IBM LinuxONE

The IBM z15 Chip is Awesome

A LinuxONE III Story

Monte Bauman

LinuxONE Technical Support Specialist

IBM Columbus

mbauman@us.ibm.com

June 2021

IBM LinuxONE

Let’s Get Started…

INTRODUCTION

Abstract

The z15 chips is awesome. What follows are 9.1 Billion reasons why.

Moore's Law used to be awesome…. and sort-of still is.

- "Atoms ain't getting smaller" … and "the speed of light ain't getting faster"
  - And hence Moore's Law is "underperforming"
- But the demand for ever bigger and faster computing systems remains
  - Consumers know what they want … they just don't know what they are asking for
- Computer scientists and engineers must meet design challenges in new ways
  - Bigger problems requires bigger thinking
- The IBM Z chip and the platforms it powers (IBM Z and LinuxONE) have been "stepping up" to the challenge and this presentation will illustrate strides made in hardware and software to make the platform excel

The numbers, they do tell a tale. This talk will examine the evolution of CMOS microprocessor technology on the mainframe platform. We will examine the intriguing relationship between MIPS and MHz, and between transistors, chips, and cores. We will observe the past and current effects of "Moore's Law", and we will look at equations (yes, equations!) suggesting how to combat computer science physics. Microprocessor engineers have quite a challenge ahead, what will they think of next?

Monte Bauman … mbauman@us.ibm.com

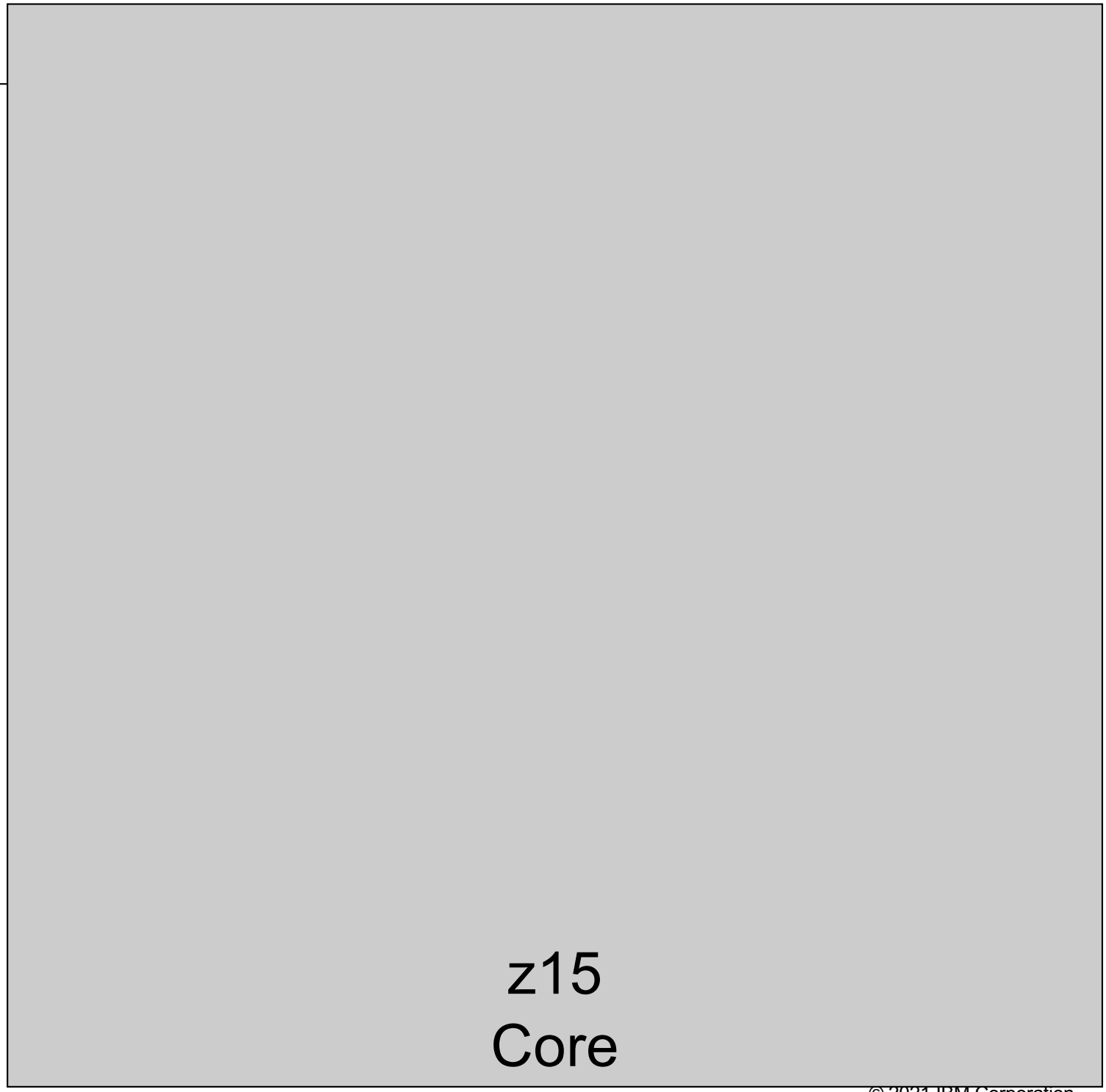
IBM Technical Support Specialist - helping clients build risk mitigated operationally excellent economic server solutions.

- Monte Bauman helps client's design and deploy risk mitigated operationally excellent super-hyperconverged server solutions using IBM LinuxONE technologies provided by IBM and by the LinuxONE partner ecosystem.

- Monte Bauman, Certified IT Specialist and z Systems Technical Support Professional, began a career in IT in 1983 hiring into the IBM Glendale Lab where he assisted a development team to build and test mid-range mainframe servers (4300s, 9370s, 9221s). Monte was given a Division Award for work on STX/370, a diagnostic operating system. Monte Bauman became a Large Systems Systems Engineer (SE) in January of 1991 in Columbus Ohio working with enterprise clients specializing in 9021 systems support, enterprise printing support (3800/3900), and MVS support. In the mid 90's Monte Bauman became a mainframe seller for a time, before returning in late 90's to the role of a regionally designated specialist large systems SE. Monte Bauman became a "new workloads" specialist at the turn of the century, technically supporting customers as they embarked upon emerging technologies on the mainframe, such as Java and Business Intelligence and Linux. The past several years Monte Bauman has been focused on LinuxONE and Fit for Purpose analysis, helping customers understand the fit of z Systems Linux solutions as well as the processes appropriate to repeatedly and effectively map software to hardware meeting IT optimization criteria for requirements and cost.

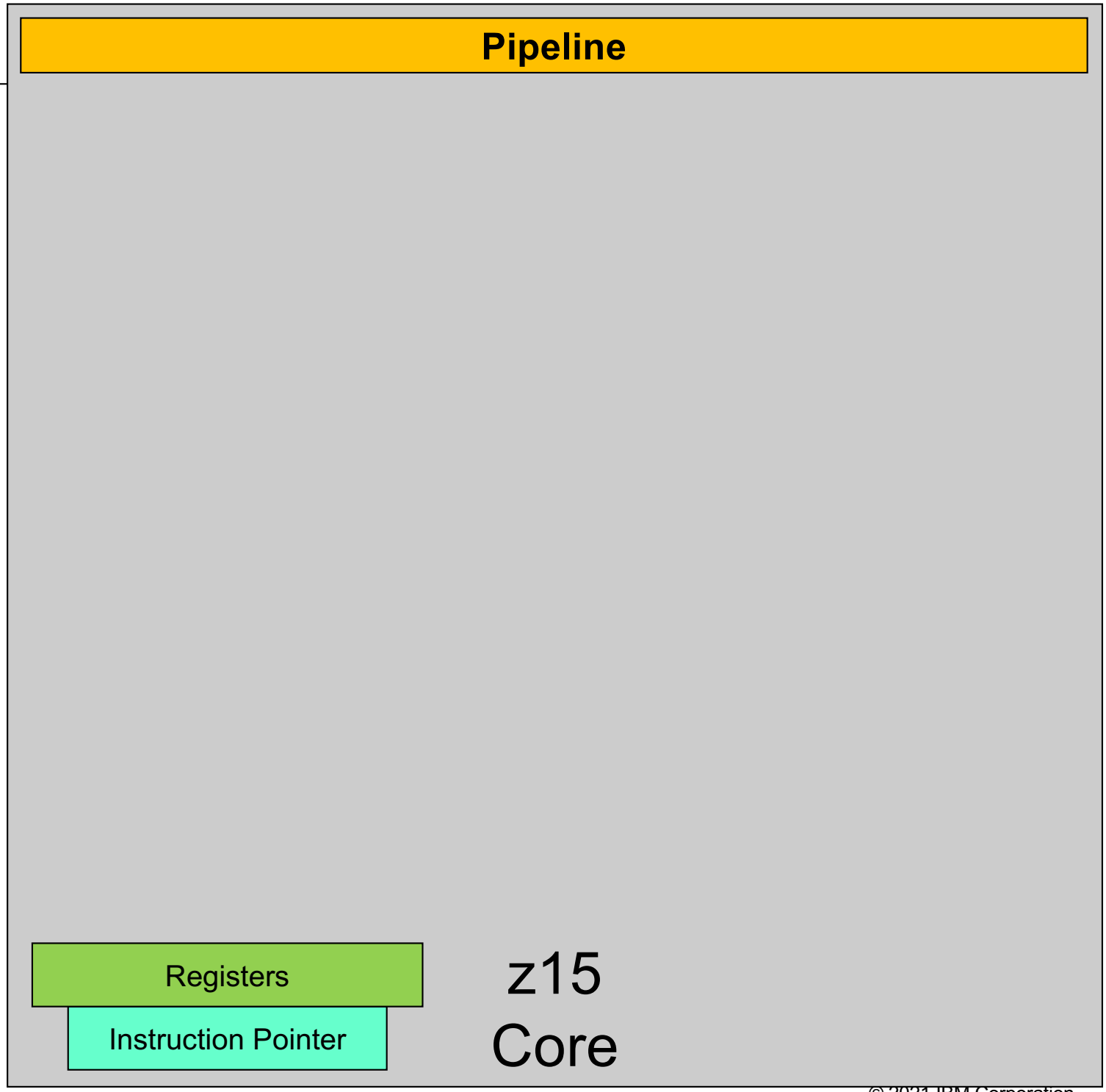
Let’s Get to the Core of this …

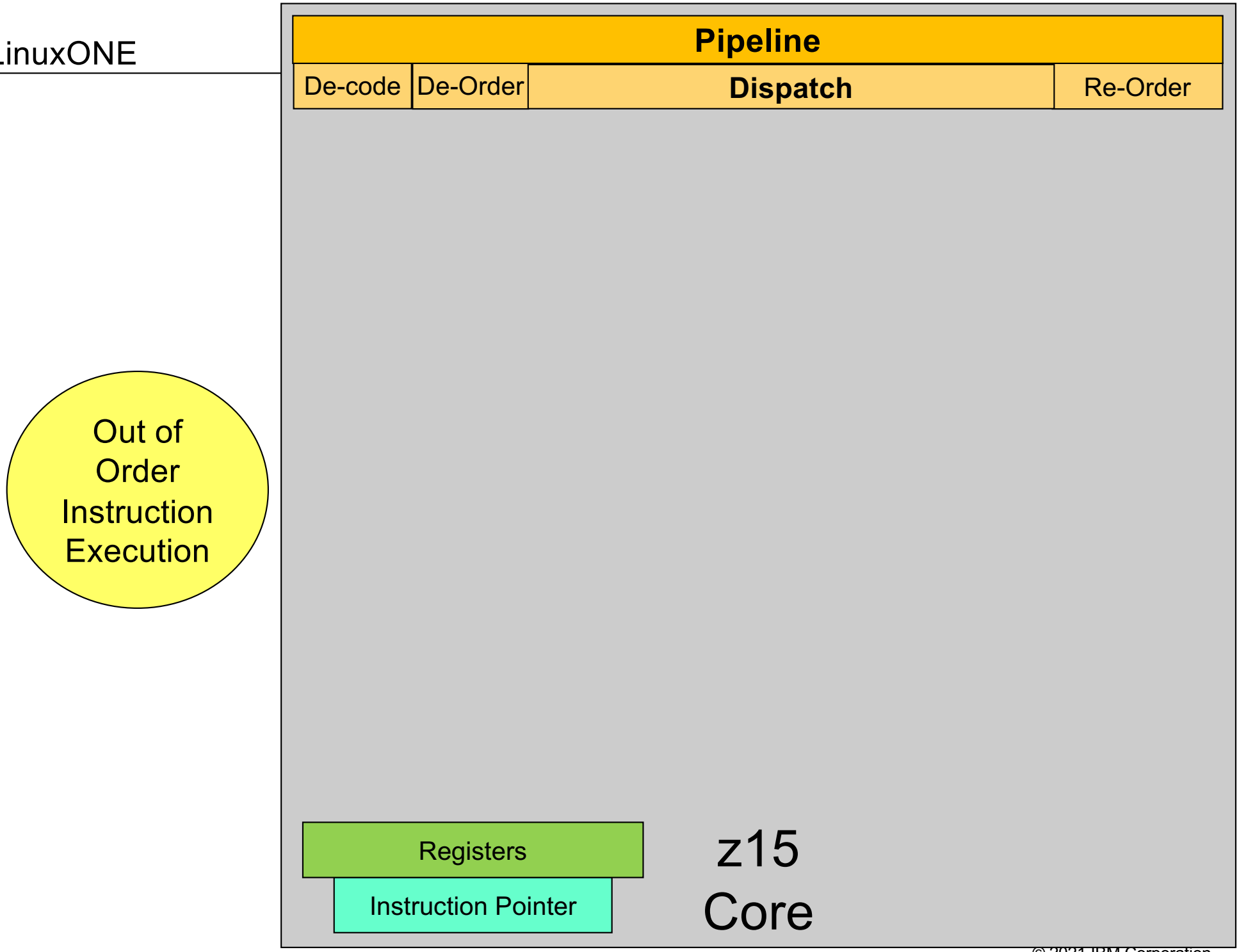
THE Z15 CORE

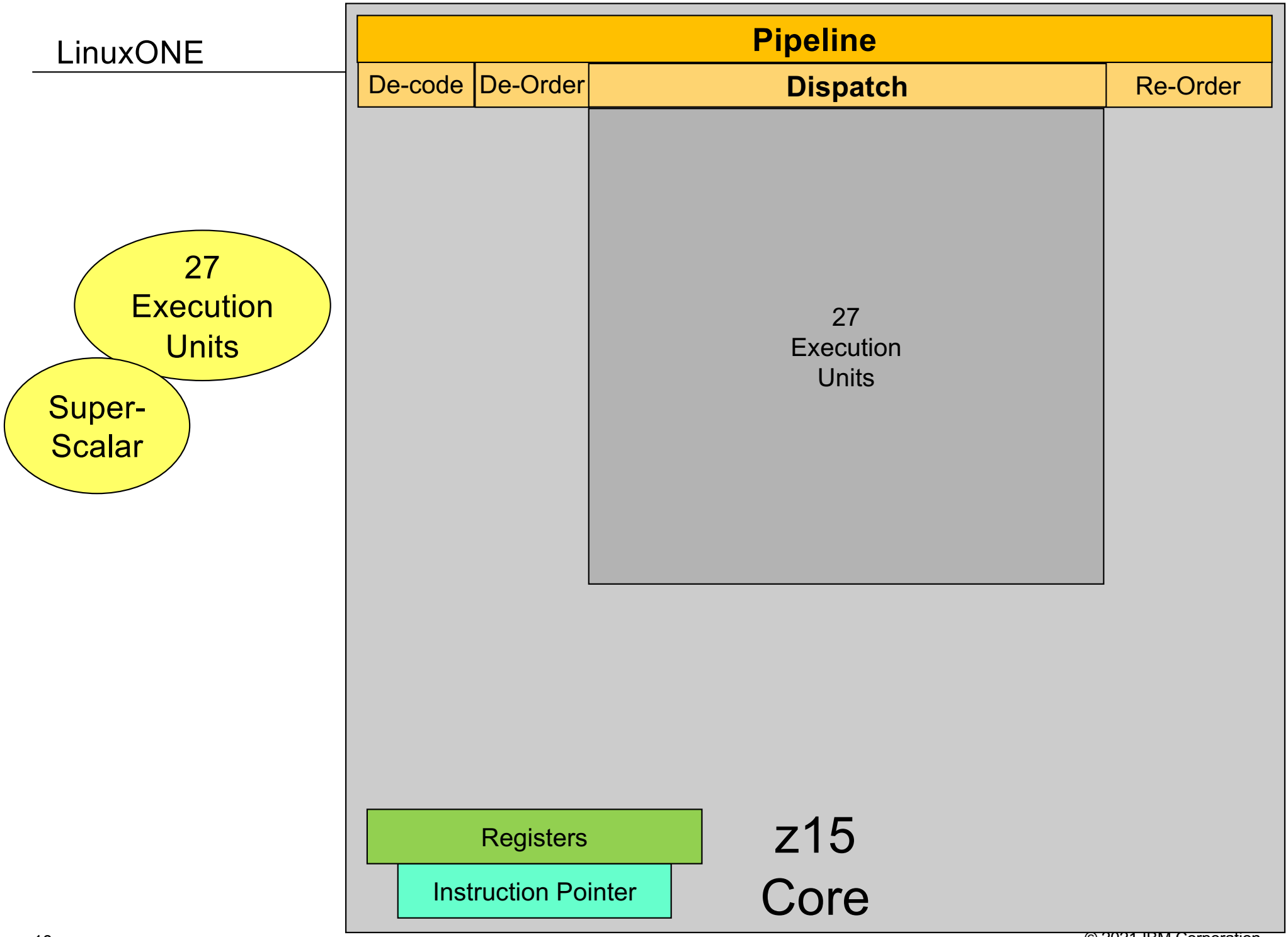


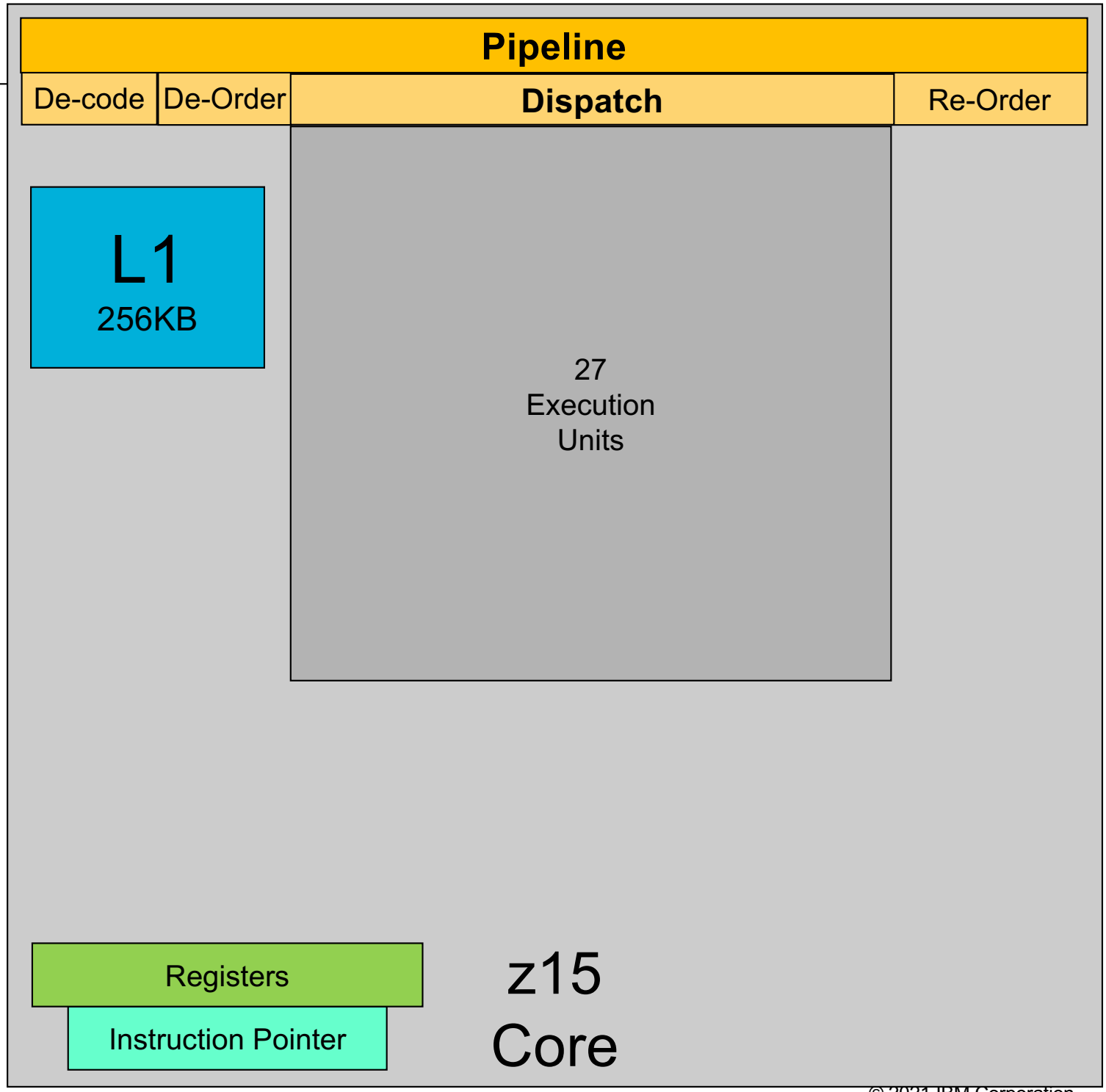
LinuxONE

5.2GHz

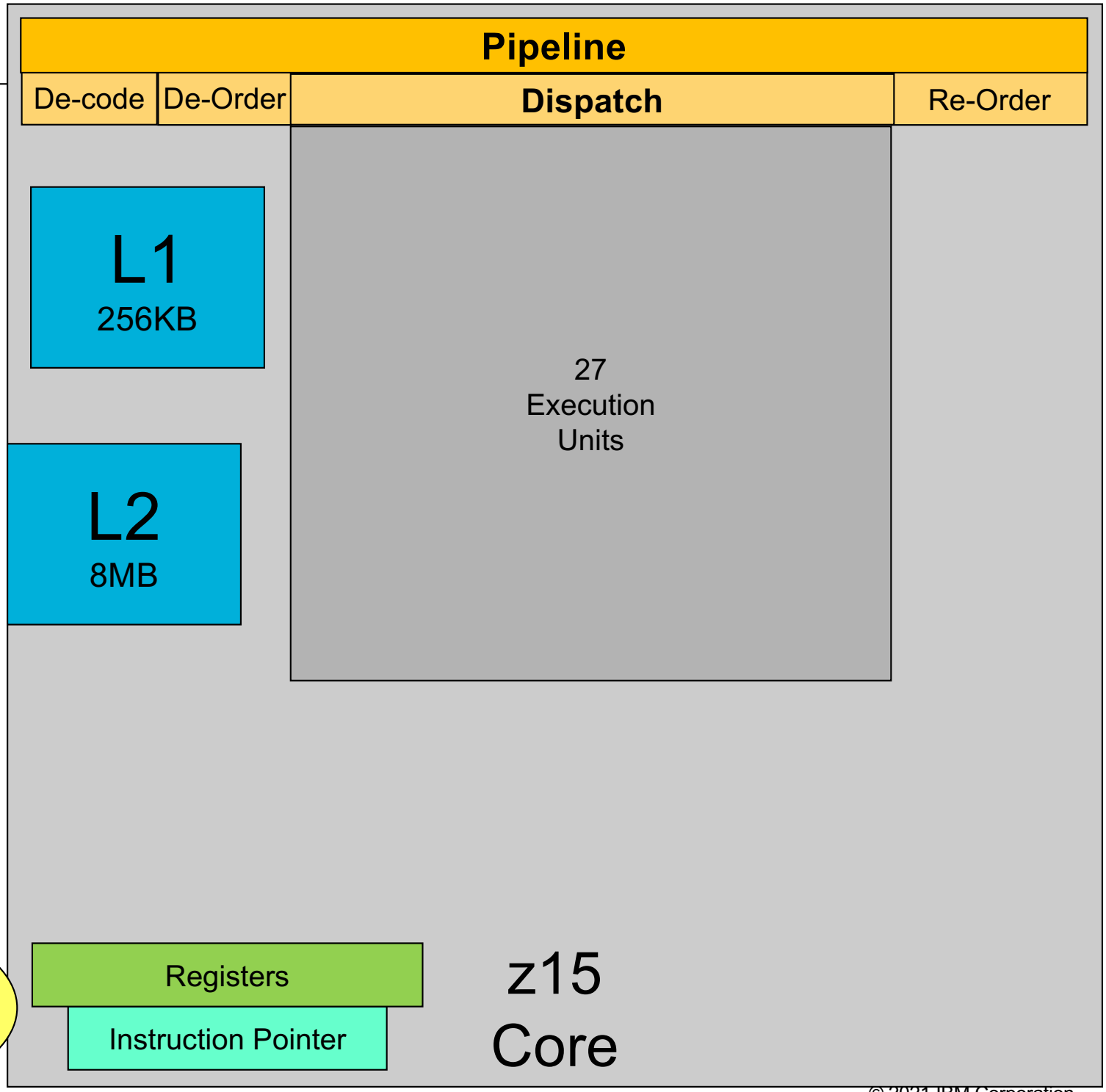








Big & Smart Cache



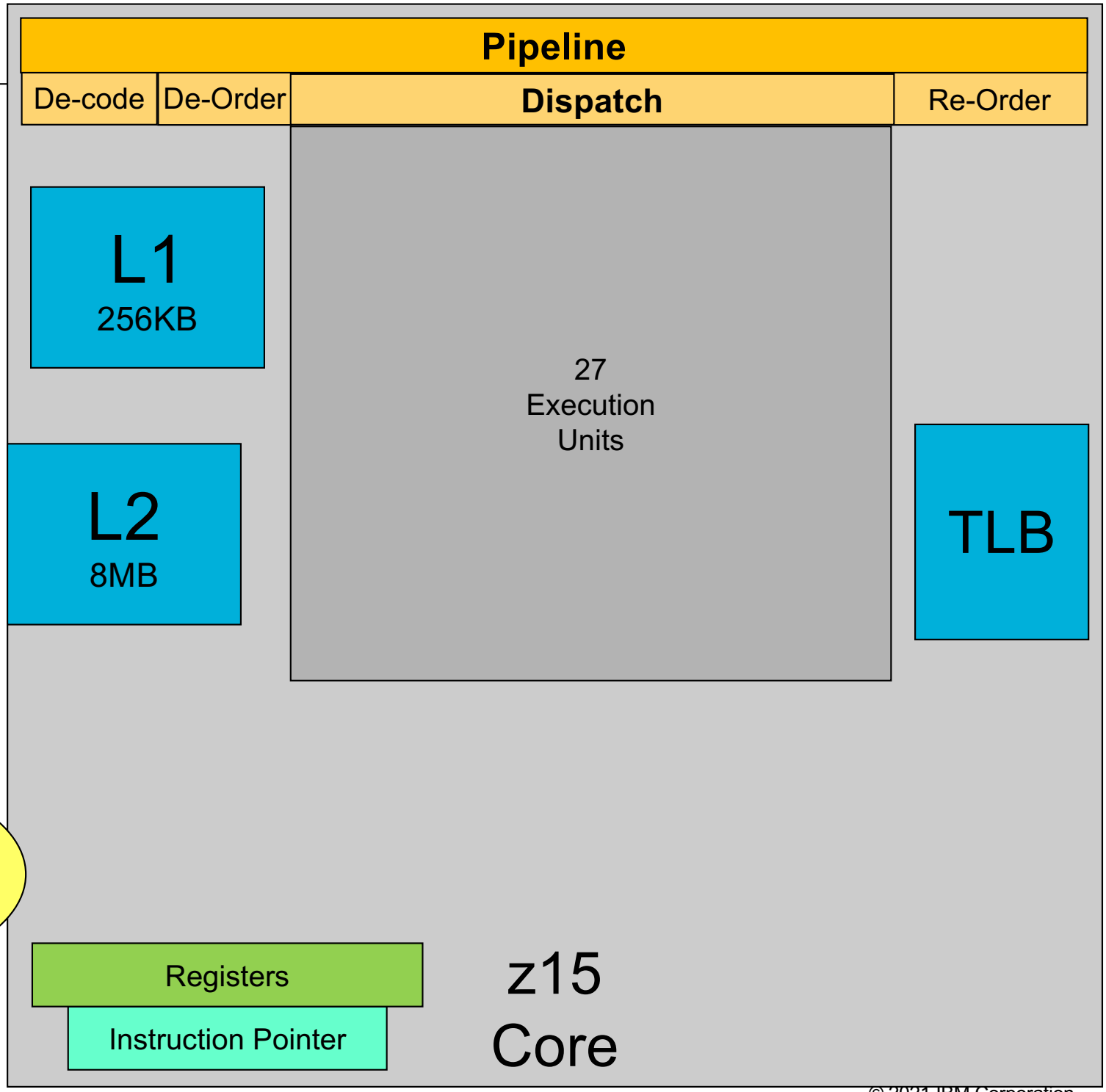
Big & Smart Cache

Big: Instruction Cache

Big: Data Cache

Smart: Sophisticated "cache associativity"

