



SOPHIA
CLUB
ENTREPRISES



Scaling an efficient flexible cloud

ARM

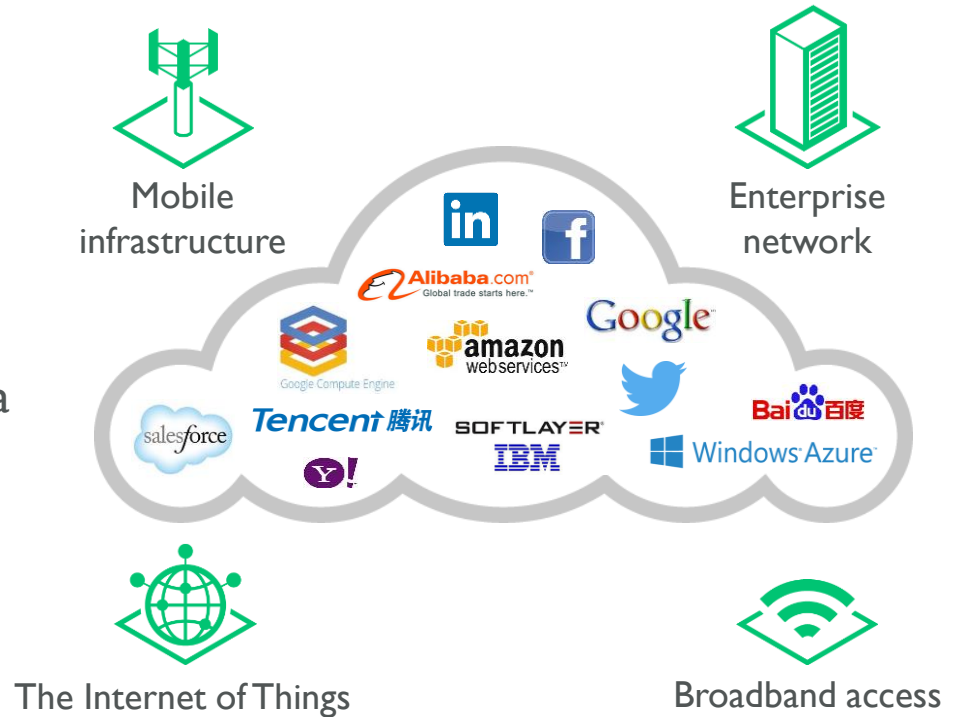
Christophe Evrard
Director Engineering
CPU Group, France

eSAME 2015
05/11/2015



Cloud Transforming Infrastructure

- Cloud sets the pace for rapid software and services evolution
- Increasing demand for intelligence at the network edge
- Evolving scale-out architectures from data center to network edge



Transition to Data and Application-Centric Networking

Rapid transformation over next three years



40%

of IoT-created data
processed at the edge of
the network

(Source: IDC)



50%

of IT networks will
become constrained

(Source: IDC)

