An Overview of Parallel Architectures

Figures, examples από

1. **Αρχιτεκτονική Υπολογιστών, Ποσοτική Προσέγγιση, J.L.Hennesy, A. Patterson**
2. **An Introduction to the Intel® QuickPath Interconnect:**

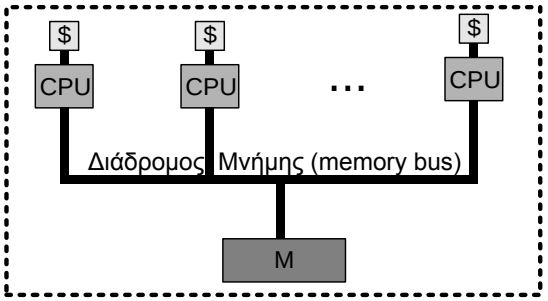
http://www.intel.com/content/www/us/en/io/quickpath-technology/quick-path-interconnect-introduction-paper.html

Συστήματα με πολλούς επεξεργαστές:
Λίγη αριθμητική

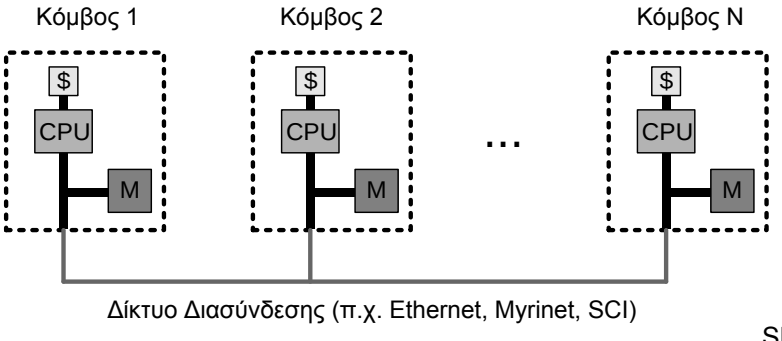
- **2-4 πυρήνες** σε προσωπικούς/φορητούς υπολογιστές και σε κινητά τηλέφωνα
- **Δεκάδες πυρήνες** σε έναν cloud server, σε μία κάρτα γραφικών, σε έναν computation accelerator
- **Εκατοντάδες/Χιλιάδες/Εκατομμύρια(!?)** πυρήνες σε ένα data center, IaaS provider, supercomputer

Συστήματα με πολλούς επεξεργαστές:
Βασικές αρχιτεκτονικές

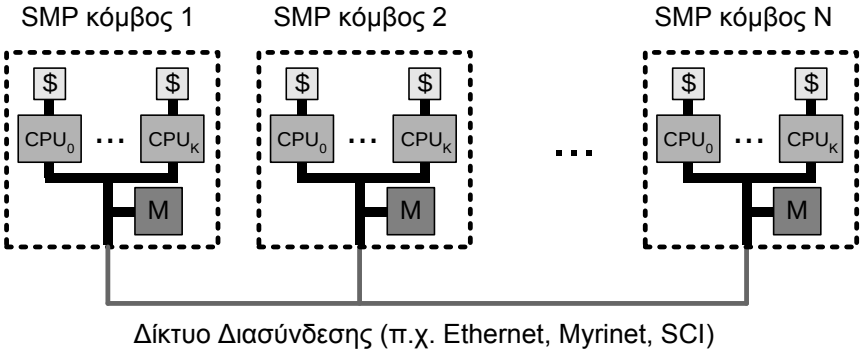
Κοινής Μνήμης



Κατανεμημένης Μνήμης



Υβριδική



Συστήματα με πολλούς επεξεργαστές:
Ζητήματα

- **Αρχιτεκτονική:**
  - Πώς επηρεάζονται οι ιεραρχίες μνημών;
  - Πώς διασυνδέονται οι επεξεργαστές;
- **Λογισμικό:**
  - Πώς προγραμματίζουμε αυτά τα συστήματα;
  - Λειτουργικό σύστημα: Πώς θα πρέπει να λειτουργεί ο χρονοδρομολογητής;
  - Πώς θα συγχρονίσουμε αποδοτικά πολλαπλά νήματα;

Συστήματα με πολλούς επεξεργαστές:
Ζητήματα

- **Αρχιτεκτονική:**
  - Πώς επηρεάζονται οι ιεραρχίες μνημών; **Προηγμένα Θέματα Αρχιτεκτονικής Υπολογιστών**
  - Πώς διασυνδέονται οι επεξεργαστές; **Σημερινό μάθημα και Συστήματα Παράλληλης Επεξεργασίας**
- **Λογισμικό:**
  - Πώς προγραμματίζουμε αυτά τα συστήματα; **Συστήματα Παράλληλης Επεξεργασίας**
  - Λειτουργικό σύστημα: Πώς θα πρέπει να λειτουργεί ο χρονοδρομολογητής; **Συστήματα Παράλληλης Επεξεργασίας**
  - Πώς θα συγχρονίσουμε αποδοτικά πολλαπλά νήματα; **Συστήματα Παράλληλης Επεξεργασίας**

Συστήματα με πολλούς επεξεργαστές:
Τεχνολογικές τάσεις

- Νόμος του Moore: Διπλασιασμός αριθμού τρανζίστορ κάθε 18 μήνες (εξακολουθεί να ισχύει)
- Dennard Scaling: Η τάση τροφοδοσίας των transistor μπορεί να μειώνεται (δεν μειώνεται πλέον με τον ίδιο ρυθμό)
- Καταναλισκόμενη ισχύς: $P = C V^2 f$ (C = count, V = Voltage, f = frequency)
- Στο άμεσο μέλλον δεν θα μπορούμε να τροφοδοτήσουμε όλα τα διαθέσιμα transistor
  - dark silicon era?

Μια ματιά στα συστήματα μεγάλης κλίμακας
Υπερυπολογιστές

- Τεράστια συστήματα με χιλιάδες/εκατομμύρια επεξεργαστές
- Χρησιμοποιούνται για επιστημονικές εφαρμογές
  - Life sciences
  - Earth Sciences
  - Engineering
  - Etc.
- Top500:
  - 2 φορές το χρόνο ανακοινώνεται η λίστα με τους 500 ισχυρότερους επεξεργαστές
    - » Supercomputing, Νοέμβριος, http://sc14.supercomputing.org/ (ΗΠΑ)
    - » ISC, Ιούνιος, http://www.isc-events.com/isc14/ (Γερμανία)

Supercomputers

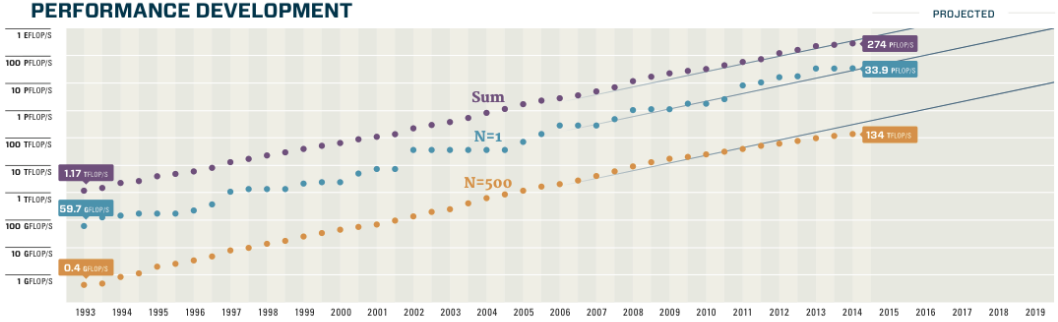
TOP500 JUNE 2014

PRESENTED BY

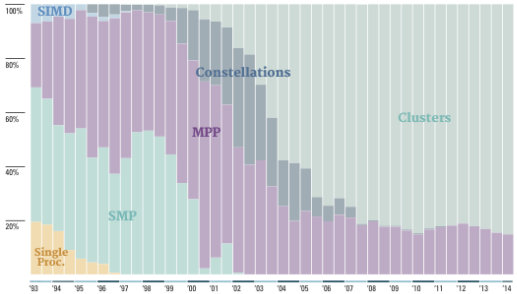
FIND OUT MORE AT www.top500.org

	NAME	SPECS	SITE	COUNTRY	CORES	RMAX PFLOP/S	POWER MW
1	Tianhe-2 (Milkyway-2)	NUDT, Intel Ivy Bridge (12C, 2.2 GHz) & Xeon Phi (57C, 1.1 GHz), Custom interconnect	NSCC Guangzhou	China	3,120,000	33.9	17.8
2	Titan	Cray XK7, Operon 6274 (16C 2.2 GHz) + Nvidia Kepler GPU, Custom interconnect	DOE/SC/ORNL	USA	560,640	17.6	8.2
3	Sequoia	IBM BlueGene/Q, Power BQC (16C 1.60 GHz), Custom interconnect	DOE/NNSA/LLNL	USA	1,572,864	17.2	7.9
4	K computer	Fujitsu SPARC64 VIIIfx (8C, 2.0GHz), Custom interconnect	RIKEN AICS	Japan	705,024	10.5	12.7
5	Mira	IBM BlueGene/Q, Power BQC (16C, 1.60 GHz), Custom interconnect	DOE/SC/ANL	USA	786,432	8.59	3.95

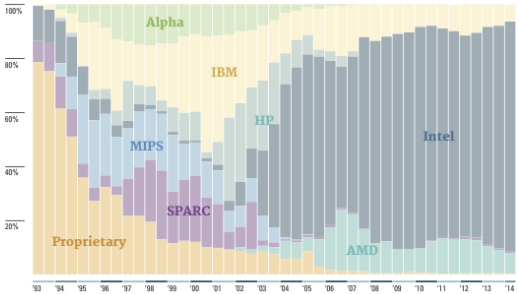
PERFORMANCE DEVELOPMENT



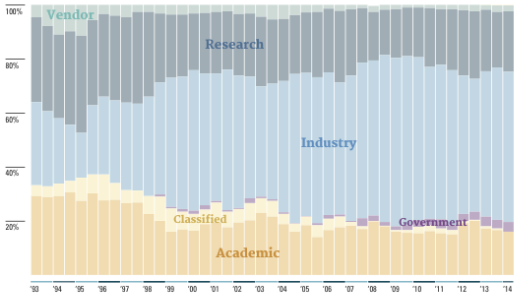
ARCHITECTURES



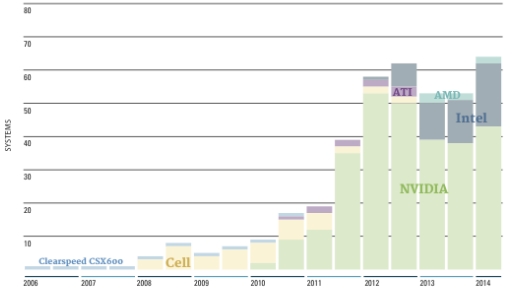
CHIP TECHNOLOGY



INSTALLATION TYPE



ACCELERATORS/CO-PROCESSORS



HPLINPACK A Portable Implementation of the High Performance Linpack Benchmark for Distributed Memory Computers ▸ MORE INFO AT http://icl.utk.edu/hpl/

Top500

www.top500.org

June 2013 List

Top 500 (June 2013 list)

Top 5

	NAME	SPECS	SITE	COUNTRY	CORES	RMAX PFLOP/S	POWER MW
1	**Tianhe-2 (Milkyway-2)**	NUDT, Intel Ivy Bridge (12C, 2.2 GHz) & Xeon Phi (57C, 1.1 GHz), Custom interconnect	NSCC Guangzhou	China	3,120,000	33.9	17.8
2	**Titan**	Cray XK7, Operon 6274 (16C 2.2 GHz) + Nvidia Kepler GPU, Custom interconnect	DOE/SC/ORNL	USA	560,640	17.6	8.2
3	**Sequoia**	IBM BlueGene/Q, Power BQC (16C 1.60 GHz), Custom interconnect	DOE/NNSA/LLNL	USA	1,572,864	17.2	7.9
4	**K computer**	Fujitsu SPARC64 VIIIfx (8C, 2.0GHz), Custom interconnect	RIKEN AICS	Japan	705,024	10.5	12.7
5	**Mira**	IBM BlueGene/Q, Power BQC (16C, 1.60 GHz), Custom interconnect	DOE/SC/ANL	USA	786,432	8.59	3.95

Top 500 (June 2013 list)

Top 5

	NAME	SPECS	SITE	COUNTRY	CORES	RMAX PFLOP/S	POWER MW
1	Tianhe-2 (Milkyway-2)	NUDT, Intel Ivy Bridge (12C, 2.2 GHz) & Xeon Phi (57C, 1.1 GHz), Custom interconnect	NSCC Guangzhou	China	3,120,000	33.9	17.8
2	Titan	Cray XK7, Operon 6274 (16C 2.2 GHz) + Nvidia Kepler GPU, Custom interconnect	DOE/SC/ORNL	USA	560,640	17.6	8.2
3	Sequoia	IBM BlueGene/Q, Power BQC (16C 1.60 GHz), Custom interconnect	DOE/NNSA/LLNL	USA	1,572,864	17.2	7.9
4	K computer	Fujitsu SPARC64 VIIIfx (8C, 2.0GHz), Custom interconnect	RIKEN AICS	Japan	705,024	10.5	12.7
5	Mira	IBM BlueGene/Q, Power BQC (16C, 1.60 GHz), Custom interconnect	DOE/SC/ANL	USA	786,432	8.59	3.95

millions of cores

Top 500 (June 2013 list)

Top 5

	NAME	SPECS	SITE	COUNTRY	CORES	RMAX PFLOP/S	POWER MW
1	**Tianhe-2 (Milkyway-2)**	NUDT, Intel Ivy Bridge (12C, 2.2 GHz) & Xeon Phi (57C, 1.1 GHz), Custom interconnect	NSCC Guangzhou	China	3,120,000	33.9	17.8
2	**Titan**	Cray XK7, Operon 6274 (16C 2.2 GHz) + Nvidia Kepler GPU, Custom interconnect	DOE/SC/ORNL	USA	560,640	17.6	8.2
3	**Sequoia**	IBM BlueGene/Q, Power BQC (16C 1.60 GHz), Custom interconnect	DOE/NNSA/LLNL	USA	1,572,864	17.2	7.9
4	**K computer**	Fujitsu SPARC64 VIIIfx (8C, 2.0GHz), Custom interconnect	RIKEN AICS	Japan	705,024	10.5	12.7
5	**Mira**	IBM BlueGene/Q, Power BQC (16C, 1.60 GHz), Custom interconnect	DOE/SC/ANL	USA	786,432	8.59	3.95

MWs of power

Top 500 (June 2013 list)

Top 5

	NAME	SPECS	SITE	COUNTRY	CORES	RMAX PFLOP/S	POWER MW
1	**Tianhe-2 (Milkyway-2)**	NUDT, Intel Ivy Bridge (12C, 2.2 GHz) & Xeon Phi (57C, 1.1 GHz), Custom interconnect	NSCC Guangzhou	China	3,120,000	33.9	17.8
2	**Titan**	Cray XK7, Operon 6274 (16C 2.2 GHz) + Nvidia Kepler GPU, Custom interconnect	DOE/SC/ORNL	USA	560,640	17.6	8.2
3	**Sequoia**	IBM BlueGene/Q, Power BQC (16C 1.60 GHz), Custom interconnect	DOE/NNSA/LLNL	USA	1,572,864	17.2	7.9
4	**K computer**	Fujitsu SPARC64 VIIIfx (8C, 2.0GHz), Custom interconnect	RIKEN AICS	Japan	705,024	10.5	12.7
5	**Mira**	IBM BlueGene/Q, Power BQC (16C, 1.60 GHz), Custom interconnect	DOE/SC/ANL	USA	786,432	8.59	3.95

What's this?