Smart: Software can provide tips to the caches

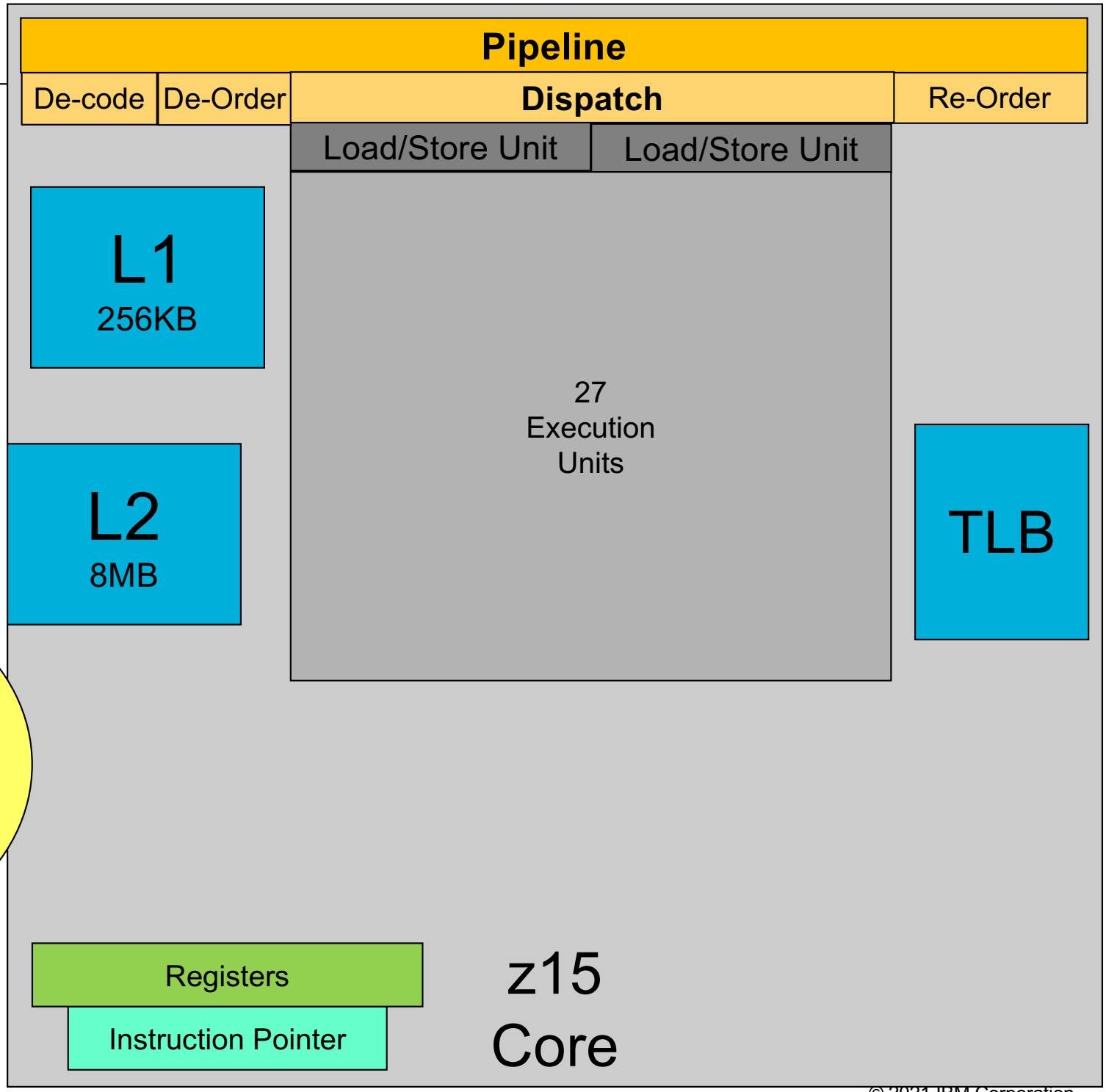


Caching Addresses

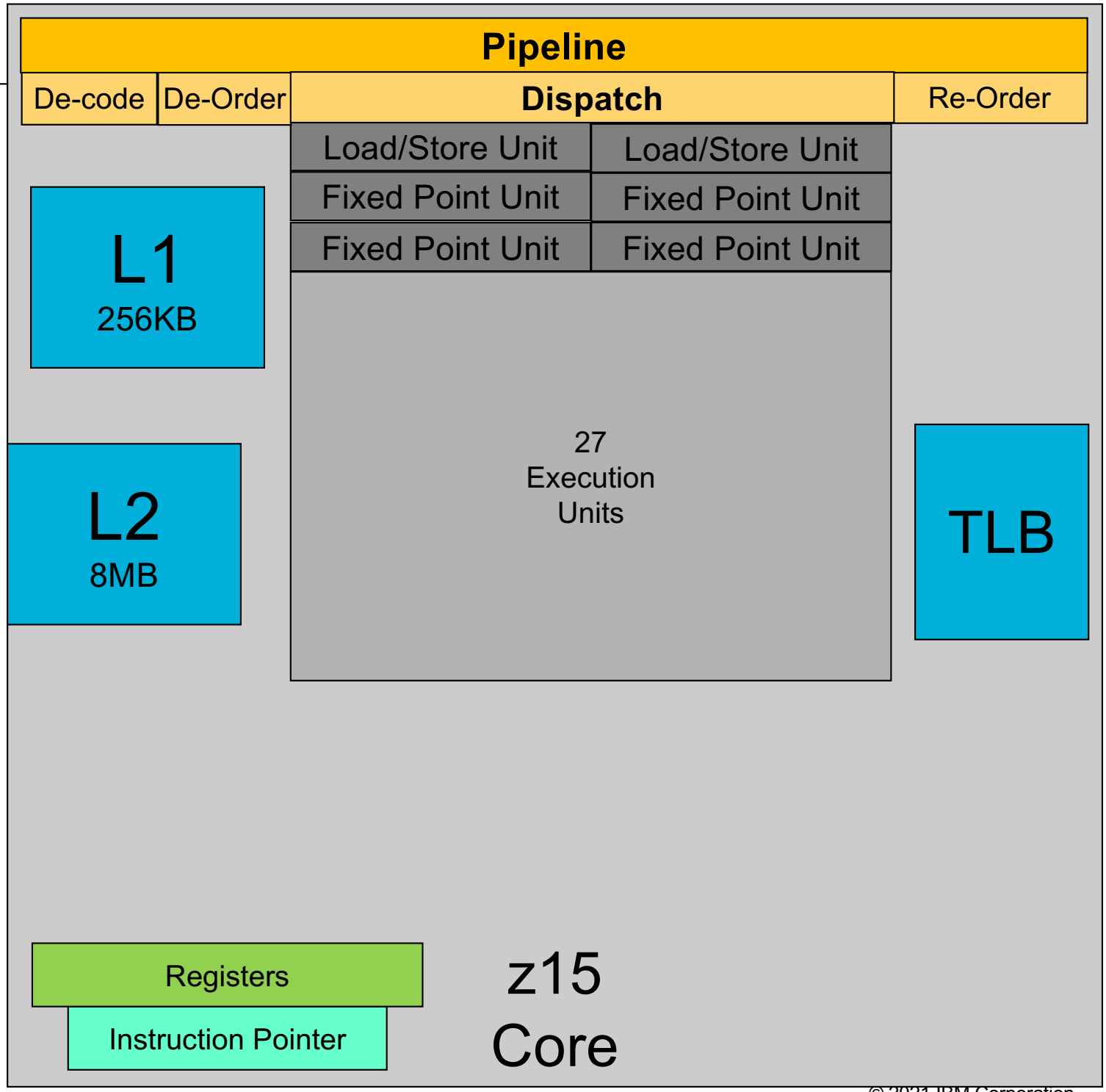
Multi-Threaded TLB

Address Translation Lookaside Buffer

Super-Scalar



Instructions can be categorized by function performed, and hardware can be optimized

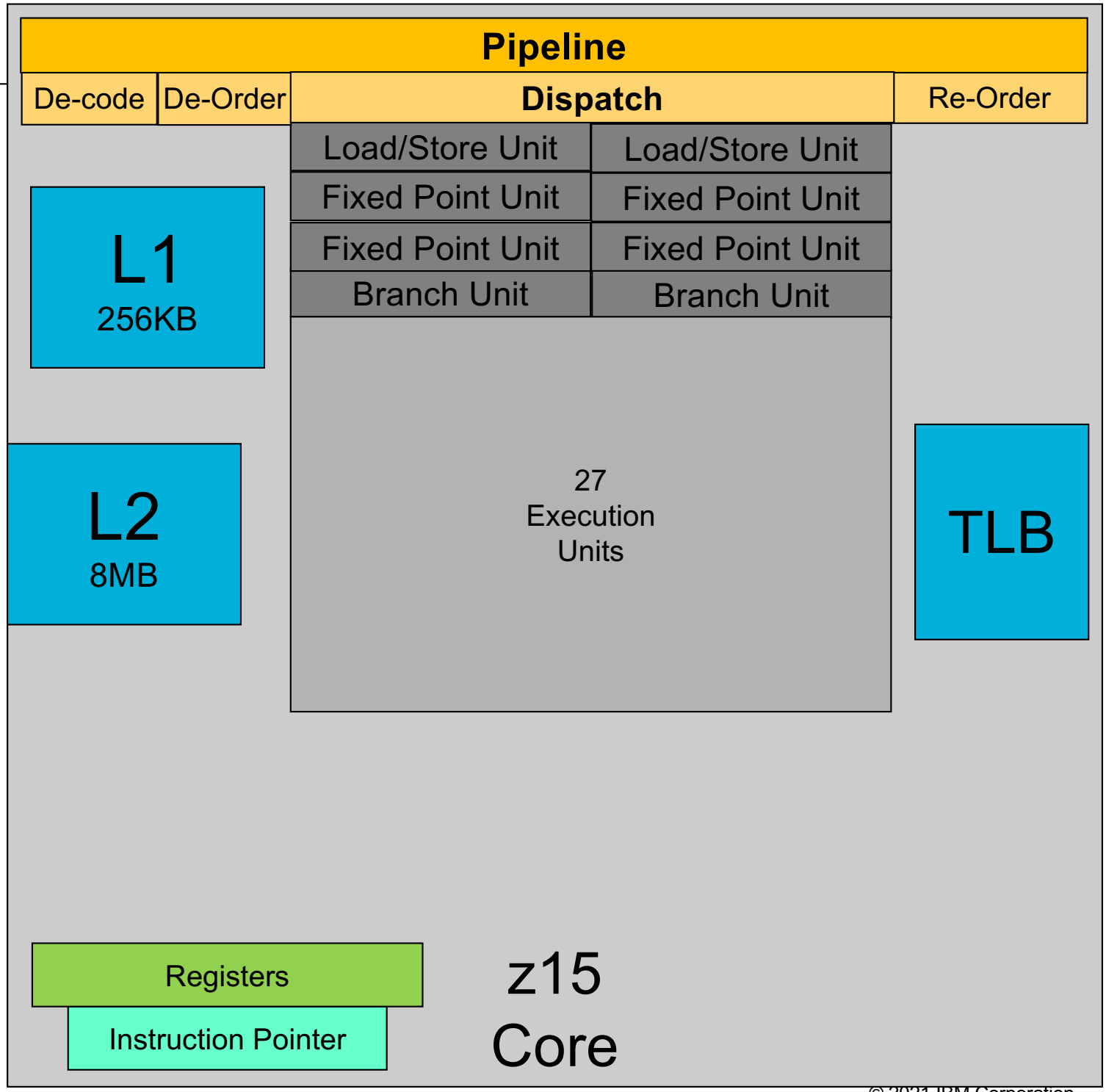


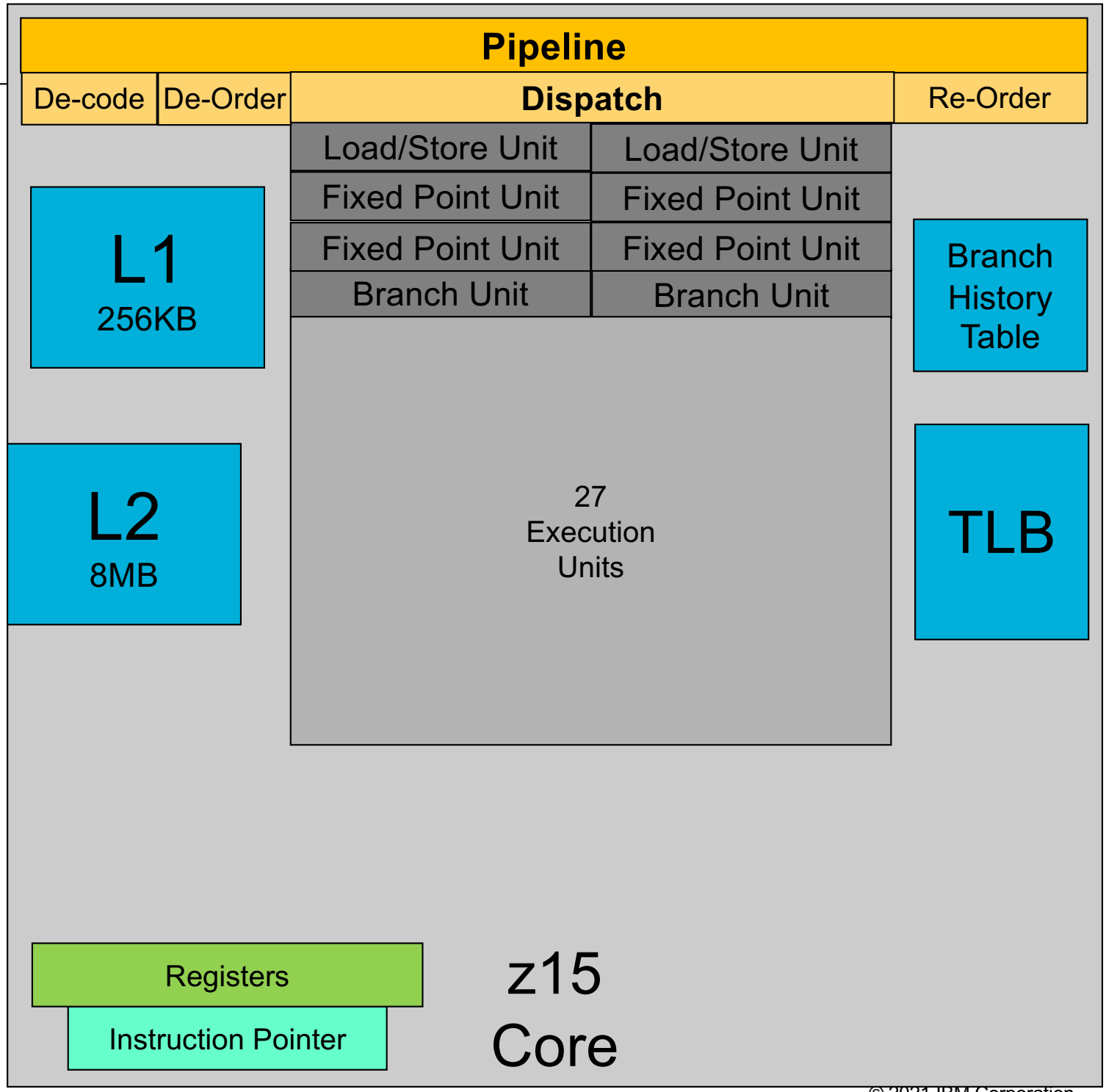
Super-Scalar

Optimize hardware for the most common instructions

Super-Scalar

Branches kill pipelines … Unless!

