

Everything You Ever Wanted to Know about Power in Routers and Switches

(and Quite a Bit More)

Lane Wigley
Technical Marketing Engineer

cisco Live !

Cisco Webex App

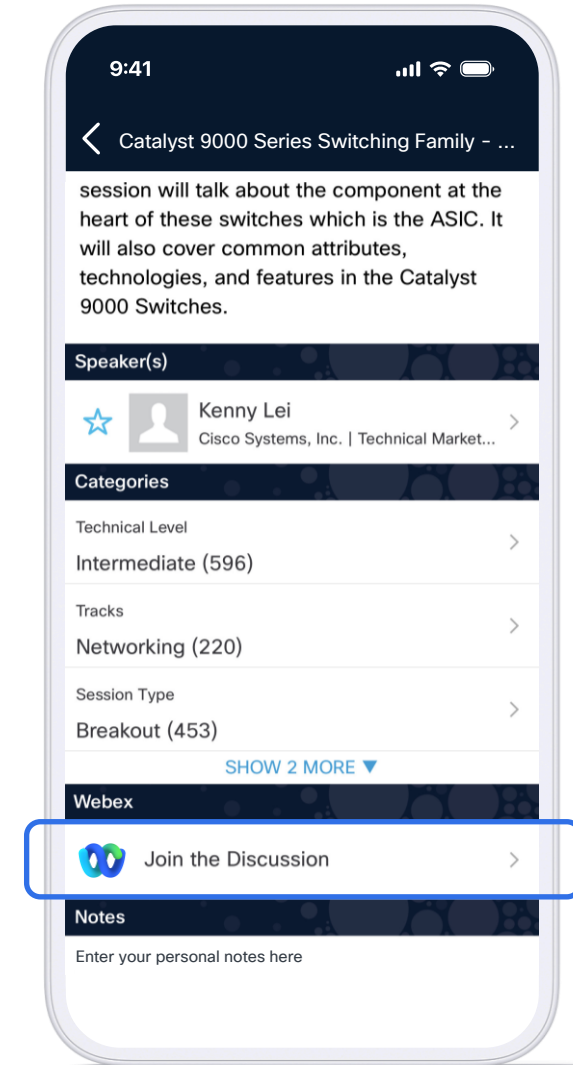
Questions?

Use Cisco Webex App to chat with the speaker after the session

How

- 1 Find this session in the Cisco Live Mobile App
- 2 Click “Join the Discussion”
- 3 Install the Webex App or go directly to the Webex space
- 4 Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until 14 November 2025.



<https://ciscolive.ciscoevents.com/ciscolivebot/#BRKGRN-2161>

Agenda

- 01 Intro
- 02 Typical / Peak / Provisioning
- 03 Power Supplies
- 04 Power Allocation
- 05 Facility Planning
- 06 Power Consumption
- 07 Examples
- 08 Power Evaluation
- 09 Power Saving Guidance

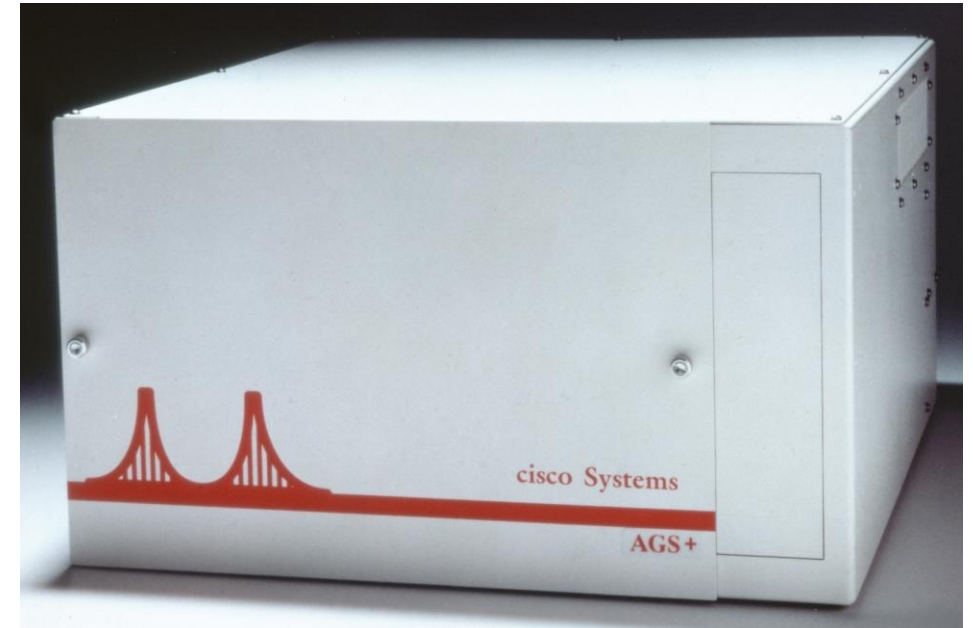
Intro

- TAC intern starting on 1995 then TAC Routing Protocols
- "ISP team", BU Critical Accounts, Technical Marketing since 2005
- Current focus on analog side of hardware
 - Power
 - Cooling (air/liquid)
 - Optics
 - Rack evolution



My Power Journey

- Focused on high end routers since forever (NCS 5500, NCS 6000, ASR 9000, GSR, 7500, 7000, AGS+?)
 - Mostly working with 8000 Series Routers today - examples use 100/400G
- Traditionally a top 5 concern for SPs
- #1 concern for largest networks today
 - Key driver of new chips and platform design
 - Extensive software optimizations
 - Power testing and modeling
- Not just CAPEX/power bill anymore
 - Sustainability
 - Opportunity cost (servers/GPUs)



