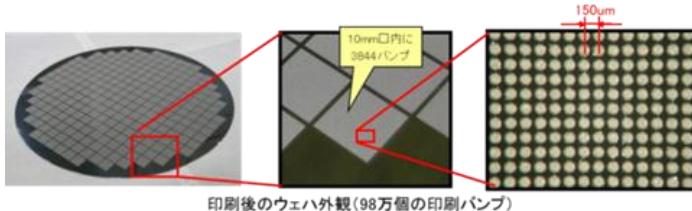
✓ Achieving 25Gbps Error-Free Operation (BER<10⁻¹²) of Optical I/O Core(Rx) for AOC



- Power Dissipation: ~2.8mW/Gbps/ch@Rx → Target: 5mW/Gbps/ch@Tx+Rx
- Developing 25Gbpsx4ch AOC (QSFP) by Mar. 2015

✓ Developing and Designing Fundamental Technology for 10Tbps Photonics-Electronics Convergence Interposer

- High Speed Transmission : Electrical Simulation Model@25Gbps
- High Density Chip Connection : 125um Flux-Free Soldering

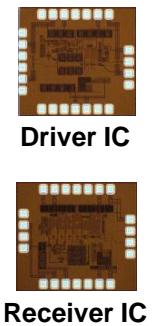


- Optical Coupling Connection : Optical Signal from Optical I/O Core to Fiber
- Cooling : Thermal Simulation and Analysis of Optical I/O Core

SSD Interface with Optical I/O Core

Demonstration of SSD w/ Optical Standard Storage Interface (SATA)

Developing Drv../ Rec. Test chip for Optical interface
SATA Signal transferred between PC and SSD

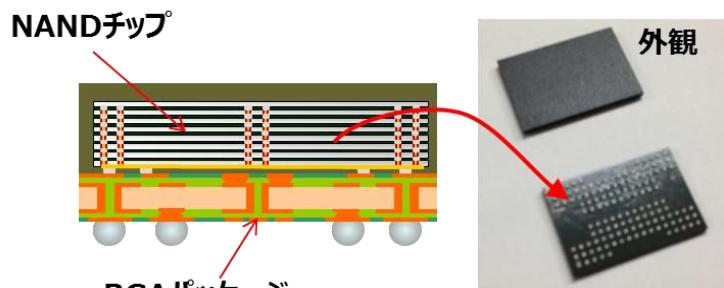


PC-CPU (ホストコントローラ)
SATAケーブル
光I/O試作品 (トランシーバIC使用)
市販SSD (SATA3)
SATA接続 リボンファイバ

Stacked NAND for SSD

3D-stacked NAND Chip (TSV)