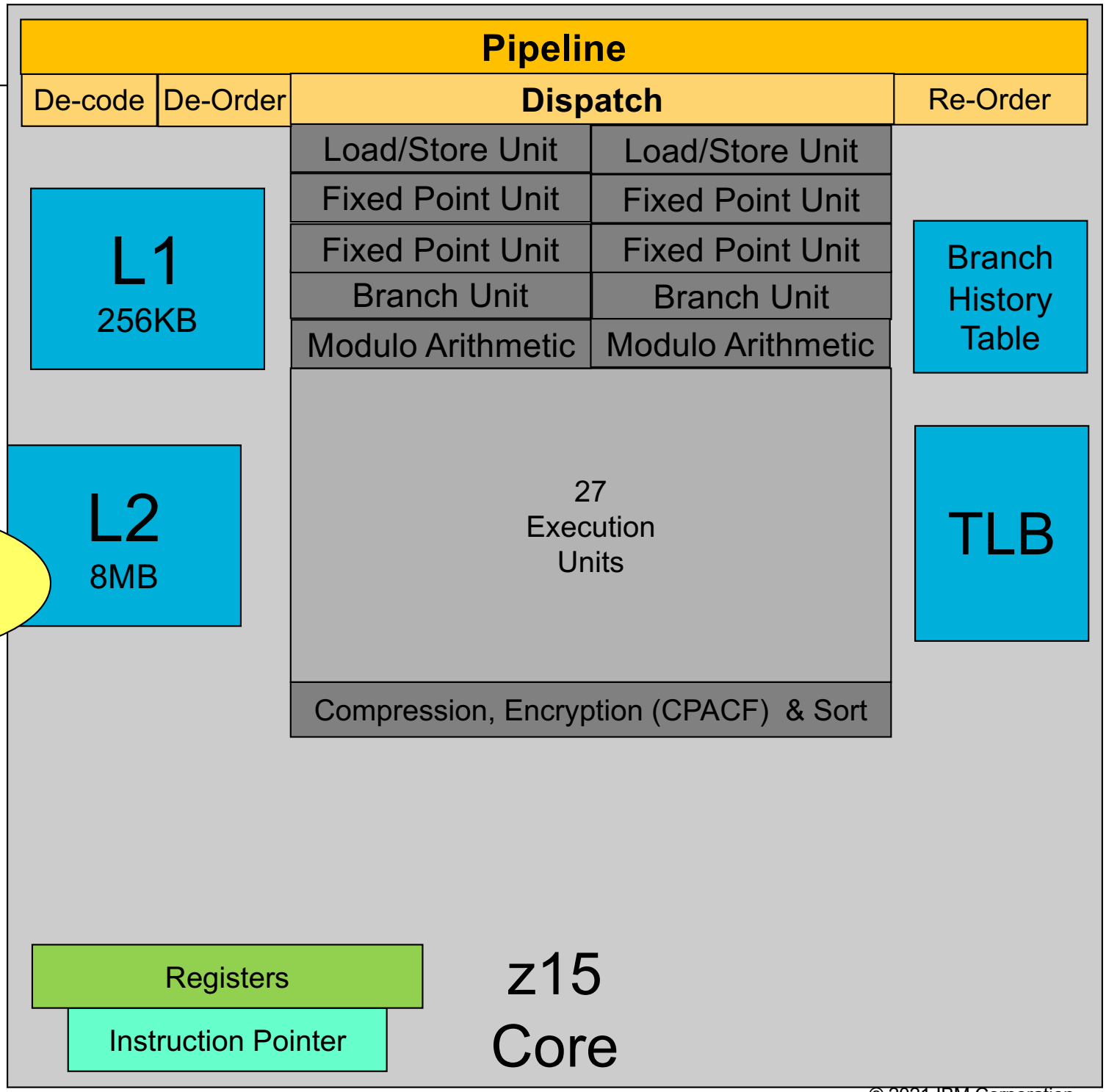




Cache Branch targets

Branch Prediction

Smartly influence caching decisions



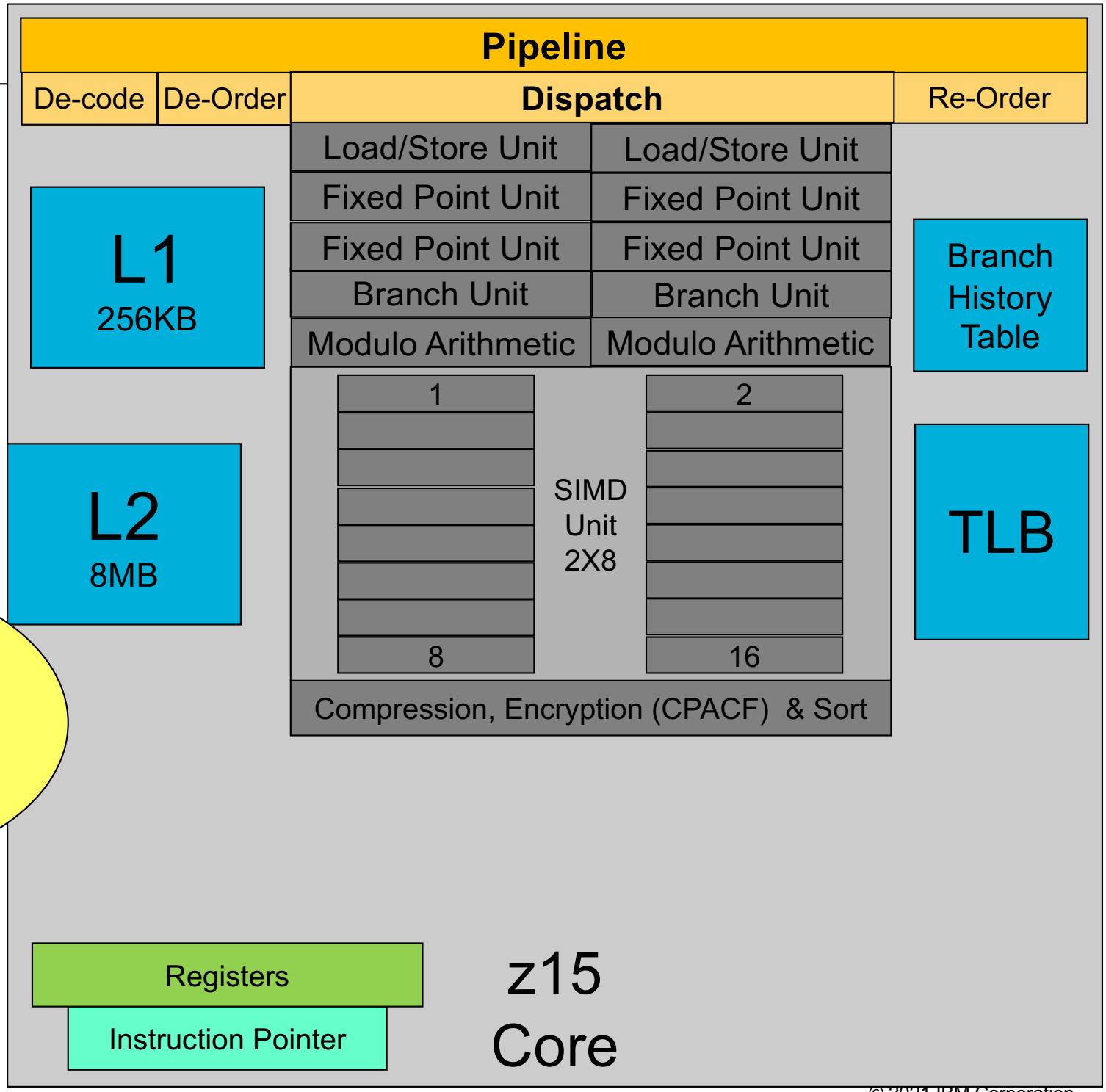
Workload Accelerators

Encryption

Compression

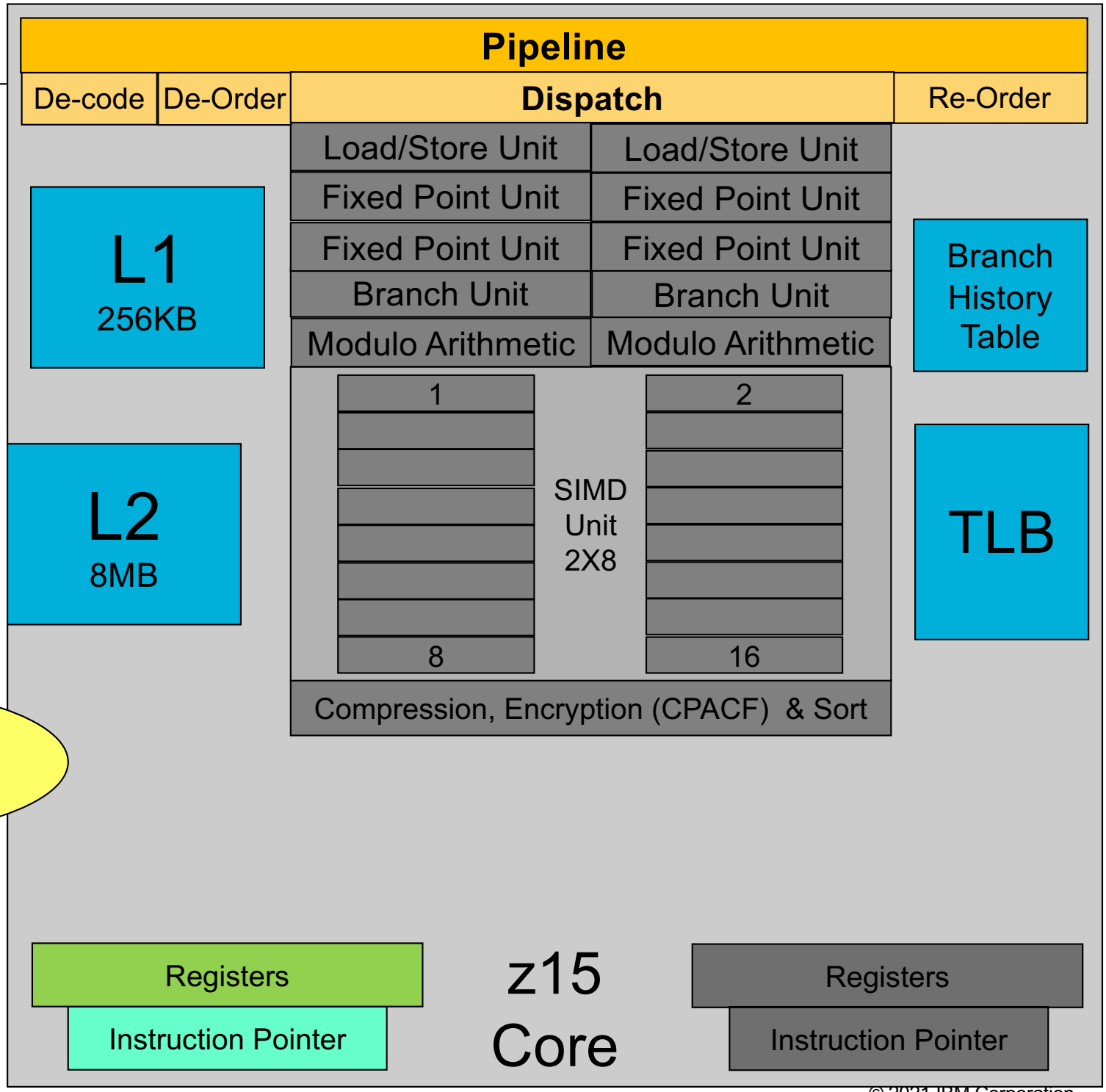
Sorting

Put tough software problems in hardware



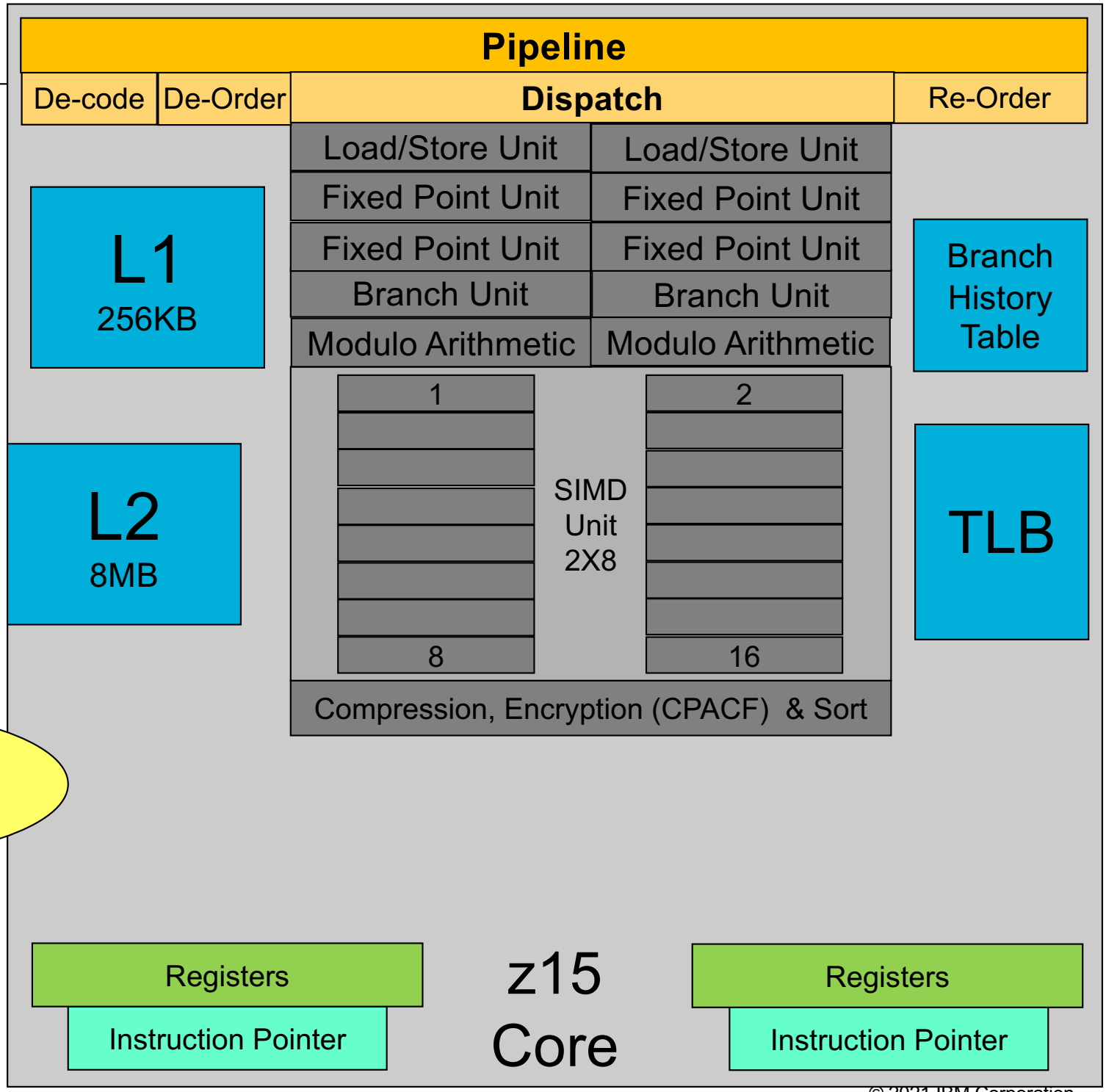
SIMD

Single Instruction Multiple Data (aka Vector Processing)



SMT1

Simultaneous Multi-Threading with 1 Thread

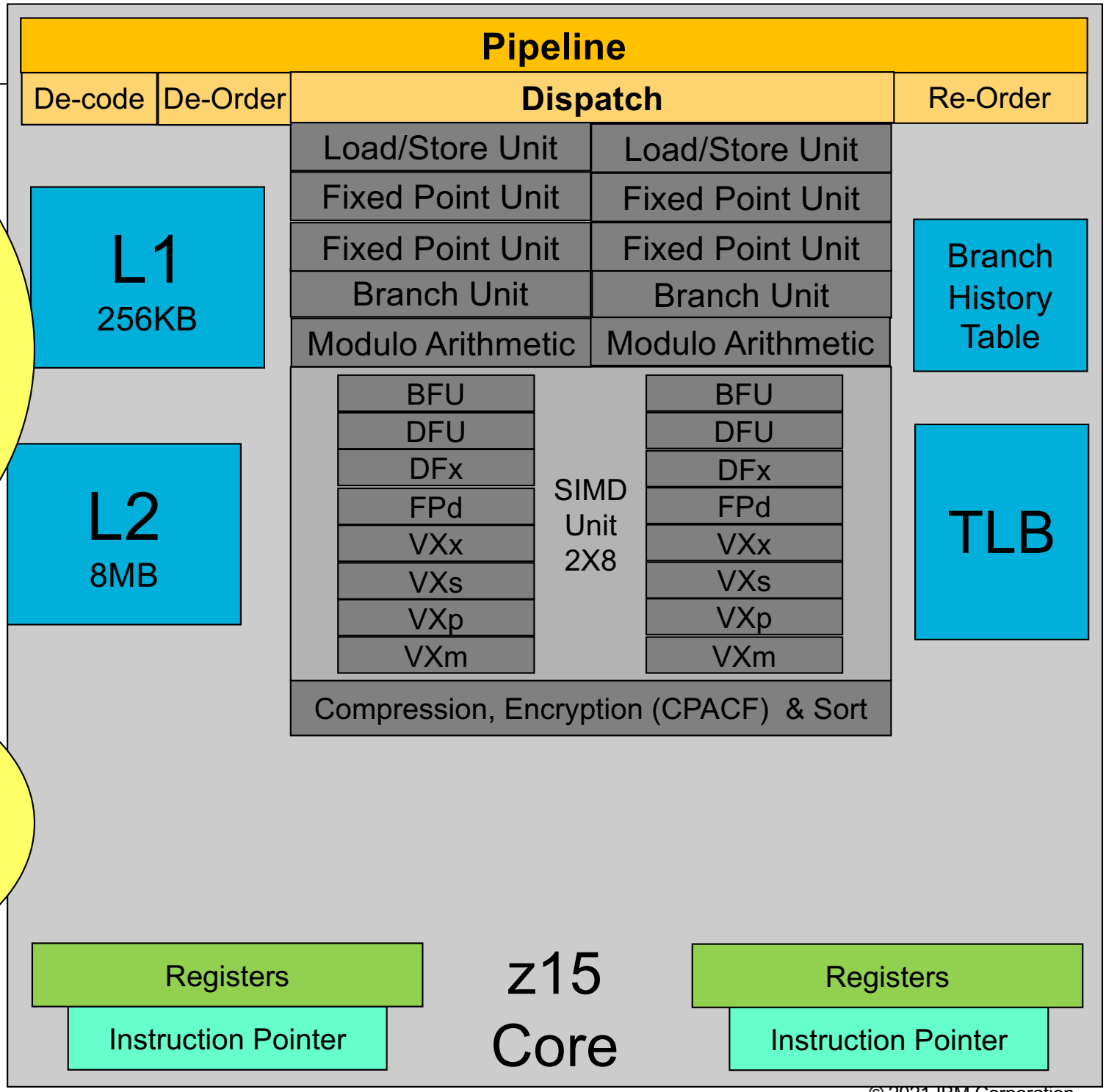


SMT2

Simultaneous Multi-Threading with 2 Threads

New Instructions: 20% more throughput for general Java workloads just from new instructions in the z15 chip architecture