Top 500 (June 2013 list)

Top 5

	NAME	SPECS	SITE	COUNTRY	CORES	RMAX PFLOP/S	POWER MW
1	**Tianhe-2 (Milkyway-2)**	NUDT, Intel Ivy Bridge (12C, 2.2 GHz) & Xeon Phi (57C, 1.1 GHz), Custom interconnect	NSCC Guangzhou	China	3,120,000	33.9	17.8
2	**Titan**	Cray XK7, Operon 6274 (16C 2.2 GHz) + Nvidia Kepler GPU, Custom interconnect	DOE/SC/ORNL	USA	560,640	17.6	8.2
3	**Sequoia**	IBM BlueGene/Q, Power BQC (16C 1.60 GHz), Custom interconnect	DOE/NNSA/LLNL	USA	1,572,864	17.2	7.9
4	**K computer**	Fujitsu SPARC64 VIIIfx (8C, 2.0GHz), Custom interconnect	RIKEN AICS	Japan	705,024	10.5	12.7
5	**Mira**	IBM BlueGene/Q, Power BQC (16C, 1.60 GHz), Custom interconnect	DOE/SC/ANL	USA	786,432	8.59	3.95

What's this?

Top 500 (June 2013 list)

Top 5

	NAME	SPECS	SITE	COUNTRY	CORES	RMAX PFLOP/S	POWER MW
1	Tianhe-2 (Milkyway-2)	NUDT, Intel Ivy Bridge (12C, 2.2 GHz) & Xeon Phi (57C, 1.1 GHz), Custom interconnect	NSCC Guangzhou	China	3,120,000	33.9	17.8
2	Titan	Cray XK7, Operon 6274 (16C 2.2 GHz) + Nvidia Kepler GPU, Custom interconnect	DOE/SC/ORNL	USA	560,640	17.6	8.2
3	Sequoia	IBM BlueGene/Q, Power BQC (16C 1.60 GHz), Custom interconnect	DOE/NNSA/LLNL	USA	1,572,864	17.2	7.9
4	K computer	Fujitsu SPARC64 VIIIfx (8C, 2.0GHz), Custom interconnect	RIKEN AICS	Japan	705,024	10.5	12.7
5	Mira	IBM BlueGene/Q, Power BQC (16C, 1.60 GHz), Custom interconnect	DOE/SC/ANL	USA	786,432	8.59	3.95

2 x Ivy Bridge processors
2 x 12 cores / node

Top 500 (June 2013 list)

Top 5

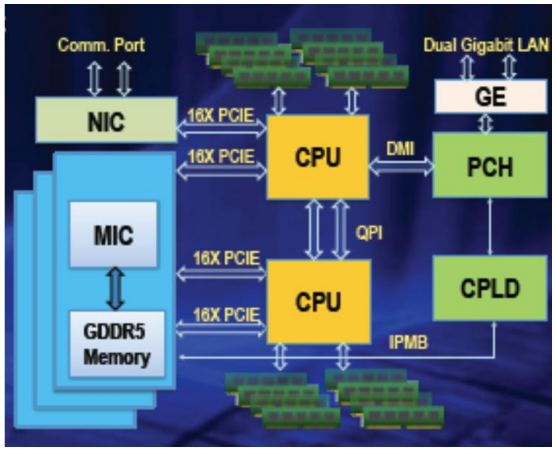
	NAME	SPECS	SITE	COUNTRY	CORES	RMAX PFLOP/S	POWER MW
1	Tianhe-2 (Milkyway-2)	NUDT, Intel Ivy Bridge (12C, 2.2 GHz) & Xeon Phi (57C, 1.1 GHz), Custom interconnect	NSCC Guangzhou	China	3,120,000	33.9	17.8
2	Titan	Cray XK7, Operon 6274 (16C 2.2 GHz) + Nvidia Kepler GPU, Custom interconnect	DOE/SC/ORNL	USA	560,640	17.6	8.2
3	Sequoia	IBM BlueGene/Q, Power BQC (16C 1.60 GHz), Custom interconnect	DOE/NNSA/LLNL	USA	1,572,864	17.2	7.9
4	K computer	Fujitsu SPARC64 VIIIfx (8C, 2.0GHz), Custom interconnect	RIKEN AICS	Japan	705,024	10.5	12.7
5	Mira	IBM BlueGene/Q, Power BQC (16C, 1.60 GHz), Custom interconnect	DOE/SC/ANL	USA	786,432	8.59	3.95

2 x Ivy Bridge processors

2 x 12 cores / node

+

Xeon Phi (MIC) accelerator

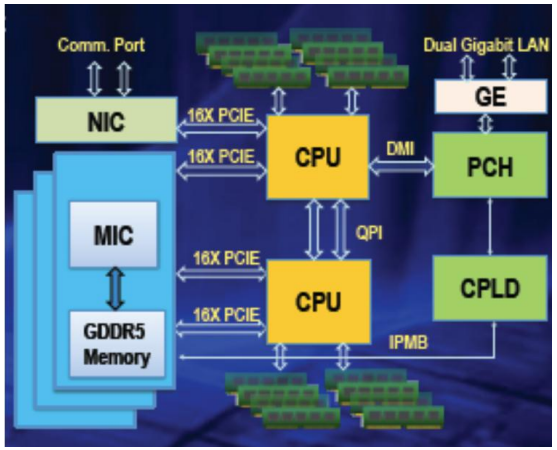


Top 500 (June 2013 list)

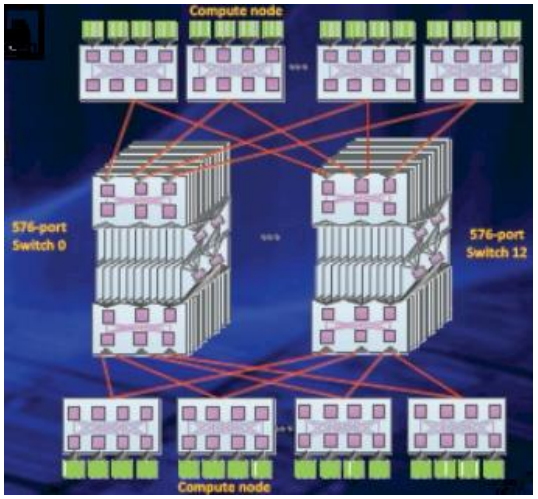
Top 5

	NAME	SPECS	SITE	COUNTRY	CORES	RMAX PFLOP/S	POWER MW
1	Tianhe-2 (Milkyway-2)	NUDT, Intel Ivy Bridge (12C, 2.2 GHz) & Xeon Phi (57C, 1.1 GHz), Custom interconnect	NSCC Guangzhou	China	3,120,000	33.9	17.8
2	Titan	Cray XK7, Operon 6274 (16C 2.2 GHz) + Nvidia Kepler GPU, Custom interconnect	DOE/SC/ORNL	USA	560,640	17.6	8.2
3	Sequoia	IBM BlueGene/Q, Power BQC (16C 1.60 GHz), Custom interconnect	DOE/NNSA/LLNL	USA	1,572,864	17.2	7.9
4	K computer	Fujitsu SPARC64 VIIIfx (8C, 2.0GHz), Custom interconnect	RIKEN AICS	Japan	705,024	10.5	12.7
5	Mira	IBM BlueGene/Q, Power BQC (16C, 1.60 GHz), Custom interconnect	DOE/SC/ANL	USA	786,432	8.59	3.95

2 x Ivy Bridge processors
2 x 12 cores / node
+
Xeon Phi (MIC) accelerator



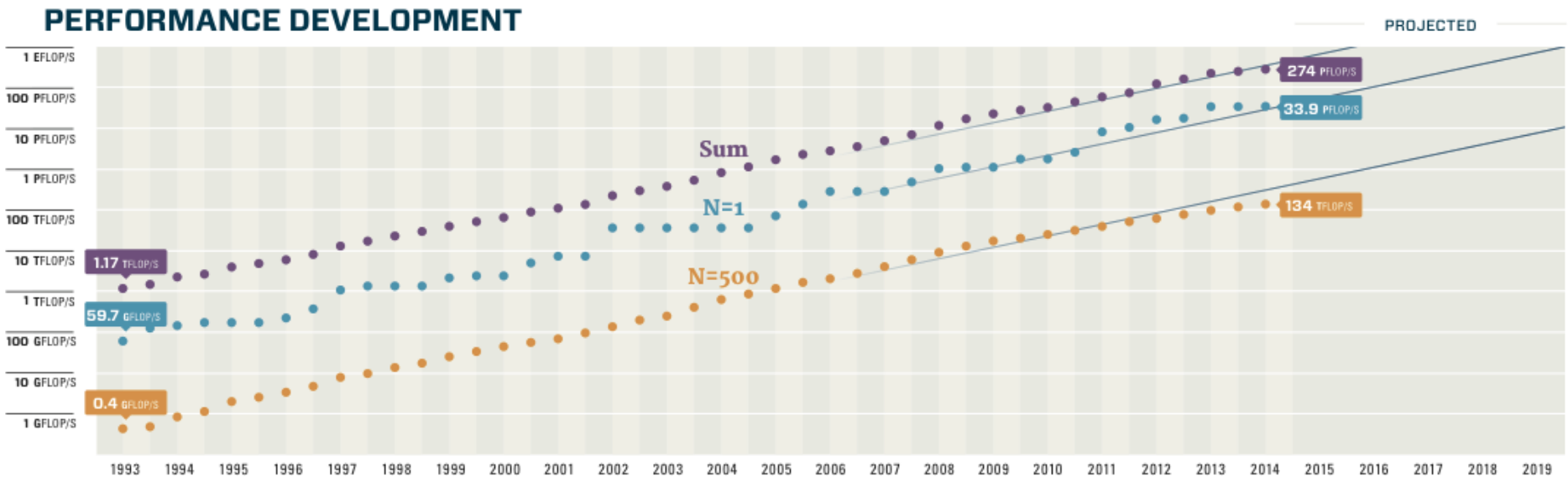
Custom interconnect (Fat- Tree topology – see next)



Images taken from: http://www.theregister.co.uk/Print/2013/06/10/inside_chinas_tianhe2_massive_hybrid_supercomputer/

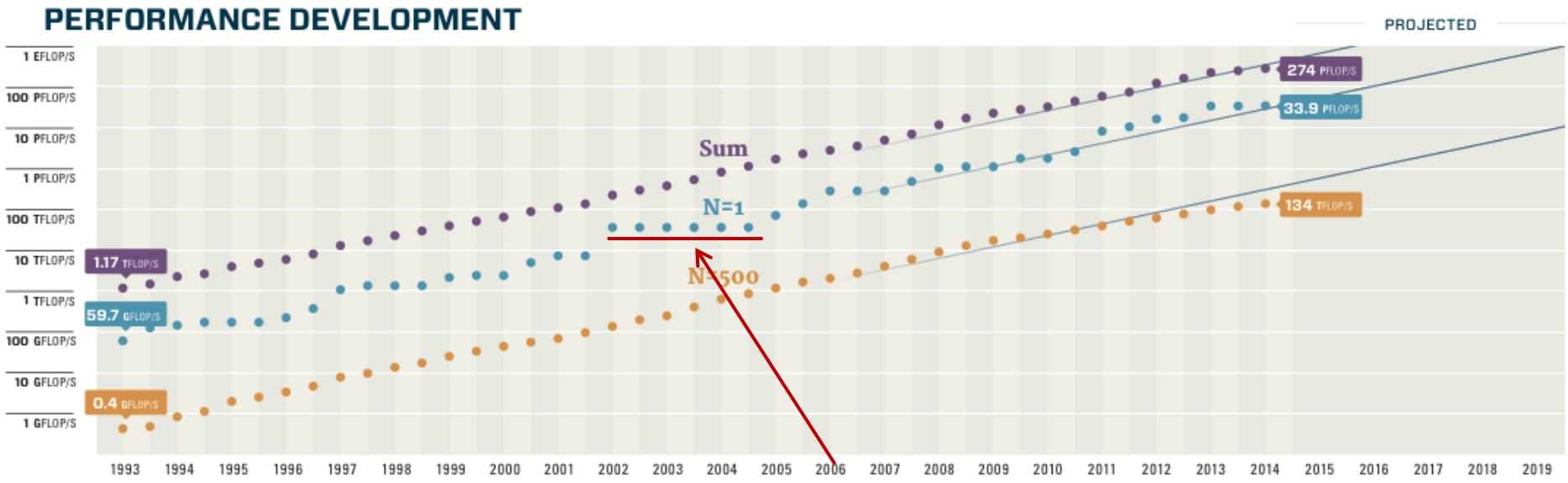
Top 500 (June 2013 list)

Performance development



Top 500 (June 2013 list)

