Photonics in computing: use more than a link for getting more than Moore

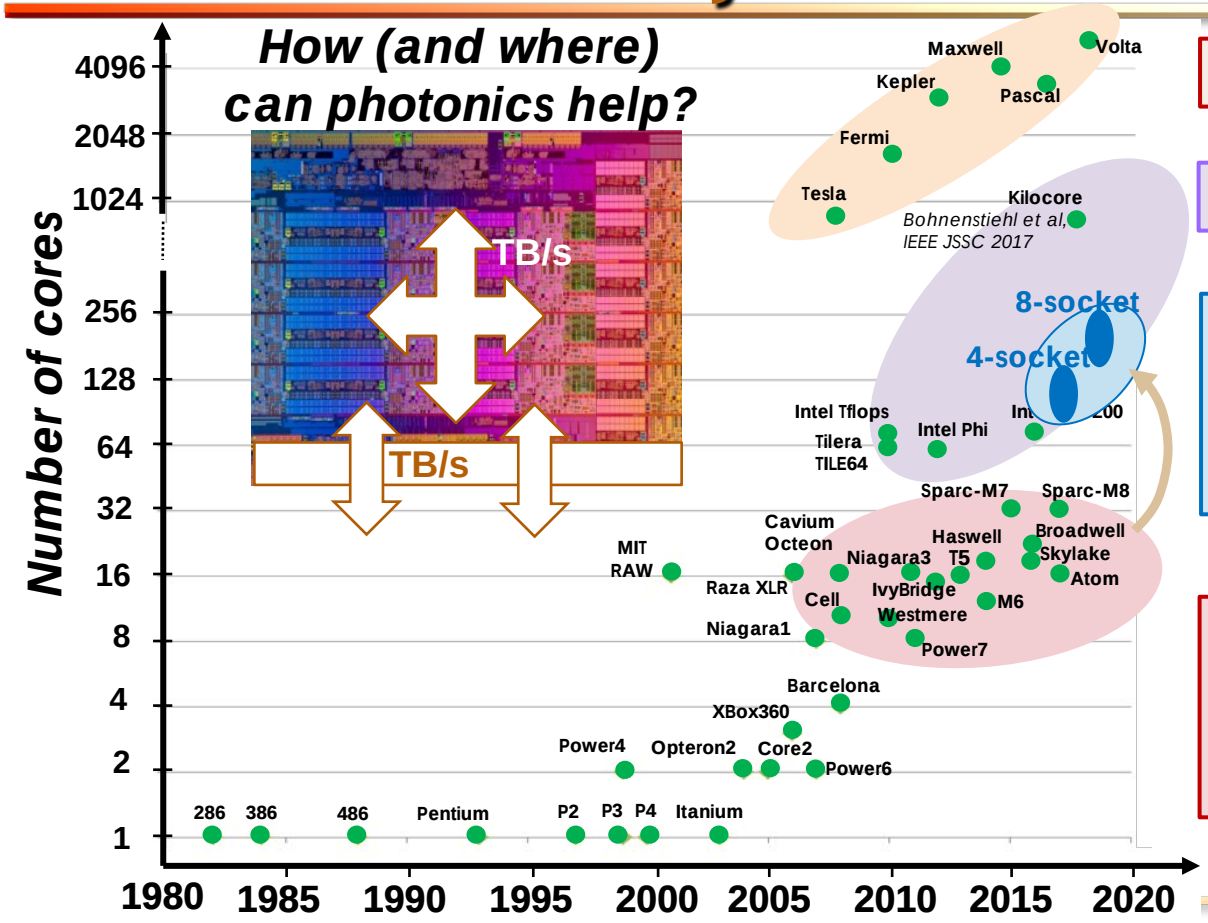
Nikos Pleros

Photonics Systems and Networks (PhosNET) research group
Dept. of Informatics, Aristotle Univ. of Thessaloniki,
Center for Interdisciplinary Research & Innovation (CIRI), Greece

ΚΕΔΕΚ
ΚΕΝΤΡΟ ΔΙΕΠΙΣΤΗΜΟΝΙΚΗΣ ΕΡΕΥΝΑΣ ΚΑΙ ΚΑΙΝΟΤΟΜΙΑΣ ΑΠΘ

PhoS-net Research Group

Multi- and many-core era



NVIDIA Cuda cores, ~2 Gflops/core

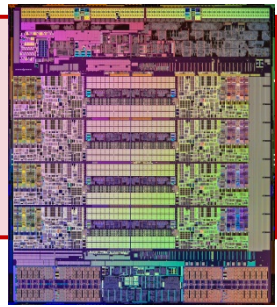
Many-core, up to 3.5 Tflops *(Xeon Phi 7290)*

- ***4S*** *(64-128 cores)*, ***up to 3.6Tflops***
- ***8S*** *(128-256 cores)*, ***up to 7.2Tflops***