IBM z Architecture alive and well … hundreds of new instructions



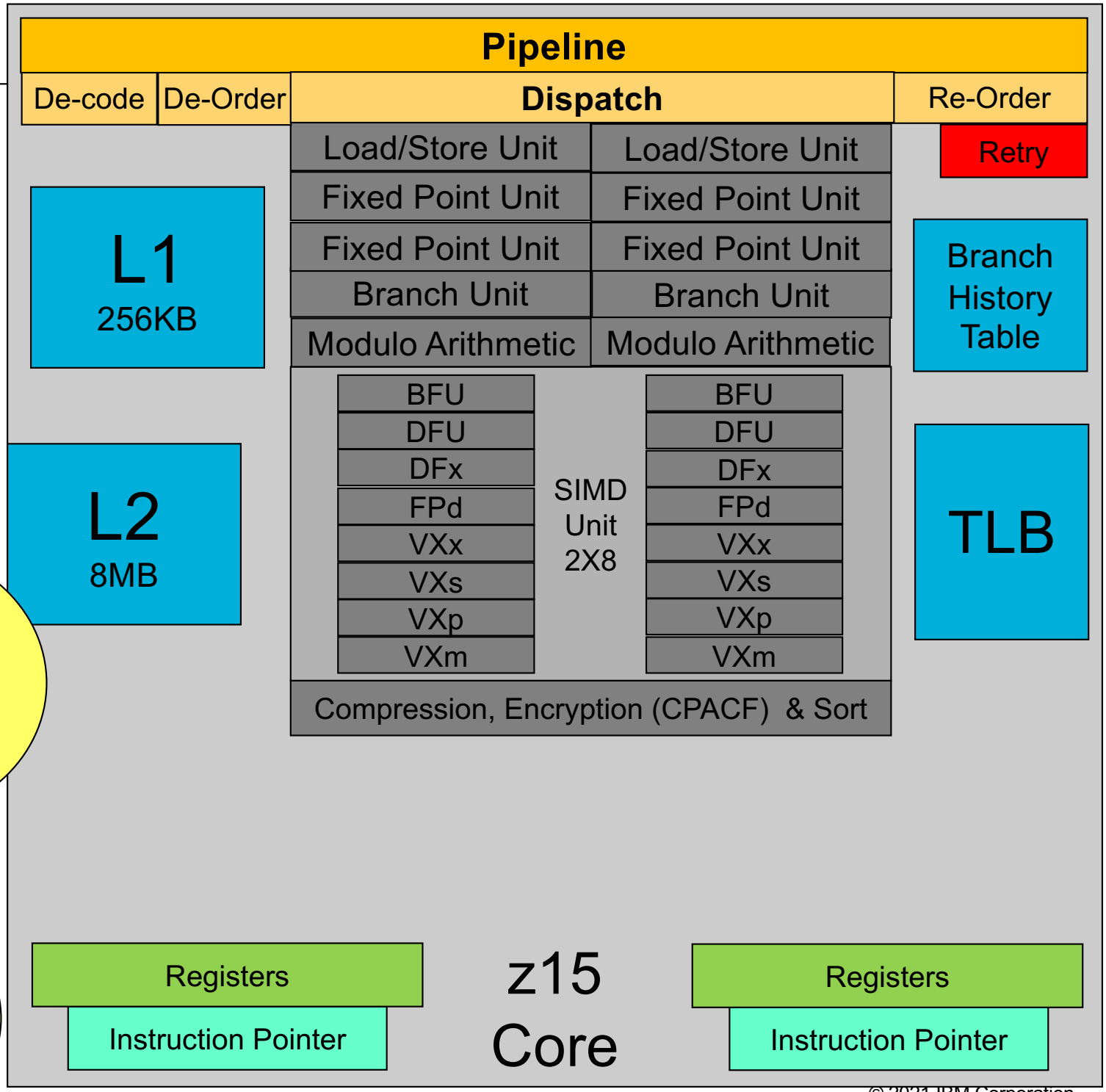
Error detection and instruction retry

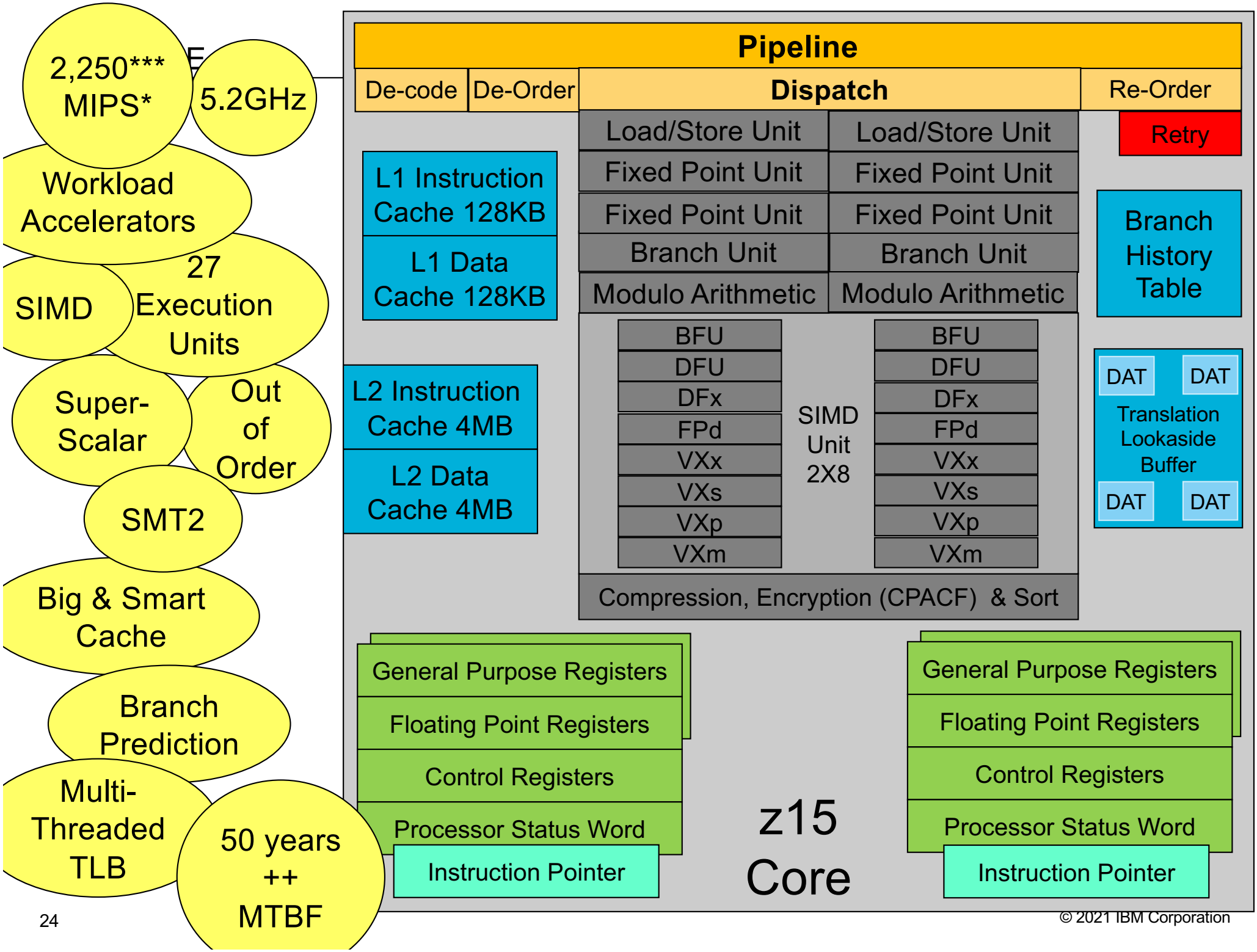
Dynamic core sparing

First failure data capture

Phone Home

100 years ++ MTBF





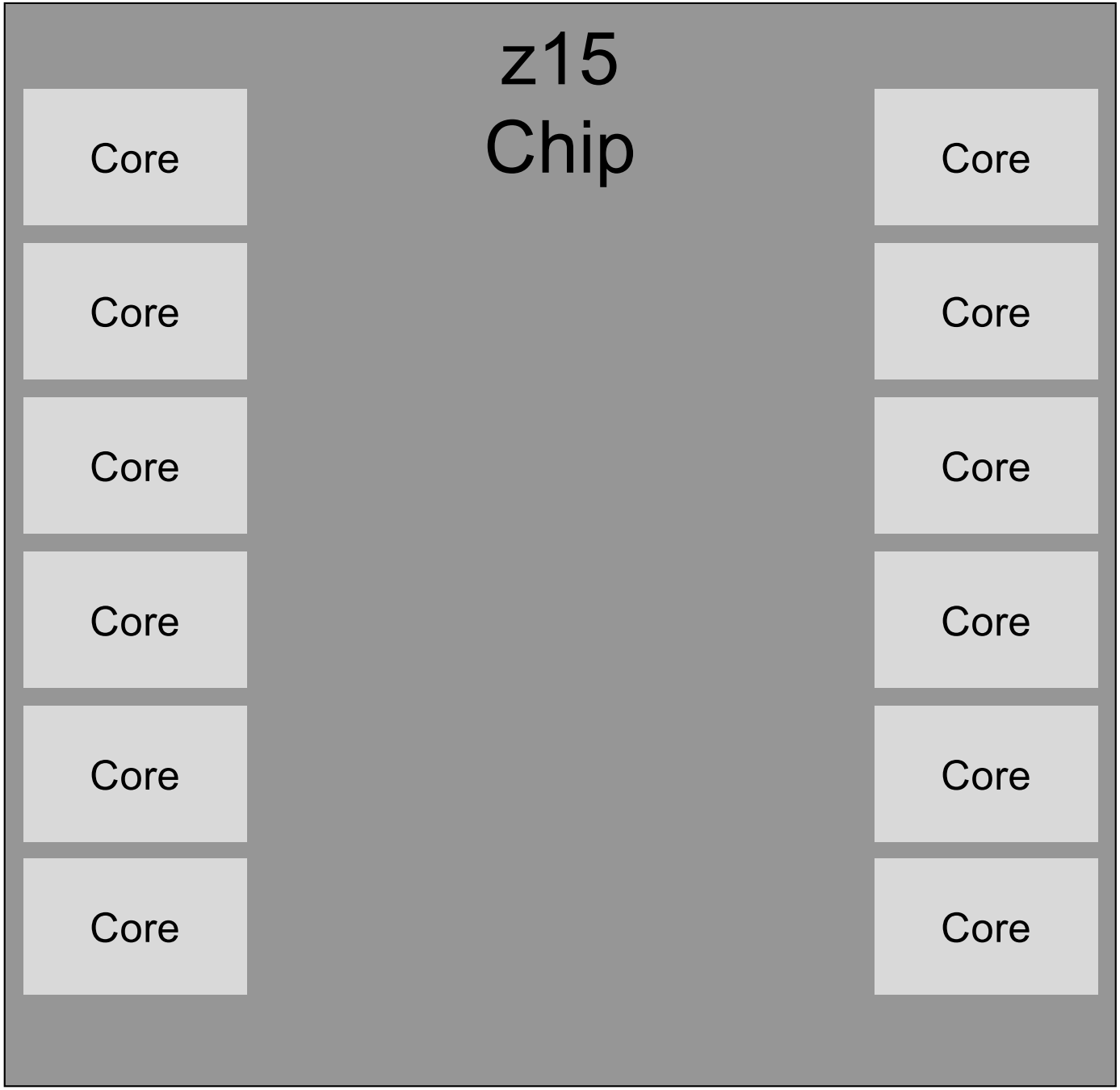
Cores to Chips

9 Billion Things to Like…

THE Z15 CHIP

9.1B
Transistors

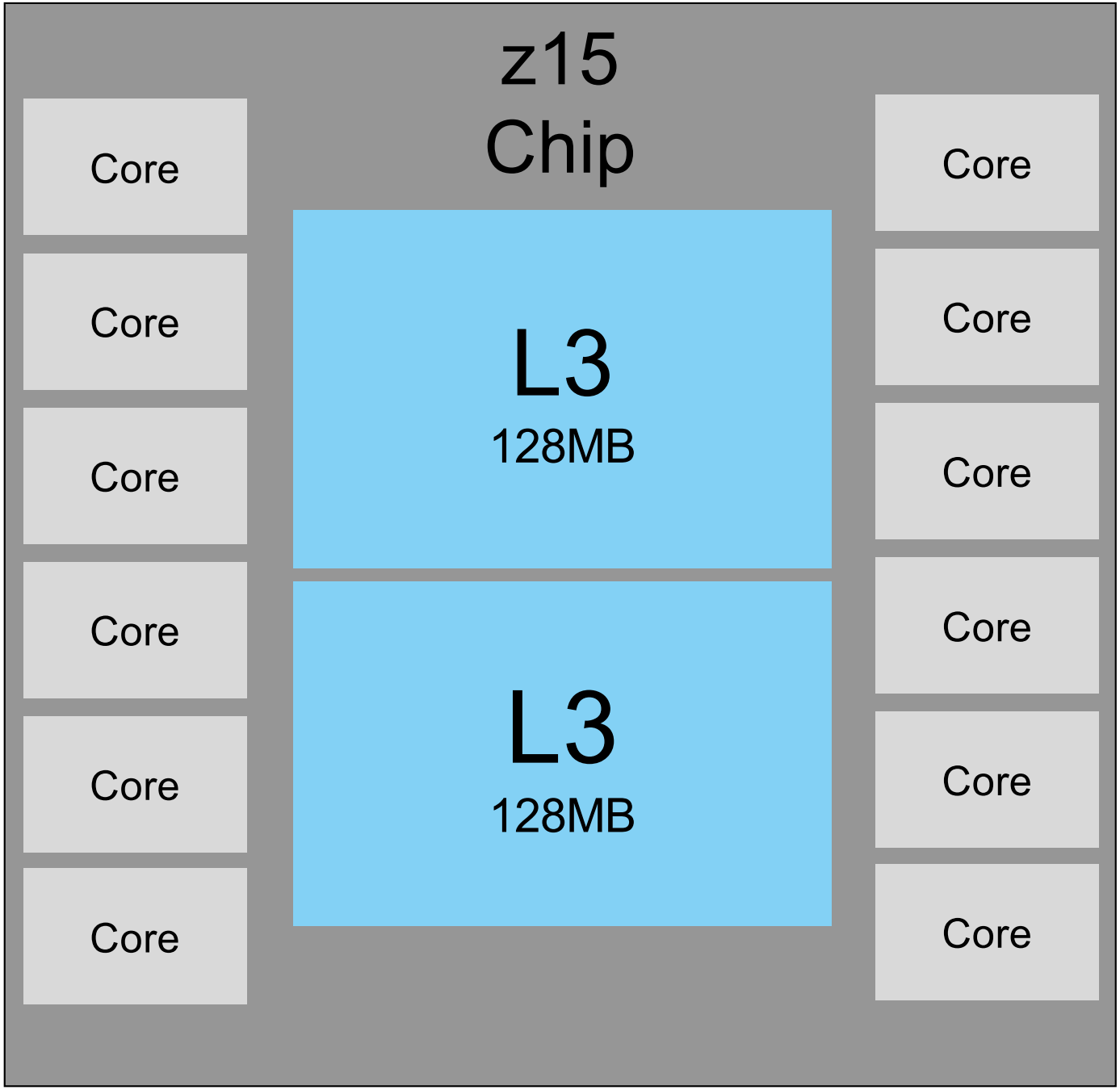
12 Cores

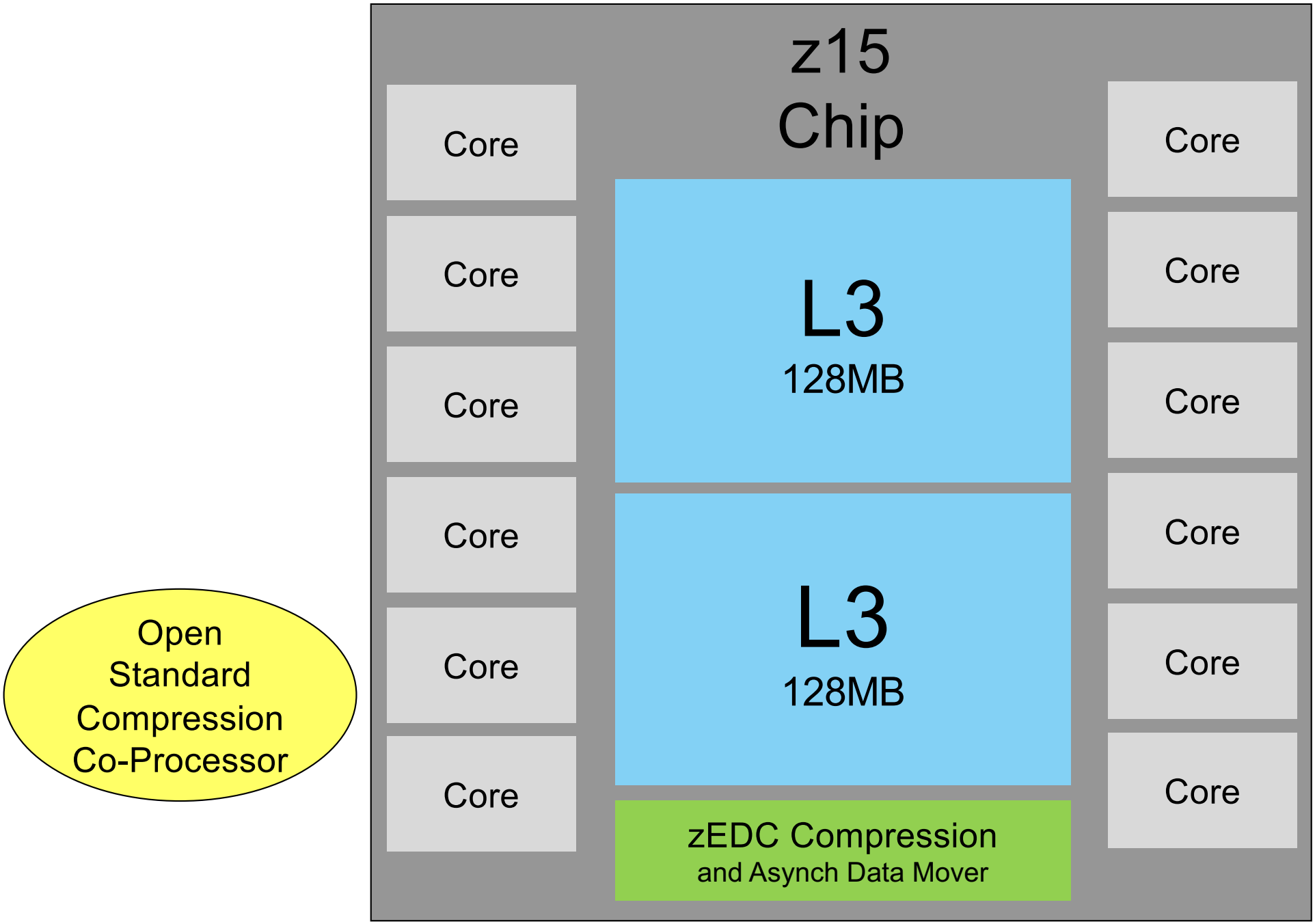


L1/L2/L3 Total 259MB cache

More (much more) big & smart cache

ECC-protected cache arrays





E

9.1B Transistors

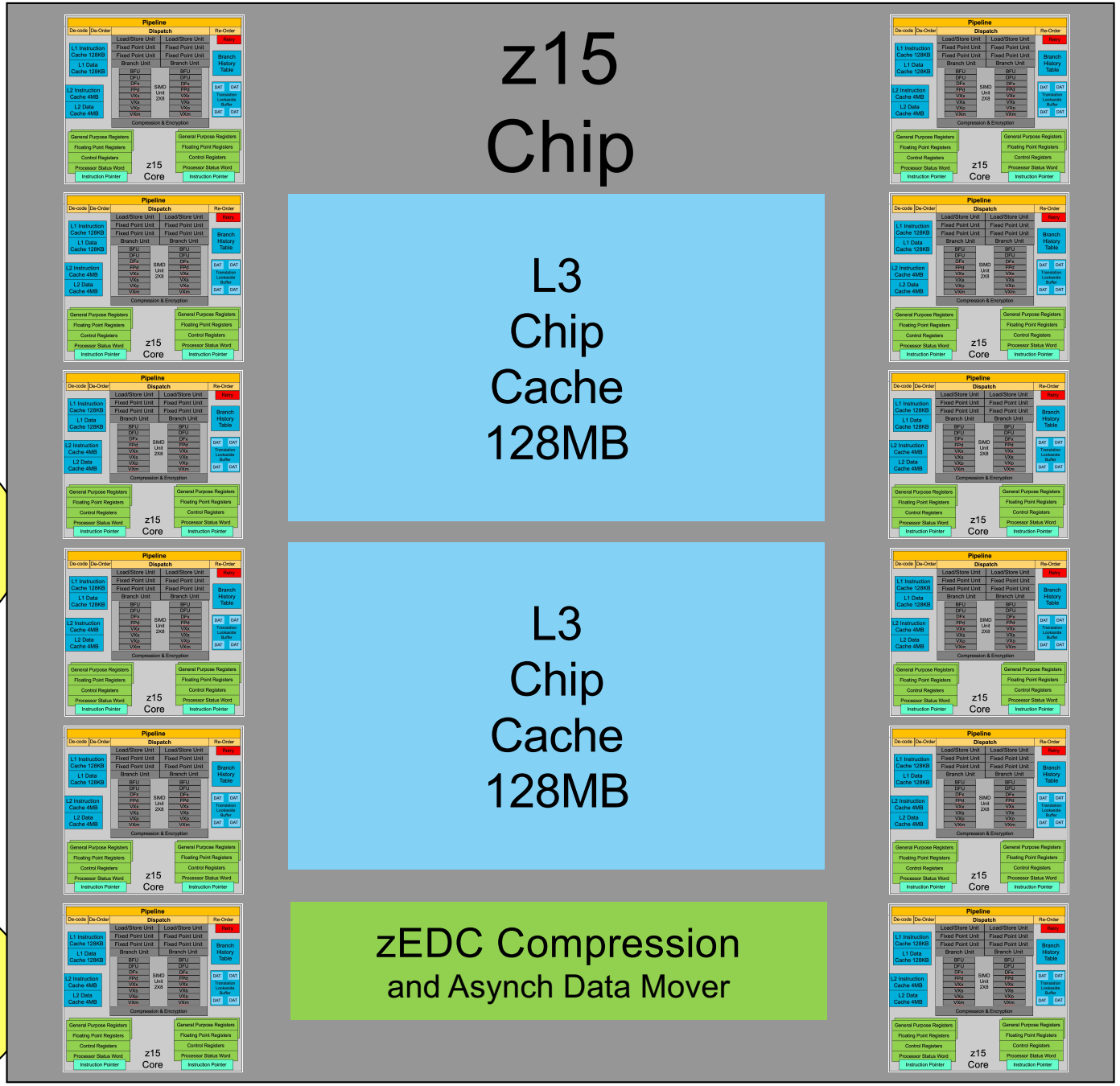
12 Cores

L1/L2/L3 Total 259MB cache

More (much more) big & smart cache

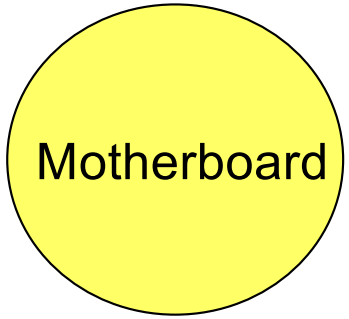
ECC-protected cache arrays

Compression Co-Processor

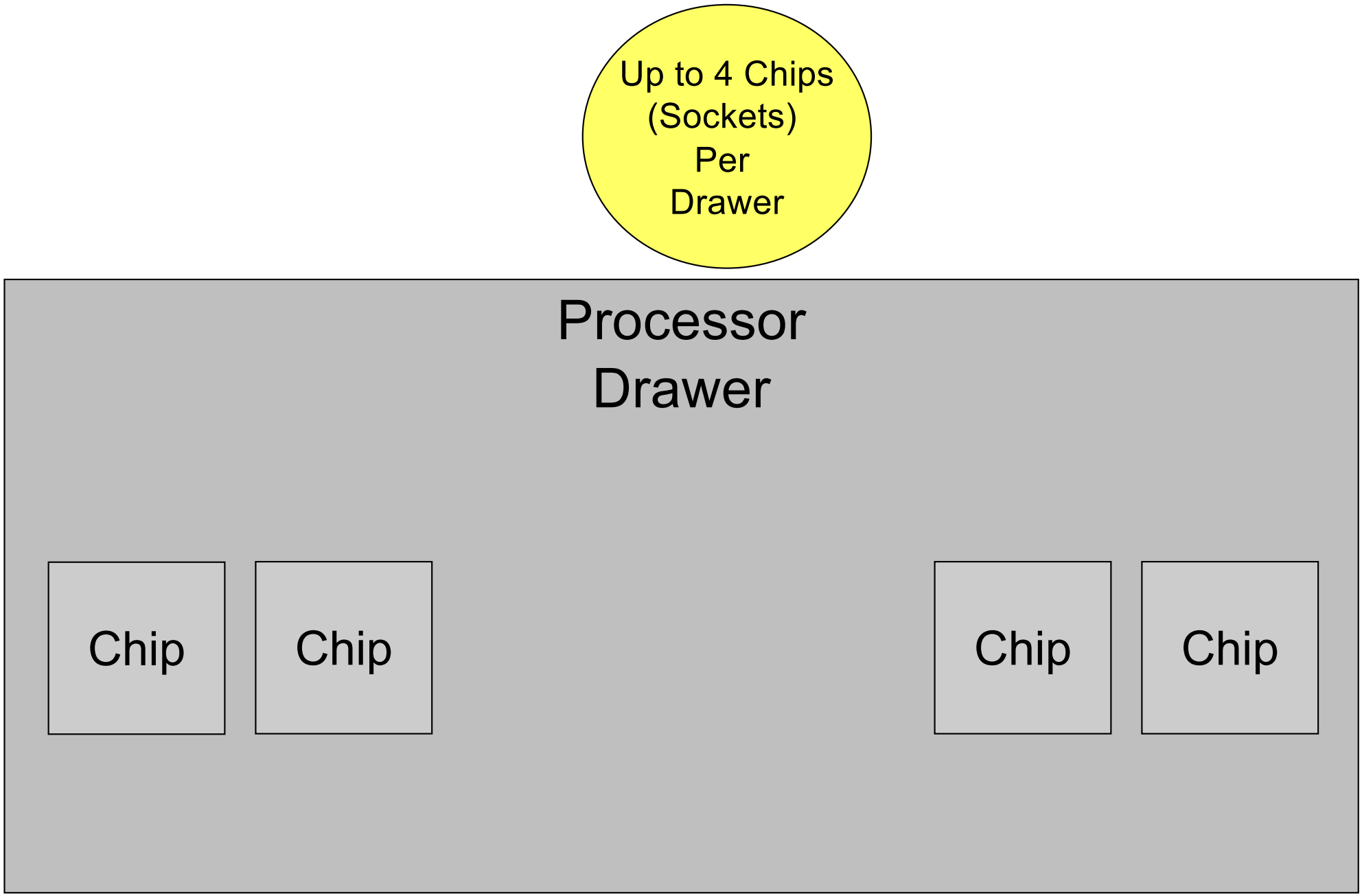


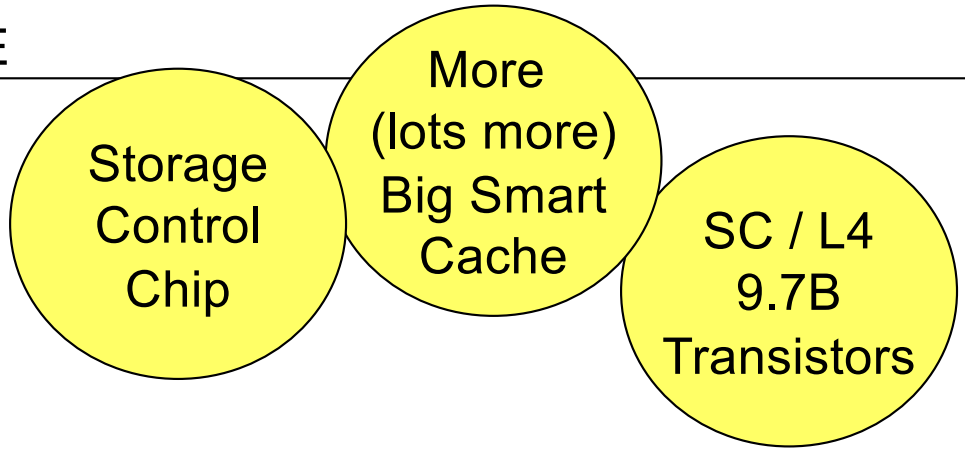
Chips to Drawers

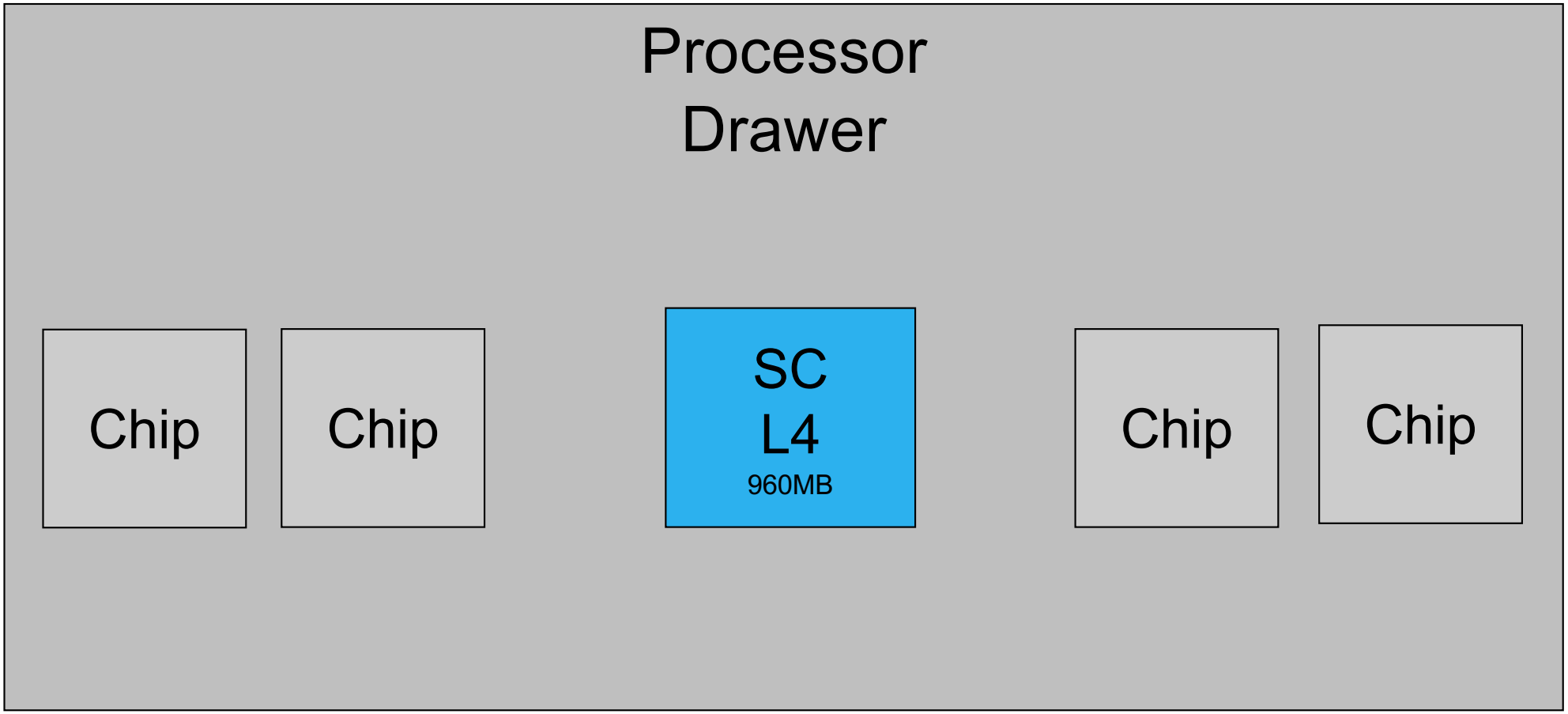
THE LINUXONE III DRAWER

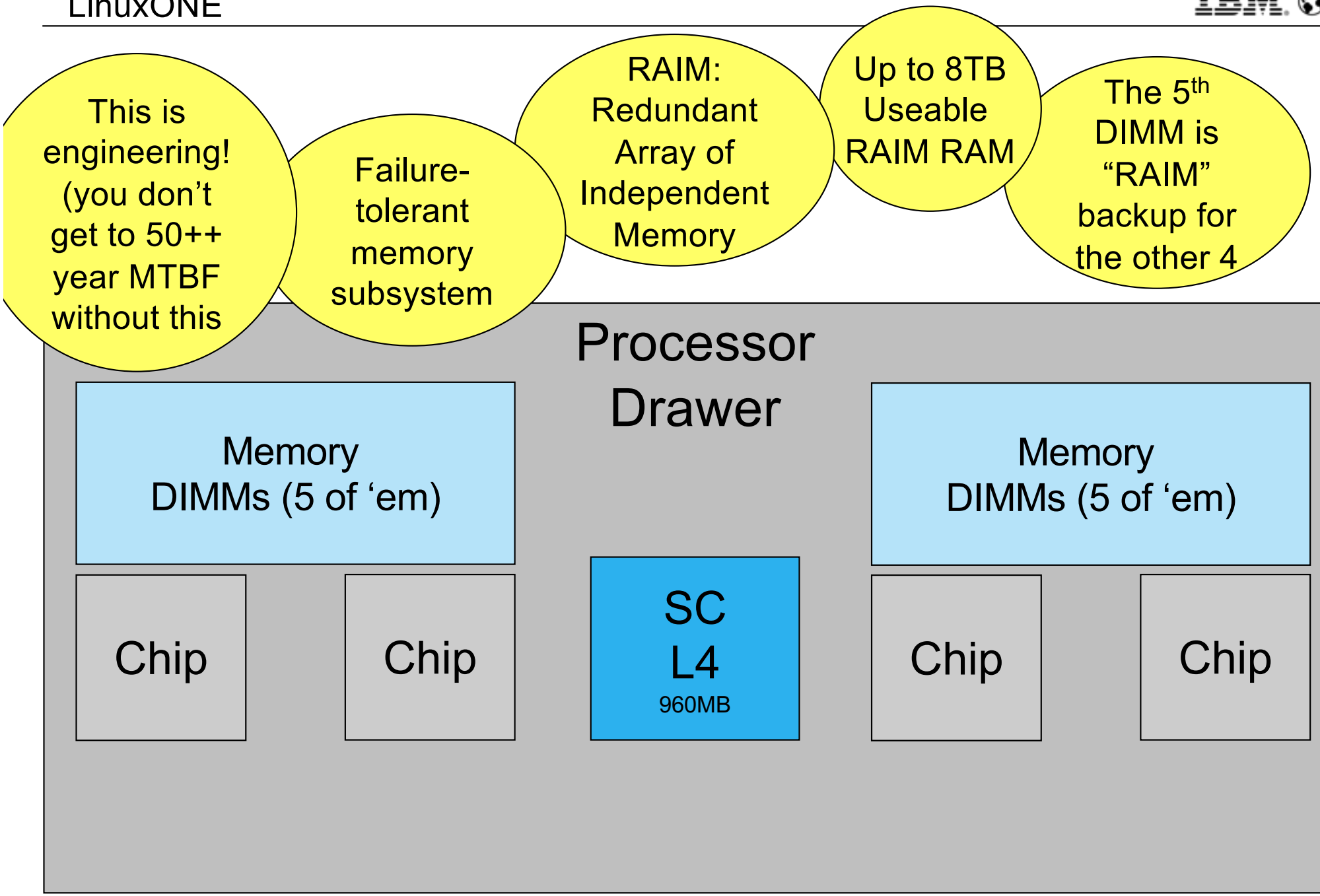


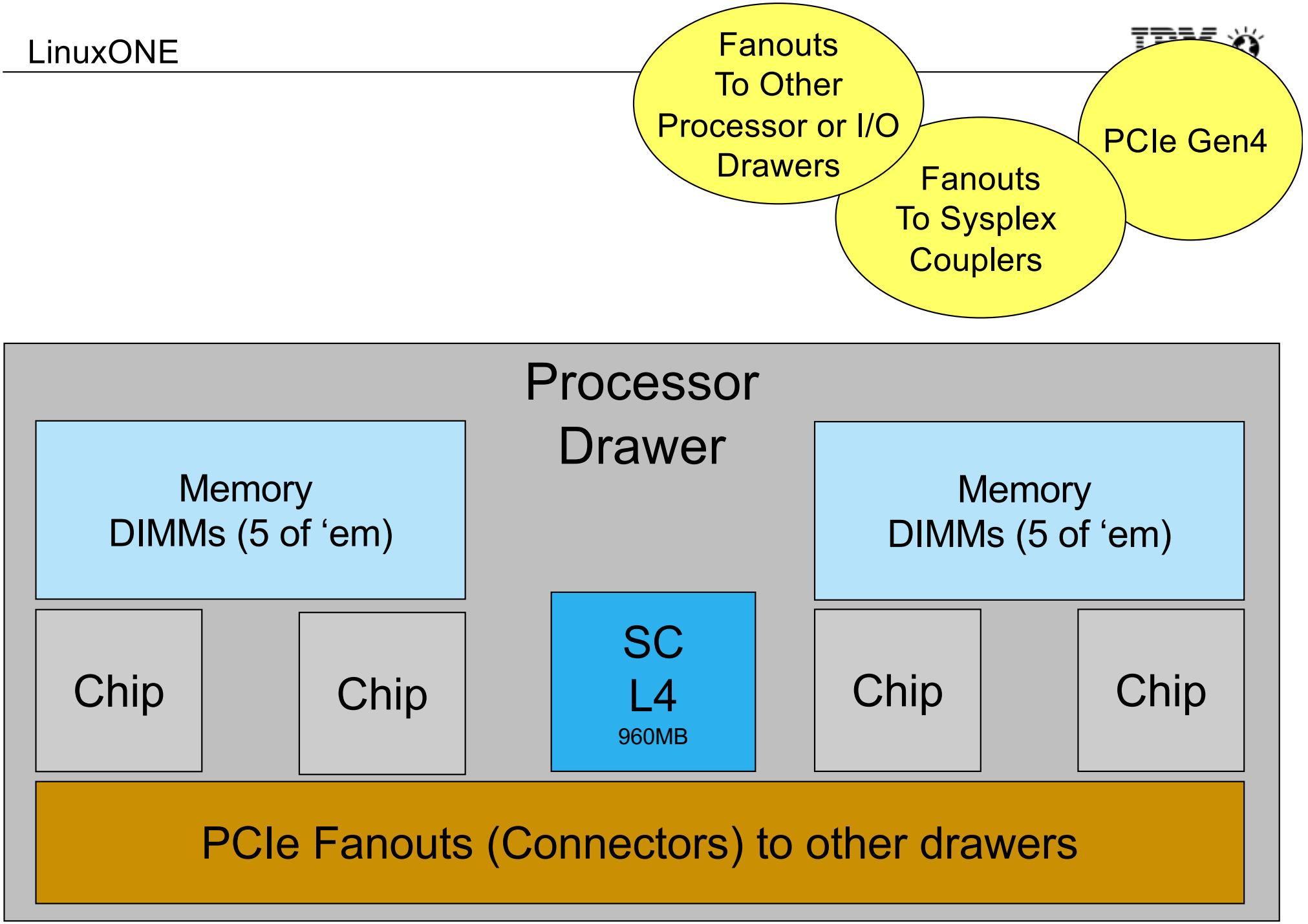
Processor
Drawer











Processor Drawer

More (lots more) Big Smart Cache

Up to 8TB Useable RAIM RAM

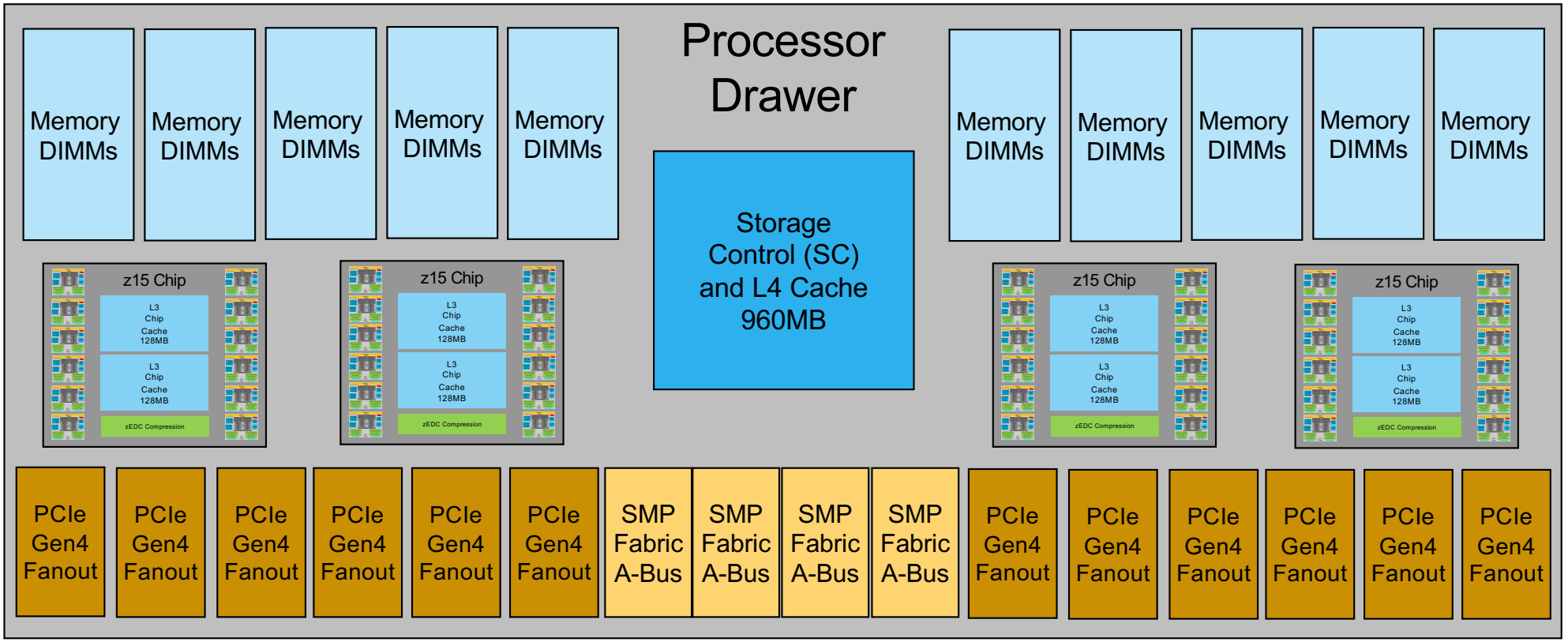
Fanouts To Other Processor and I/O Drawers

Motherboard

SC / L4 9.7B Transistors

4 Chips Per Drawer

Fanouts To Sysplex Couplers



Drawers to CEC

CEC: Central Electronics Complex

The Processor Nest

THE LINUXONE III CEC

The Big Guy