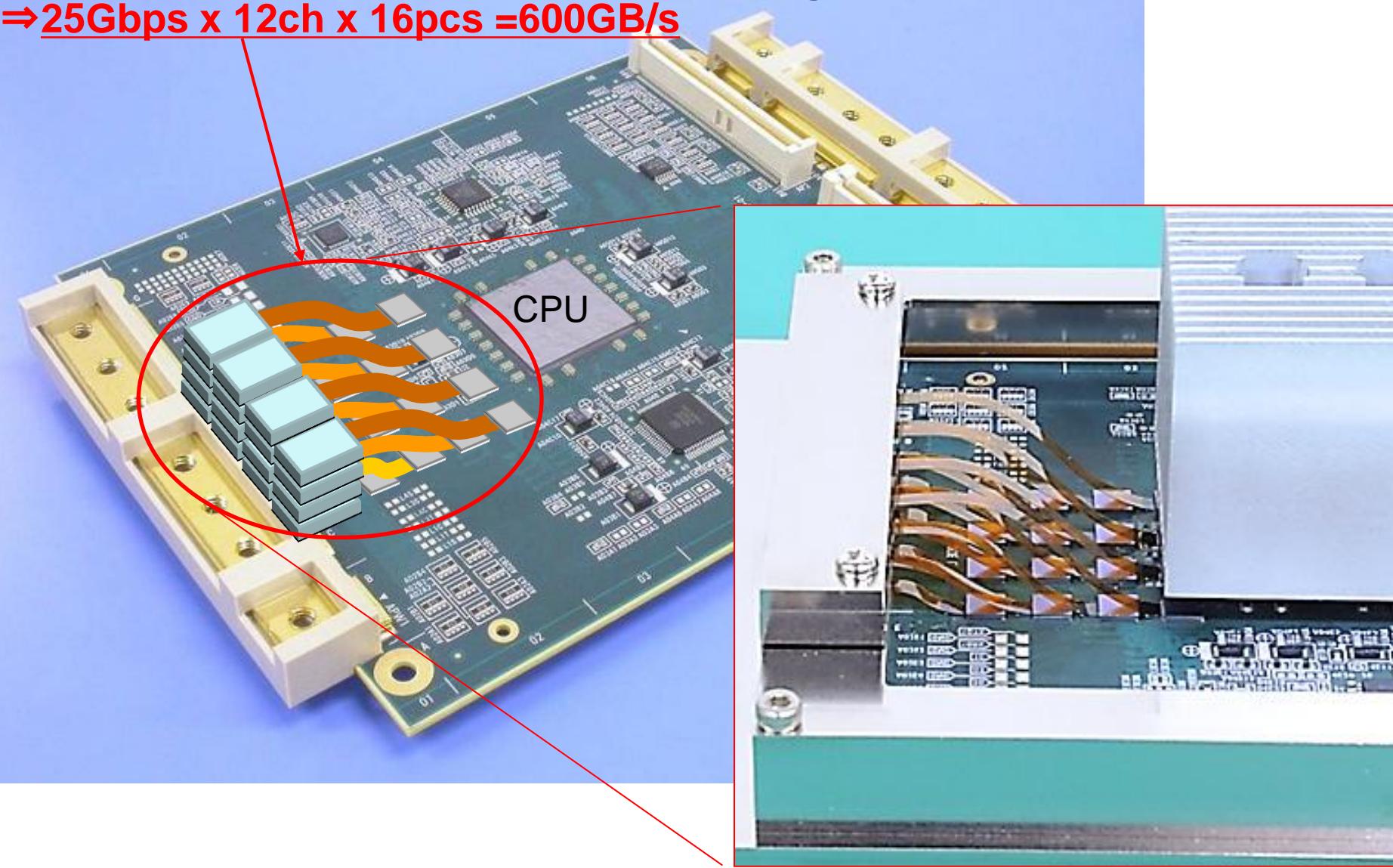
Developing 3D-NAND test chip packaged in BGA
Operating memory block



- SATA Gen3 Out-of-Band signal/DATA transfer succeeded
- Operating read/write/delete sequence on 3D-stacked NAND for SSD using TSV technology

3D-stacked NAND with Optical SATA interface for SSD

Optical I/O Core beside CPU : Electrical Sig. convert to Optical
→25Gbps x 12ch x 16pcs =600GB/s