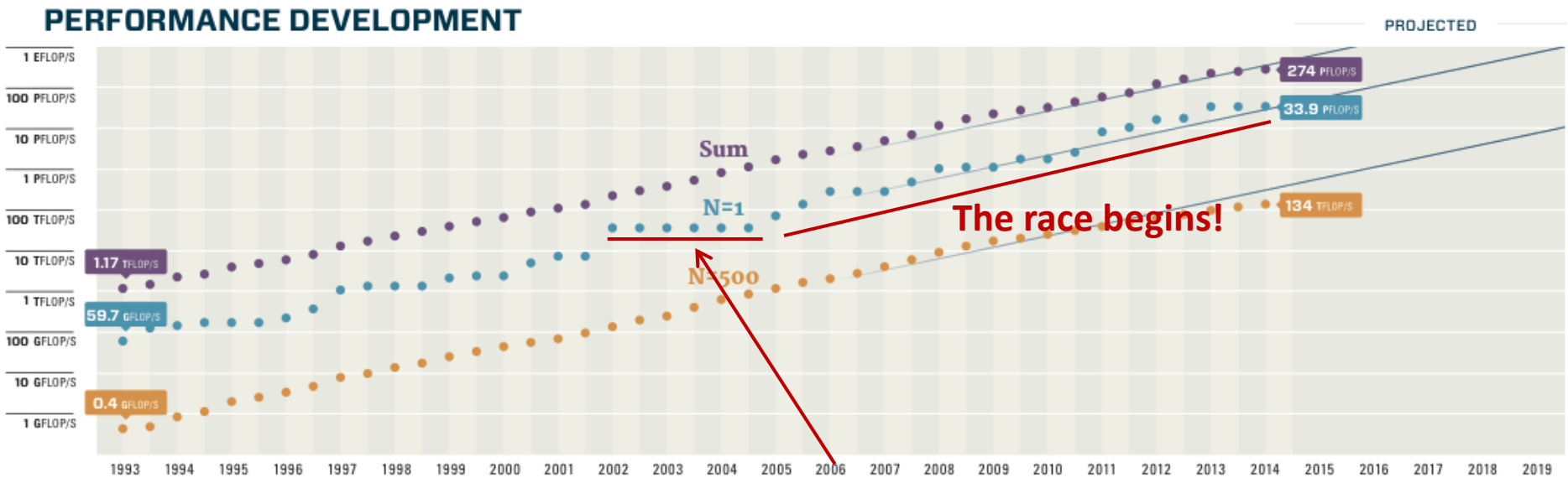
Performance development



3 years in Top1!
Earth Simulator
Japan

Top 500 (June 2013 list)

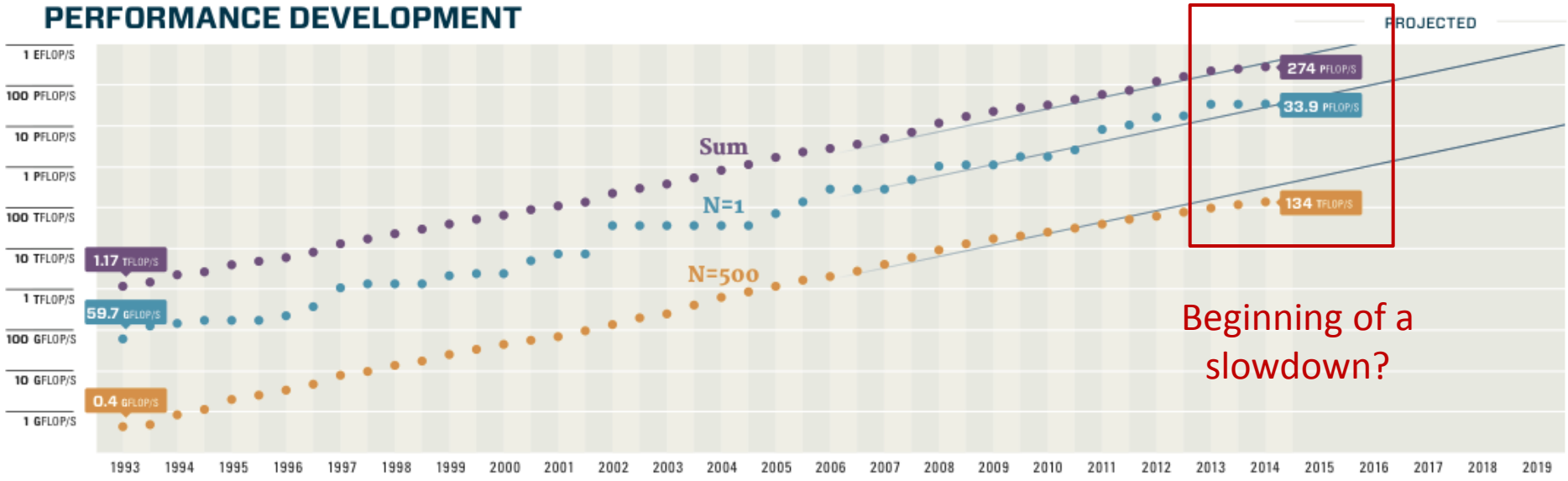
Performance development



3 years in Top1!
Earth Simulator
Japan

Top 500 (June 2013 list)

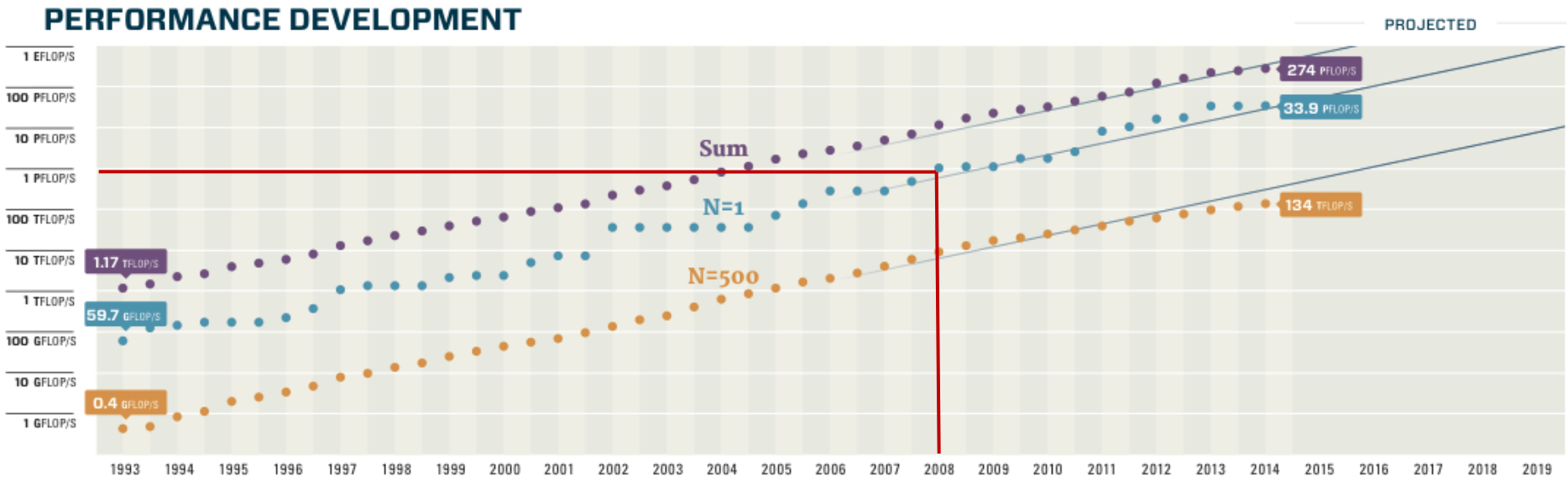
Performance development



Beginning of a slowdown?

Top 500 (June 2013 list)

Performance development



Petaflop barrier

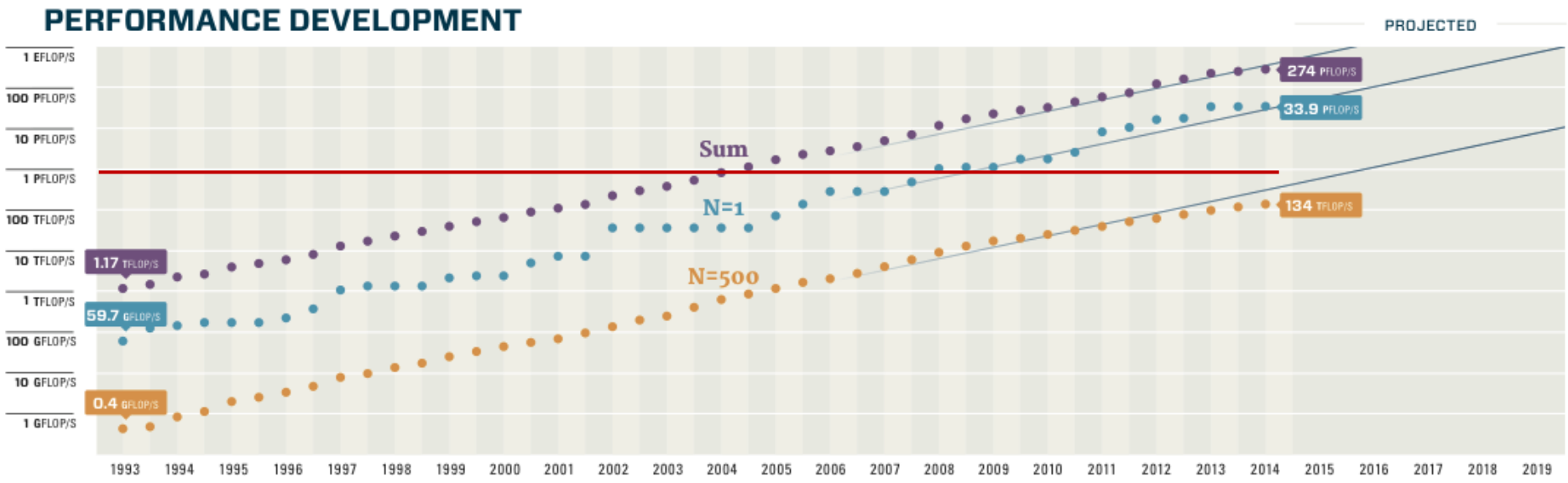
Roadrunner

Los Alamos National Laboratory, USA

Decommisioned 31st March 2013 ☹

Top 500 (June 2013 list)

Performance development



Why?

Roadrunner would be still high in Top500 (rank 40!)

Top 500 (November 2012 list)

Performance development

Rank	Site	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
21	Information Technology Center, The University of Tokyo Japan	**Oakleaf-FX** - PRIMEHPC FX10, SPARC64 IXfx 16C 1.848GHz, Tofu interconnect Fujitsu	76800	1043.0	1135.4	1177
22	DOE/NNSA/LANL United States	**Roadrunner** - BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz, Voltaire Infiniband IBM	122400	1042.0	1375.8	2345
23	University of Edinburgh United Kingdom	**DiRAC** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM	98304	1035.3	1258.3	493

Top 500 (November 2012 list)

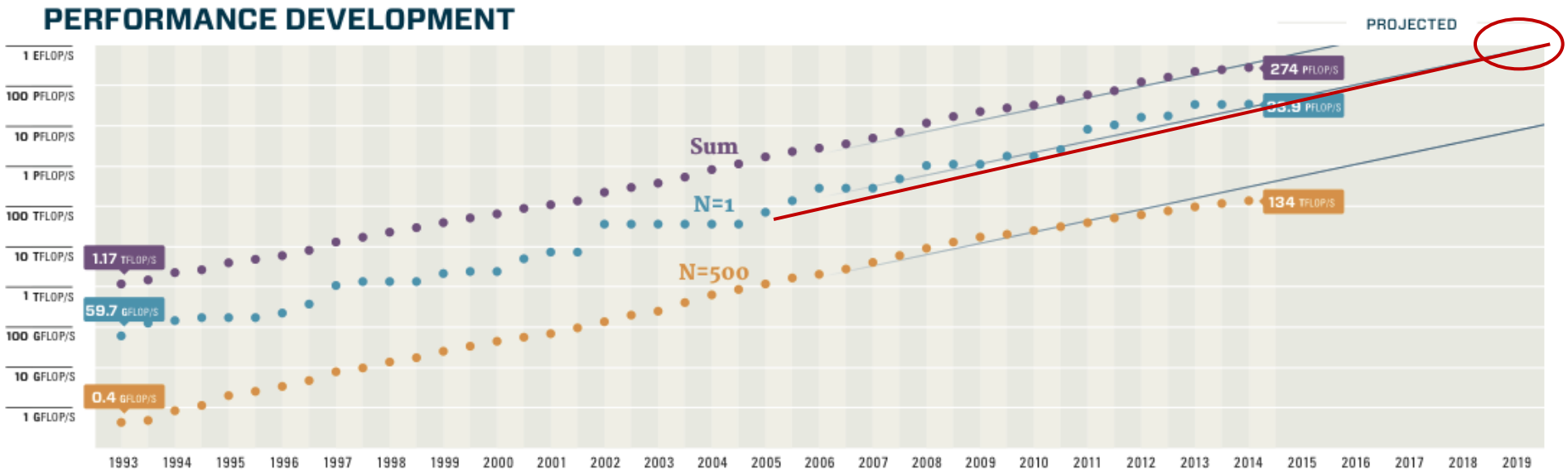
Performance development

Rank	Site	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
21	Information Technology Center, The University of Tokyo Japan	**Oakleaf-FX** - PRIMEHPC FX10, SPARC64 IXfx 16C 1.848GHz, Tofu interconnect Fujitsu	76800	1043.0	1135.4	1177
22	DOE/NNSA/LANL United States	**Roadrunner** - BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz, Voltaire Infiniband IBM	122400	1042.0	1375.8	2345
23	University of Edinburgh United Kingdom	**DiRAC** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM	98304	1035.3	1258.3	493

Low power efficiency!
(Flop/Watt)

Top 500 (June 2013 list)

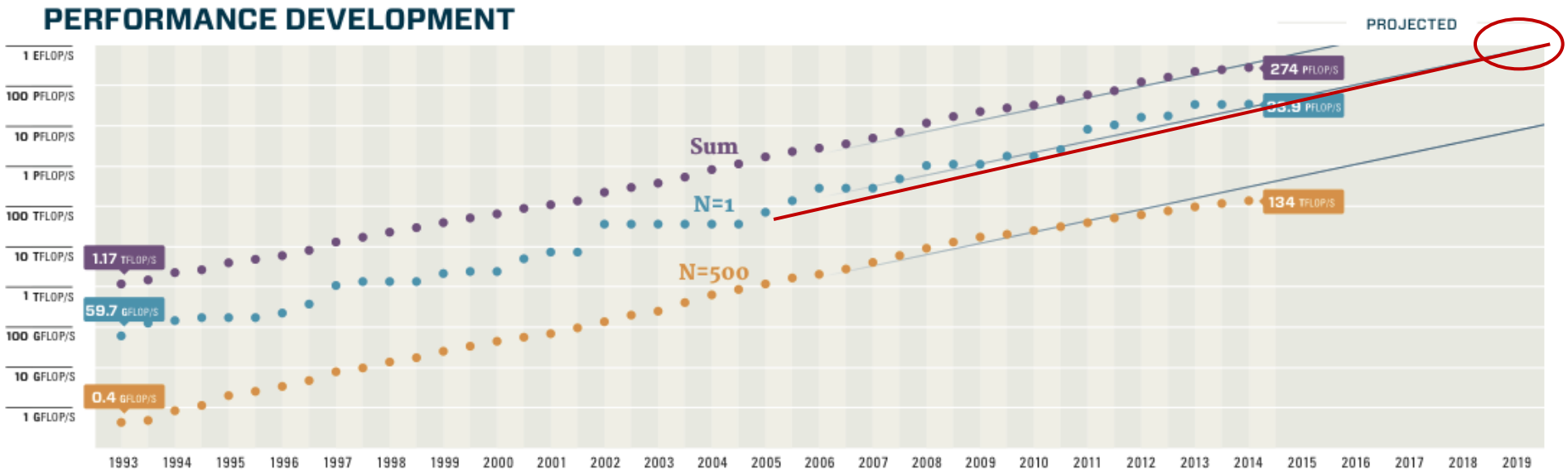
Performance development



Shall we reach “Exaflop computing” by the end of this decade?