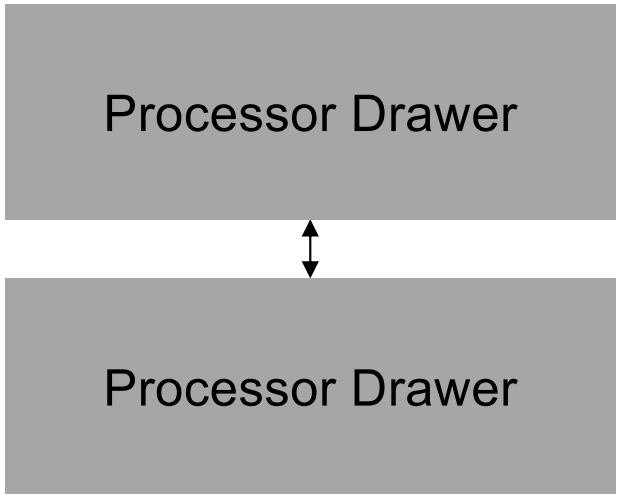
THE LINUXONE III LT1 CEC

LT1 Model	Drawers	Cores	Memory (RAIM)
Max34	1	1 to 34	Up to 8TB

Processor Drawer

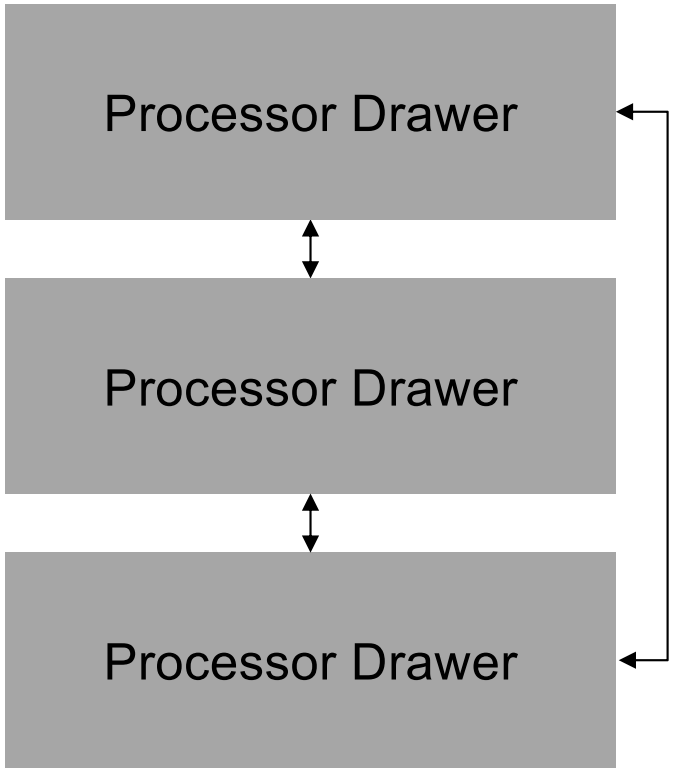
Start at 512GB RAIM

LT1 Model	Drawers	Cores	Memory (RAIM)
Max71	2	1 to 71	Up to 16TB

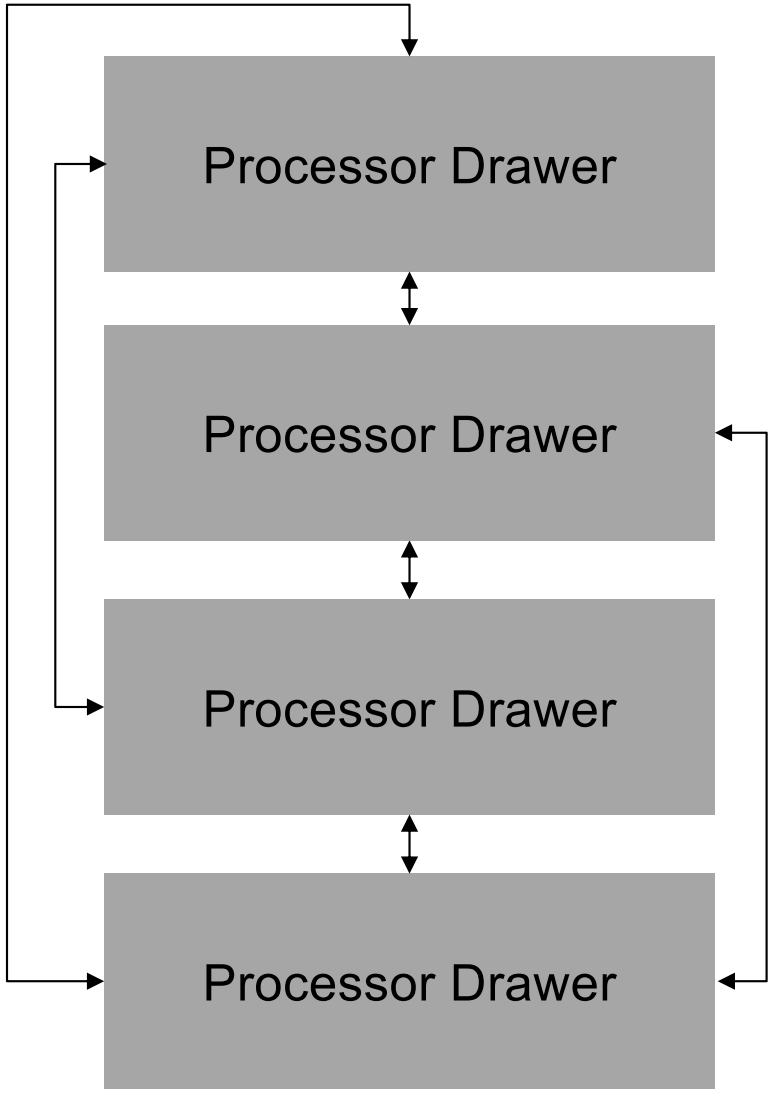


Fully Interconnected SMP Fabric

LT1 Model	Drawers	Cores	Memory (RAIM)
Max108	3	1 to 108	Up to 24TB

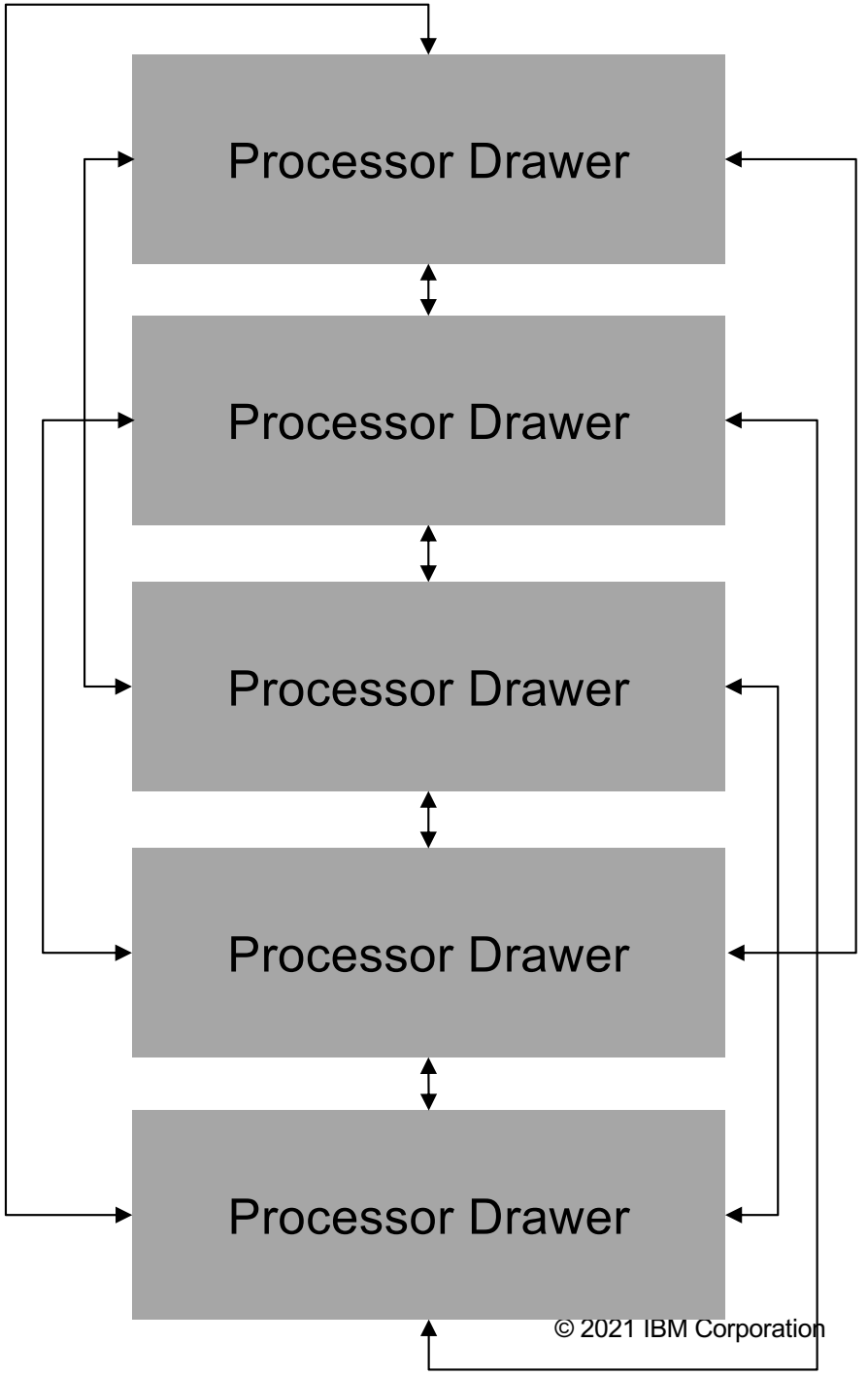


LT1 Model	Drawers	Cores	Memory (RAIM)
Max145	4	1 to 145	Up to 32TB



LT1 Model	Drawers	Cores	Memory (RAIM)
Max190	5	1 to 190	Up to 40TB

Scale to 200,000 MIPS



CEC: LinuxONE III LT1

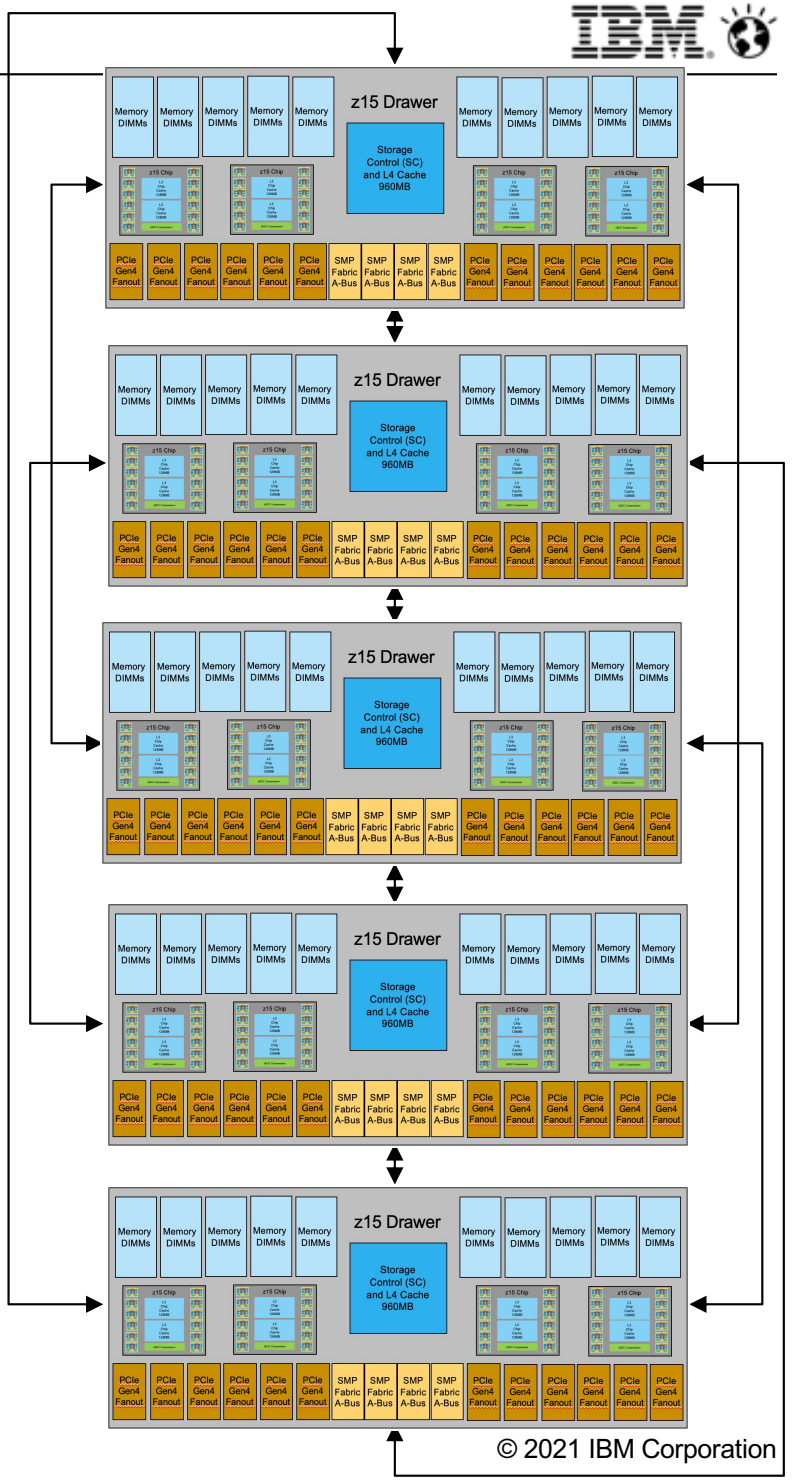
LT1 Model	Drawers	Cores	Memory (RAIM)
Max34	1	1 to 34	Up to 8TB
Max71	2	1 to 71	Up to 16TB
Max108	3	1 to 108	Up to 24TB
Max145	4	1 to 145	Up to 32TB
Max190	5	1 to 190	Up to 40TB

1 to 190 Cores

512GB to 40TB RAIM

Scale to 200,000 MIPS

Fully Interconnected SMP Fabric

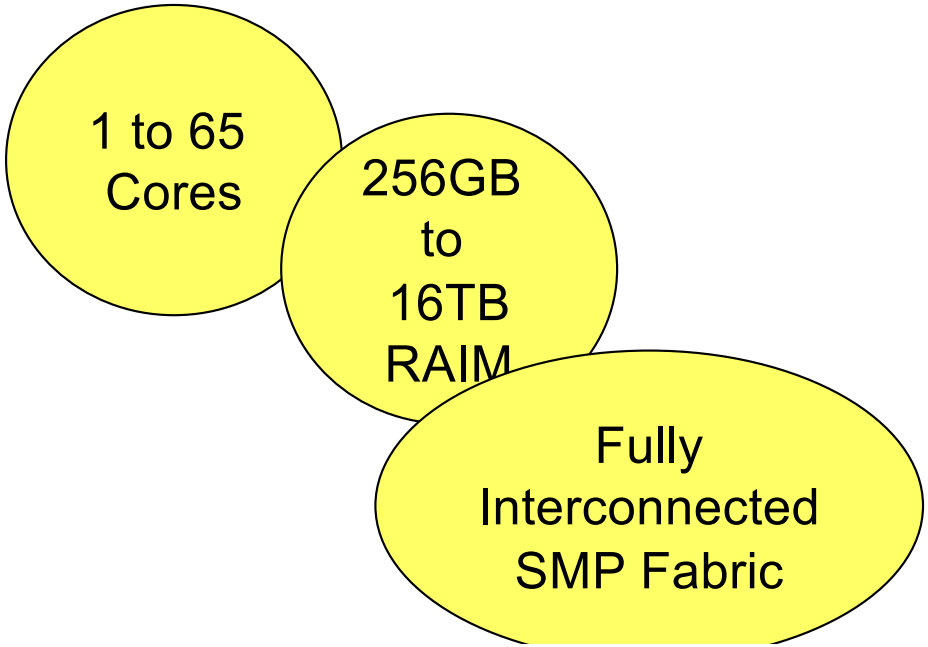


The Little Guy

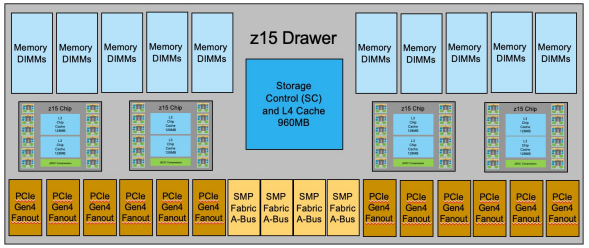
THE LINUXONE III LT2 CEC

CEC: LinuxONE III LT2

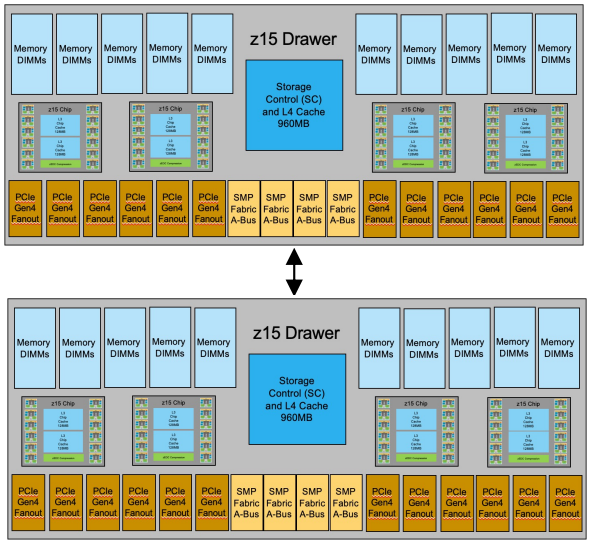
LT2 Model	Drawers	Cores	Memory (RAIM)
Max04	1	1 to 4	Up to 2TB
Max13	1	1 to 13	Up to 4TB
Max21	1	1 to 21	Up to 4TB
Max31	1	1 to 31	Up to 8TB
Max65	2	1 to 65	Up to 16TB



Max04 Max13 Max21 Max31



Max65



Rack’em and Stack’em

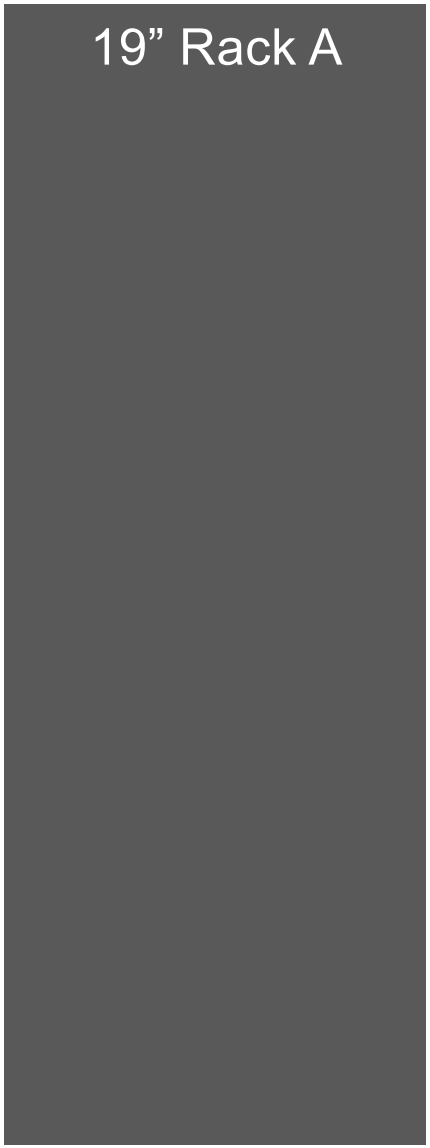
THE LINUXONE III RACK(S)