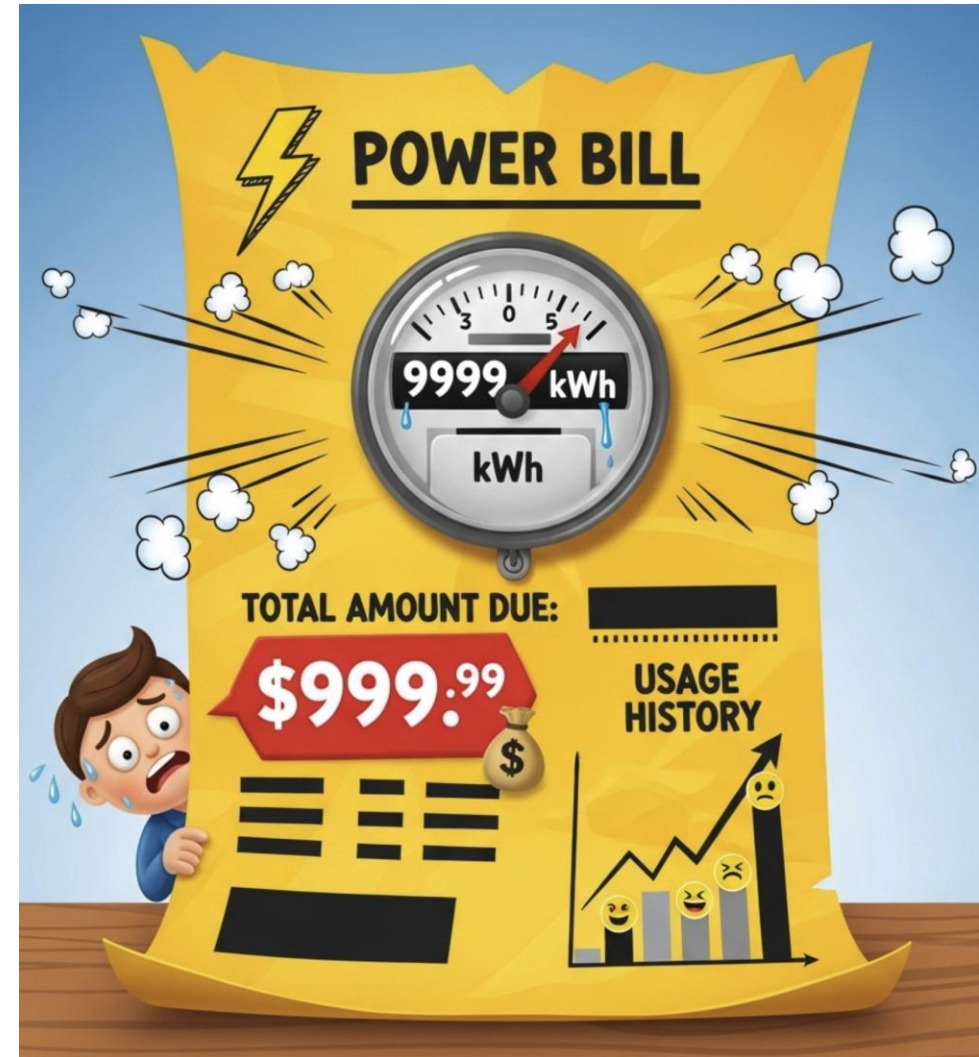
“The Power Numbers”

Typical (Power Bill)

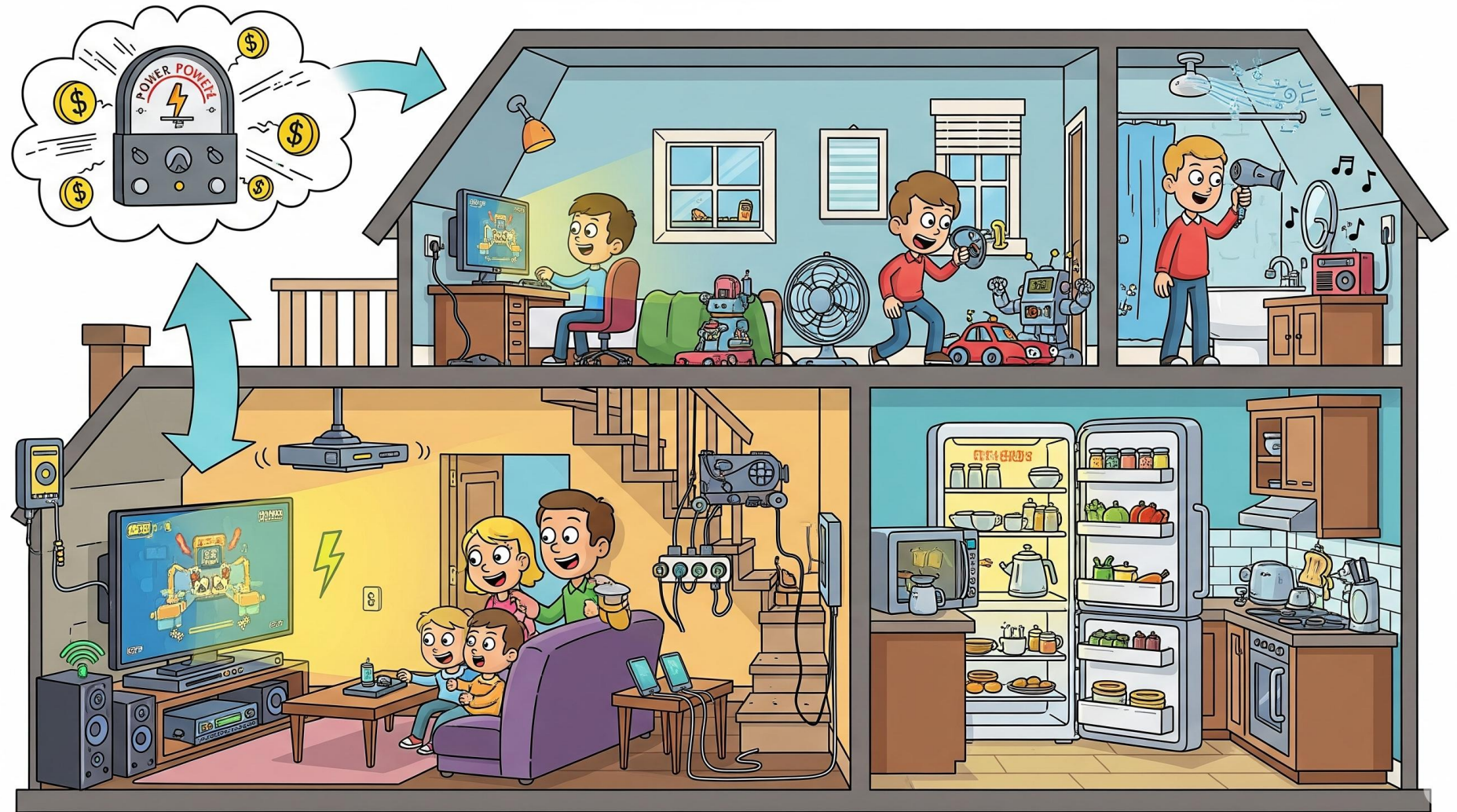
Peak (Circuit Breakers)