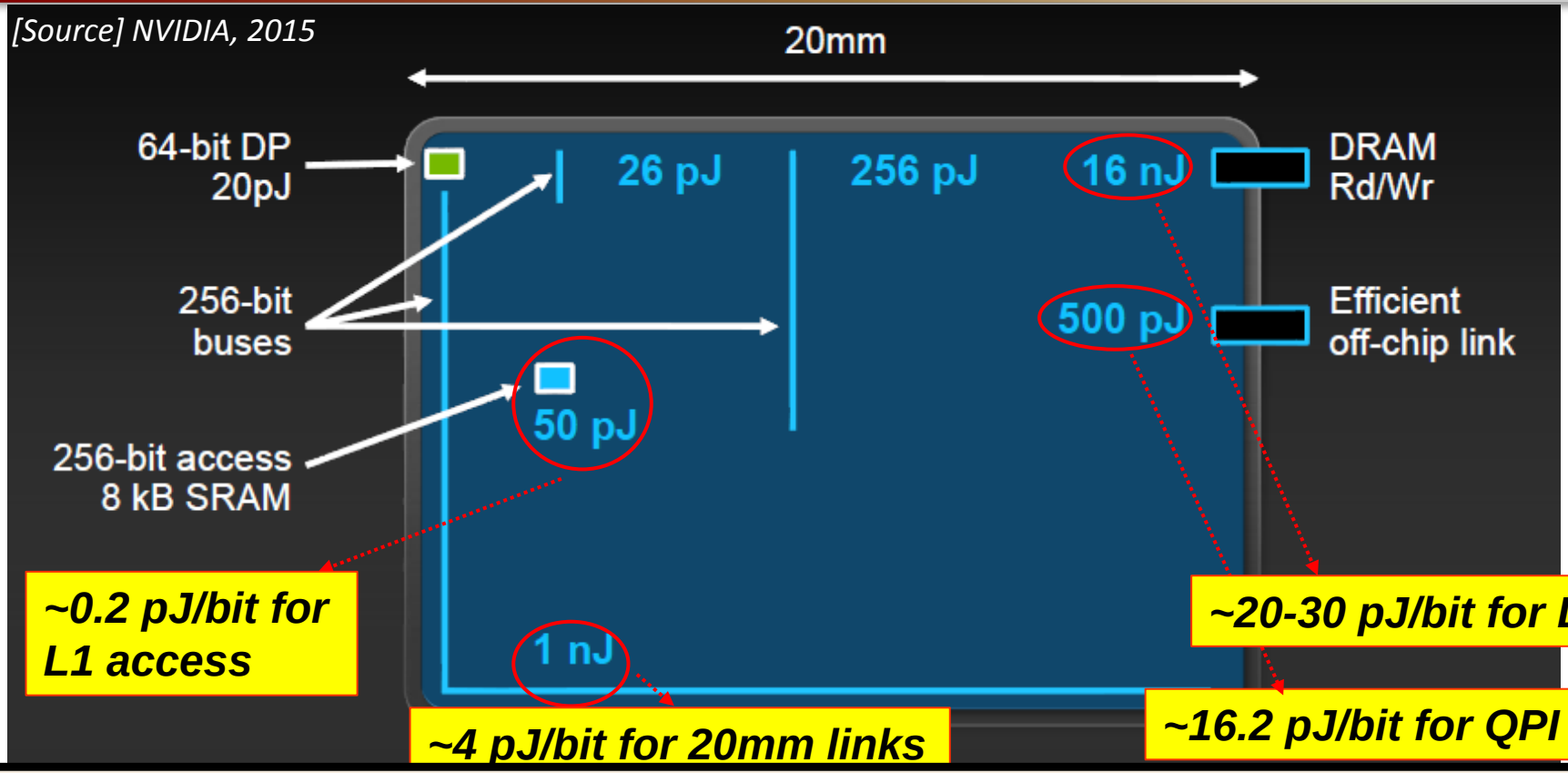
High-perf multicore

up to 0.9Tflops *(E5-2599v4)*

up to 32 cores



Energy

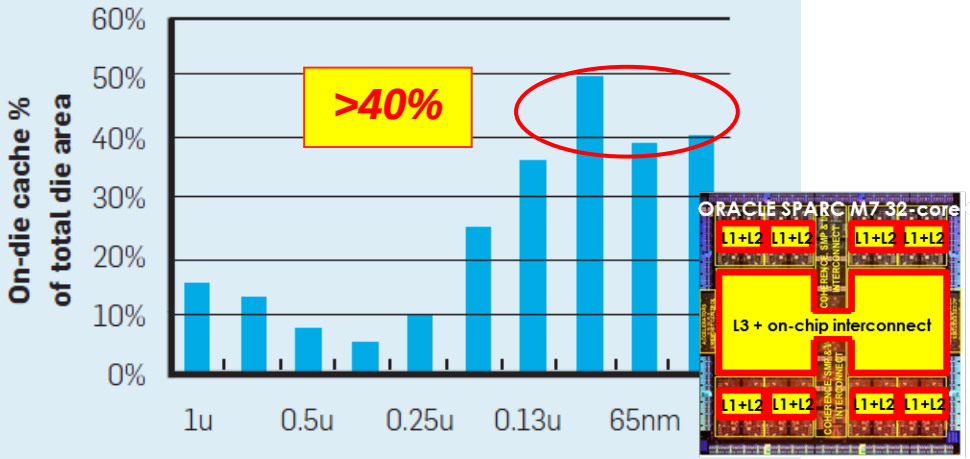


- ***Fetching operands costs more than computing on them!***

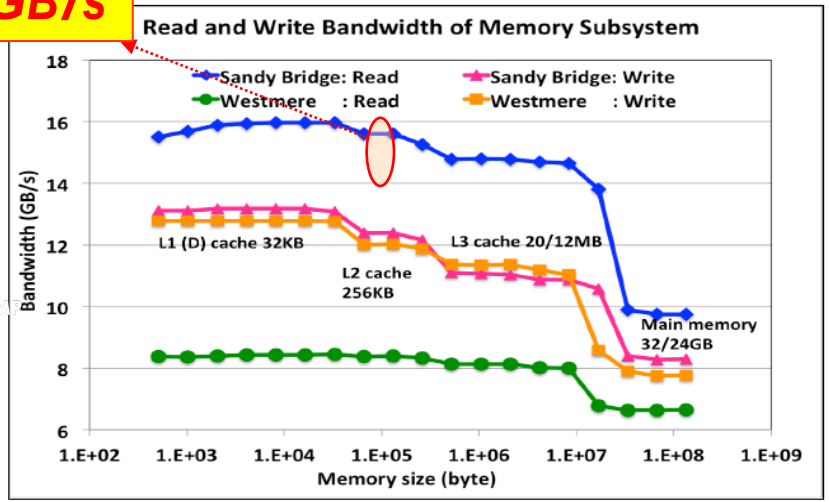
Die Area & Memory BW

Evolution of on-die caches in high-perf CMPs

[Source: S. Borkar and A.A.Chien, Com. ACM, 2011]