Following…

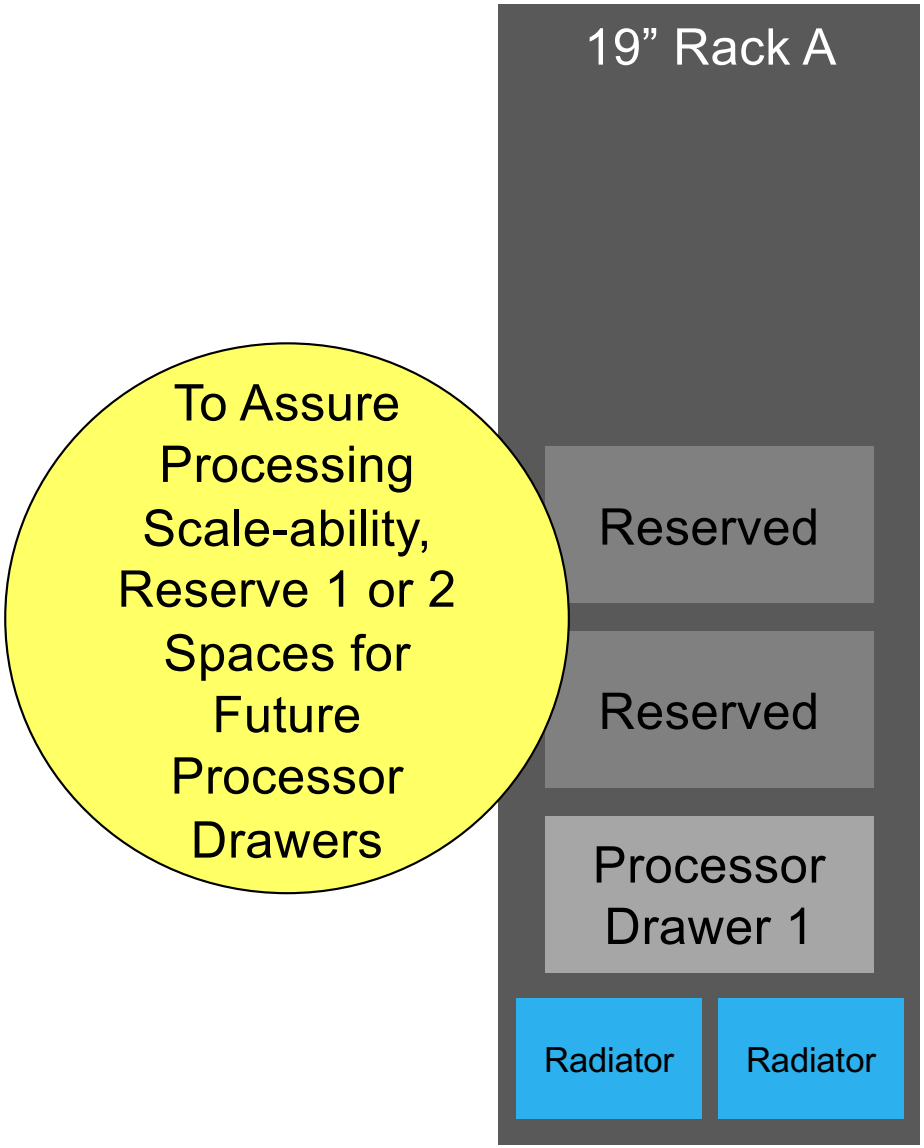
- Following LinuxONE III LT1 server rack-buildout scenarios
  - and several LinuxONE III LT2 rack scenarios

- They are examples
- They illustrate key concepts
- There are many more

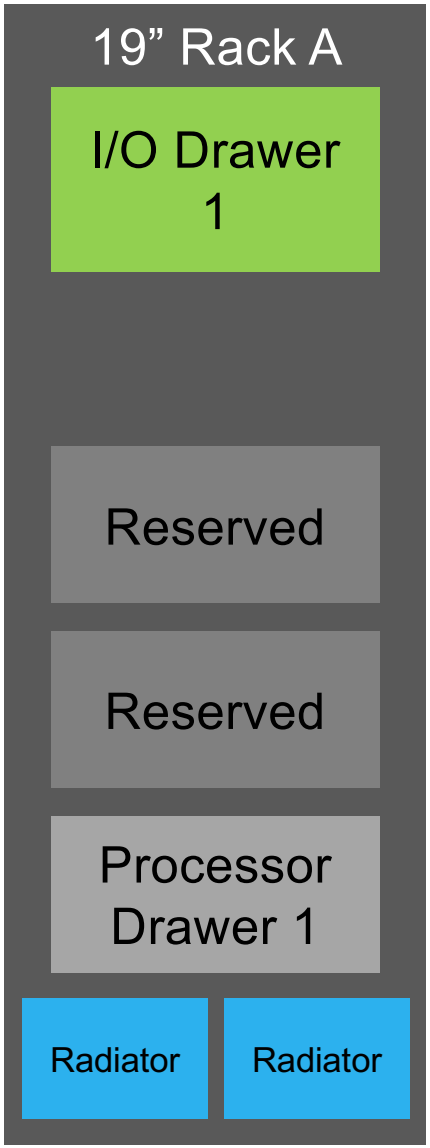
LinuxONE III LT1 Racks: Scenario 2 – iPDU Powered CPU Heavy



LinuxONE III LT1 Racks: Scenario 2 – iPDU Powered CPU Heavy

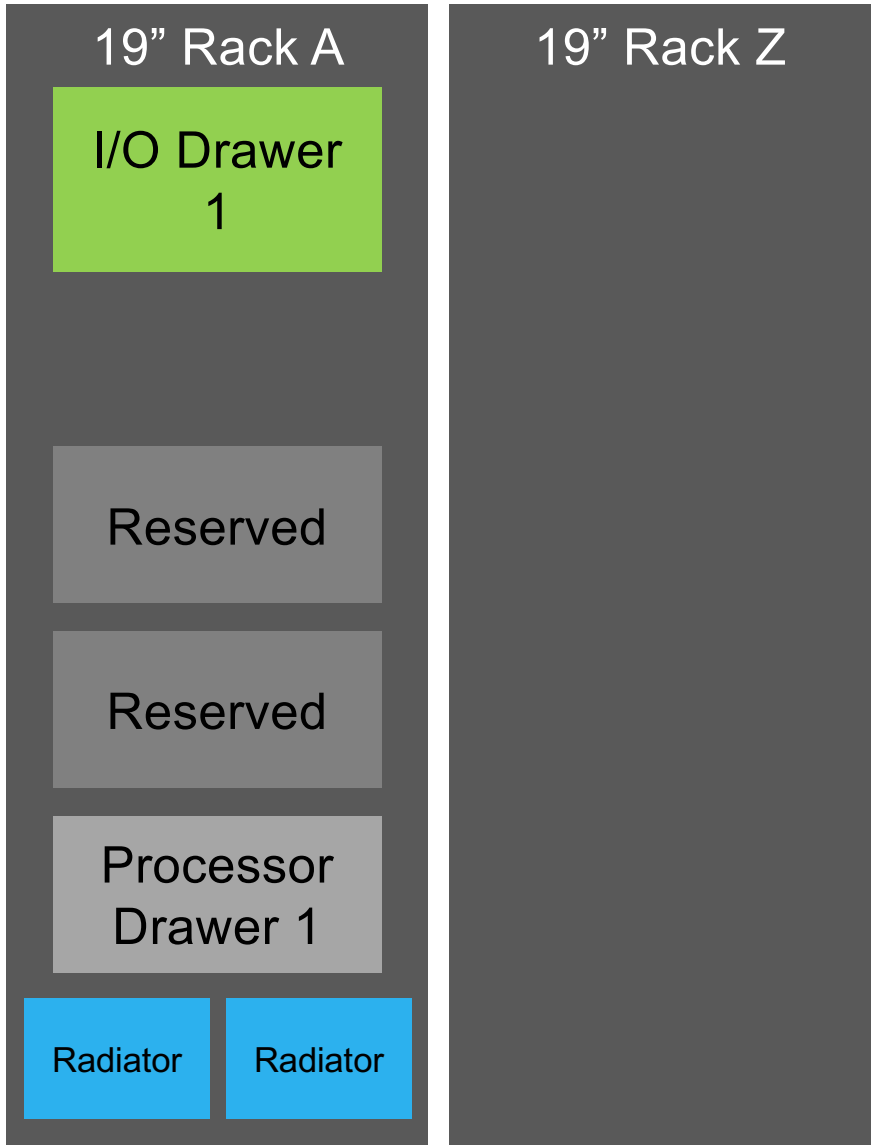


LinuxONE III LT1 Racks: Scenario 2 – iPDU Powered CPU Heavy

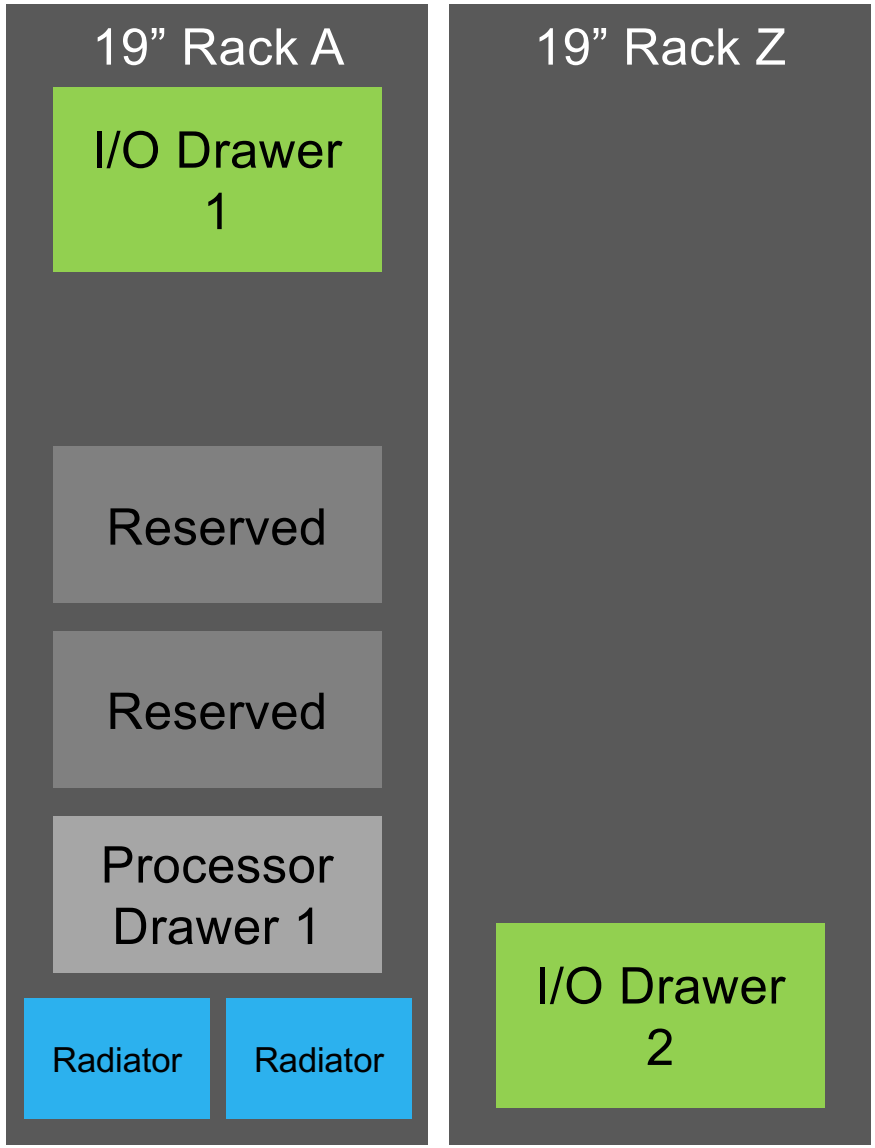


Everything that happens next CAN BE non-disruptive

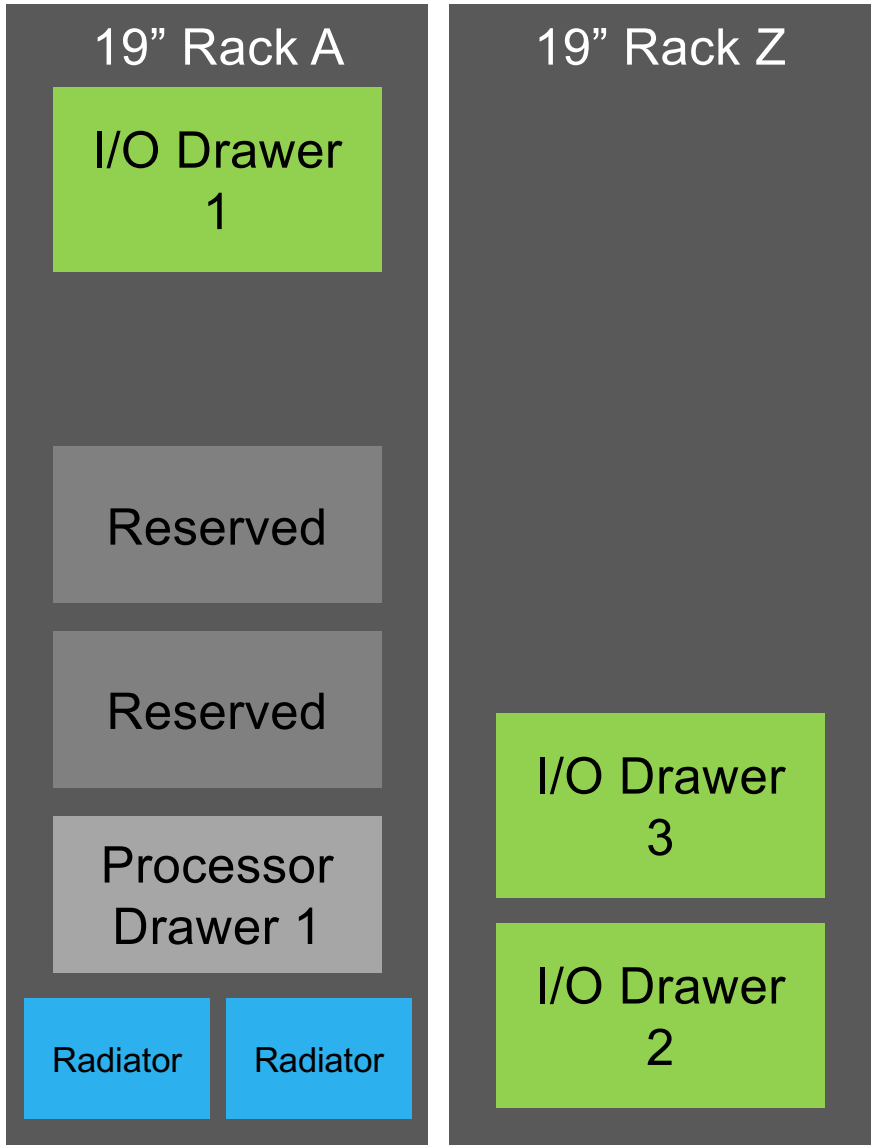
LinuxONE III LT1 Racks: Scenario 2 – iPDU Powered CPU Heavy



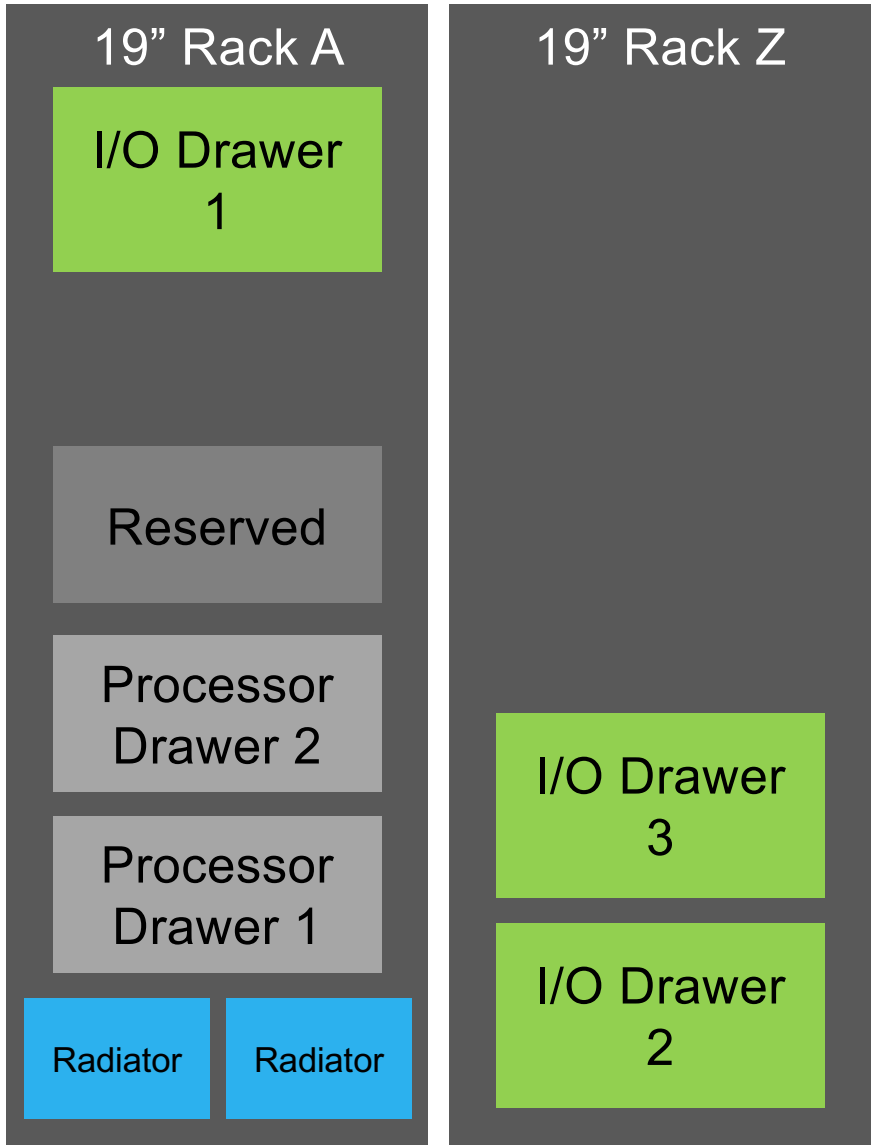
LinuxONE III LT1 Racks: Scenario 2 – iPDU Powered CPU Heavy



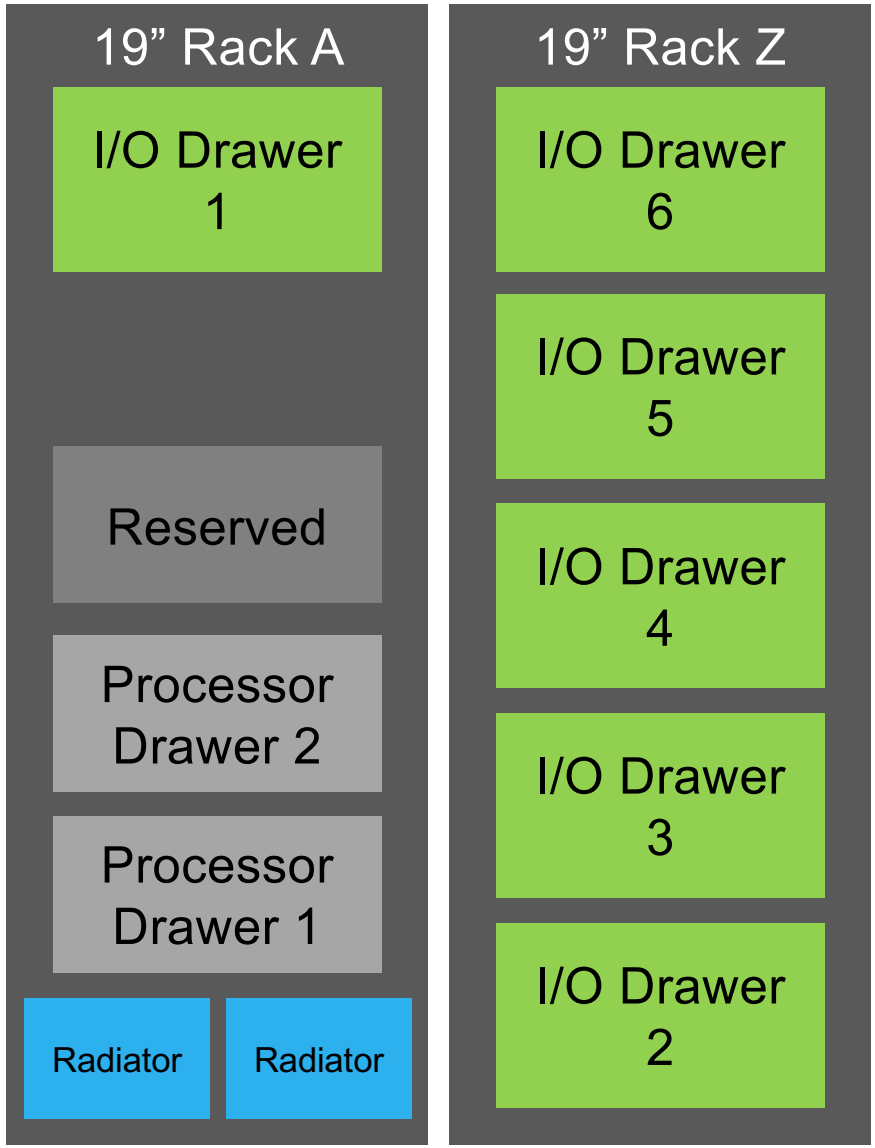
LinuxONE III LT1 Racks: Scenario 2 – iPDU Powered CPU Heavy



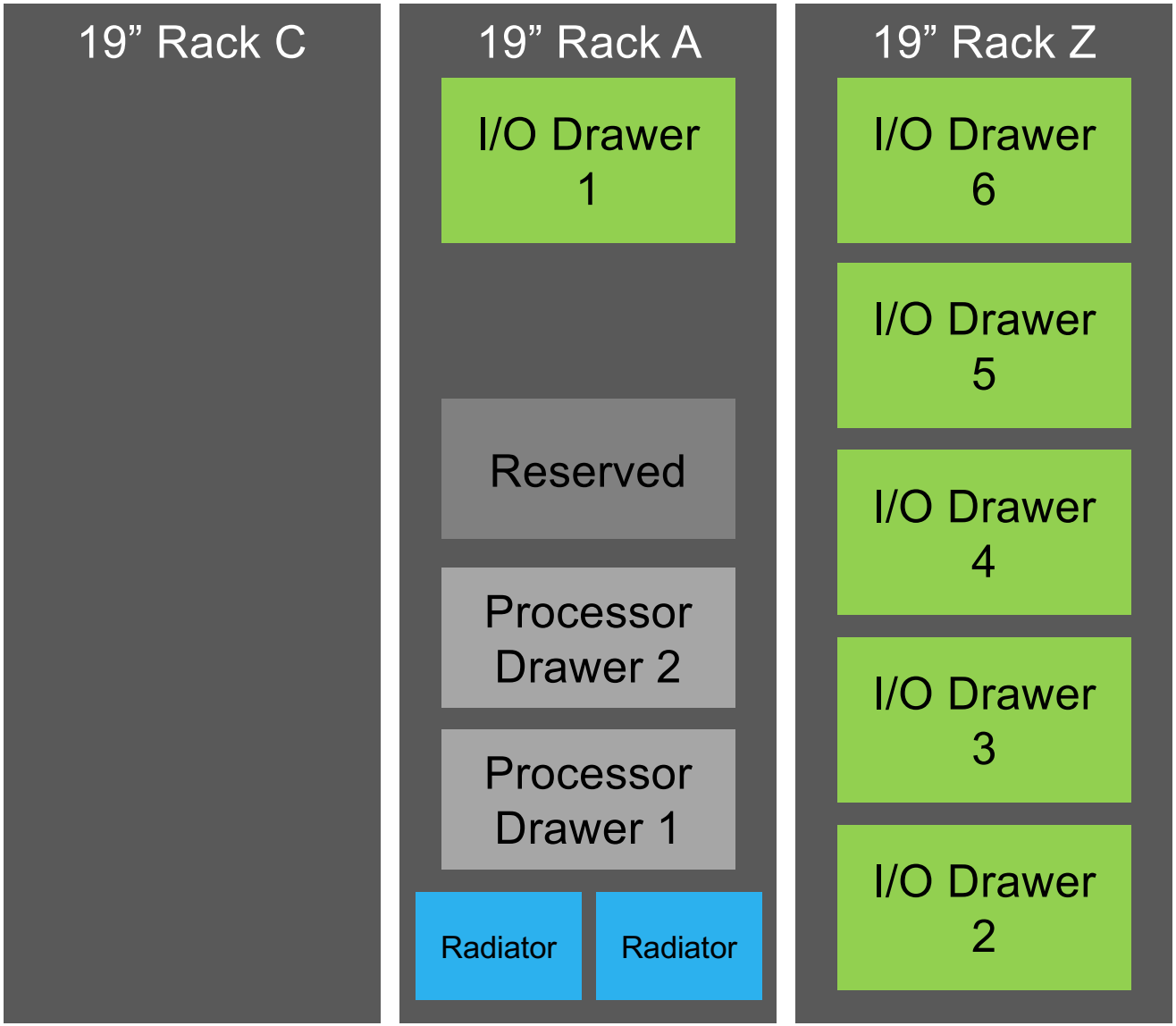
LinuxONE III LT1 Racks: Scenario 2 – iPDU Powered CPU Heavy