Provisioning (Peak w/ Redundancy)

Facility Planning (e.g., HVAC)



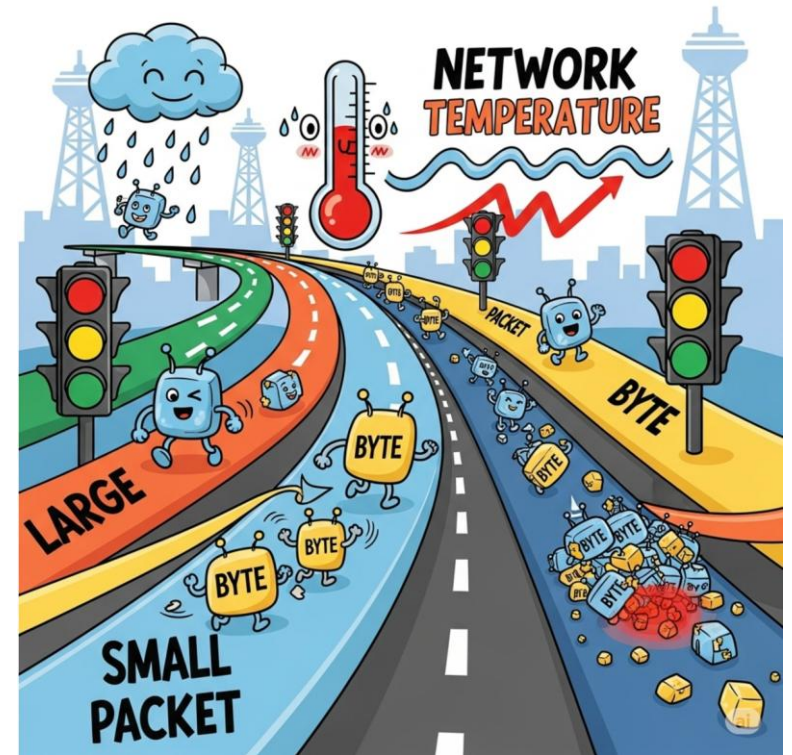
Typical Power



How Well Do You Know Your Network?

Answer for your most important router/switch type

- Traffic load over a full month (24x7x30 all interfaces)
 - Q: Show of hands – do you know within 10%?
 - Q: Hands down if it's more than 50%, 40%, 30%, 20%, 10%
- Average packet size
 - Q: Raise of hands – average size within 10%
 - Q: >1000B, 999-800B, 799-500B, <500B
- Do bursts matter?
- Facility temperature
 - Average 22-27C
 - How often do they exceed 35C/40C?



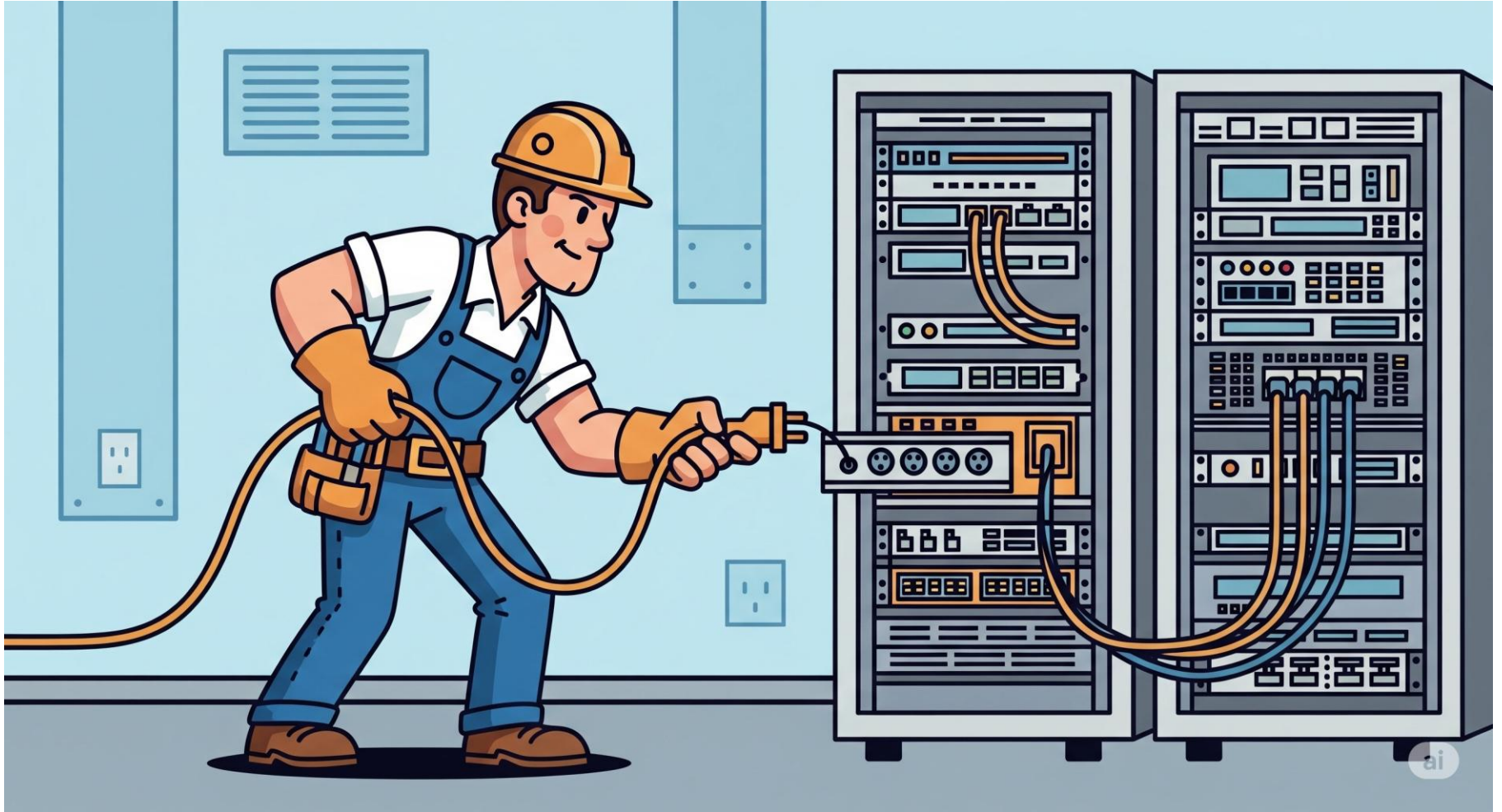
Peak Power



Peak = Everything Everywhere All at Once

- CPUs
- Memories
- ASICs (small packets @ 100% on all ports)
- Max fans
- Optics
- All cards, all ports
- Must consider component variation & temp