<20GB/s

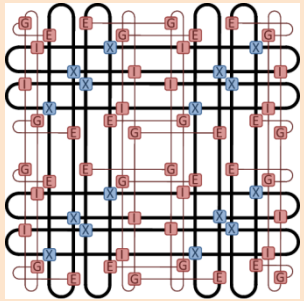


Sandy Bridge vs Westmere

- ✗ ***Up to 40% of die area consumed by NoC & caches – but still <20GB/s L1 bandwidth***
- ✗ ***Small no# of high-perf cores fit in 400mm² area***

The optical NoC escape-way

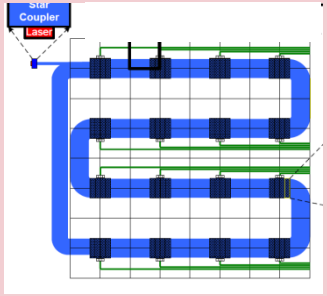
Photonic torus,
(Columbia, A. Shachamet al., 2008)



switch
- 4x4 Si PSE
- 144 in total
- 10mw

TxRx
- 40Gb/s
- 0.2pJ/bit
- 24 λ's

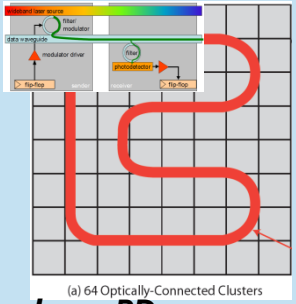
CORONA *(HP, D.Vantrease et al, 2008)*



mod
- Si-ring
- 10Gb/s

PD
- SiGe ring
- 1fF
- no TIA

ATAC 1000-core
(G.Kurian et al, PACT 2010)