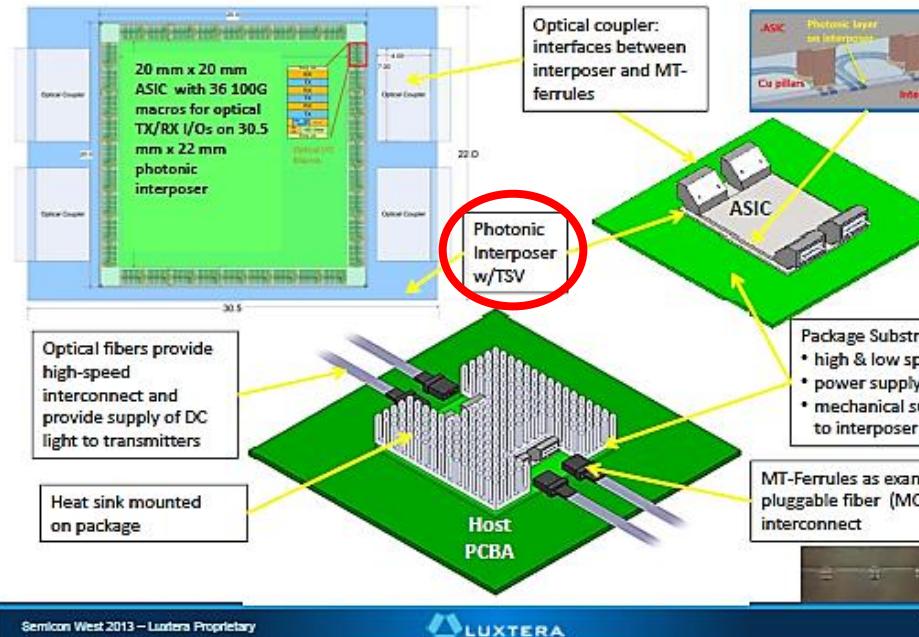


Photonics with Electronics

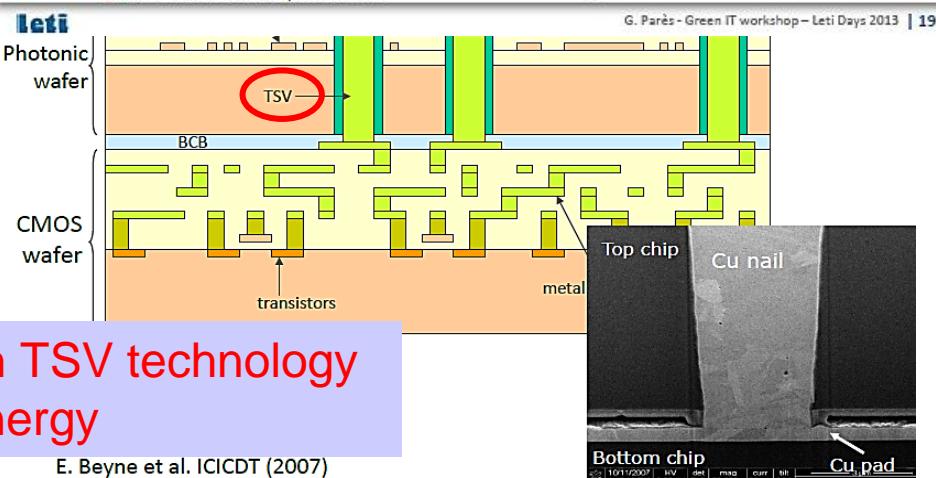
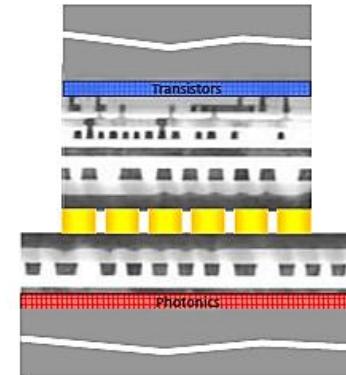
Photonics + Through Silicon Via

Page 25

Example: Interposer/ASIC Floor Plan & Packaging



How to integrate photonics with electronics ?



- ✓ Photonics Electronics Convergence with TSV technology
⇒ High Sustained Performance, Low Energy

E. Beyne et al. ICICDT (2007)

*from LUXTERA, Leti, IMEC Report



PHOTONICS RESEARCH GROUP

imec 32

➤ Photonics and 3D-LSI(TSV) Technology are Killer Application

- Huge Potential to achieve a major breakthrough on Performance and Function
- IBM, Intel : for High End Product , Apple : for Consumer Product



RAID Storage
Video Capture, DSP
Mobile Storage



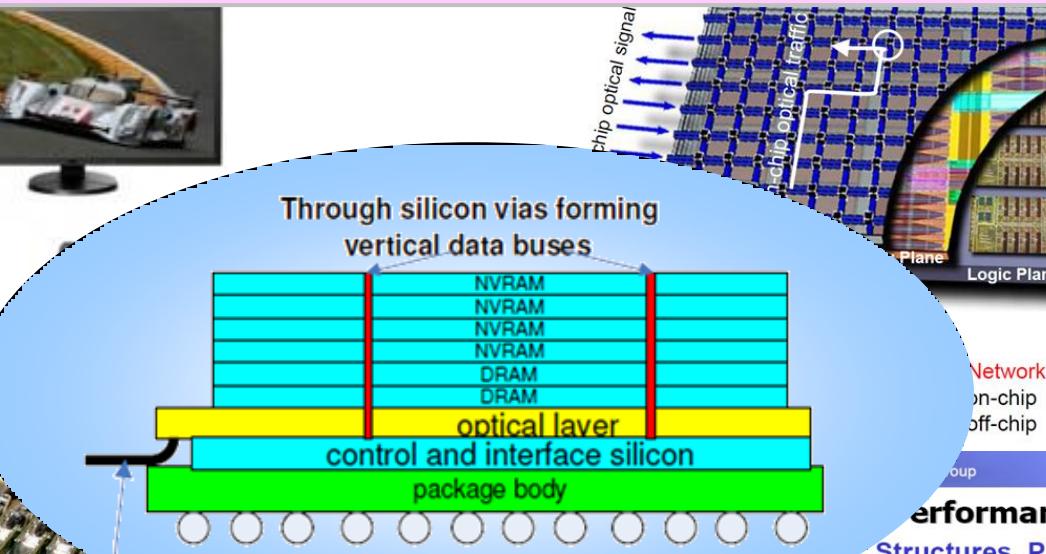
Optical Interface on Board Level



Fiber connections



From DARPA Report



36 "Cell" chip (~300 cores)