Provisioning

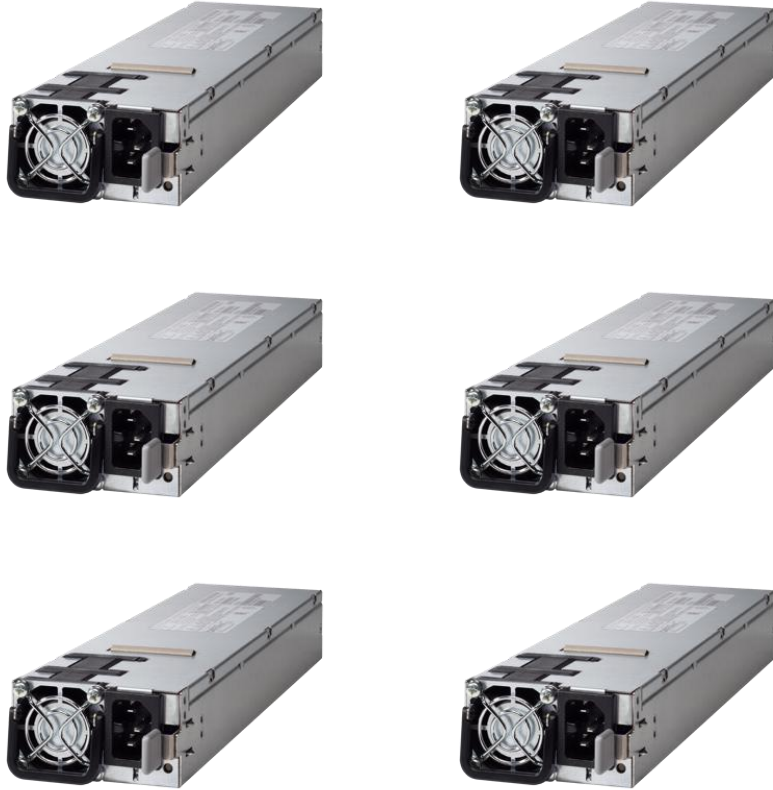


Approaches to Populating Power Supplies

- Supply failure – “N+1 protection”
- Feed failure – “N+N protection”
- Both – “Double failure”
- Max PSUs – “Job protection”
 - Avoid need to add power with future upgrades
 - Avoid exhausted facility budget