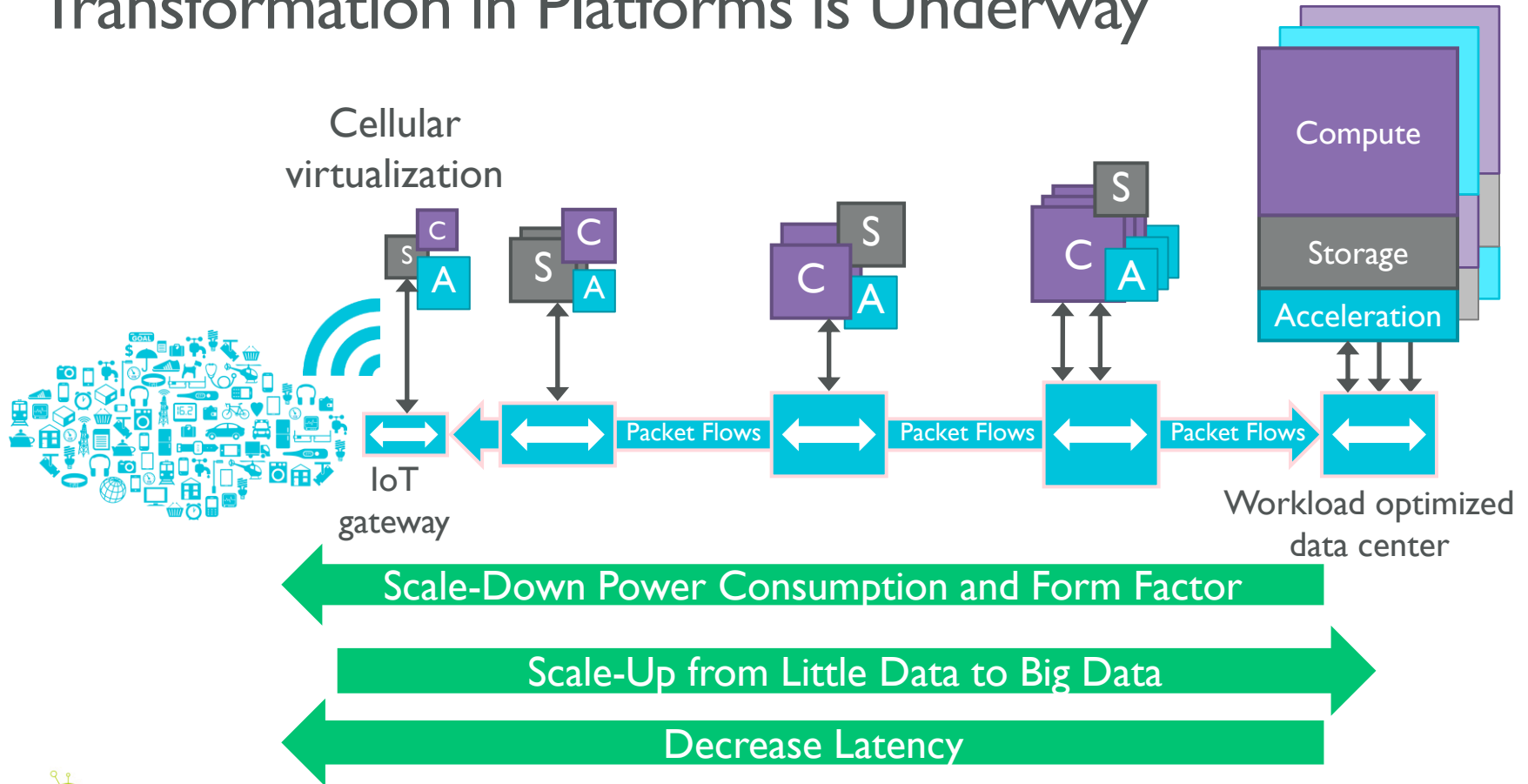


55%

of all internet traffic will be
through distributed content
delivery networks

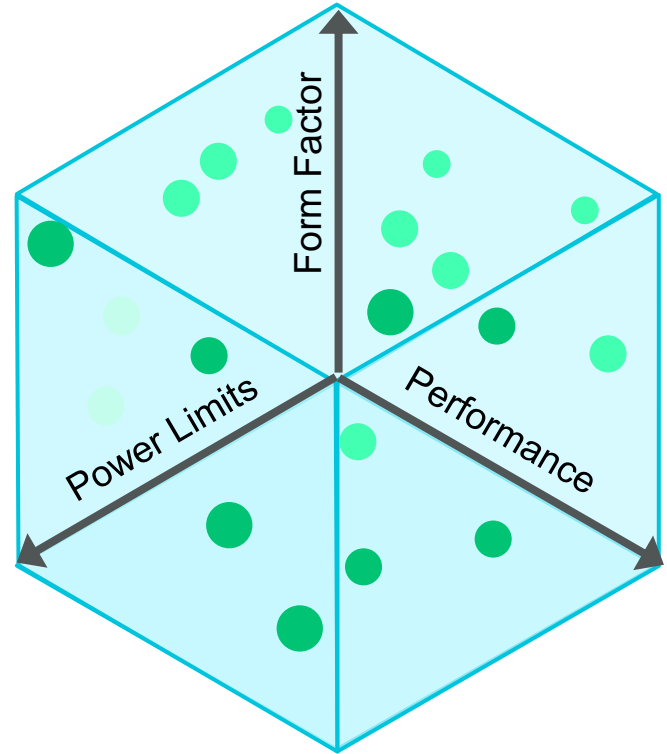
(Source: Cisco)