mod
- 50um2
- 1Gb/s
- 25fJ/bit

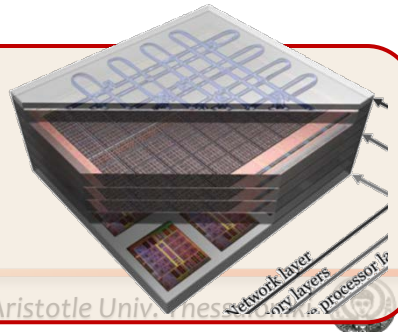
PD
- >1 A/W, 20um2
- 1GHz
- 1fF, no TIA

- ***>1000-core CMPs***
- ***assumed 10-40Gb/s***
- ***20-200 fJ/b OEO***

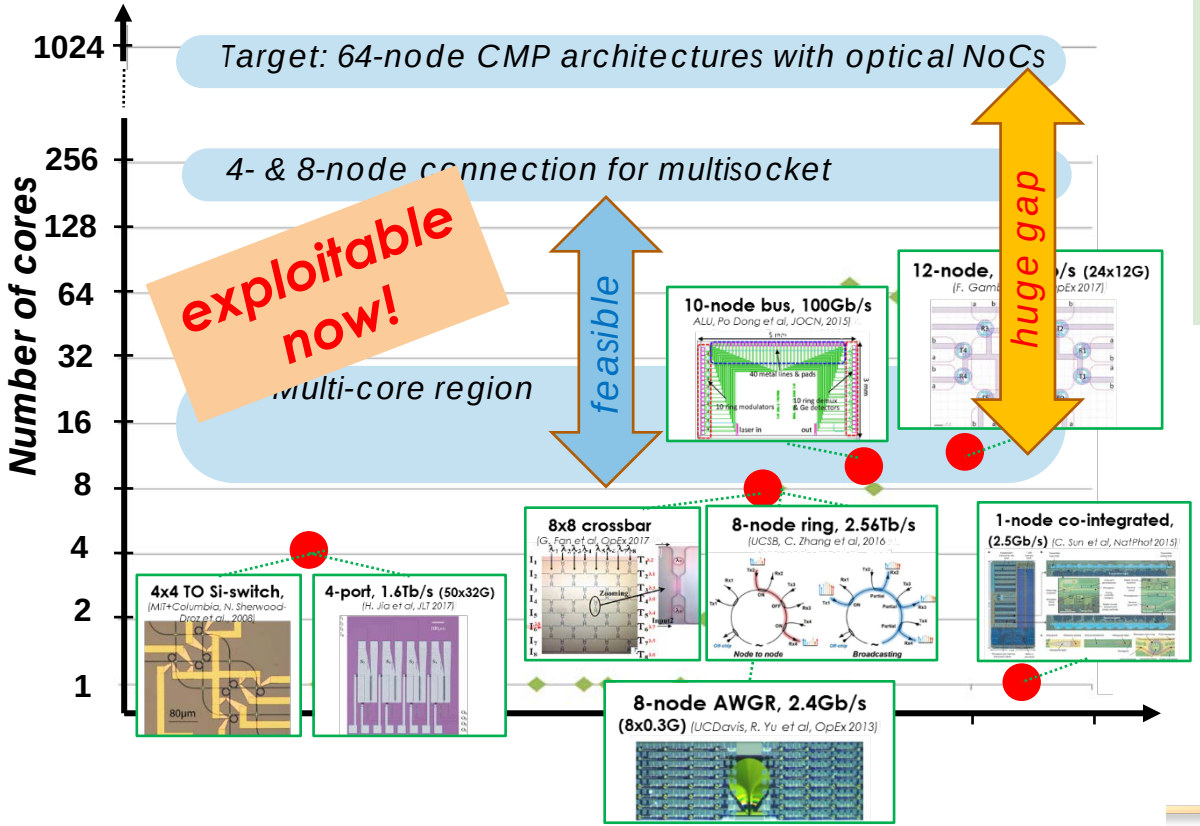
2008 — 2010

Open Problems

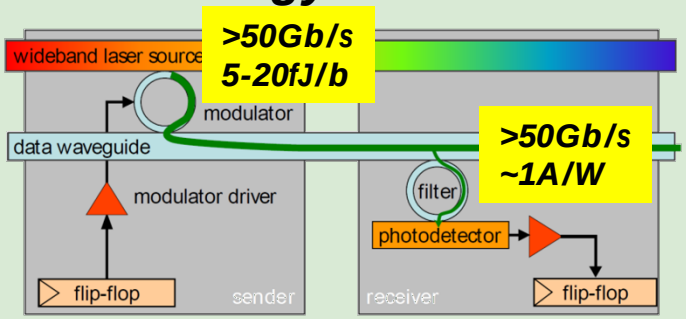
- ***How to co-integrate 1000's of photonic structures!***
- ***3D SiP still far away!***



The optical NoC technologies



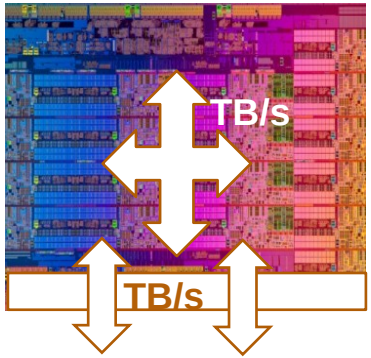
Si-technology is almost here!



BUT, after 10 years

- ***Limited pNoC connectivity***
- ***co-integration only recently realized at 2.5G***

How (and where) can photonics help

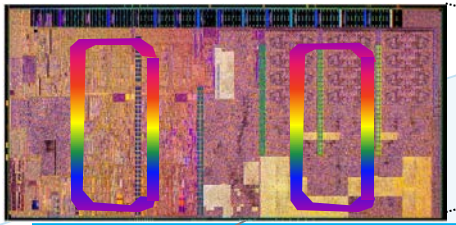


- ✓ NoCs for **high-perf multicores** instead of lower-perf manycores:
  - ✓ focus on easier co-integration technology
  - ✓ Disintegration for overcoming die-area constraints and yield modular and flexible CMPs

- ✓ Adapt the pNoC know-how **in multi-socket interconnects** for scaling into high-perf manycore setups:
  - ✓ at a much lower energy than electronics
  - ✓ and an increased number of sockets!

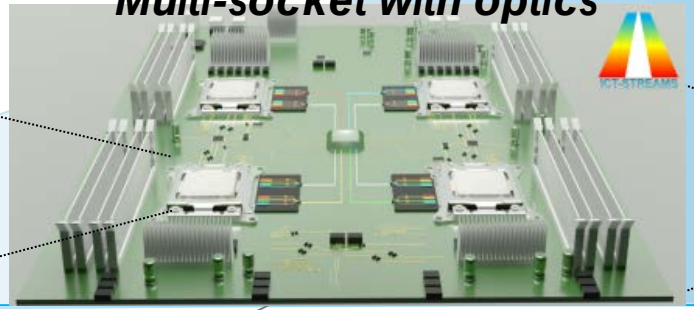
Our work @PhosNET

Chip-scale

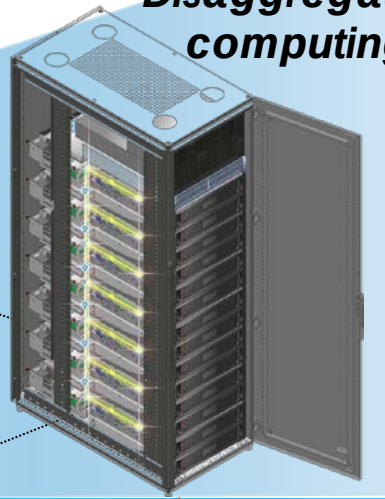


hierarchy

Multi-socket with optics



Disaggregated computing

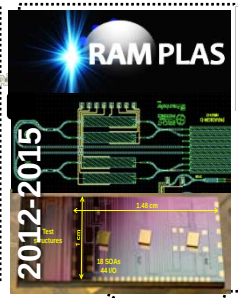


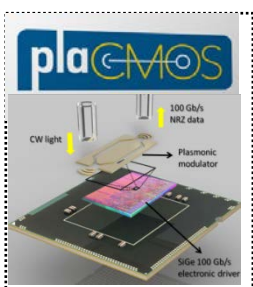
- ✓ ***CMOS plasmonics:*** *easier co-integration of Tx/Rx interfaces*
- ✓ *Deploy* ***ultra-fast optical cache memories***
- ✓ ***Disintegrate:*** *macro-chips!*

- ✓ ***Use Silicon photonics*** *for:*
  - ✓ *scaling to 16 sockets*
  - ✓ *lower by 75% the QPI pJ/b*
  - ✓ *Improve 5x the QPI BW*

- ✓ ***Use silicon photonics*** *for low-latency high-port switching: Hipoλaos*
- ✓ ***Deployed OptoHPC*** *simulator*