Top 500 (June 2013 list)

Performance development



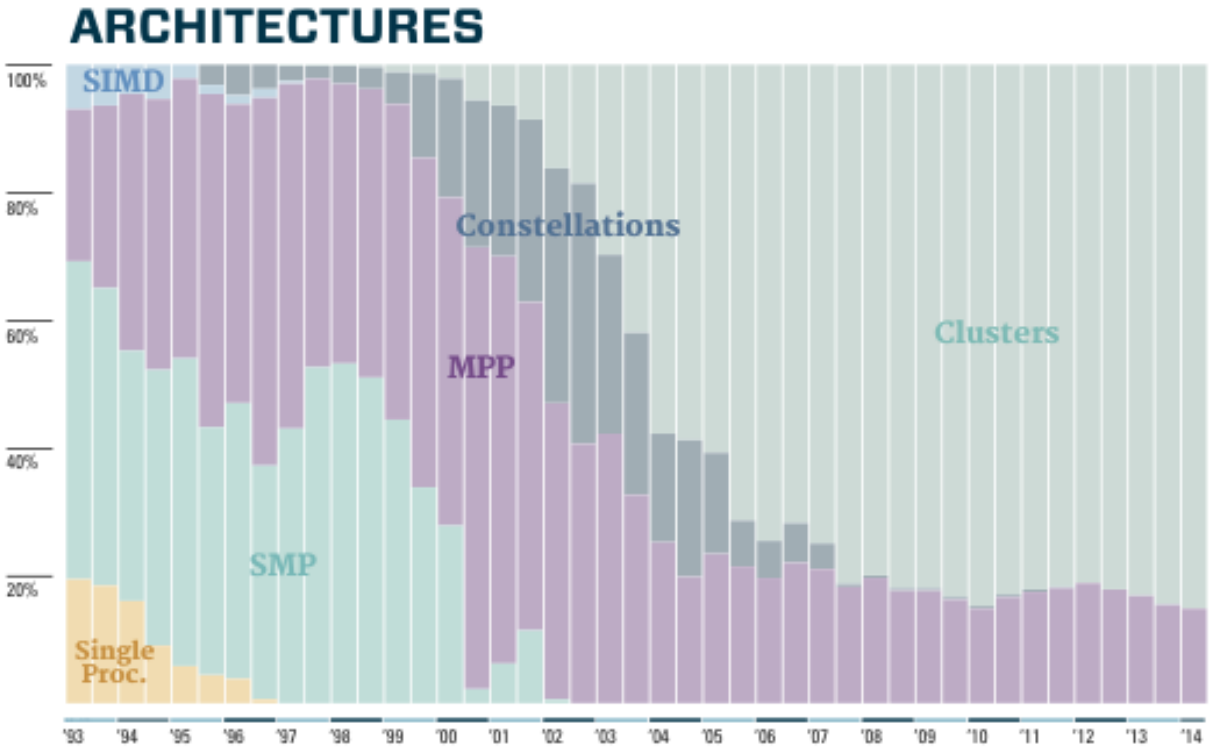
Three major problems:

a) power

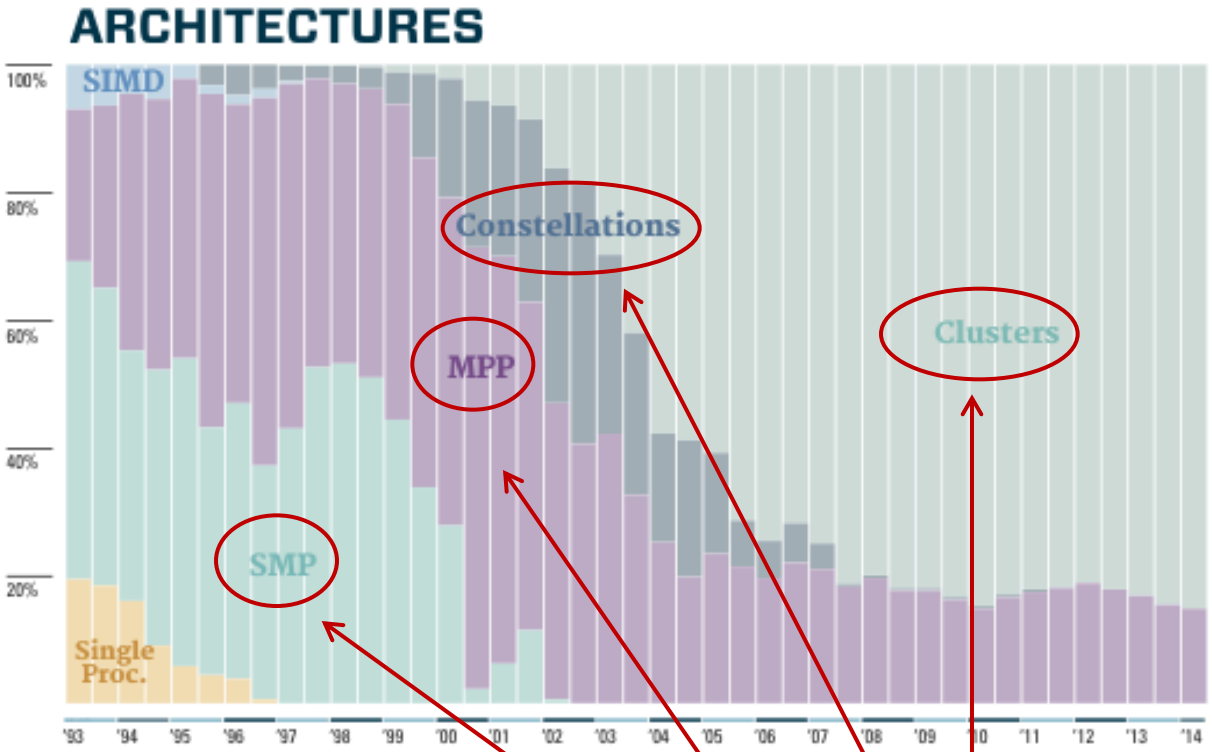
b) data movements

c) resilience

Top 500 (June 2013 list)
Architecture



Top 500 (June 2013 list)
Architecture

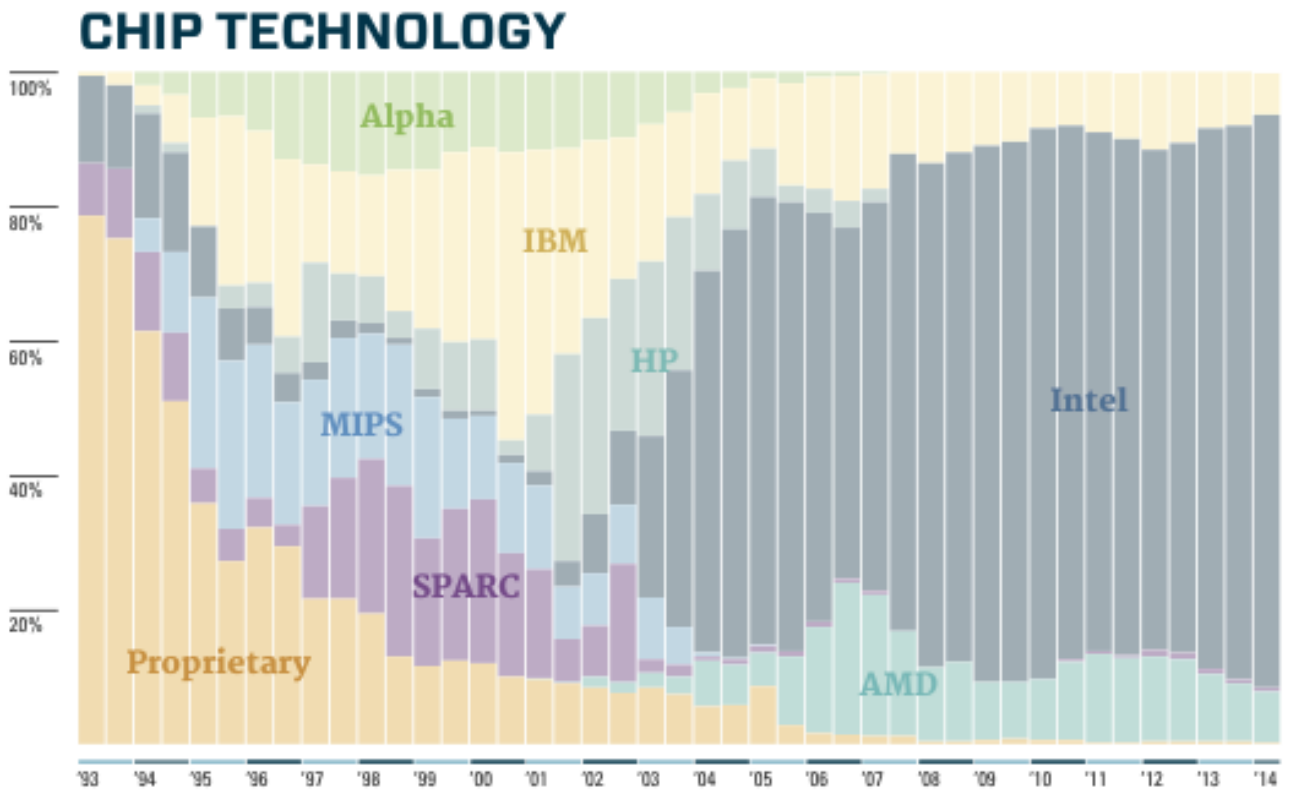


Top 500 (June 2013 list)
Architecture

- **SMP**: Symmetric Multiprocessor
- **Constellation**: Συλλογή από διαφορετικά συστήματα
- **MPP**: Massively Parallel Processing
  - Μαζικά παράλληλα συστήματα
  - Βασίζονται σε ειδικά κατασκευασμένα (custom made) στοιχεία
    - » Δίκτυα διασύνδεσης (κατά κύριο λόγο)
    - » Επεξεργαστικές μονάδες
  - Π.χ. Blue Gene/Q, Cray XK7
  - Καταλαμβάνουν τις υψηλότερες θέσεις του Top500
  - Χαμηλότερη κατανάλωση ενέργειας
  - Υψηλότερο κόστος
- **Clusters:** Συστοιχίες συστημάτων
  - Όλα τα στοιχεία τους είναι ήδη εμπορικά διαθέσιμα
  - Δίκτυα διασύνδεσης: Infiniband, 10G Ethernet, Gbit Ethernet

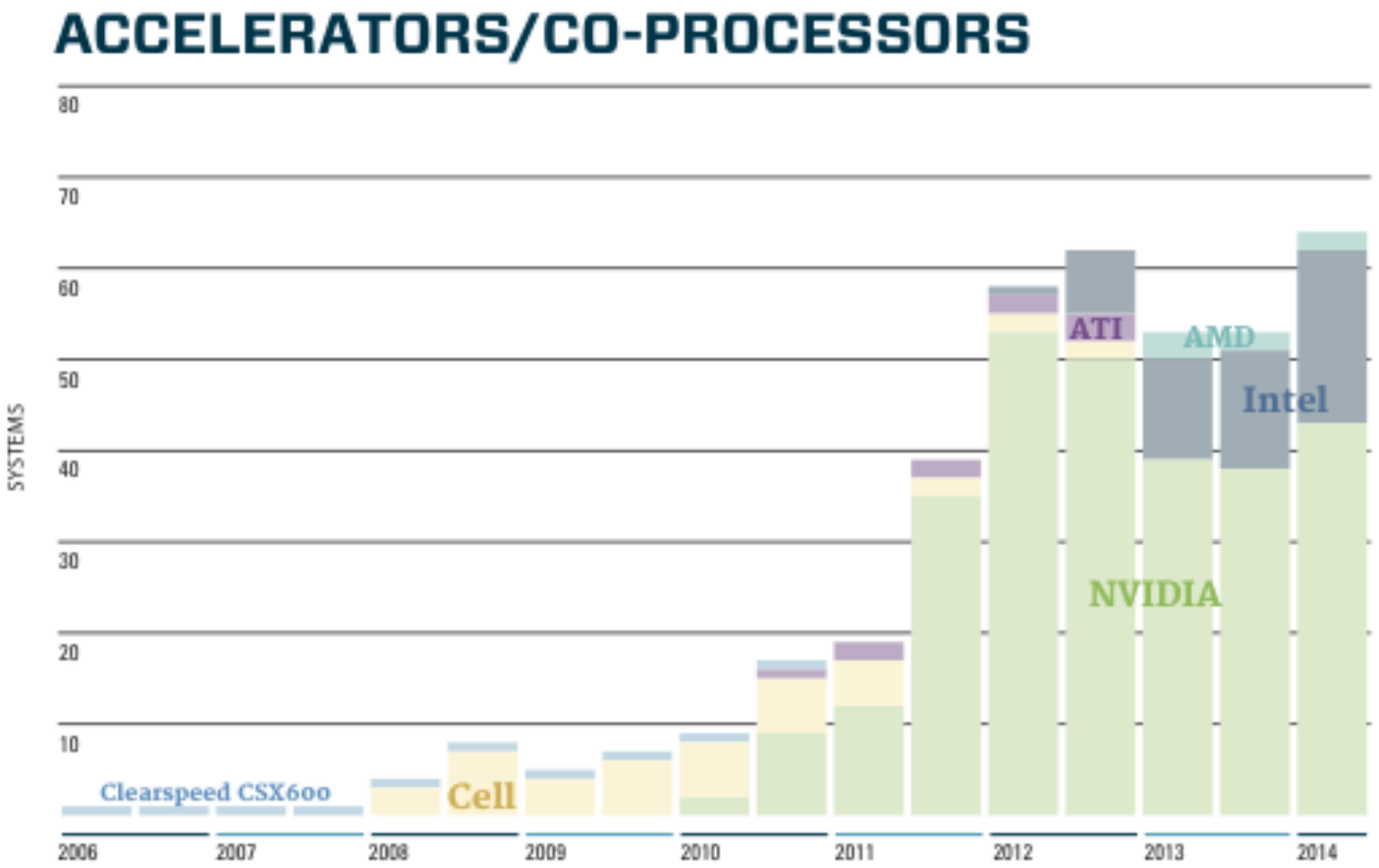
Top 500 (June 2013 list)