Transformation in Platforms is Underway



Physical design constraints

- Multiple Design Points across:
 - Power limits
 - Space and form-factor limits
 - Raw performance and latency needs



Physical design constraints

Telco Servers

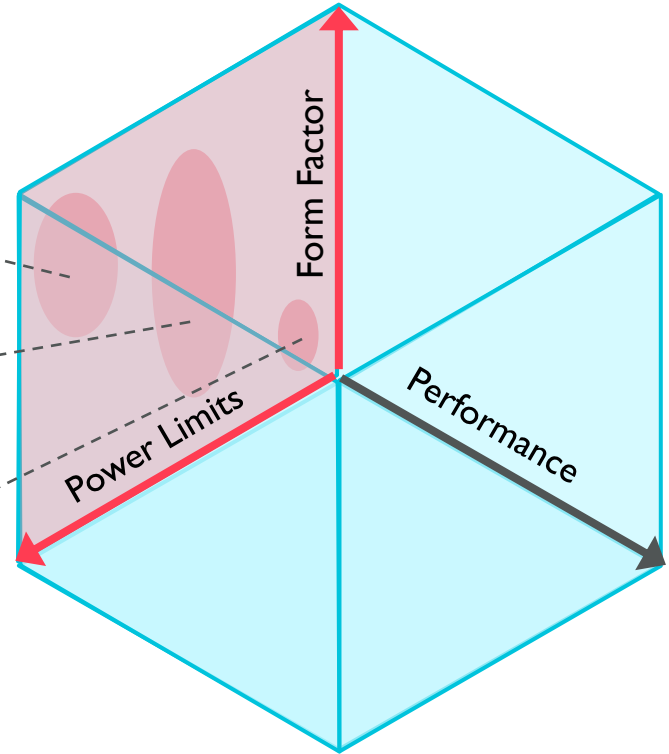
- 1.4KW 1U - 2x Xeon E5, 4HDD
- 2.8KW 4U - 4x Xeon E7, 8HDD