Our work @PhosNET



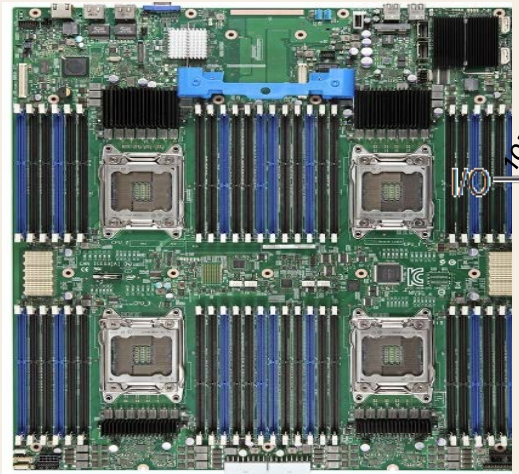


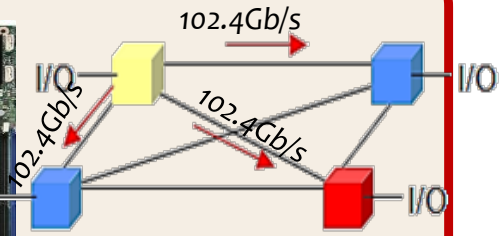
2010 2012 2016 2018

- ✓ *First concepts for plasmonics in NoCs*
- ✓ *Ultra-fast Optical RAMs and caches*
- ✓ *16-socket compute with 12.8Tb/s*
- ✓ *OptoHPC simulator software*
- ✓ *Hipoλaos 1024-port switch for disaggregation*
- ✓ *to CMOS plasmonics*
- ✓ *To 200G CMOS plasmonic mod*

Multisocket processors…"glueless"

4-Socket



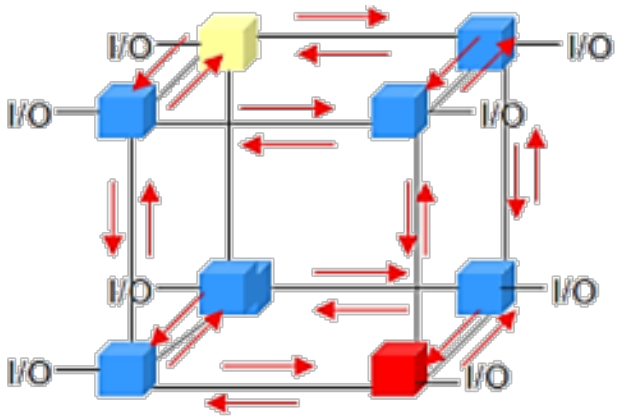


Intel® QuickPath Interconnect

- ✓ *4 sockets P2P*
- ✓ *9.6 Gb/s*
- ✓ *38.4 GB/s BW (307.2 Gb/s) or 102.4 Gb/s /direction*
- ✓ *60ns-240ns Latency*
- ✓ *16 pJ/bit link*

- ✓ ***120 high-perf processors directly connected !***
- ✓ ***Low-latency, low-power***

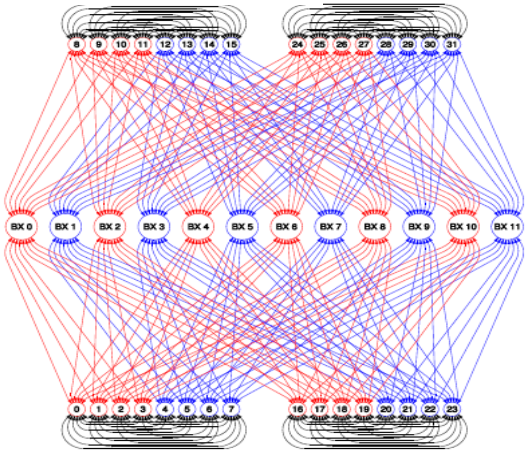
8-Socket



- ✓ ***240 high-perf processors!***
- ✗ ***up to 2-hop***
- ✗ ***Increased latency***

Going beyond 8 sockets?

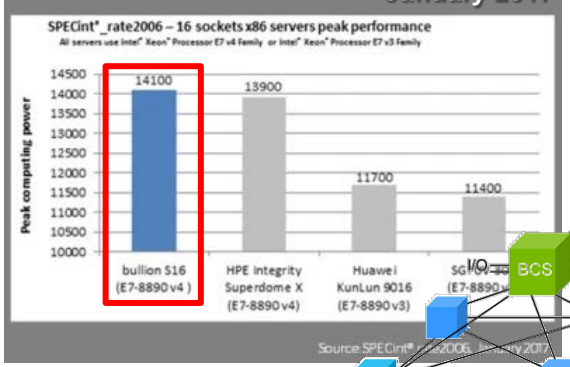
ORACLE



Bixby (BX) switches

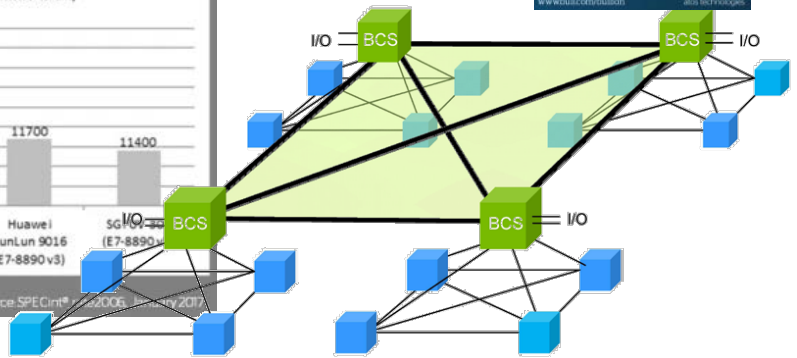
bullion™ high-end server beats worldwide records