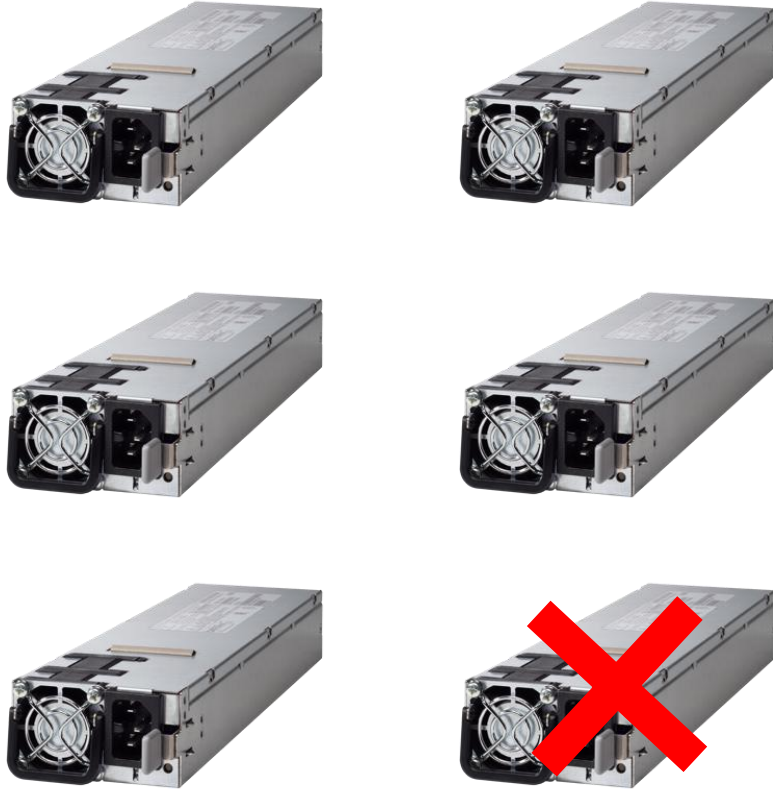
2 kW Power Supplies – No Redundancy



$$6 \times 2 \text{ kW} = \mathbf{12 \text{ kW}}$$

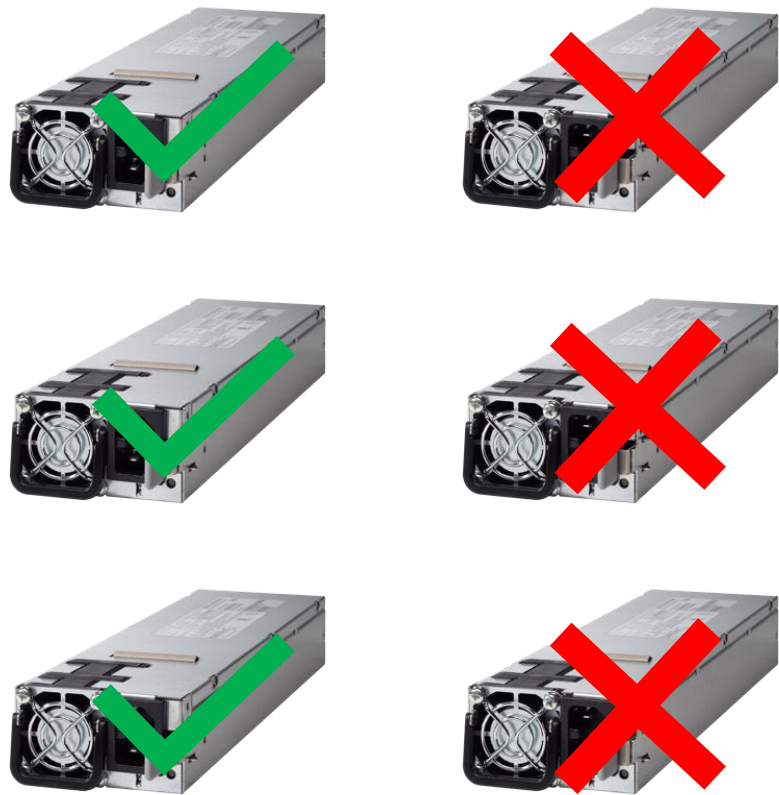
Available for allocation
“Global power pool”

Supply (N+1) Redundancy



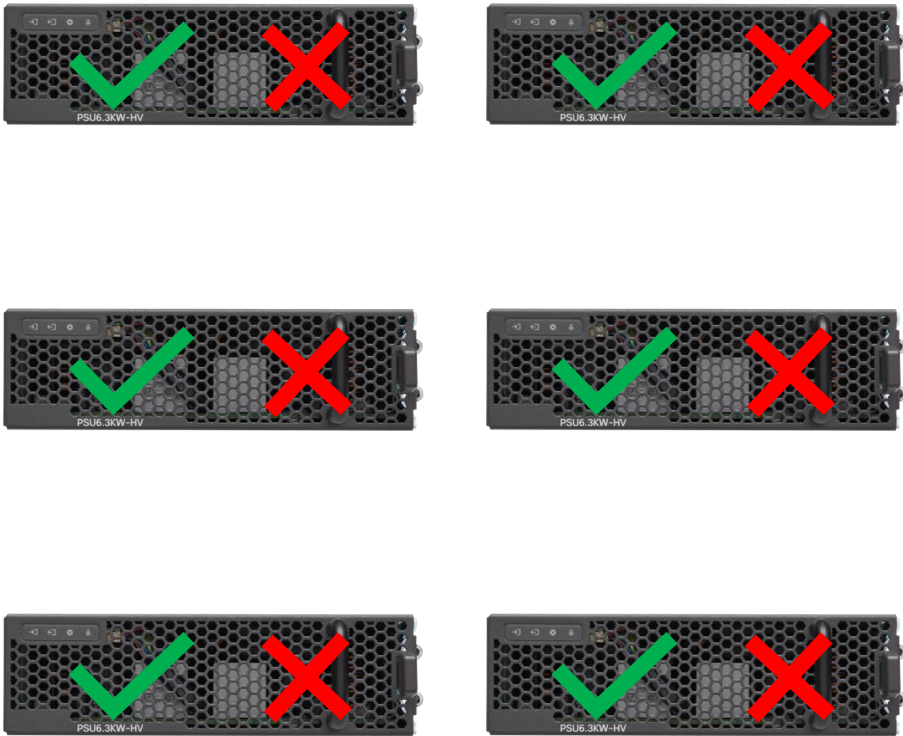
$$(6-1) \times 2 \text{ kW} = \mathbf{10 \text{ kW}}$$