Architecture – Chip technology



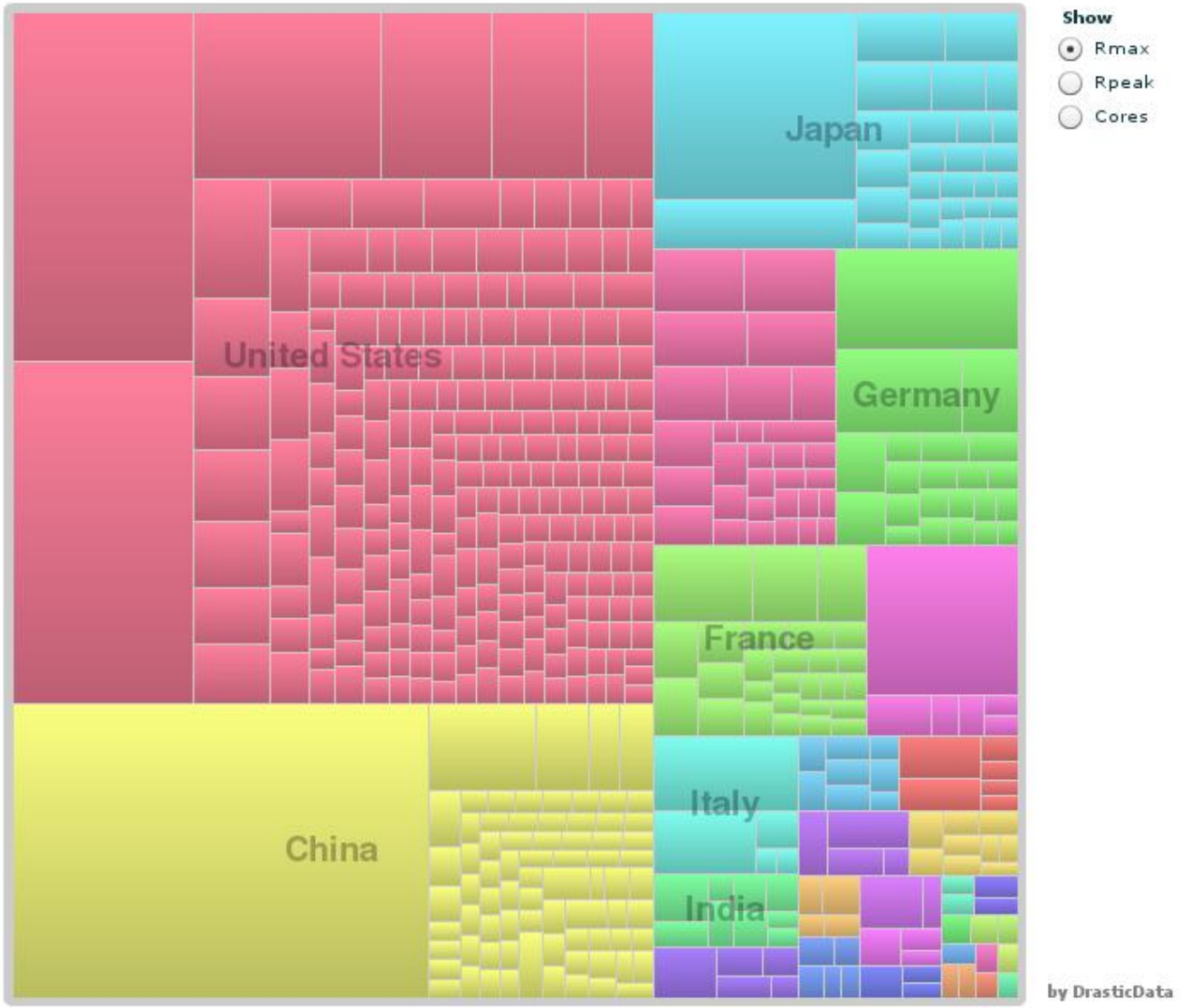
Top 500 (June 2013 list)

Architecture – The accelerator trend



Top 500 (June 2013 list)

Countries



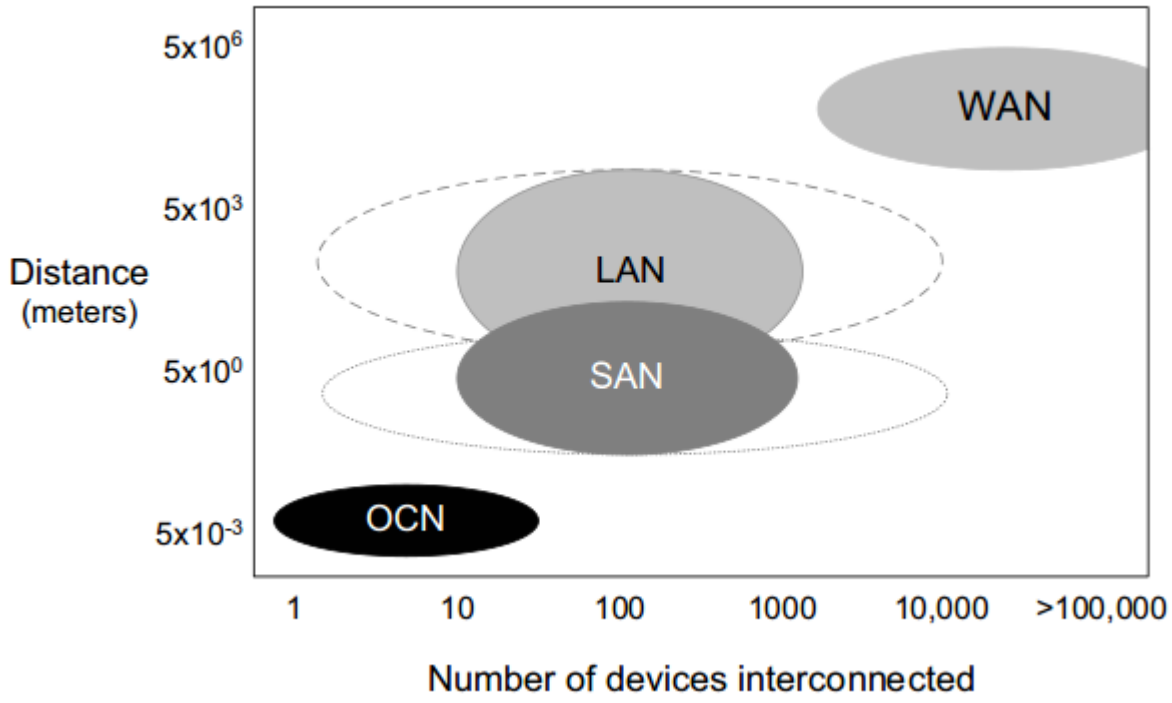
Δίκτυα διασύνδεσης

- Διασυνδέουν δομικές μονάδες ενός σύνθετου συστήματος
- **On-Chip Network (OCN) or Network-on-Chip (NoC):**
  - Caches
  - Processing cores
  - CMPs.
- **System/Storage Area Networks (SAN):**
  - Επεξεργαστές με μονάδες μνήμης
  - Υπολογιστές μεταξύ τους
  - Υπολογιστές με συσκευές αποθήκευσης
- **Local Area Networks (LAN):**
  - Υπολογιστές σε ένα τοπικό δίκτυο
- **Wide Area Networks (WAN):**
  - Υπολογιστές σε οποιοδήποτε σημείο του πλανήτη

Δίκτυα διασύνδεσης

- Διασυνδέουν δομικές μονάδες ενός σύνθετου συστήματος
- **On-Chip Network (OCN) or Network-on-Chip (NoC):**
  - Caches
  - Processing cores
  - CMPs.
- **System/Storage Area Networks (SAN):**
  - Επεξεργαστές με μονάδες μνήμης
  - Υπολογιστές μεταξύ τους
  - Υπολογιστές με συσκευές αποθήκευσης
- **Local Area Networks (LAN):**
  - Υπολογιστές σε ένα τοπικό δίκτυο
- **Wide Area Networks (WAN):**
  - Υπολογιστές σε οποιοδήποτε σημείο του πλανήτη

Δίκτυα διασύνδεσης



Κρίσιμες μετρικές για την αξιολόγηση ενός δικτύου διασύνδεσης

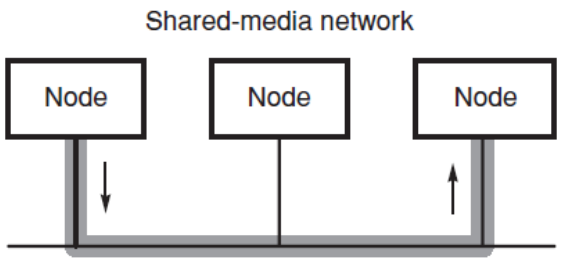
- **Επίδοση:**
  - **Latency**: Χρόνος που απαιτείται για να φτάσει το πρώτο byte πληροφορίας από τον αποστολέα στον παραλήπτη
  - **Bandwidth:** Ο ρυθμός με τον οποίο μεταδίδεται η πληροφορία
- **Κόστος:**
  - Αριθμός ports στα switches
  - Αριθμός switches
  - Αριθμός συνδέσεων
- **Επεκτασιμότητα:** Η δυνατότητα του δικτύου να υποστηρίξει επέκταση σε μεγαλύτερο αριθμό διασυνδεόμενων μονάδων

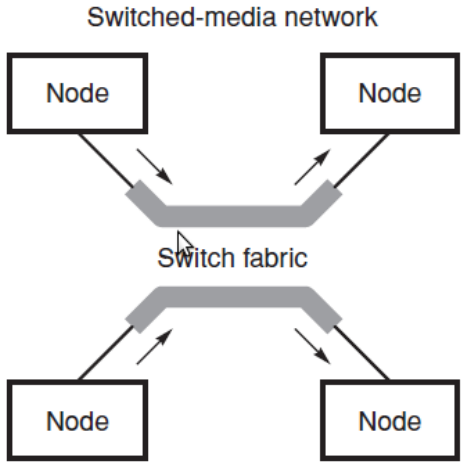
Χαρακτηριστικά συνδεσμολογιών

- **Βαθμός κόμβου (node degree) d:** αριθμός συνδέσμων σε ένα κόμβο
  - πρέπει να είναι
    - » μικρός (λόγω κόστους)
    - » σταθερός (για επεκτασιμότητα)
- **Διάμετρος δικτύου D:** μέγιστο ελάχιστο μονοπάτι μεταξύ δύο οποιωνδήποτε κόμβων
  - Όσο μικρότερη, τόσο καλύτερη η χειρότερη περίπτωση επικοινωνίας
- **Εύρος τομής (bisection width) b:** ο ελάχιστος αριθμός ακμών που κόβουμε, χωρίζοντας το δίκτυο στα δύο
  - Αποτελεί ένα καλό δείκτη του μέγιστου εύρους ζώνης επικοινωνίας σε ένα δίκτυο

Κατηγορίες δικτύων

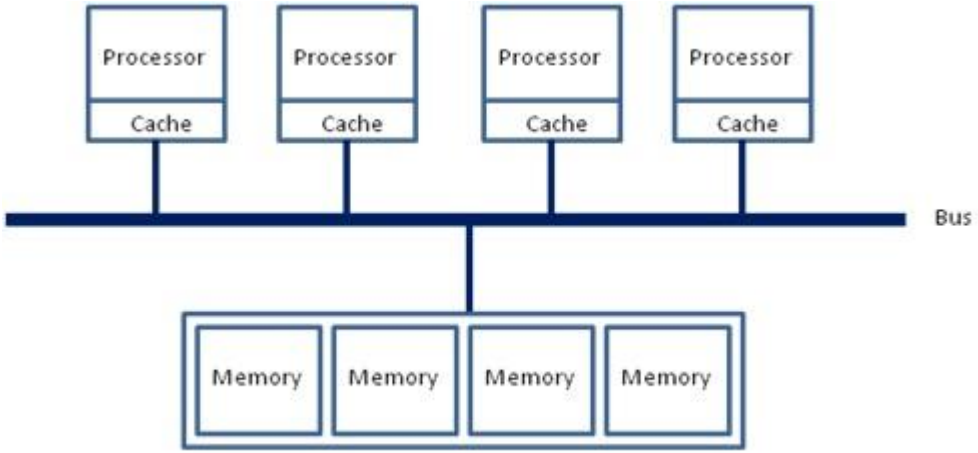
- **Shared-media networks:** Το μέσο είναι διαμοιραζόμενο από όλους τους κόμβους, π.χ.
  - Δίαυλος (bus) σε μονοεπεξεργαστικά και πολυεπεξεργαστικά συστήματα
  - Το παραδοσιακό Ethernet
- **Switched-media networks:** Υπάρχουν διακοπτόμενα μονοπάτια που μπορούν να υποστηρίξουν την ταυτόχρονη επικοινωνία ανάμεσα σε διαφορετικά ζεύγη κόμβων



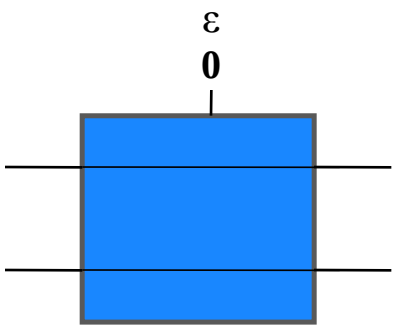


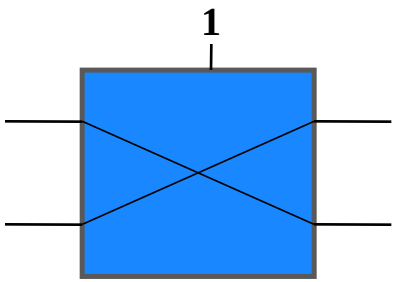
Διάδρομος (Bus)