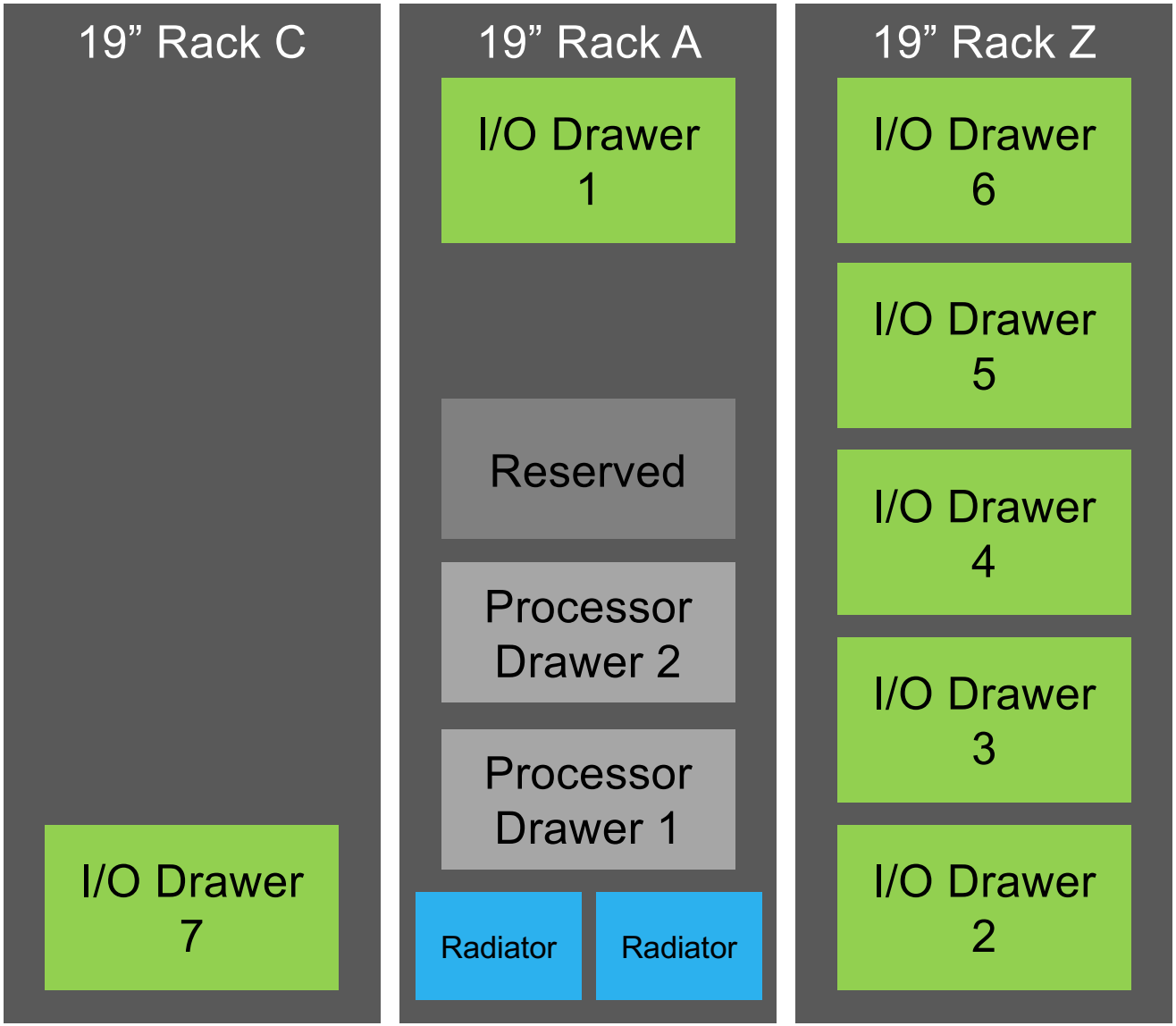
LinuxONE III LT1 Racks: Scenario 2 – iPDU Powered CPU Heavy



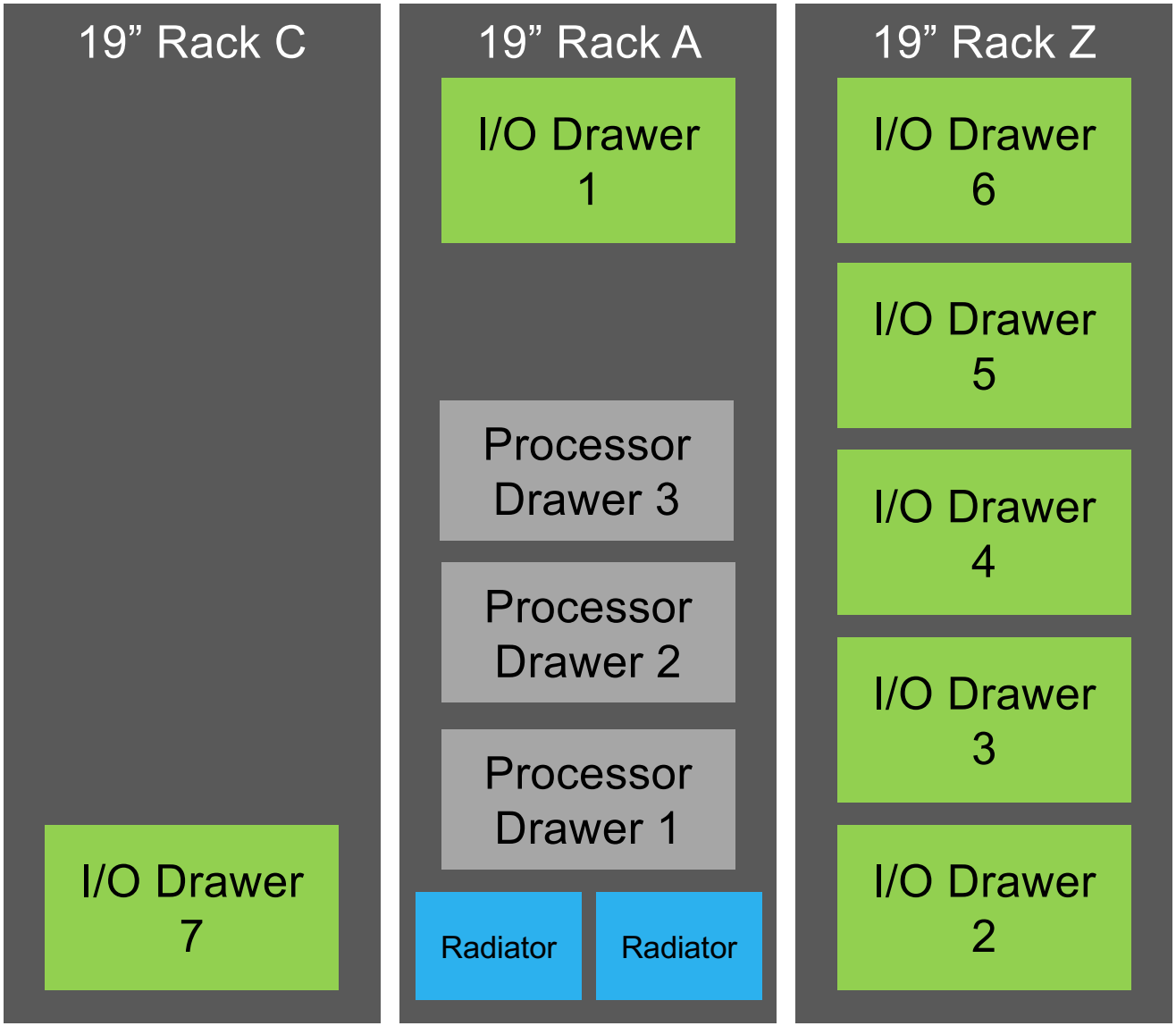
LinuxONE III LT1 Racks: Scenario 2 – iPDU Powered CPU Heavy



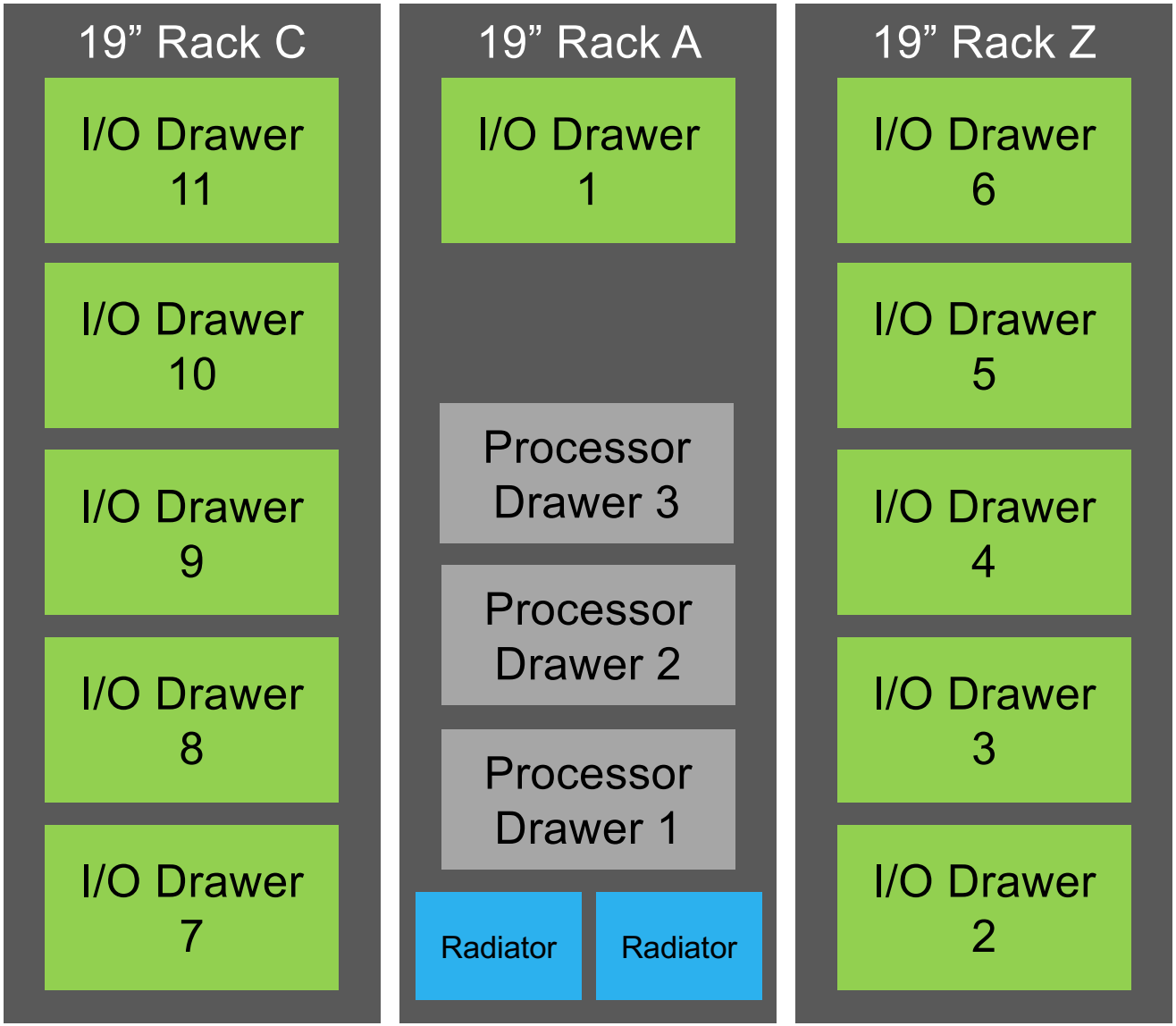
LinuxONE III LT1 Racks: Scenario 2 – iPDU Powered CPU Heavy



LinuxONE III LT1 Racks: Scenario 2 – iPDU Powered CPU Heavy



LinuxONE III LT1 Racks: Scenario 2 – iPDU Powered CPU Heavy

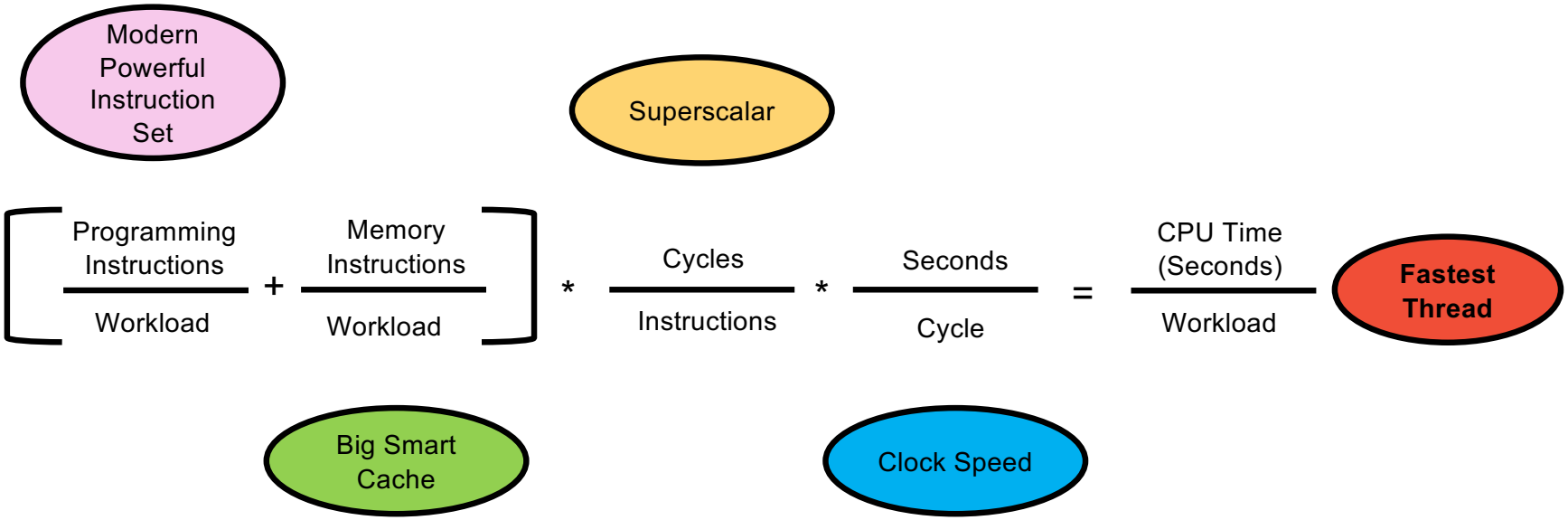


Max’d Out Config

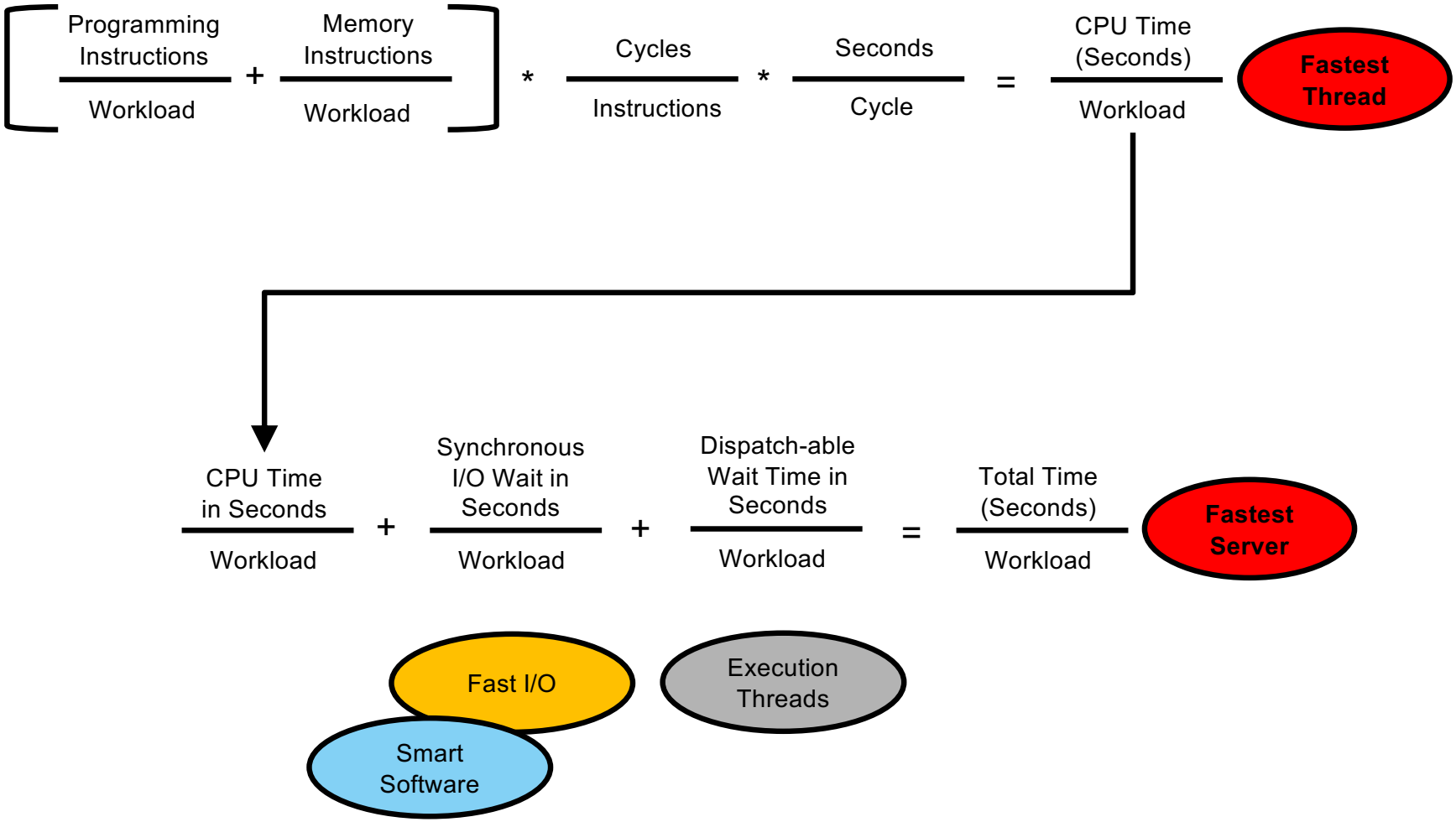
I'll Have Fries with That!

THE SPECIAL SAUCE

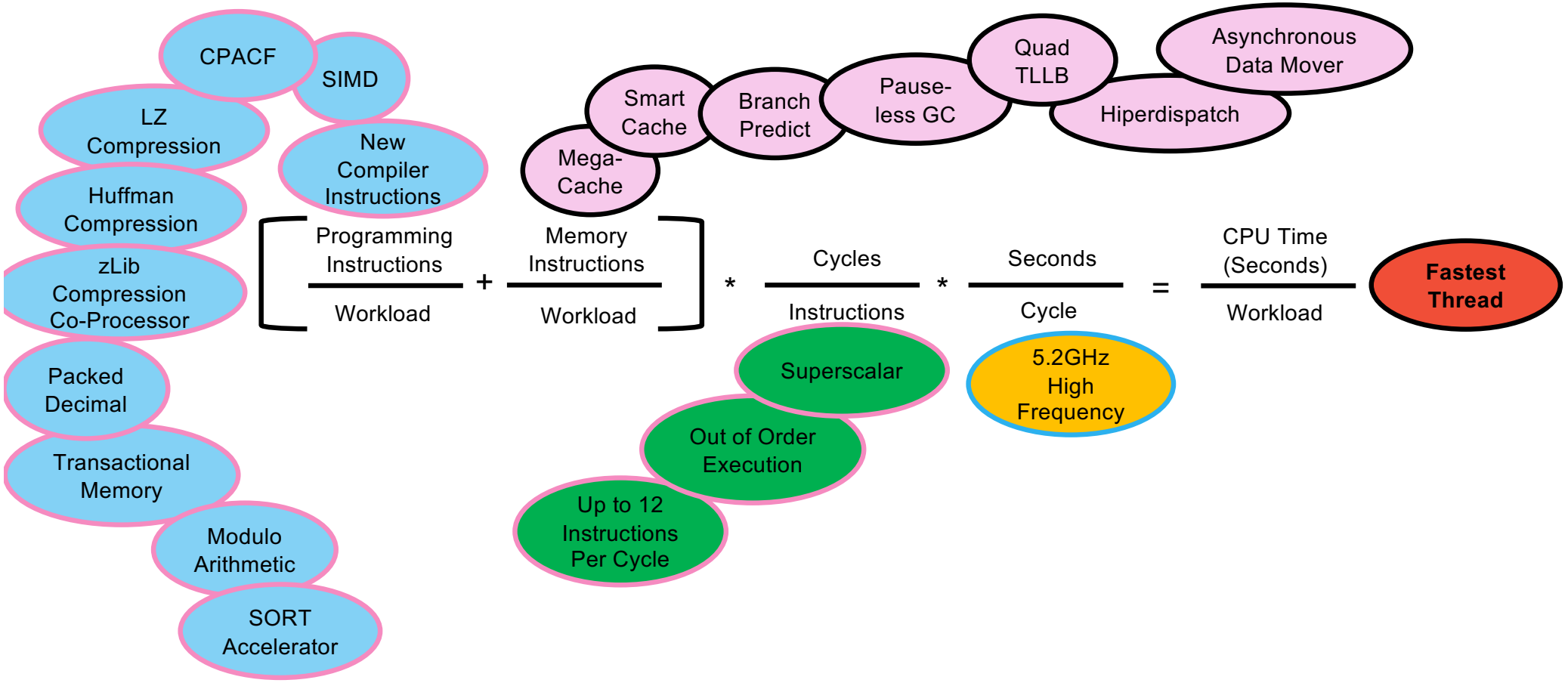
Computer Science Core Performance



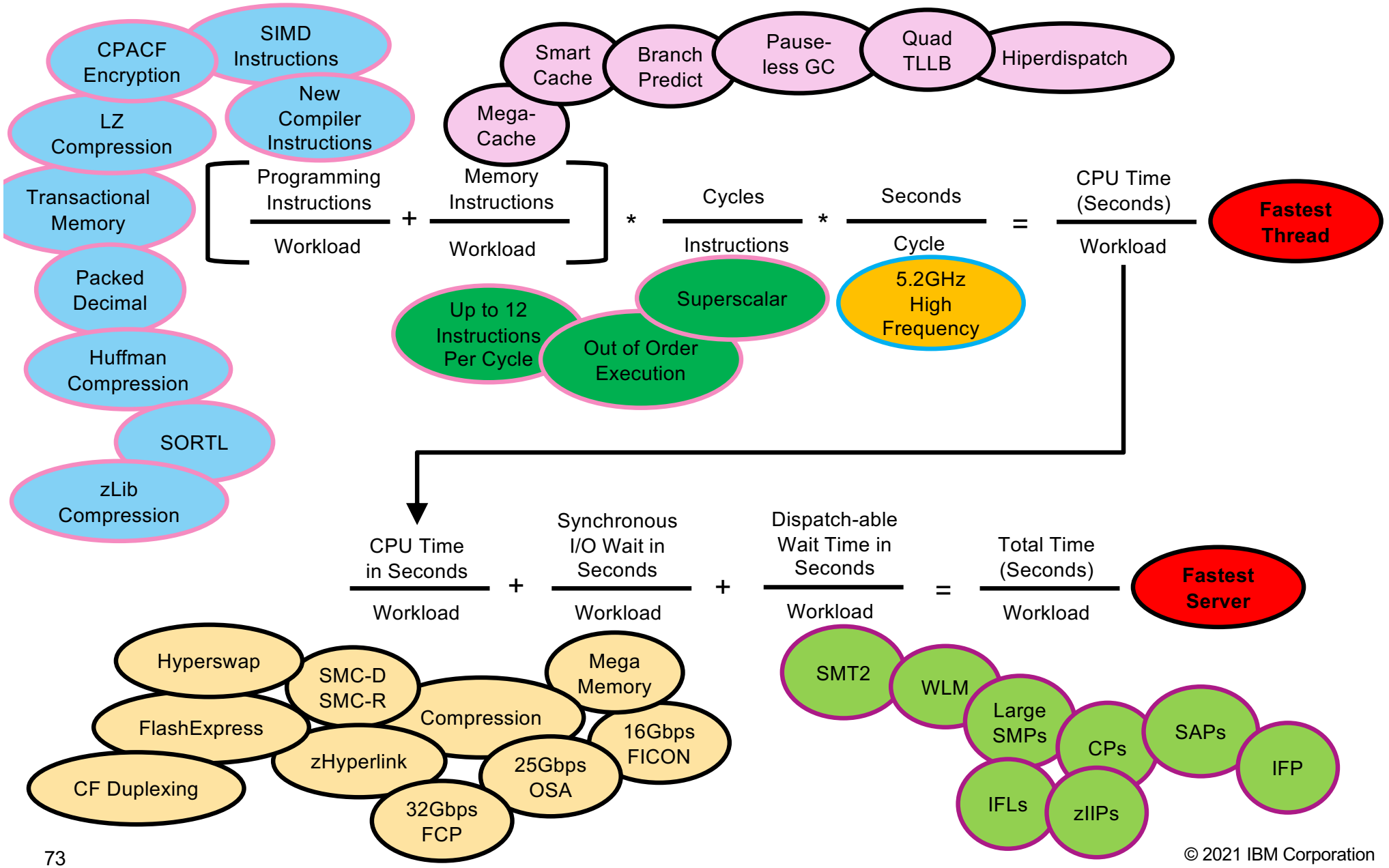
Computer Science Workload Performance



Computer Science Core Performance



Computer Science Workload Performance



MIPS

- Millions of Instructions Per Second
- Mileage In-between Processing Systems
- Meaningless Information Per Second
- Meaningless Information Propagated by Sales-reps

LSPR to the Rescue

- https://www-01.ibm.com/servers/resourcelink/lib03060.nsf/pages/lsprindex?OpenDocument
- Large Systems Performance Reference
  - The IBM Large System Performance Reference (LSPR) ratios represent IBM's assessment of relative processor capacity in an unconstrained environment for the specific benchmark workloads and system control programs specified in the tables. Ratios are based on measurements and analysis.
- A Capacity Planning Resource
- A Real(ish) set of Benchmarks for Mainframes
  - Almost certainly the best set of benchmarks in the industry
- zPCR – z Processor Capacity Reference – an Open Tool Full of LSPR Data & More
  - https://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS1381

CMOS
MIPSTORY

Historical Look At MIPS

Year	3rd Party $K/MIPS	MIPS per CORE	MIPS per CHIP	MIPS per BOARD (book or drawer or mcm)	Single Image MIPS	Notes	CORES per CHIP
1994	$ 30.00	15	15	61	61	9672 G1	1
1995	$ 20.00	22	22	186	186	9672 G2	1
1996	$ 13.00	45	45	357	357	9672 G3	1
1997	$ 9.00	63	63	446	446	9672 G4	1
1998	$ 5.50	117	117	1,069	1,069	9672 G5 Turbo	1
1999	$ 3.50	176	176	1,606	1,606	9672 G6 Turbo	1
2000	$ 2.50	235	235	765	3,061	z900 1Cs	1
2001	$ 2.40	235	235	765	3,061		1
2002	$ 2.20	278	278	1,091	3,804	z900 2Cs	1
2003	$ 2.00	443	876	1,715	11,391	z990	1
2004	$ 2.00	443	876	1715	11391		1
2005	$ 2.00	443	876	1715	11391		1
2006	$ 1.80	600	2,317	8,271	23,716	z9	2
2007	$ 1.50	600	2,317	8,271	23,716		2
2008	$ 1.40	990	3,701	10,177	43,426	z10	4
2009	$ 1.30	990	3,701	10,177	43,426		4
2010	$ 1.00	1,280	4,833	16,289	68,410	z196	4
2011	$ 1.00	1,280	4,833	16,289	68,410		4
2012	$ 1.00	1,650	9,097	26,604	103,699	zEC12	6
2013	$ 1.00	1,650	9,097	26,604	103,699		6
2014	$ 1.00	1,650	9,097	26,604	103,699		6
2015	$ 1.00	1,906	13,627	43,269	154,904	z13	8
2016	$ 1.00	1,906	13,627	43,269	154,904		8
2017	$ 1.00	2,020	17,783	50,521	195,496	z14	10
2018	$ 1.00	2,020	17,783	50,521	195,496		10
2019	$ 1.00	2,232	23,414	58,197	240,718	z15	12