January 2017



Bull

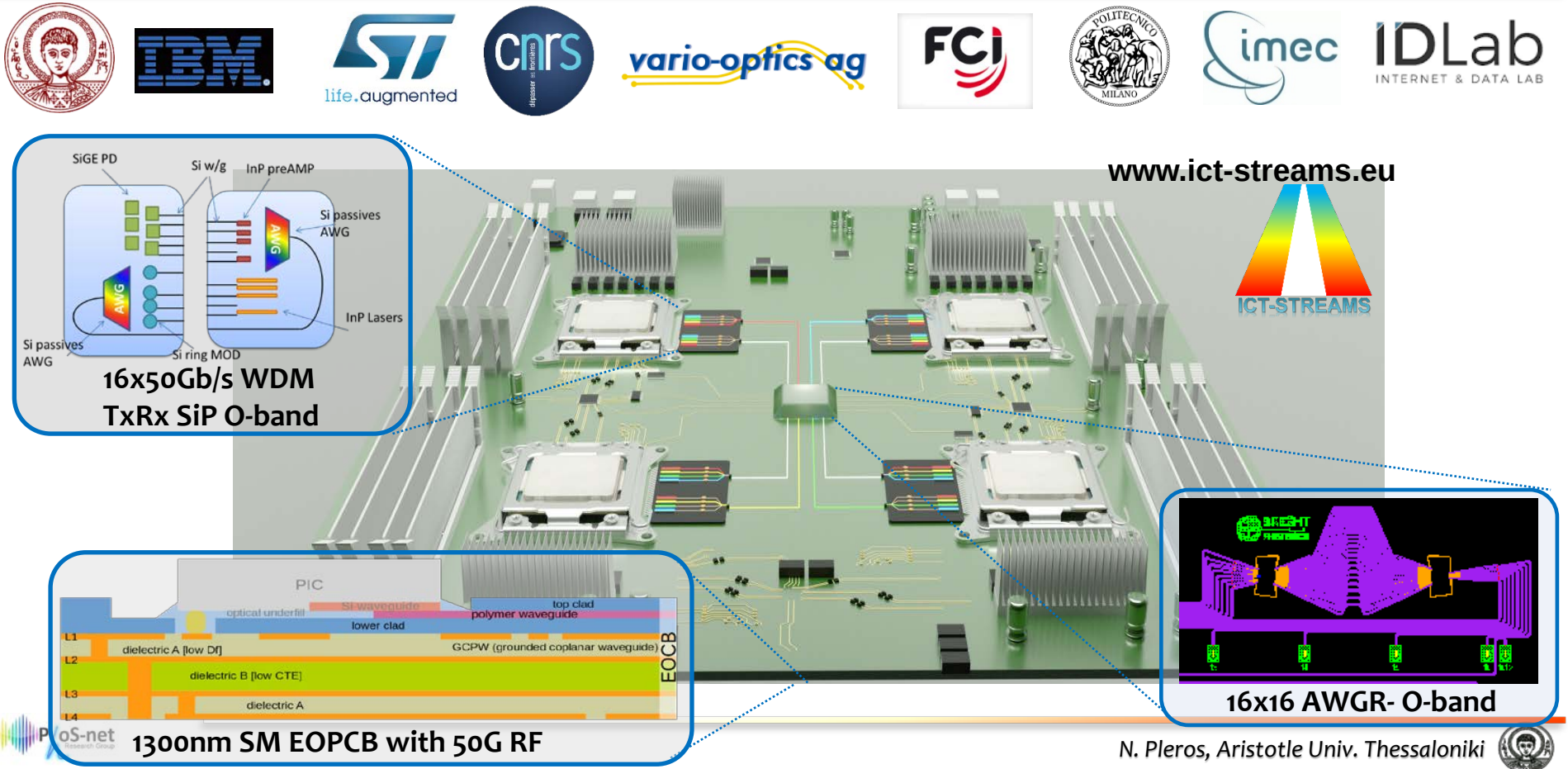
bullion



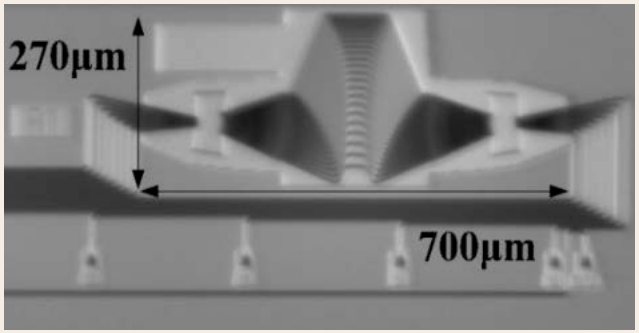
enormous compute power: >14TFlops

- ✓ *only through extra switch stage*
- ✓ *...but latency increases dramatically*
- ✓ *65% of QPI BW wasted for cache coherency updates*

The ICT-STREAMS multisocket interconnect



Si-based 8x8 AWGR & RM technology

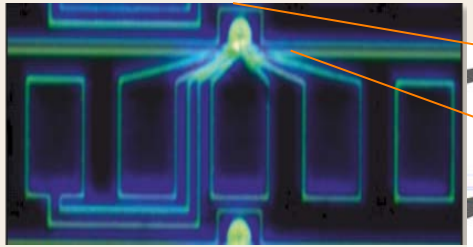


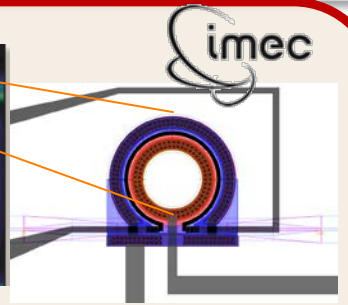
- ***8x8 AWGR O-band***
- ***10nm channel spacing***