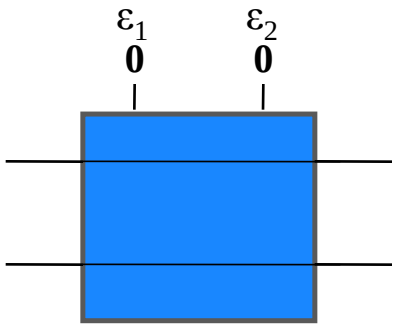
- Παραδοσιακός τρόπος διασύνδεσης σε ένα NoC
- Απλή υλοποίηση με χαμηλό κόστος
  - Data, address, control buses
  - Διαιτησία (Arbitration)
- Υποστηρίζει εύκολα broadcast και multicast
- Εύκολη υλοποίηση cache coherence με snooping
- **Αλλά:** δεν είναι επεκτάσιμος (τυπικά λίγες δεκάδες στοιχείων)
  - Περιορισμένο συνολικό bandwidth
  - Δυσκολία στη διαιτησία

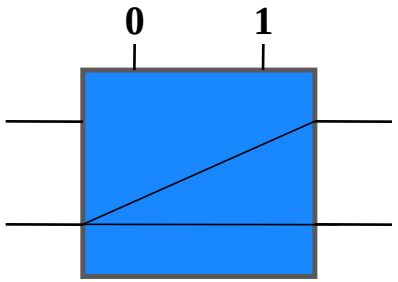


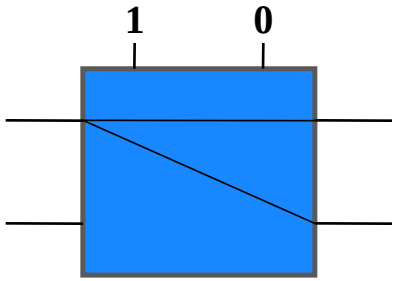
Διακόπτες

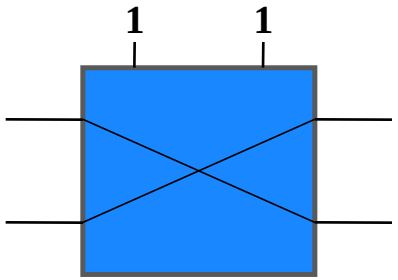








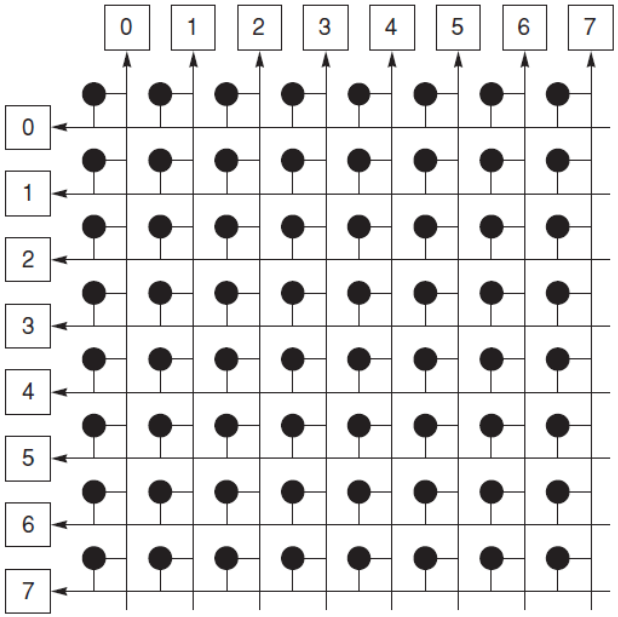




Centralized Switched Networks

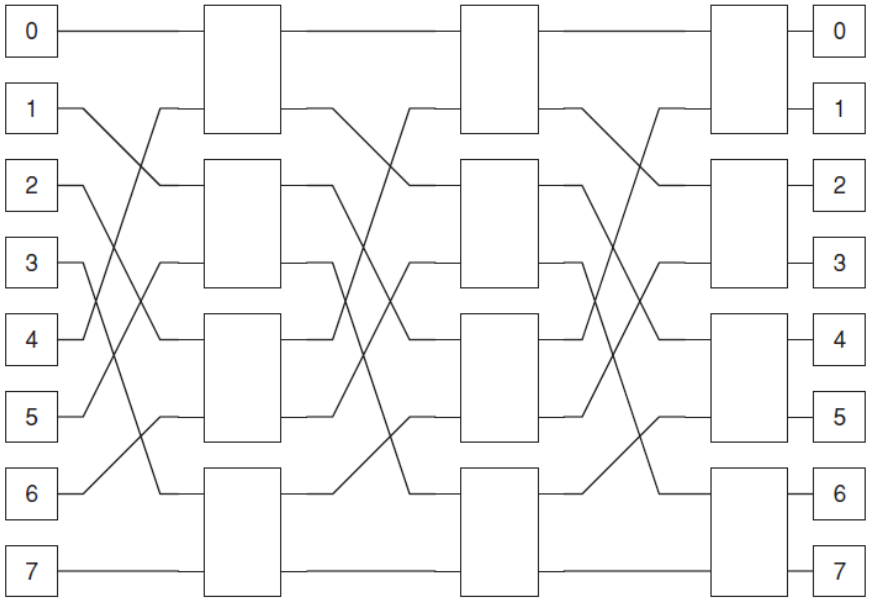
Crossbar Switch

- Απλούστερη, ταχύτερη αλλά και ακριβότερη λύση για τη διασύνδεση N στοιχείων
- Απαιτεί $\mathbf{N^2}$ διακόπτες, δεν κλιμακώνει λόγω κόστους
- Χρησιμοποιείται σε NoC για τη διασύνδεση λίγων δεκάδων στοιχείων



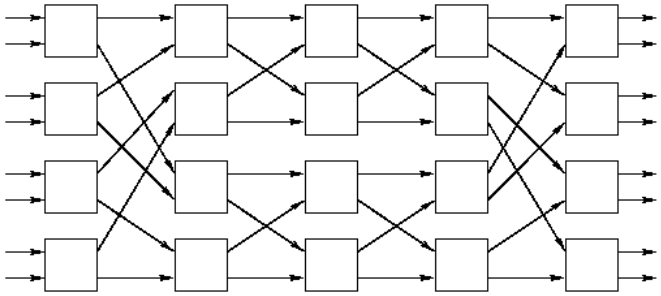
Centralized Switched Networks
Multistage Interconnection Networks

- Διασυνδέουν N στοιχεία με τη χρήση πολυεπίπεδων διακοπτών
- Αν χρησιμοποιηθούν k × k διακόπτες, χρειάζονται $\log_k N$ στάδια με N/k διακόπτες ανά στάδιο (σύνολο **$N/k \log_k N$** διακόπτες)
- Ανάλογα με τη διασύνδεση των διακοπτών έχουν προκύψει διαφορετικά δίκτυα που ανταποκρίνονται σε διαφορετικά patterns επικοινωνίας

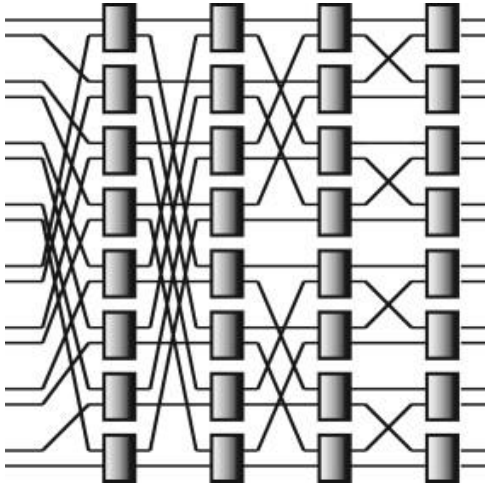


Centralized Switched Networks

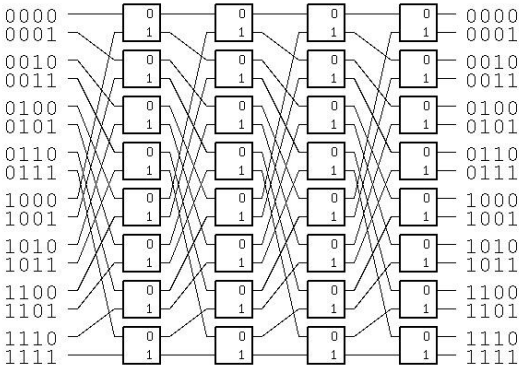
Multistage Interconnection Networks



Benes network



Butterfly network

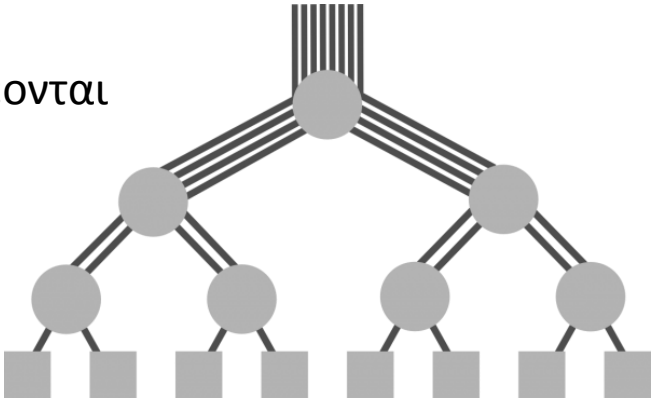


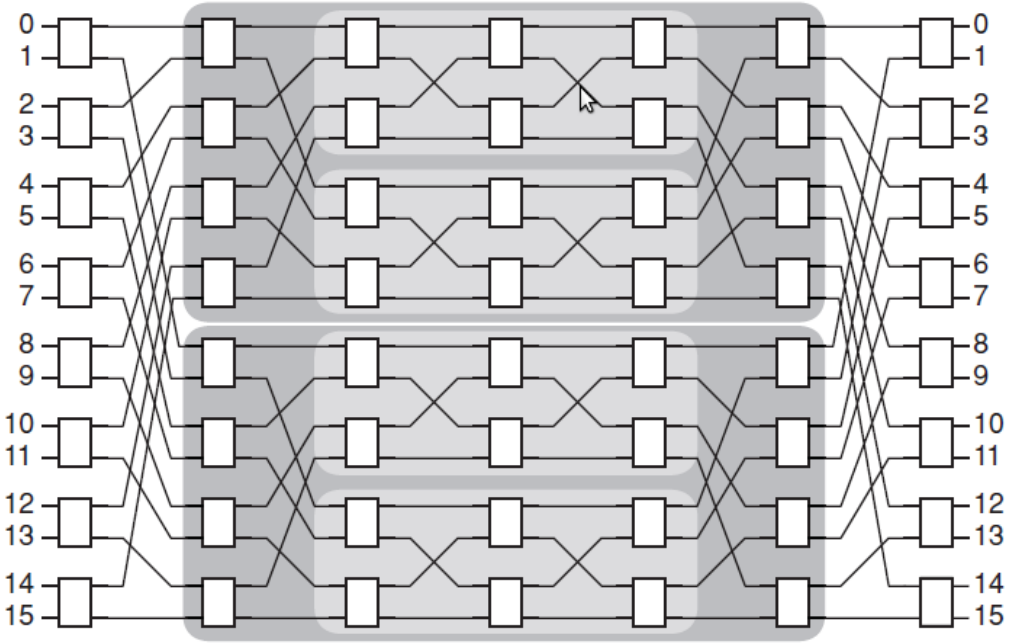
Omega network

Centralized Switched Networks

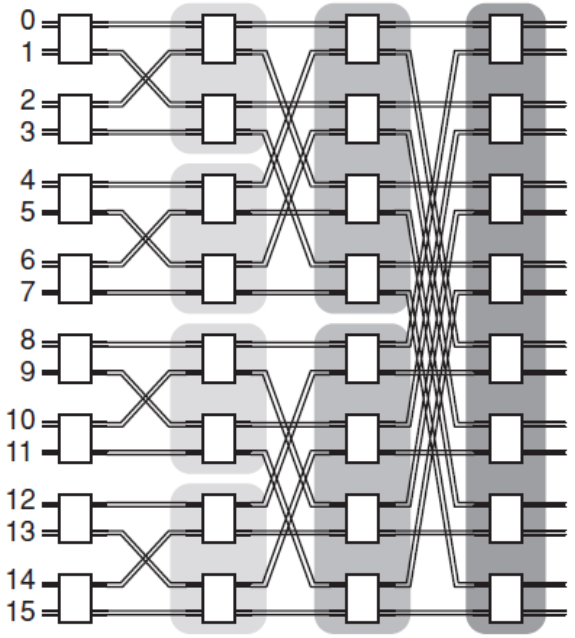
Fat tree

- Τα φύλλα του δέντρου είναι τα στοιχεία που διασυνδέονται
- Οι εσωτερικοί κόμβοι είναι διακόπτες
- Χρησιμοποιείται κατά κόρον σε SANs και κυρίως σε Supercomputers (Infiniband, Myrinet, κλπ)





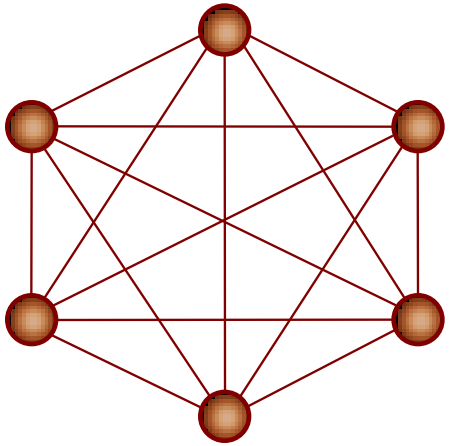
Benes network



Folded Benes network

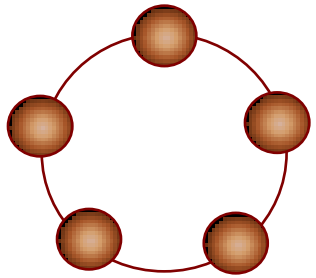
Distributed Switched Networks
Fully connected

- N κόμβοι
- N(N-1)/2 σύνδεσμοι
- Βαθμός κόμβου d=N-1
- Διάμετρος D=1
- Εύρος τομής $b=(N/2)^2$
- Είναι συμμετρικό



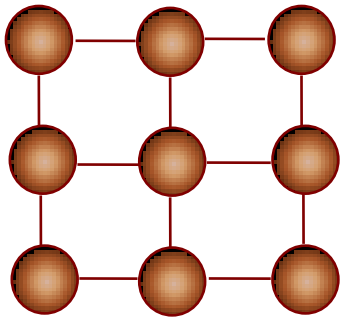
Distributed Switched Networks
Ring

- Ν κόμβοι
- Ν σύνδεσμοι
- Βαθμός κόμβων d=2
- Διάμετρος: Ν/2
- Εύρος τομής b=2
- Είναι συμμετρικό