Feed (N+N) Redundancy



Single Feed PSU | 3 x 2 kW = **6 kW**

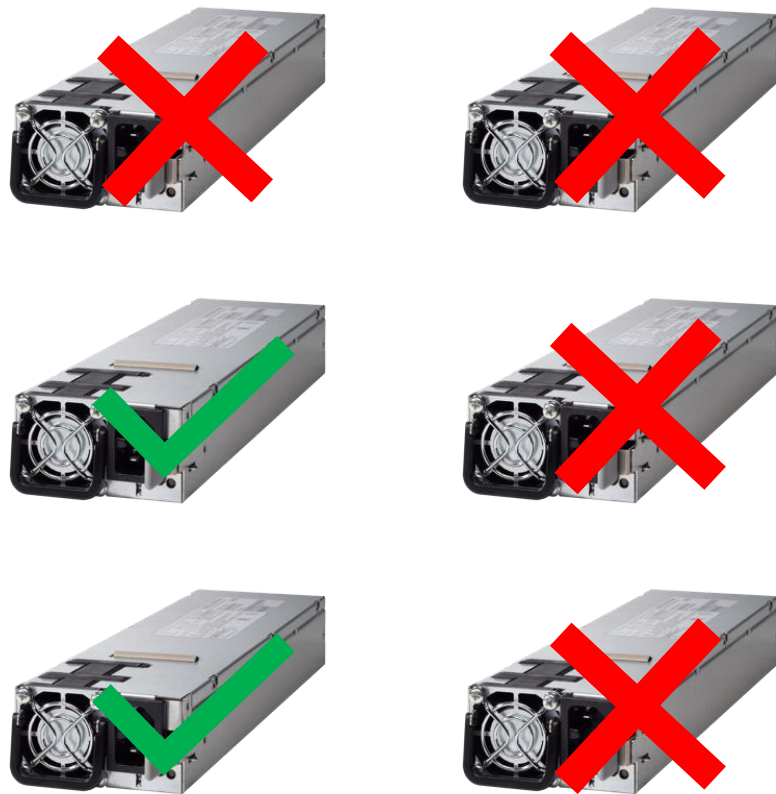
or



Dual Feed PSU | 6 x 1 kW = **6 kW**

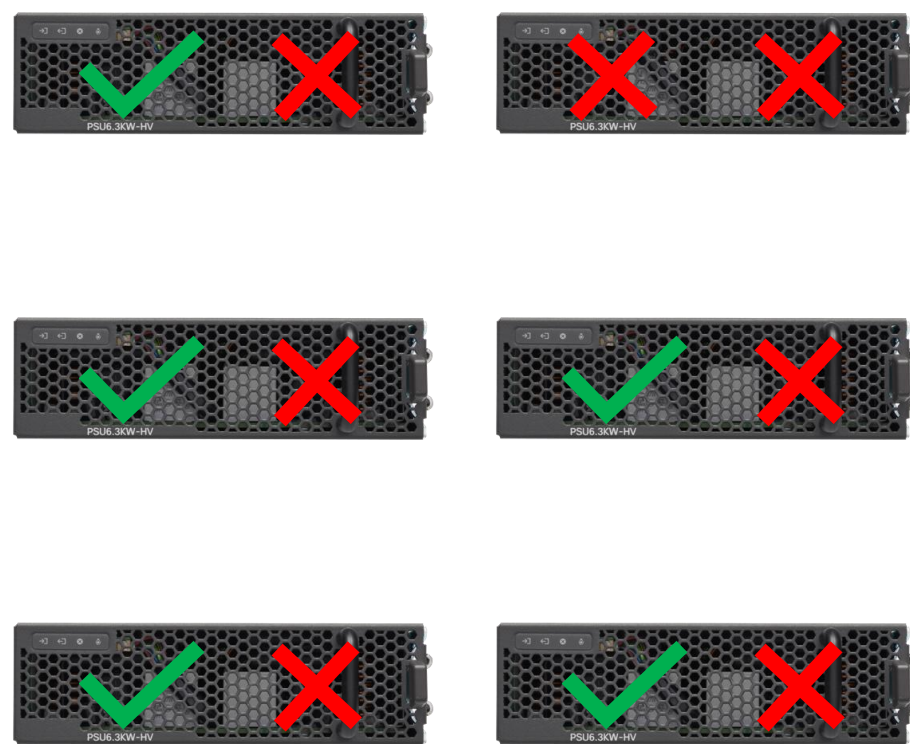
Feed (N+N) & Supply Redundancy (Double Failure)

How many provision for double failure in multi-PSU systems?



$2 \times 2 \text{ kW} = 4 \text{ kW}$

or



$5 \times 1 \text{ kW} = 5 \text{ kW}$

Power Allocation Values

- Represent Peak Power
- Component variation
 - Boards, Chips, and Optics
- Platforms have a range of granularity
 - Assume worst case optics or actual optics present?
 - Assume all modular slots are full?

Allocation Approaches – Line Card Example

Traditional Approach – assume worst-case optics (usually 10km, but all about same power)



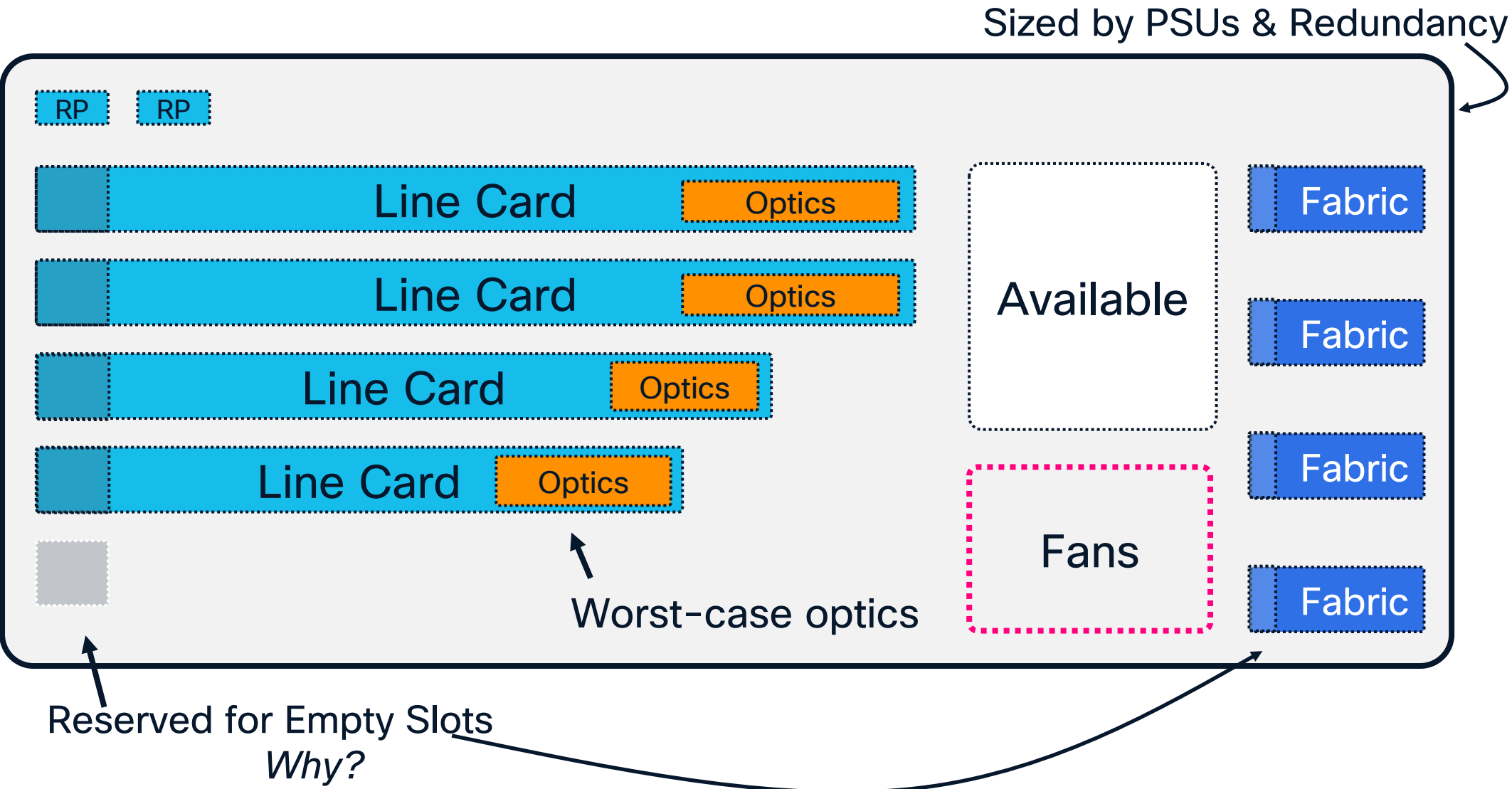
What if we support 36x 25W 400G ZR+ coherent but use 10W FR?