Telco Equipment

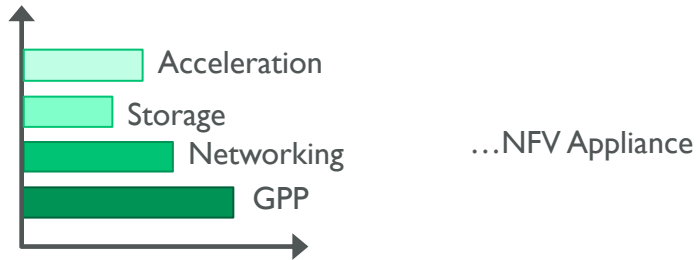
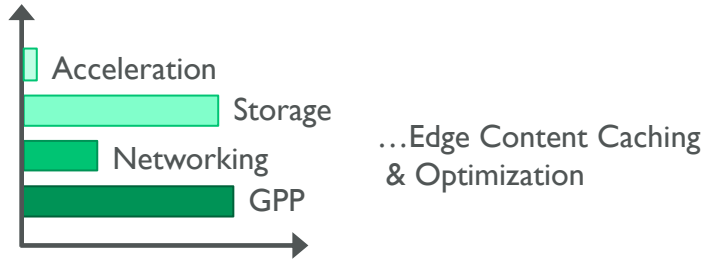
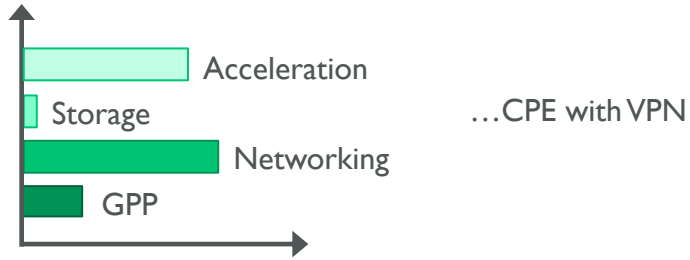
- 350W 1U – 10Gbps routing - NPU, PLD, GPP
- 2.8KW 4U - 320Gbps routing -- NPU, PLD, GPP
 - ~200W 40Gbps Blades
- 150W BLADE – DSP for LTE – DSP, PLD, NPU

Outdoor Equipment

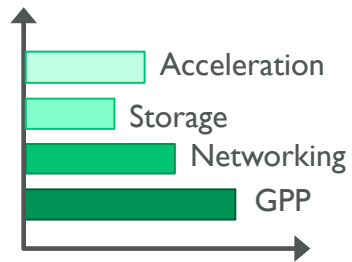
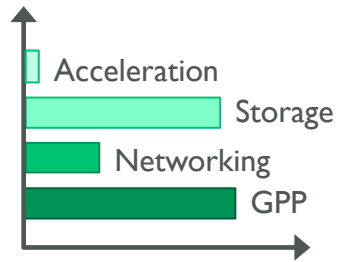
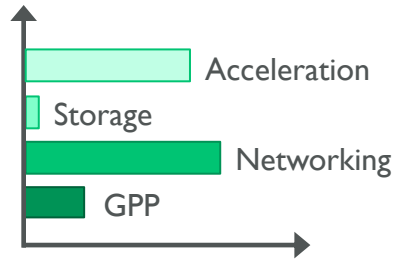
- 15W - PoE Access - Dedicated SoC
- 20W – RRH – Dedicated SoC