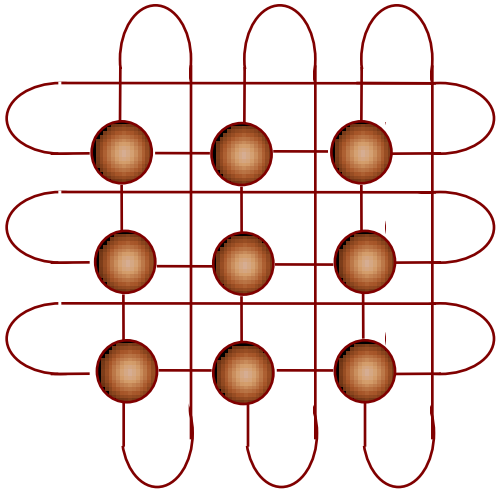
Distributed Switched Networks
Mesh

- $N=n^k$ κόμβοι
- k-διάστατο mesh με n κόμβους ανά διεύθυνση
- βαθμός κόμβου d=2k
- διάμετρος δικτύου D=k(n-1)
- Για ένα 2-διάστατο mesh:
  - $N=n^2$ κόμβοι
  - $2N-2n=2n^2-2n$ σύνδεσμοι
  - Βαθμός εσωτερικών κόμβων d=4
  - Διάμετρος D=2(n-1)
  - Εύρος τομής b=n
  - Δεν είναι συμμετρικό



Distributed Switched Networks
Torus

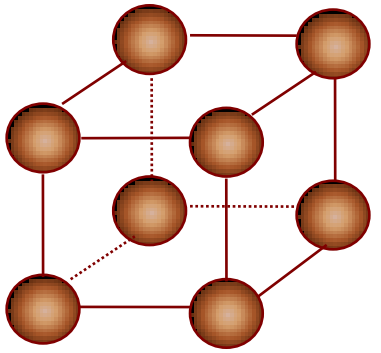
- Υποδιπλασιάζεται η διάμετρος σε σχέση με το Mesh
- για έναν n×n δυαδικό torus (k=2):
  - $N=n^2$ κόμβοι
  - 2N σύνδεσμοι
  - βαθμός κόμβου d=4
  - Διάμετρος $D = 2\left\lfloor \frac{n}{2} \right\rfloor$
  - Εύρος τομής 2n
  - Είναι συμμετρικό

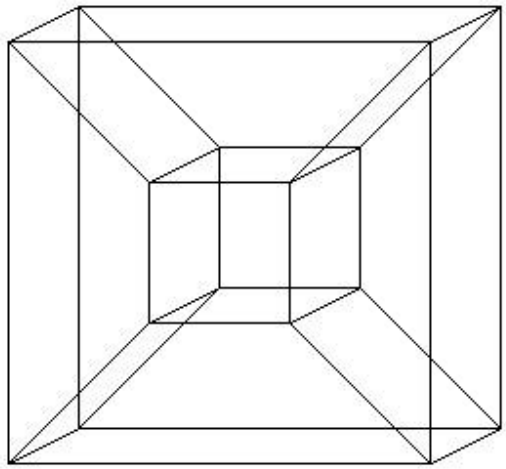


Distributed Switched Networks

Hypercube

- $N=2^n$ κόμβοι
- $nN/2$ σύνδεσμοι
- Βαθμός κόμβου $d=n$
- Διάμετρος $D=n$
- Εύρος τομής $b=N/2$
- Είναι συμμετρικό
- Άμεσος προσδιορισμός διαδρομής





Dragonfly topology (Aries interconnect)

- Ιεραρχικό δίκτυο
- Πλήρες δίκτυο ανάμεσα στις ομάδες
- Οποιαδήποτε επιλογή εντός της ομάδας (προτείνεται butterfly)

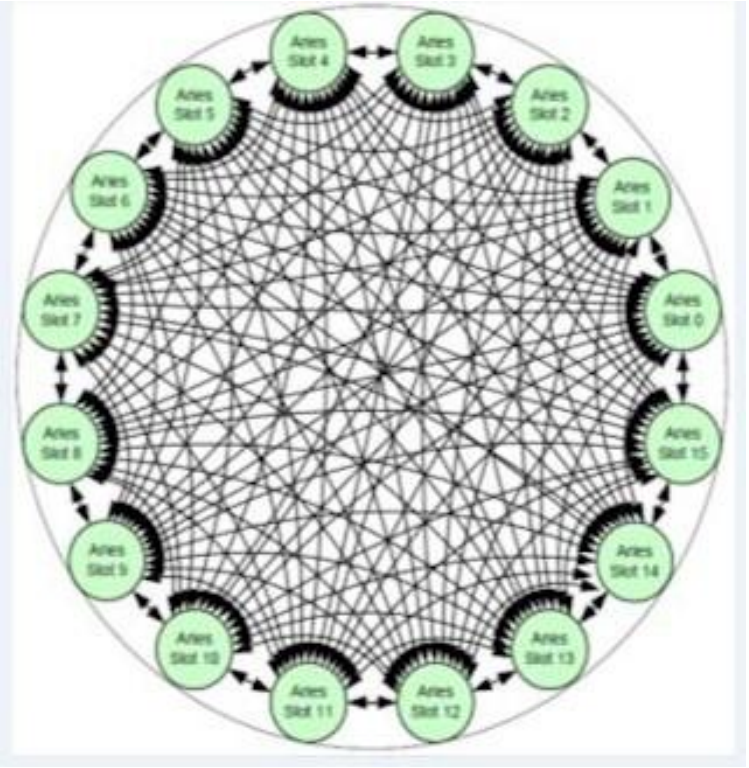
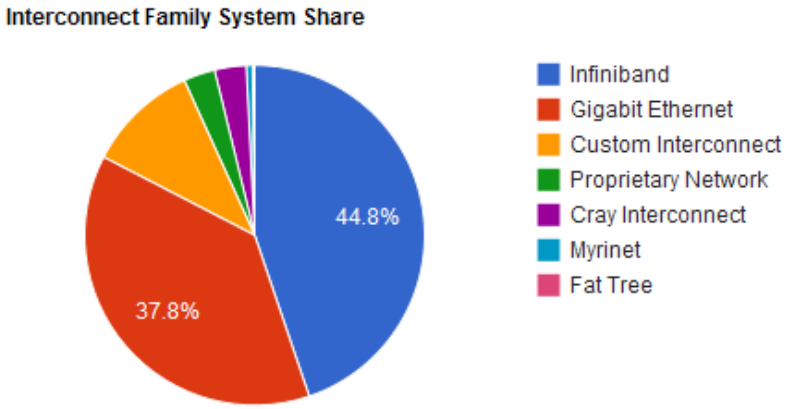
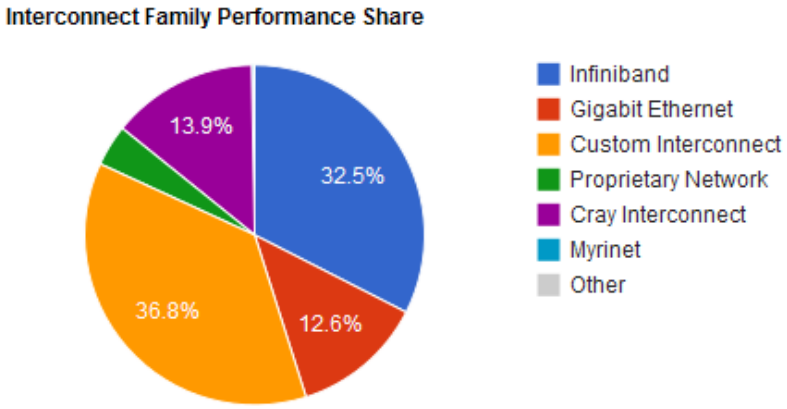


Image taken from:
http://www.theregister.co.uk/Print/2012/11/08/cray_cascade_xc30_supercomputer/

Δίκτυα διασύνδεσης στους υπερυπολογιστές
Top500, November 2012

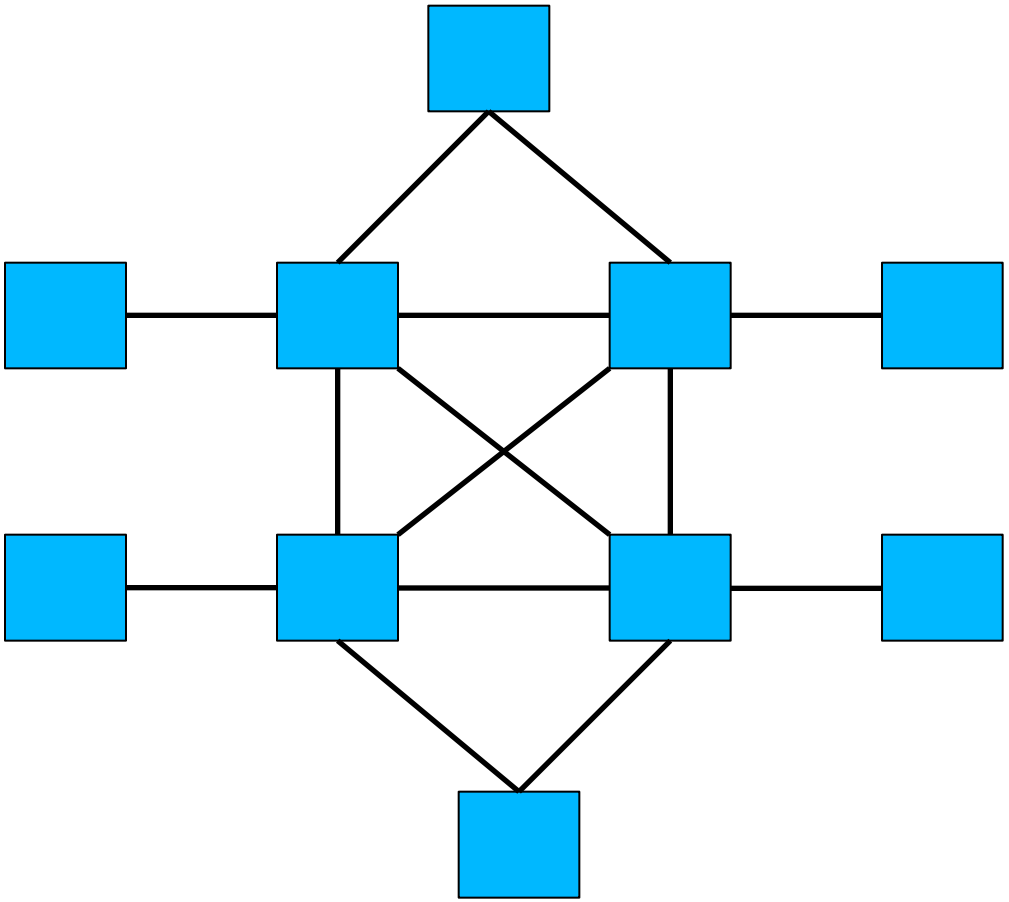




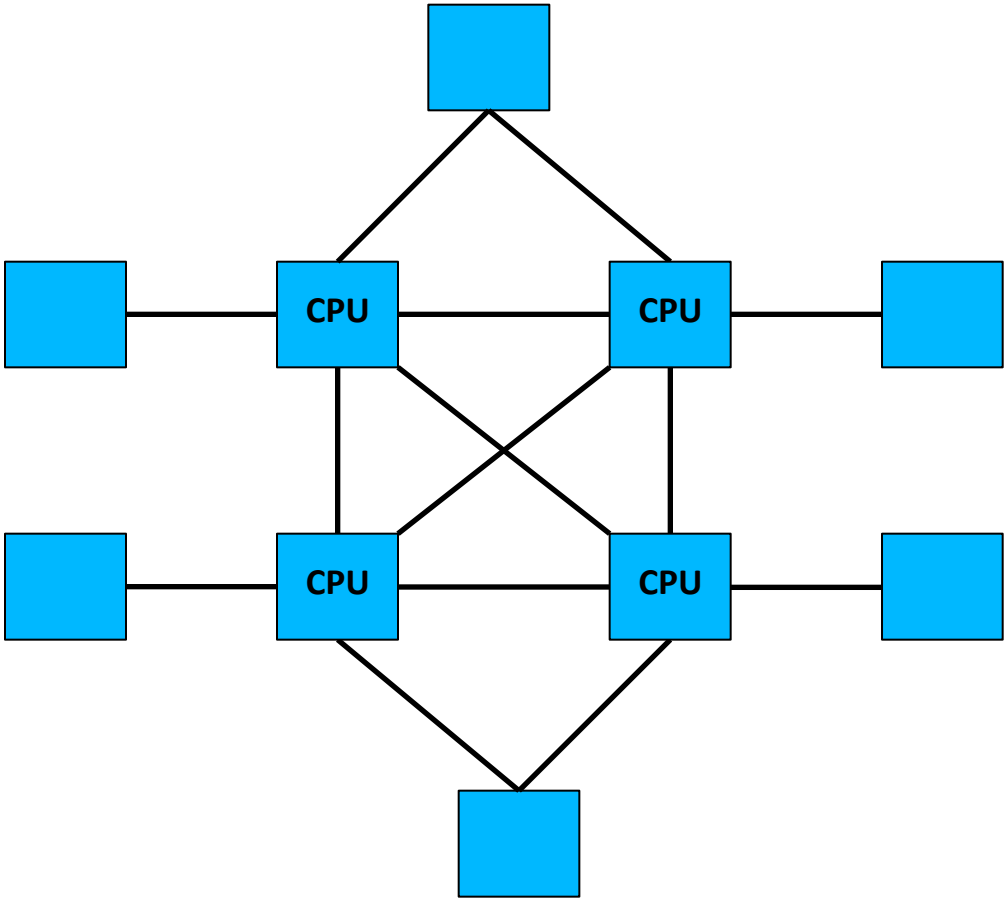
Supercomputers

- **Tianhe-2:** Fat tree
- **BlueGene/Q :** 5D torus
- **BlueGene/P :** binary tree, 3D torus
- **K computer:** 6D torus
- **Infiniband configuration:** Fat tree
- **Cray Gemini interconnect:** 3D torus
- **Cray Aries interconnect:** Dragonfly
- **Historical note (1987):** Connection Machine CM-2, 8192 nodes, hypercube