For a very large chassis, 9 kW over-allocation – 10 extra power supplies w/ N+N !!!

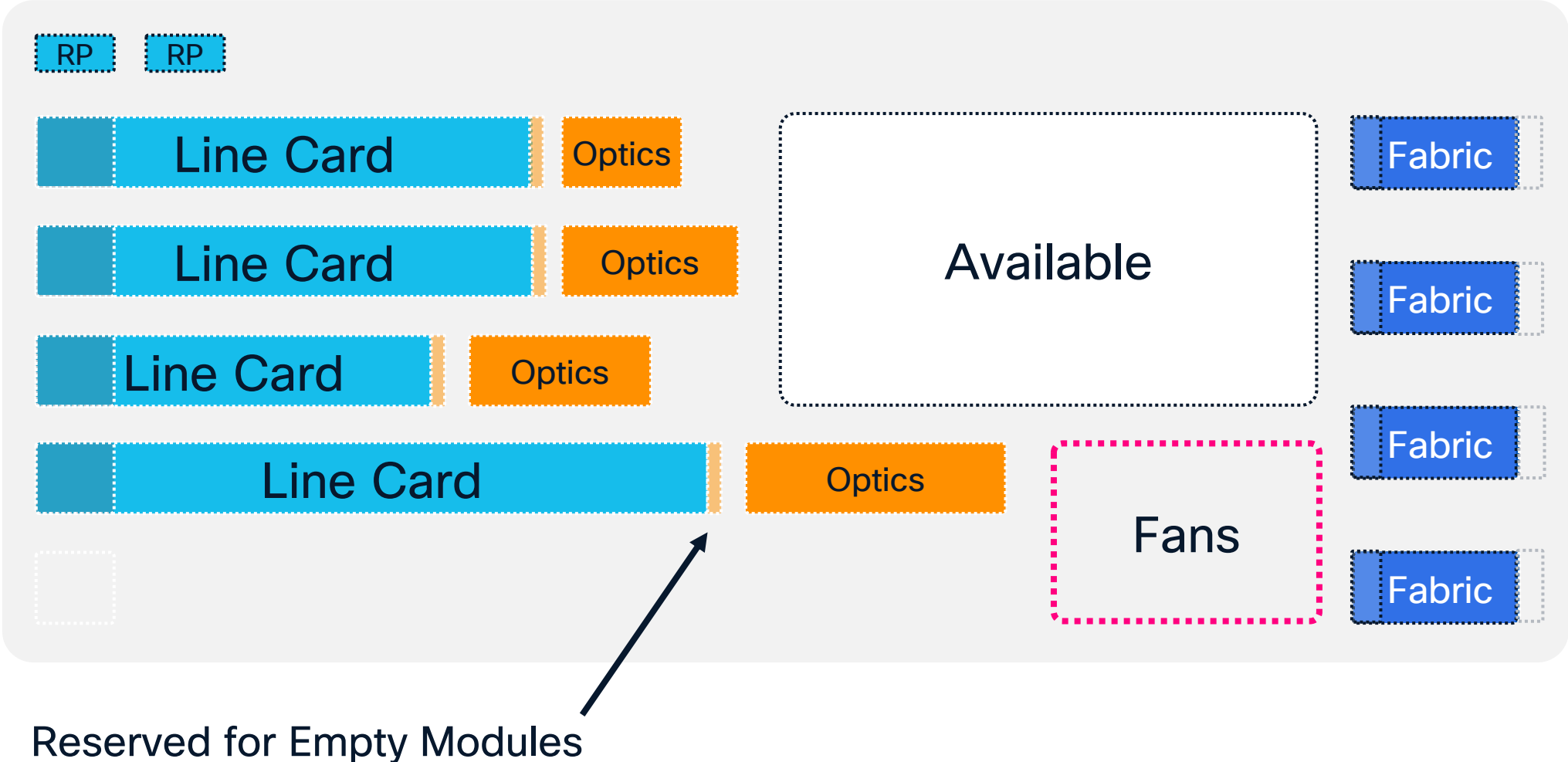
Better approach – optics (and other stuff) allocation on demand