System level study:
IBM, Columbia, Cornell, UCSB

Co-PIs:
Jeff Kash (IBM)
Keren Bergman (Columbia)
Yuri Vlasov (IBM)

Photonic layer is not only connecting various cores, but also routes the traffic

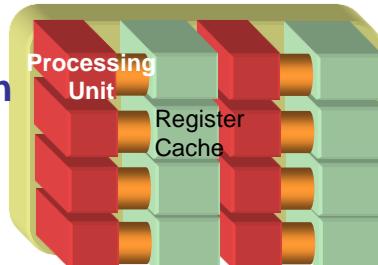
Performance Through Innovation

3D-LSI is contained
IBM Technology Innovation



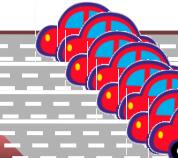
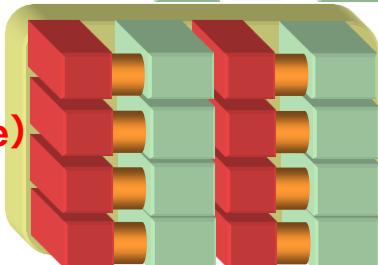
IBM has developed TSV technology for DRAM on the POWER chip within two years@2012
V.P of Semiconductor Research Center : Gary Patton

Performance Extension
(Peak Performance)