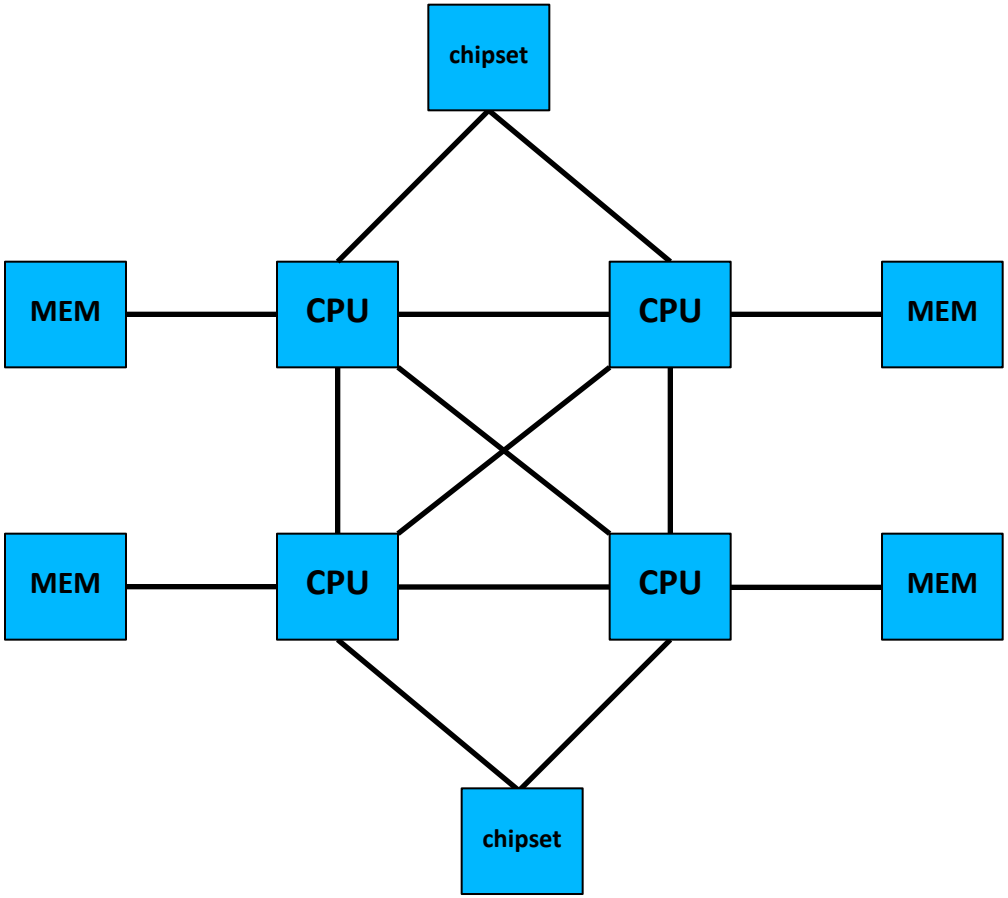
Δίκτυα διασύνδεσης



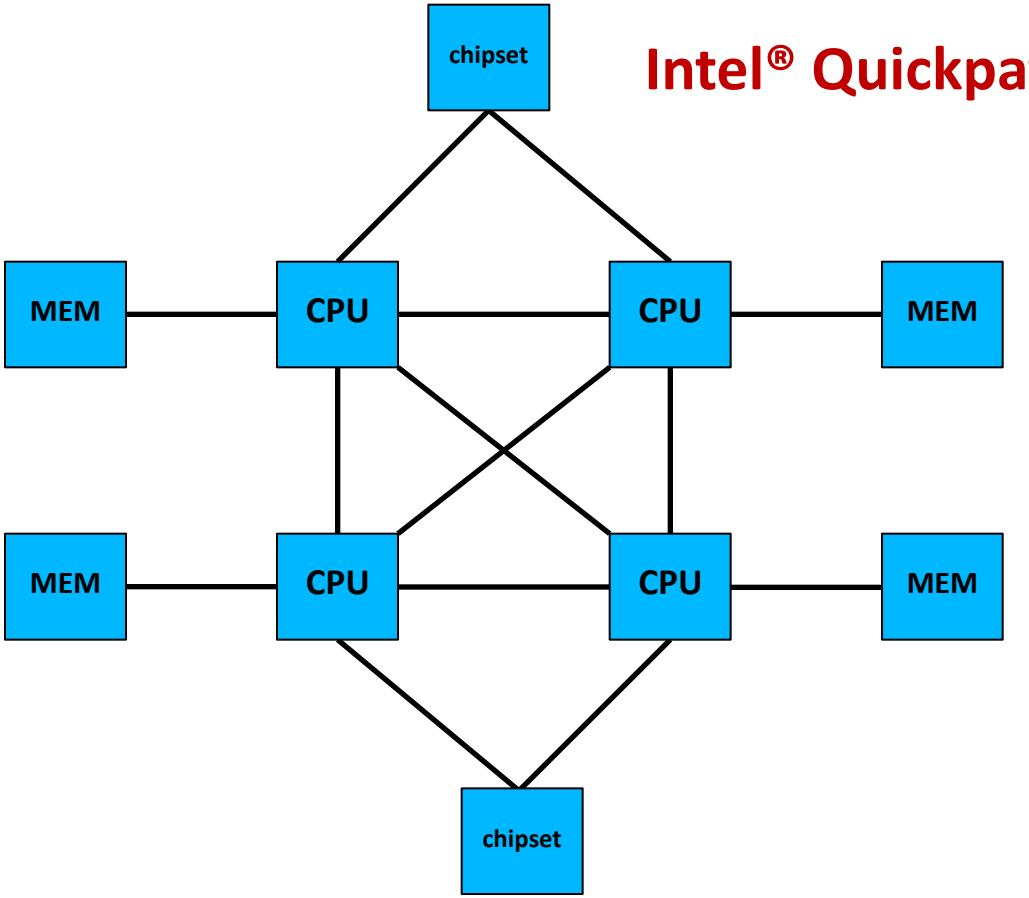
Δίκτυα διασύνδεσης



Δίκτυα διασύνδεσης



Δίκτυα διασύνδεσης



Intel® Quickpath Inteconnect

Intel® Quickpath Inteconnect

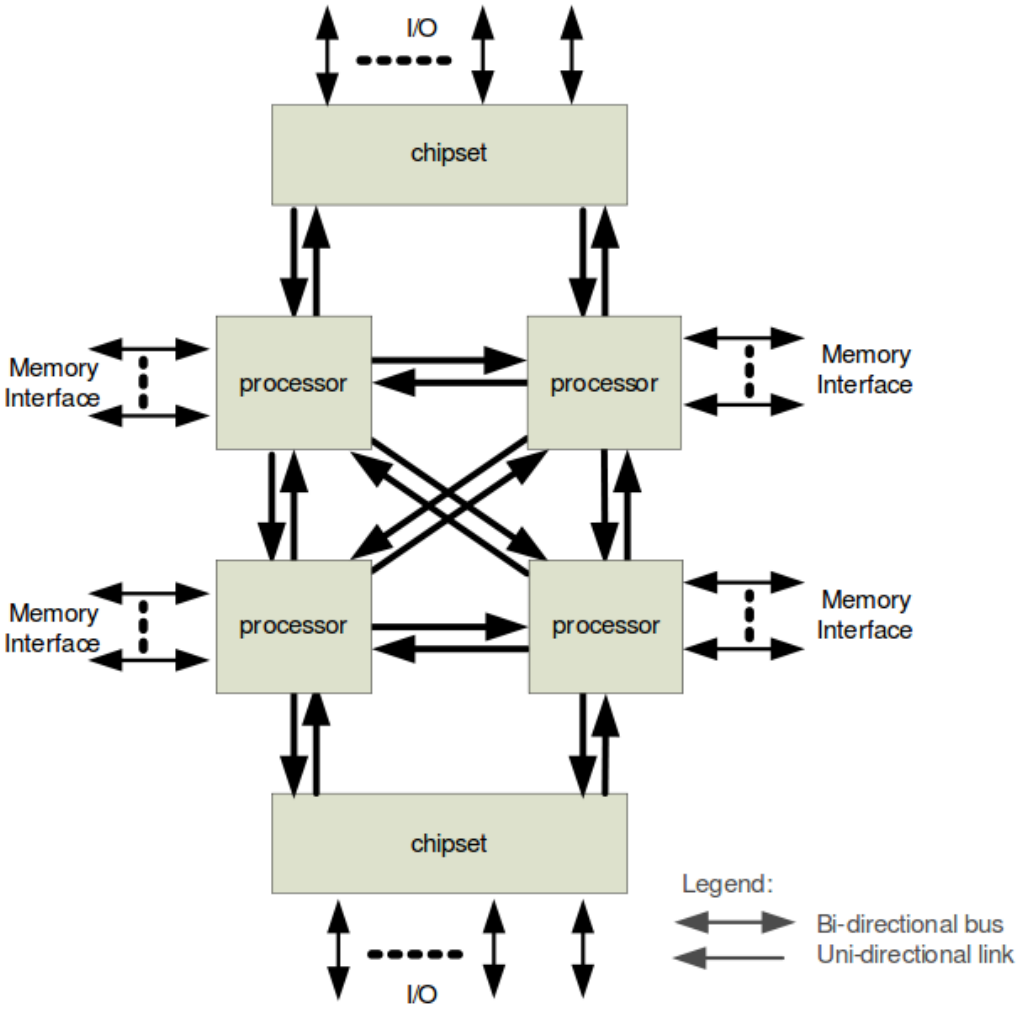


Image taken from: An Introduction to the Intel® QuickPath Interconnect:
http://www.intel.com/content/www/us/en/io/quickpath-technology/quick-path-interconnect-introduction-paper.html

Λίγη διαφήμιση…

- Συστήματα Παράλληλης Επεξεργασίας (9ο Εξάμηνο)
- Αντικείμενο μαθήματος:
  - Αρχιτεκτονικές Παράλληλων Συστημάτων
  - Δίκτυα διασύνδεσης
  - Μοντελοποίηση επίδοσης
  - Σχεδιασμός παράλληλων προγραμμάτων
  - Υλοποίηση παράλληλων προγραμμάτων
  - Συγχρονισμός
  - Χρονοδρομολόγηση
  - Εφαρμογές

Λίγη διαφήμιση…

- Εργαστηριακές ασκήσεις:
  - Προγραμματισμός για συστήματα μεγάλης κλίμακας (Message Passing Interface – MPI)
  - Προγραμματισμός για πολυπύρηνα συστήματα (OpenMP, Cilk, TBBs)
  - Προγραμματισμός για κάρτες γραφικών (CUDA) και επιταχυντές (Xeon Phi)
  - Εκτέλεση εφαρμογών σε:
    - 8-node, Intel Xeon Cluster (16 cores, 32 threads)
    - 32-node, 8-core Intel Harpertown Cluster (256 cores)
    - 24-core Intel Dunnigton
    - 32-core/64-thread Intel Sandy Bridge
    - Fermi GPUs
    - Intel Xeon Phi

Ερευνητικές περιοχές στο CSLab
(και πεδία διπλωματικών εργασιών)

- Αρχιτεκτονική υπολογιστών
  - CMPs
    - *Improving Reliability and Efficiency of Performance Monitoring in Linux*
    - *Characterizing Thread Placement and Thread Priorities in the IBM POWER7 Processor*
  - Ιεραρχίες και τεχνολογίες μνημών
    - Cache organization, partitioning, memory link compression
    - *Μελέτη και Υλοποίηση Αλγορίθμων Καταμερισμού της Διαμοιραζόμενης Ιεραρχίας Μνήμης σε Πολυπύρηνες Αρχιτεκτονικές*
    - *A study of a dynamic placement policy in a NUCA cache*
    - *Memory-Link Compression to overcome the Memory Wall*

Ερευνητικές περιοχές στο CSLab
(και πεδία διπλωματικών εργασιών)

- Παράλληλος προγραμματισμός και εφαρμογές
  - Transactional memory
    - » *Μελέτη και Αξιολόγηση του TSX στους Haswell επεξεργαστές της Intel: Παραλληλοποίηση Red Black Tree*
    - » *Extension of a Recognition, Mining and Synthesis Benchmark Suite for Transactional Memory*
    - » *Αλγόριθμοι διάσχισης γράφων σε σύγχρονες πολυπύρηνες αρχιτεκτονικές*
  - Παραλληλοποίηση και βελτιστοποίηση εφαρμογών
    - » Επιστημονικοί υπολογισμοί, Αλγόριθμοι γράφων, Δομές δεδομένων, κλπ
    - » *Τεχνικές Βελτιστοποίησης για παράλληλο λογισμικό μεγάλης κλίμακας*
    - » *SparseX: Βιβλιοθήκη για τον πολλαπλασιασμό αραιού πίνακα με διάνυσμα σε πολυπύρηνες αρχιτεκτονικές*
    - » Σε CMPs, SMTs, NUMA, Accelerators (GPUs, Xeon Phi), large-scale systems
    - » *Υλοποίηση και Βελτιστοποίηση του Αλγορίθμου Smith - Waterman σε Πολυπύρηνους Επεξεργαστές και Πολυνηματικούς Επεξεργαστές Γραφικών*
    - » *Implementation and optimization of algorithms on multicore and manycore architectures*

Ερευνητικές περιοχές στο CSLab
(και πεδία διπλωματικών εργασιών)

- Παράλληλα συστήματα
  - Συστήματα χρόνου εκτέλεσης για παράλληλο προγραμματισμό / Ανάθεση πόρων σε διεργασίες
  - Δρομολόγηση εργασιών σε πολυπύρηνες αρχιτεκτονικές
    - *Power Aware Scheduling on Multicore Systems*
    - *Δρομολόγηση παράλληλων εφαρμογών σε πολυπύρηνα συστήματα*
    - *Αλγόριθμοι δρομολόγησης εφαρμογών σε επίπεδο υλικού για πολυνηματικές αρχιτεκτονικές*
  - Μοντέλα πρόβλεψης επίδοσης
    - *Parallel program scaling prediction*

Ερευνητικές περιοχές στο CSLab
(και πεδία διπλωματικών εργασιών)

- Λογισμικό συστήματος
  - Βελτιστοποίηση επικοινωνίας / Εικονικές μηχανές
    - *V4Vsockets: Μηχανισμός αποδοτικής ενδο-επικοινωνίας εικονικών μηχανών χαμηλής επιβάρυνσης*
    - *Πειραματική αποτίμηση και βελτιστοποίηση της επικοινωνίας εικονικών μηχανών που συνυπάρχουν στο ίδιο φυσικό μηχάνημα*
  - Διαμοιρασμός επιταχυντών σε περιβάλλον νέφους
    - *Κατανεμημένο Σύστημα Διαχείρισης Εργασιών Απομακρυσμένης Εκτέλεσης Κώδικα Για Επιταχυντές Γραφικών Σε Συστοιχίες Υπολογιστών*