9 CMOS Generations of IBM Z CORES and CHIPS and Drawers

1	z/OS-2.3 LSPR	zPCR V9.3	
2	LSPR Multi-Image Capacity Ratios		
3	IBM Z IFL CPs		
4	Values are representative of z/VM, KVM,		
5	Capacity basis: 2827-701 @ 1,650.0 MIPS		
6	Capacity for z/OS on z10 and later process		
8	Processor	Features	
9			Low
z900	z900 Turbo		
CHIP	2064-2Cx I1	1W IFL	278
BOARD	2064-2Cx I4	4W IFL	1,091
CEC	2064-2Cx I15	15W IFL	3,804
z990	z990/300		
CORE	2084-3xx I1	1W IFL	443
CHIP	2084-3xx I2	2W IFL	876
BOARD	2084-3xx I4	4W IFL	1,715
CEC	2084-3xx I32	32W IFL	11,391
z9 EC	z9 EC/700		
CORE	2094-7xx I1	1W IFL	600
CHIP	2094-7xx I4	4W IFL	2,317
BOARD	2094-7xx I16	16W IFL	8,271
CEC	2094-7xx I54	54W IFL	23,716
z10 EC	z10 EC/700		
CORE	2097-7xx I1	1W IFL	990
CHIP	2097-7xx I4	4W IFL	3,701
BOARD	2097-7xx I12	12W IFL	10,177
CEC	2097-7xx I64	64W IFL	43,426
z196	z196/700		
CORE	2817-7xx I1	1W IFL	1,280
CHIP	2817-7xx I4	4W IFL	4,833
BOARD	2817-7xx I15	15W IFL	16,289
CEC	2817-7xx I80	80W IFL	68,410
zEC12	zEC12/700		
CORE	2827-7xx I1	1W IFL	1,650
CHIP	2827-7xx I6	6W IFL	9,097
BOARD	2827-7xx I20	20W IFL	26,604
CEC	2827-7xx I101	101W IFL	103,699
z13	z13/700		
CORE	2964-7xx I1	1W IFL	1,906
CHIP	2964-7xx I8	8W IFL	13,627
BOARD	2964-7xx I30	30W IFL	43,269
CEC	2964-7xx I141	141W IFL	154,904
z14	z14/700		
CORE	3906-7xx I1	1W IFL	2,020
CHIP	3906-7xx I10	10W IFL	17,783
BOARD	3906-7xx I33	33W IFL	50,521
CEC	3906-7xx I170	170W IFL	195,496
z15	z15/700		
CORE	8561-7xx I1	1W IFL	2,232
CHIP	8561-7xx I12	12W IFL	23,414
BOARD	8561-7xx I34	34W IFL	58,197
CEC	8561-7xx I190	190W IFL	240,718

Core: 278 > 443 > 600 > 990 > 1280 > 1650 > 1906 > 2020 > 2232
Chip: 278 > 876 > 2317 > 3701 > 4833 > 9097 > 13627 > 17783 > 23414
Server: 3804 > 11,391 > 23,716 > 43,426 > 68,410 > 103,699 > 154,904 > 195,496 > 240,718

IBM Z Chip Capacity Metrics

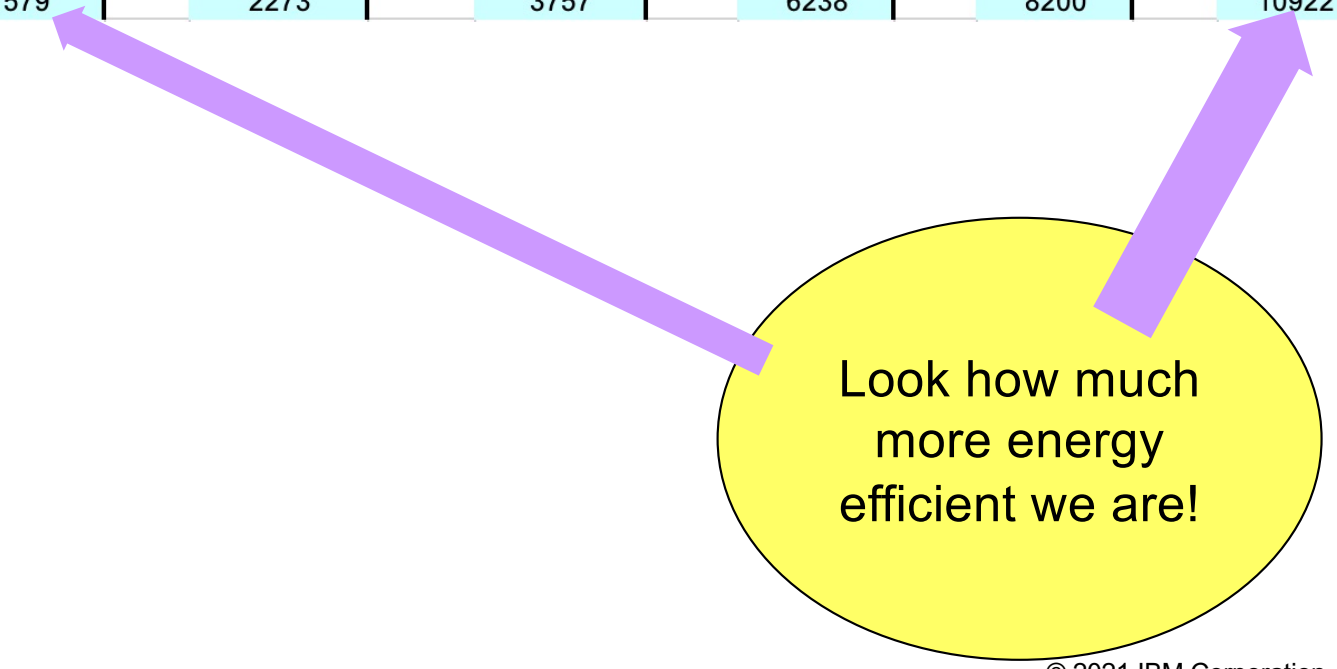
		2097	z10	2817	z196	2827	zEC12	2964	z13	3906	z14	8561	z15
Release Date			1Q 2008		3Q 2010		3Q 2012		Q1 2015		3Q 2017		3Q 2019
SMT-1			n/a		n/a		n/a		**1**		**1**		**1**
Max-MIPS/Core	**(MIPS)**		**990**		**1,280**		**1,650**		**1,906**		**2,020**		**2,232**
Max-MIPS/Chip	(MIPS)		3,701		4,833		9,097		13,627		17,783		23,414
Max-MIPS/Drawer	(MIPS)		10,177		16,289		26,604		43,269		50,521		58,197
Max-MIPS/CEC	**(MIPS)**		**43,426**		**68,410**		**103,699**		**154,904**		**195,496**		**240,718**
SMT-2			n/a		n/a		n/a		2		2		2
SMT-2 Factor			n/a		n/a		n/a		20%		25%		27.50%
Max-MIPS/Core	**(MIPS)**		**990**		**1,280**		**1,650**		**2,287**		**2,525**		**2,846**
Max-MIPS/Chip	(MIPS)		3,701		4,833		9,097		16,352		22,229		29,853
Max-MIPS/Drawer	(MIPS)		10,177		16,289		26,604		51,923		63,151		74,201
Max-MIPS/CEC	**(MIPS)**		**43,426**		**68,410**		**103,699**		**185,885**		**244,370**		**306,915**
Max-MIPs/Core/GHz SMT1	(MIPS)		225		246		300		381		388		429
Max-MIPs/Core/GHz SMT2	(MIPS)		225		246		300		457		486		547

Look how much more work we can do!

Look how much more work we do per cycle!!!

IBM Z Chip Technology Metrics

		2097 z10	2817 z196	2827 zEC12	2964 z13	3906 z14	8561 z15
Release Date		1Q 2008	3Q 2010	3Q 2012	Q1 2015	3Q 2017	3Q 2019
Max-GHz	(GHz)	4.4	5.2	5.5	5.0	5.2	5.2
Chip Transistors	(Billion)	1.00	1.40	2.75	3.99	6.10	9.10
Max Cores per Chip		4	4	6	8	10	12
Max Chips per MCM or Drawer		5	6	6	6	6	4
Max MCMs or Drawers per CEC		4	4	4	4	4	5
Max Chips per CEC		20	24	24	24	24	20
Max Core per CEC		80	96	144	192	240	240
Net Cache/Core	(M)	5.6	15.7	20.8	41.6	30.3	49.6
PU Chip Transistors	(B)	1.00	1.40	2.75	3.99	6.10	9.10
SC Chip Transistors	(B)	1.6	1.5	2.75	7.1	9.7	9.7
MAX KWs		27.5	30.1	27.6	29.8	29.8	28.1
Max-MIPS/CEC/KW	(MIPS)	1579	2273	3757	6238	8200	10922



Moore’s Law

- https://en.wikipedia.org/wiki/Moore%27s_law

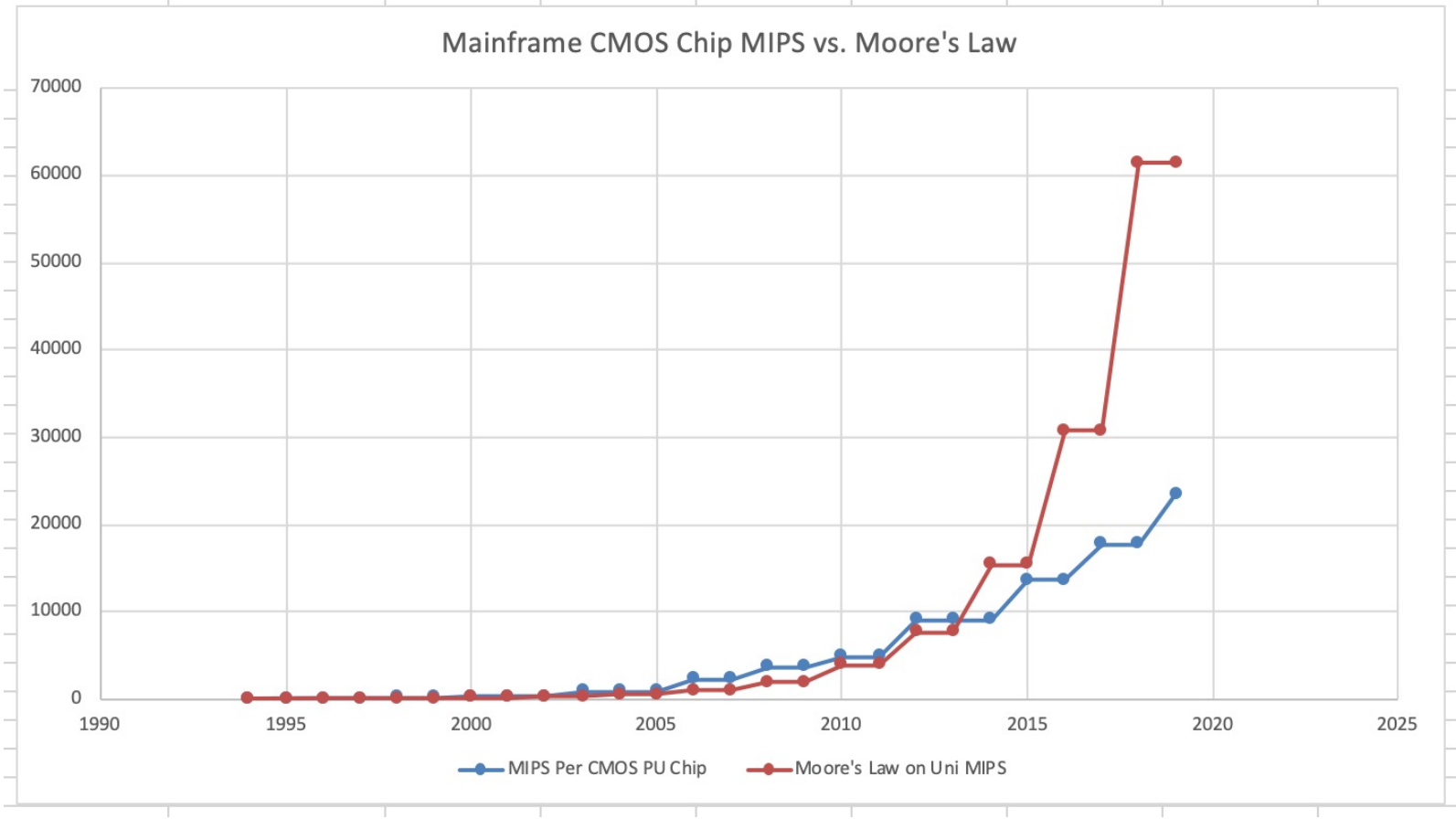
- **Moore's law** is the observation that the number of transistors in a dense integrated circuit doubles about every two years. The observation is named after Gordon Moore, the co-founder of Fairchild Semiconductor and CEO of Intel, whose 1965 paper described a doubling every year in the number of components per integrated circuit[2] and projected this rate of growth would continue for at least another decade.[3] In 1975,[4] looking forward to the next decade,[5] he revised the forecast to doubling every two years.[6][7][8]

MOoRE ON MOORE'S LAW for MORE

- n April 2005, Gordon Moore stated in an interview that the projection cannot be sustained indefinitely: "It can't continue forever. The nature of exponentials is that you push them out and eventually disaster happens." He also noted that transistors eventually would reach the limits of miniaturization at atomic levels:
- In terms of size [of transistors] you can see that we're approaching the size of atoms which is a fundamental barrier, but it'll be two or three generations before we get that far—but that's as far out as we've ever been able to see. We have another 10 to 20 years before we reach a fundamental limit. By then they'll be able to make bigger chips and have transistor budgets in the billions.[100]

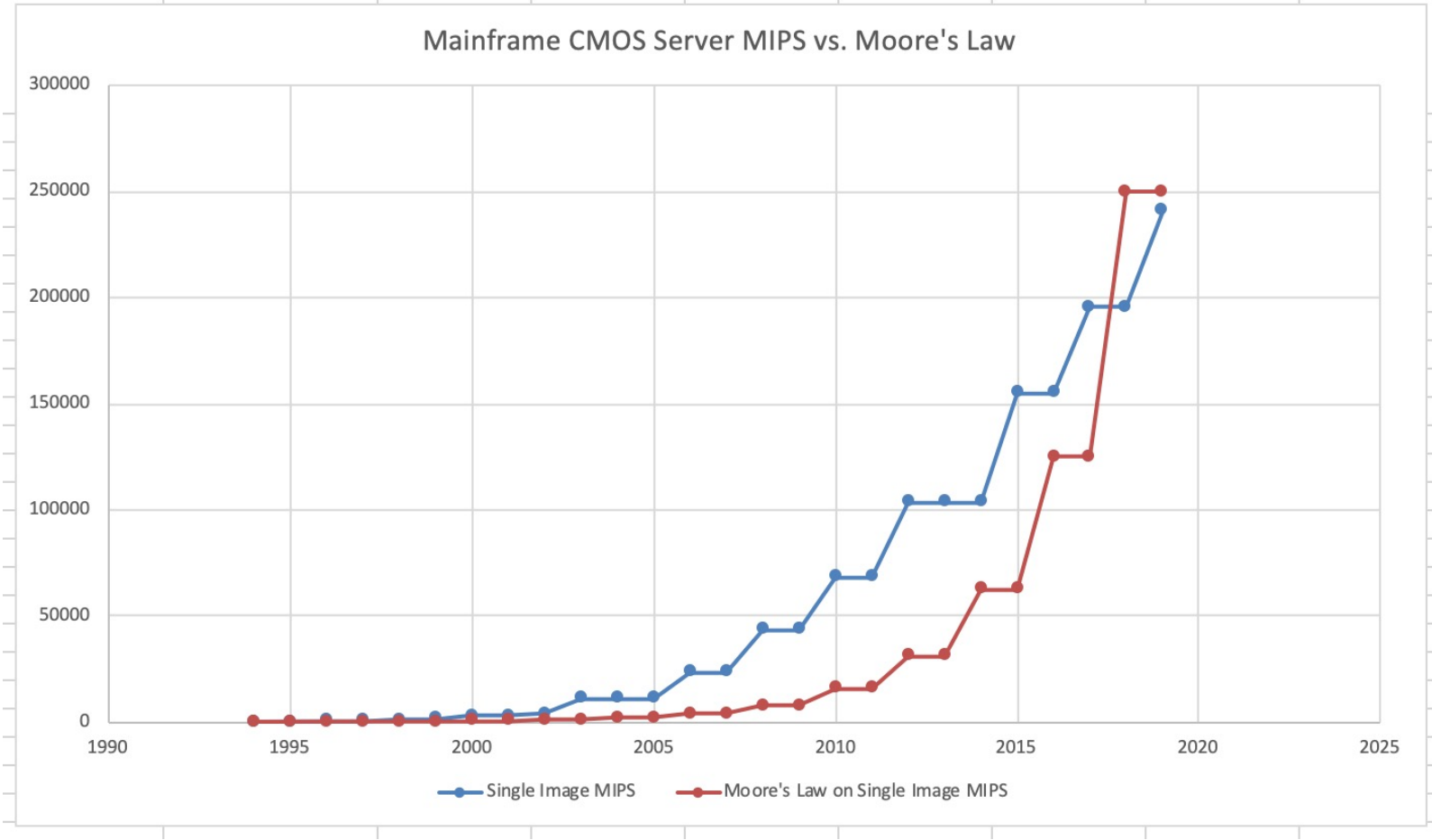
IBM Z Chip Chip MIPS vs. Moore's Law

Year	MIPS Per CMOS PU Chip	Moore's Law on Uni MIPS
1994	15	15
1995	22	15
1996	45	30
1997	63	30
1998	117	60
1999	176	60
2000	235	120
2001	235	120
2002	278	240
2003	876	240
2004	876	480
2005	876	480
2006	2317	960
2007	2317	960
2008	3701	1920
2009	3701	1920
2010	4833	3840
2011	4833	3840
2012	9097	7680
2013	9097	7680
2014	9097	15360
2015	13627	15360
2016	13627	30720
2017	17783	30720
2018	17783	61440
2019	23414	61440



IBM Z Chip Powered Platform MIPS vs. Moore's Law

Year	Single Image MIPS	Moore's Law on Single Image MIPS
1994	61	61
1995	186	61
1996	357	122
1997	446	122
1998	1069	244
1999	1606	244
2000	3061	488
2001	3061	488
2002	3804	976
2003	11391	976
2004	11391	1952
2005	11391	1952
2006	23716	3904
2007	23716	3904
2008	43426	7808
2009	43426	7808
2010	68410	15616
2011	68410	15616
2012	103699	31232
2013	103699	31232
2014	103699	62464
2015	154904	62464
2016	154904	124928
2017	195496	124928
2018	195496	249856
2019	240718	249856



The End is Near

THE END IS INEVITABLE

IBM z15 Chip and LinuxONE III Holistic Performance

Scale Out

- Execute up to **1 trillion secure web transactions per day** on a LinuxONE III server
- Scale-out to **2.4 million Docker containers** in a LinuxONE III
- Consolidate **100 x86 running https** to 1 LinuxONE III
- **20%** TCO reduction over 5 years

Competitive Results

- Up to **2.3x more throughput** running pgBench 9.6.1 read-only benchmark on **PostgreSQL** 11.1 vs. compared x86
- Up to **2.1x more throughput** running YCSB 0.15 (write-heavy) benchmark on **MongoDB** 4.0.6 vs. compared x86
- Up to **2.8x more throughput** running Cassandra-stress benchmark on ScyllaDB 2.3.1 vs. compared x86 platform
- Up to **2.6x more throughput** running DayTrader 7 benchmark on **WAS Liberty** vs. compared x86

IBM Private Cloud and SSC

- Run **2.4x more AcmeAir containers per core** deployed on ICP on LinuxONE III vs. compared x86
- Scale-up AcmeAir on SSC4ICP to 12 IFLs with **79% scaling efficiency** on a LinuxONE III LPAR

Data Compression

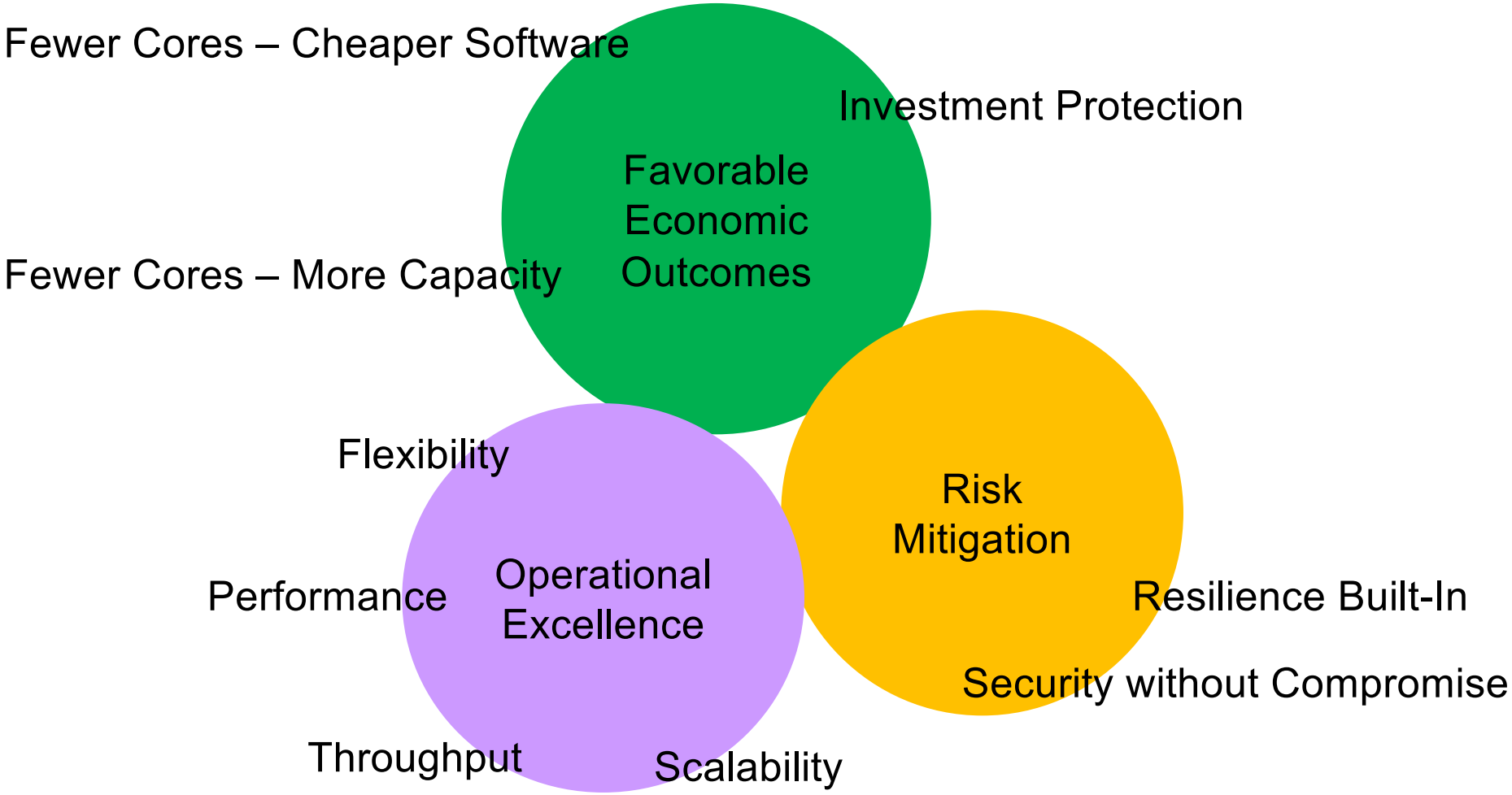
- Compress up to **275 GB web data per second** on a LinuxONE III server
- Up to **2.7x lower latency** **and** up to **1.8x less CPU** at **2.6x less network bandwidth** running NGINX web server on LinuxONE III using Integrated zEDC to **compress before encryption** (up to **30x lower latency** and up to **28x less CPU** when compared to software compression)
- Up to **27x faster** **Db2 LUW database backup** using **30x less CPU time** using Integrated zEDC vs. compared x86 with software compression. (up to **20x faster** using **23x less CPU time** vs. LinuxONE Emperor II)
- Up to **6.5x faster** **MongoDB database dump** using **5.5x less CPU time** using Integrated zEDC vs. compared x86 with software compression. (up to **5.4x faster** using **5.4x less CPU time** vs. LinuxONE Emperor II)
- **3.2x faster** MongoDB Dump on **encrypted** btrfs on LinuxONE III using Integrated Accelerator for zEDC compression versus compared x86 with software compression

Instant Recovery

- Start **100 KVM guests**, in parallel, **2.1x faster** on LinuxONE III LPAR versus using a compared x86

Note: Client results will vary
Please see approved claims for details and disclaimers

The z15 Chip is Awesome – and so are LinuxONE III Servers!



The Last Page

Questions?
Comments?
Requests?

THANK YOU for your Time

mbauman@us.ibm.com