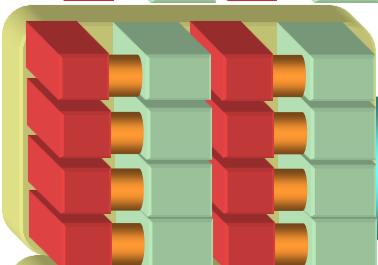
Band Width+

Better Data Flow rate
(Sustained Performance)
⇒ for Memory Wall



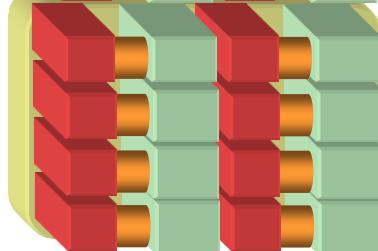
Power Consumption-

Low Power Consumption
⇒ for Power wall



Main
Memory

Higher Sustained Performance
Power Efficiency
⇒ for Memory Wall & Power wall



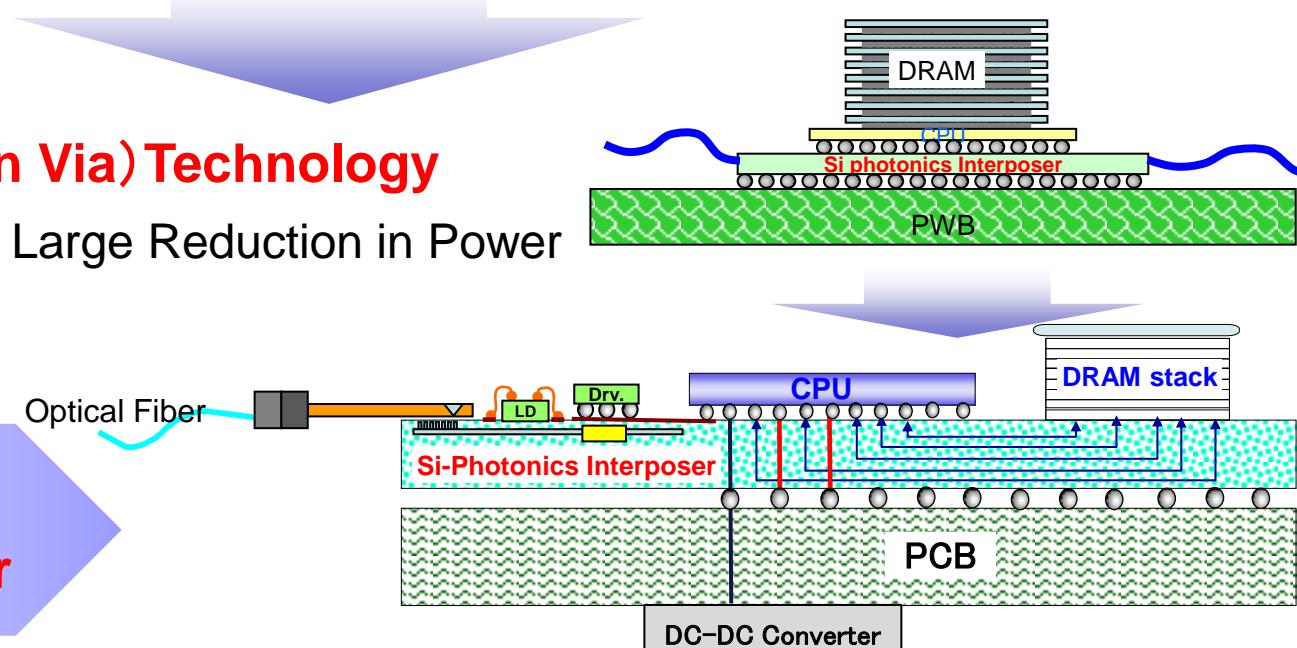
3D-LSI with TSVs

✓ Photonics + Through Silicon Vias
→ LSI Paradigm Shift

✓ High Speed & High Reliability Optical I/F with Silicon Photonics

⇒ Low Power, Small Photonic Modulator & Detector

- High Reliability & High Efficiency (Low Power) Light Source
- Low Loss Optical Connection (Wave Guide, Optical Fiber Connection)
- Small Low Power Photonic Modulator & Detector



✓ TSV (Through Silicon Via) Technology

⇒ Bandwidth, Latency, Large Reduction in Power

· 3D-LSI ⇒ 2.5D

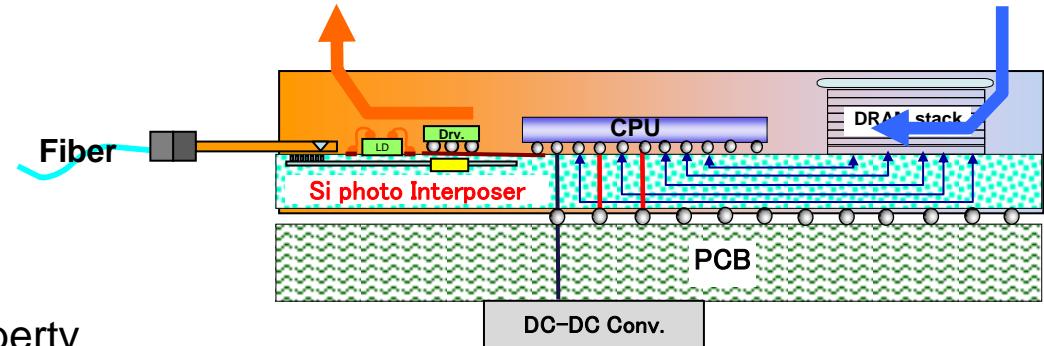
Photonics + TSV
⇒ Ultimate Server

1chip-Data Center : Integrating Photonics +CPU+ DRAM

✓ 1Chip-Data Center Module Assembly

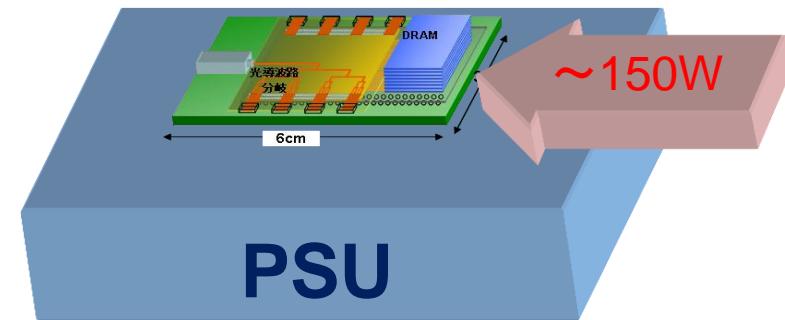
⇒ Advanced Liquid Cooling for CPU-Memory-Optical Fabric

- Reliability
- Thermal Management
- High Density Soldering Connection
- Different Thermal Expansion
- Hermetic Seal
- Cooling Liquid Performance and Property



✓ Energy (Power) Concentration to 2.5D-LSI

- ⇒ Noise Management
- ⇒ Power
- ⇒ Power Supply (1chip-DC~System Level)
- ⇒ PSU size must be fitted to 1Chip-DC



✓ Test-Known Good Die

- E-Fuse
- Test(Individual, Total)

Avoid falling into the Death Valley of Technology
merging to the Conventional Technologies



This research is partly supported by New Energy and Industrial Technology Development Organization (NEDO)