Workload & platform diversity



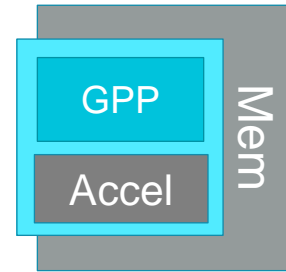
Workload & platform diversity



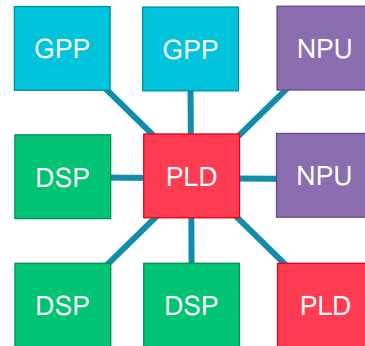
...CPE with VPN

...Edge Content Caching & Optimization

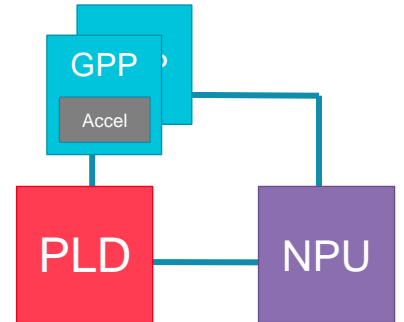
...NFV Appliance



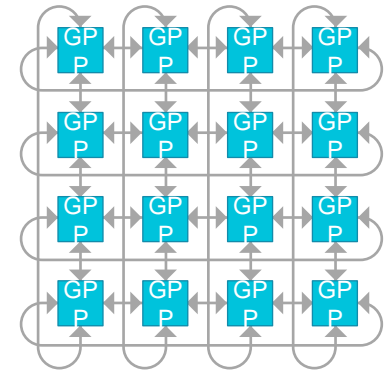