OFC2018

S.Pitris et al, Th.2.A.1 Poster Ses., OFC18

S. Pitris et. al., to appear at OpEx

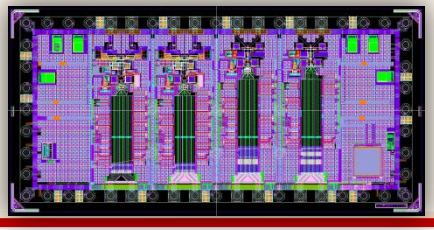




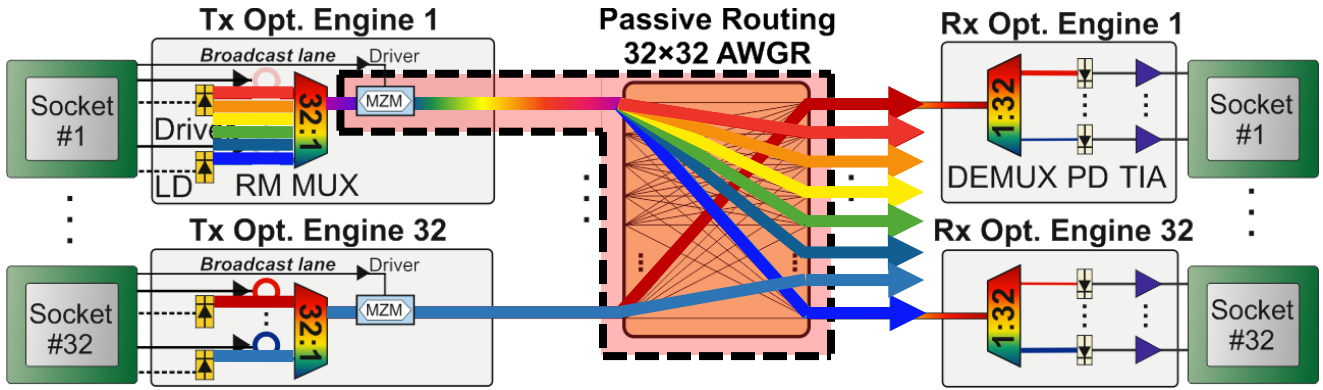
~20 fF 2.3V Vpp @25 Gb/s

- ***26.5 fJ/bit @ 25Gbps***
- ***2.4pj/b expected when driven by IC driver (in fab)***

ST 55 nm SiGe BiCMOS

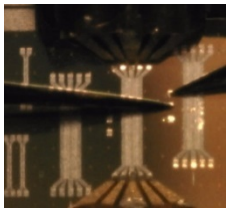


Improving energy in cache coherency updates



S.Pitris et al, Th.1.G.5, OFC2018

Si-pho Mach-Zehnder Modulator



D. Perez-Galacho et al, Opt. Express 25 (10), 2017

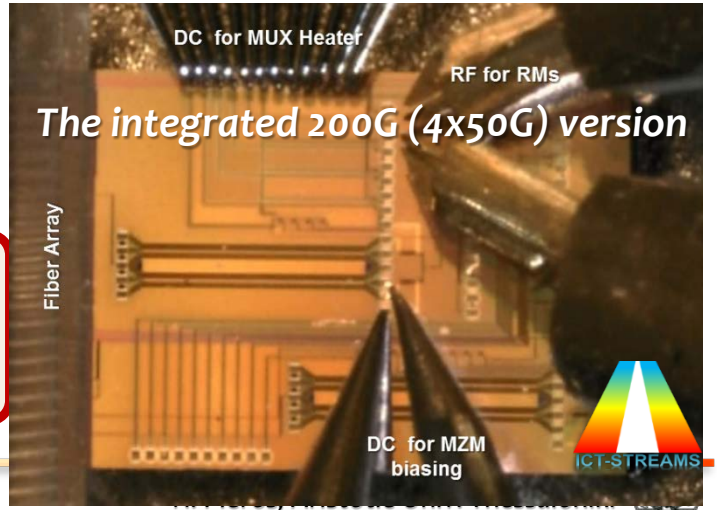
Unicasting — *Enable ➔ 1 LD + 1 RM + 1 PD/TIAs*

Broadcasting — *Enable ➔ 32 LDs + 32 ~~RMs~~ + 32 PD/TIAs*

1 MZM

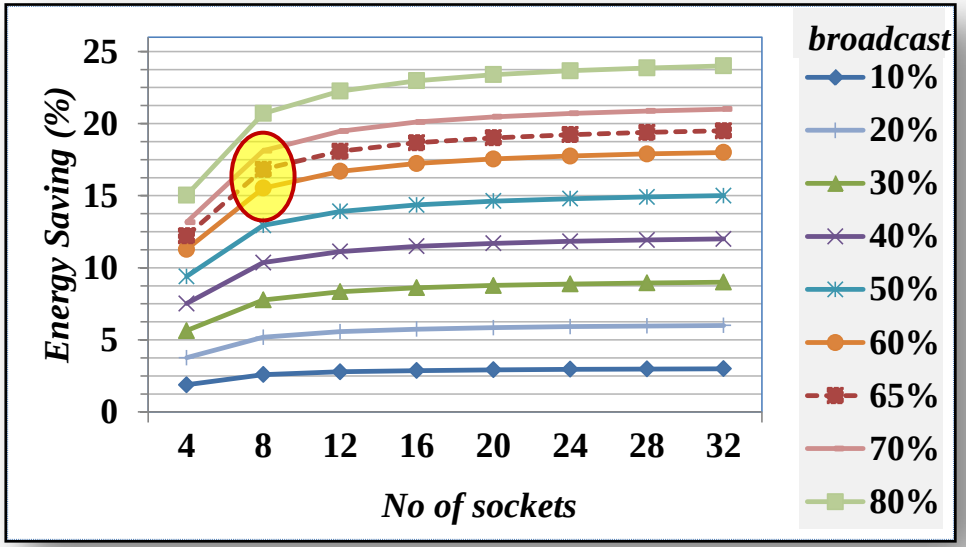
Broadcast-friendly topology:

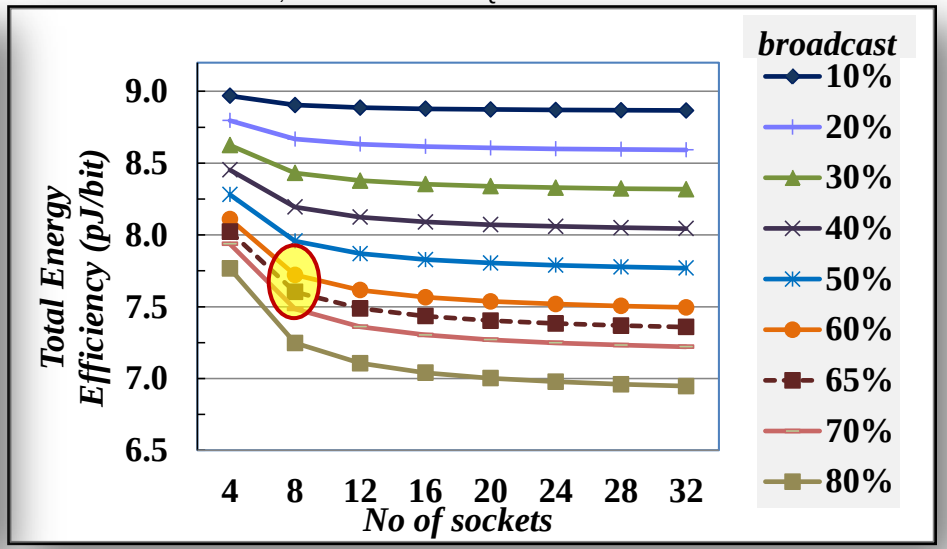
- *Use MZM for simultaneous modulation*
- *RMs in transparency mode*



The energy gain

C. Mitsolidou et al, submitted at JSTQE





✓ ***>15% energy gain*** *for >8-sockets & 65% broadcasted traffic*

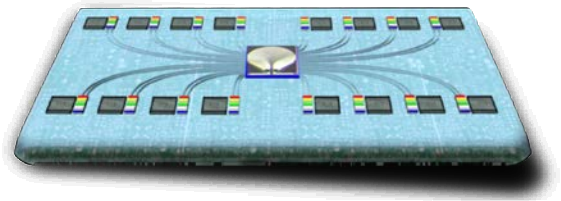
*…**solely** through the broadcast-friendly architecture*

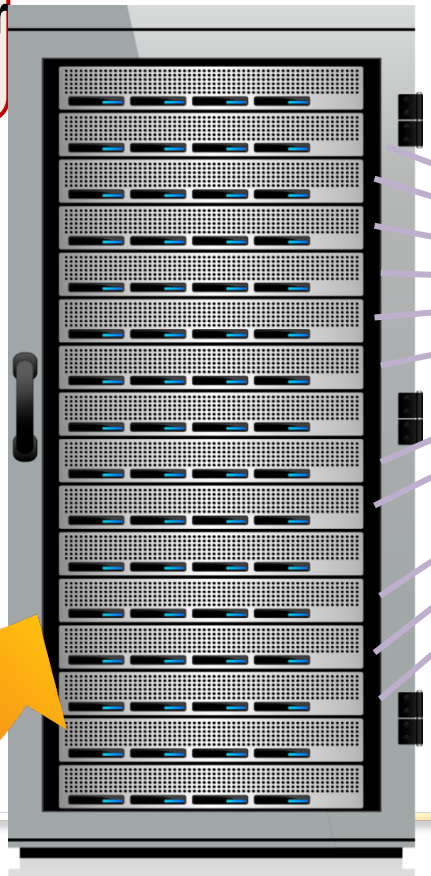
✓ ***~7.5pJ/bit*** *for >8-sockets & 65% broadcasted traffic (25G)*

✓ ***>50% energy savings compared to QPI***

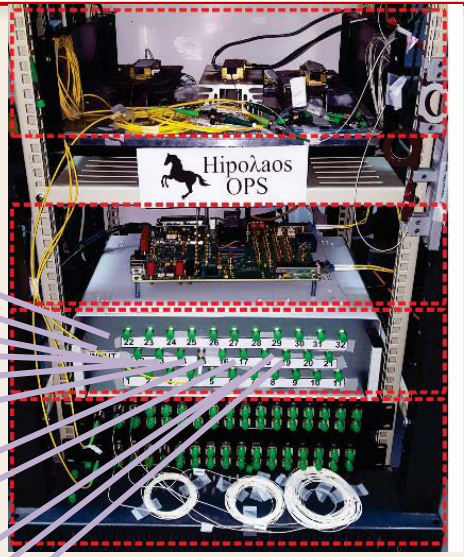
Disaggregate powerful multisocket boards

Release a powerful setup for disaggregated computing





Scaling up to Rack-level



- ***Hipολaos OPS: 256x256 up to 1024x1024 ports***
- ***sub-μsec latency***

N. Terzenidis et al, to appear at OpEx

Conclusions

- ✓ *Connectivity and 3D co-integration the limits for Si-Pho pNoCs in manycore… but not for multicore!*
- ✓ *Form powerful manycore setups through:*
  - ✓ *Si-Pho for >16-socket compute nodes (~480 cores)*
    - ✓ *directly linked, with 12.8T*
    - ✓ *<50% energy over QPI*
    - ✓ *adapting to cache coherency-induced broadcasting*
  - ✓ *Disintegrated macro-chip layouts via off-chip optical cache memories*
- ✓ *Disaggregated computing through powerful Si-Pho multi-socket boards connected over a low-latency 256x256 Hipoλaos OPS*

Acknowledgements

PHOTONICS PUBLIC PRIVATE PARTNERSHIP

ICT-STREAMS

www.ict-streams.eu

L3MATRIX
LARGE, LOW POWER AND LOW COST DATA CENTRES

http://l3matrix.eu

PLASMO

www.plasmofab.eu

Acknowledgements

Greek colleagues and PhD students who stayed in Greece despite the financial crisis, “guarding Thermopylae” (φυλάττω Θερμοπύλες)

ΚΕΔΕΚ

now with University of BRISTOL

now with cnrs

THANK YOU !

www.oi-summerschool.eu

25 - 29 June 2018

COSMICC STREAMS

3rd European Summerschool on

PICs for Optical Interconnects

Thessaloniki, Greece

Jointly co-organized by H2020 projects

ICT-STREAMS

COSMIC

http://phos-net.csd.auth.gr/