...Wireless-Access-point / CPE



...Base Station / Modem

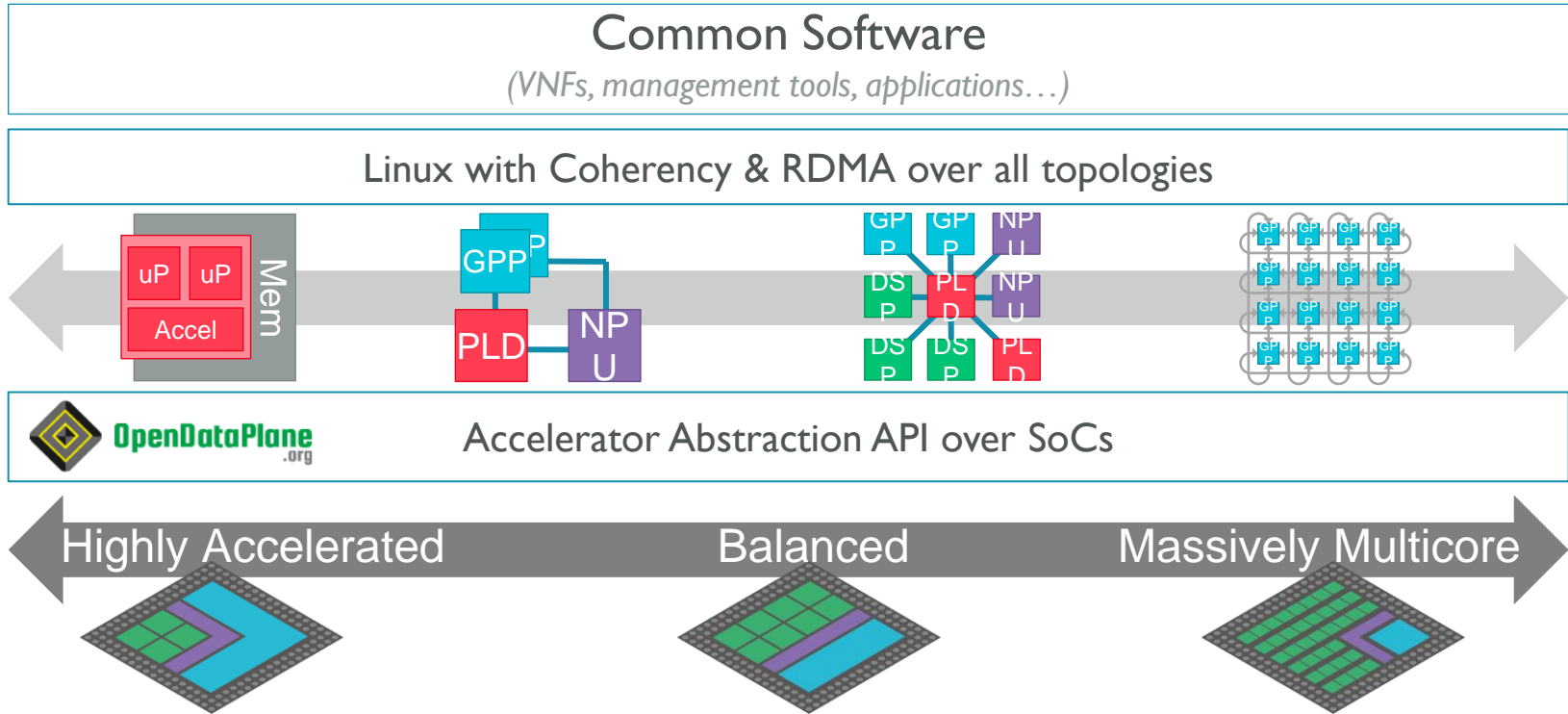


...Switch/Router Blade



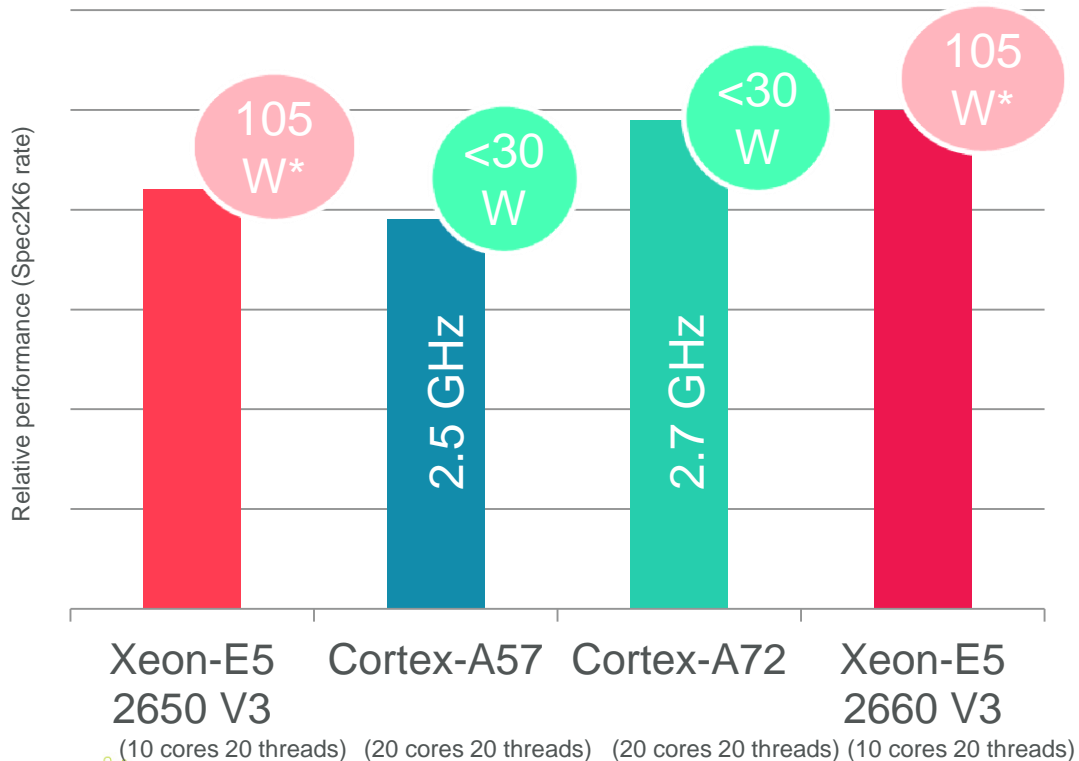
...Scale Out / Telecom Server

Software consistency



Maximizing Throughput Density: Per mm², per Watt

20-Thread Workload



ARM solution benefits:

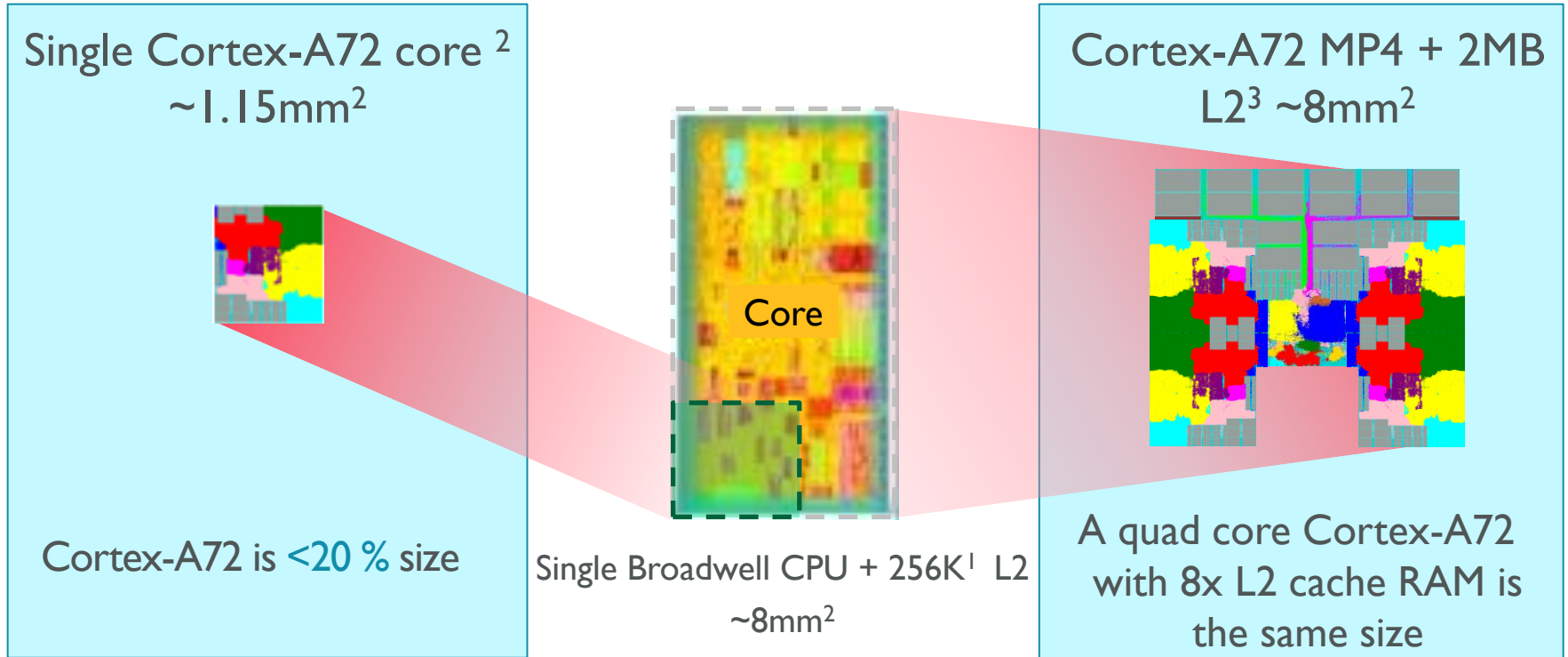
- Less than 1/3rd the power for equivalent performance*
- Allows power headroom for specialized computing or greater thread density

Comparison for equivalent number of threads

- Platforms used:
 - Xeon-E5 2660 10C20T platform (measured)
 - Xeon-E5 2650 10C20T platform (measured)
 - Gcc compiler v4.9 with -o3 flag
 - TDP rating source: ark.intel.com
- Estimated result on example 20C ARM Cortex platforms with CCN-508, 28MB total L2+L3 cache
 - per-core measurements on RTL with relevant memory system
 - GCC compiler v4.9 with -o3 flag
 - Scaled to 20T based on modelled and empirical results
 - Power estimated in 16nm based on ARM internal implementations for entire CPU+interconnect complex including 20xCPU, CCN-508, L2+L3 caches
 - Actual results on silicon platforms may vary

* A portion of Intel TDP power will be consumed by IO, The Cortex-A72 and Cortex-A57 estimates exclude IO power

Cortex-A72: Ideal for Dense Compute Environments



Summary

- There is a transformation in networking and infrastructure
- Energy efficient scalable network optimized solutions are required
- Momentum around ARM architecture rapidly increasing in infrastructure
- ARM delivers higher efficiency in terms of performance per mm per watt for tomorrow's greener computing