Allocating Power from Global Pool – Simple Model



Allocating Power – Advanced Model



Facility Planning

Comparison of Parameters

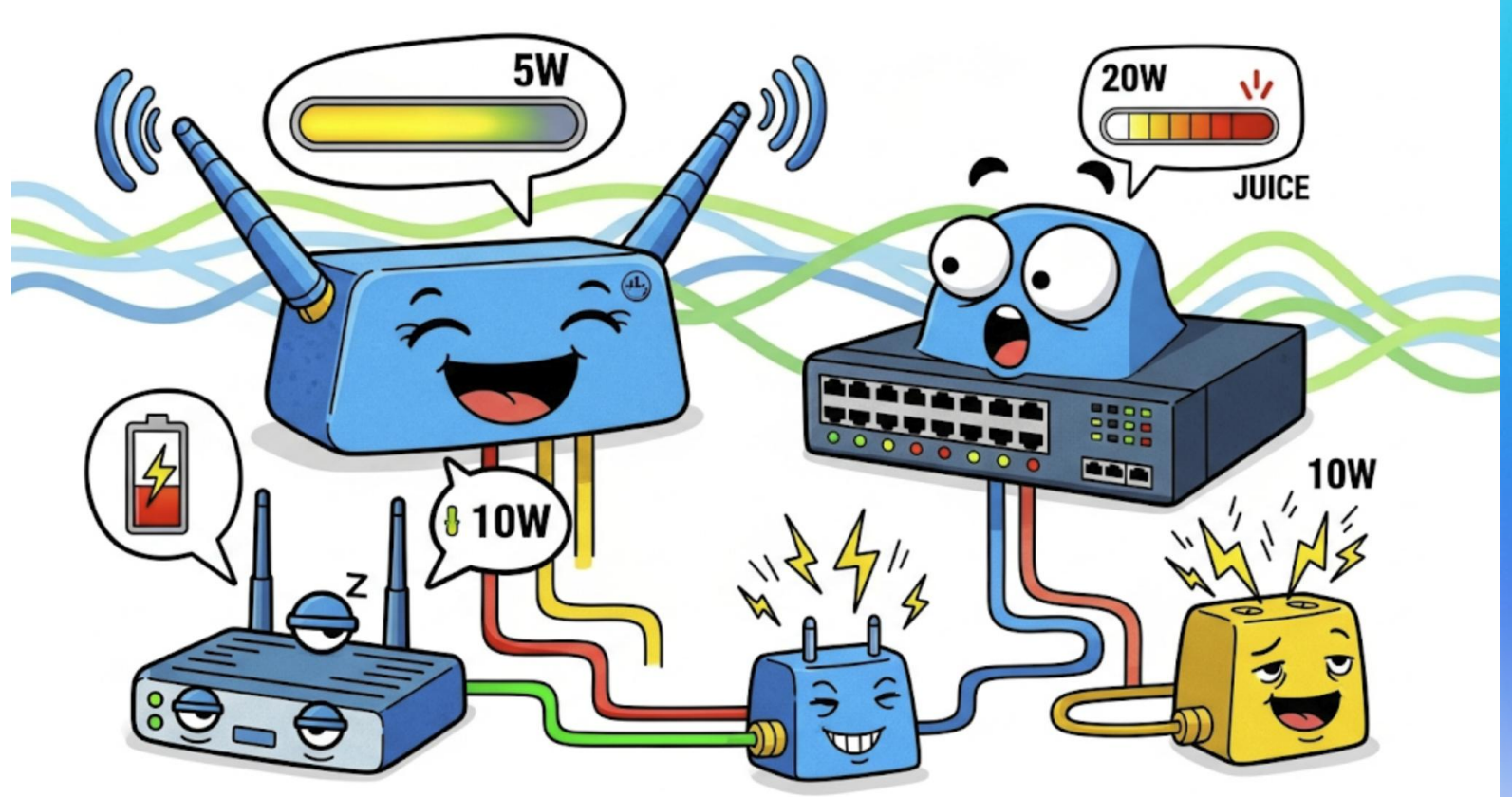
	Typical	Facility Planning	Peak
Traffic	1000B / 25%	800B / 75%	200-700B* / 100%
Payload	Random	Random	Random
CPU/DRAM	Light	Light	Max
Silicon/Board variation	Average	Average	Worst-case
Optics variation	Average	Average	Worst-case
Ambient temp (depends on platform)	22-27C @ Sea Level	40C	40C/50C/55C/65C

Summary

	Impacts	Notes
Typical	Power bill (operator-owned) Scope 2 emissions	Varies by application optics/load/temperature
Peak	Circuit breakers Battery backup Allocation requirements	Worst-case for every component
Redundancy	Number & size of power supplies	N+1, N+N, etc.
Facility Planning	HVAC Rack spacing Room-level power capacity	

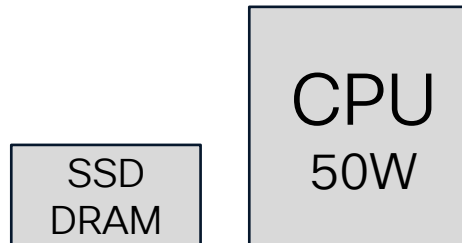
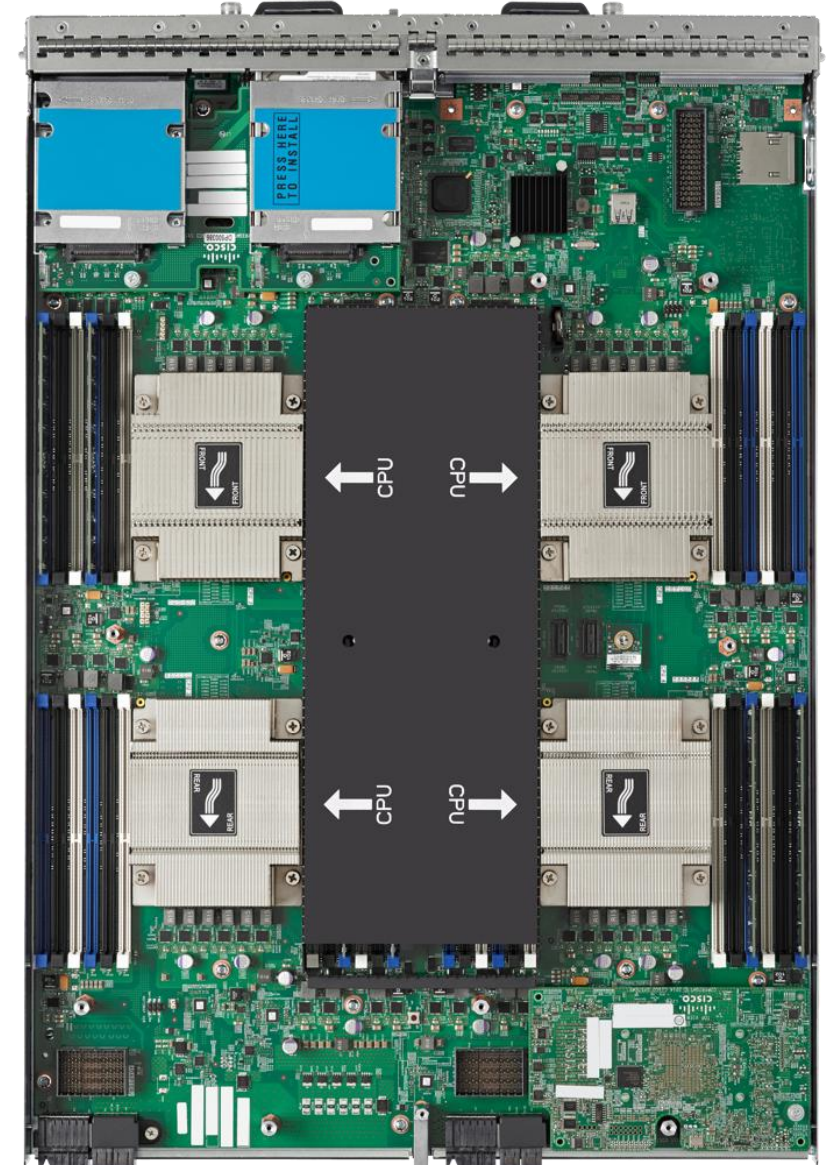
Power Consumption in Systems

CPU/DRAMs
NPUs
Fans
Optics
Small components
Analog
Power supplies
Examples



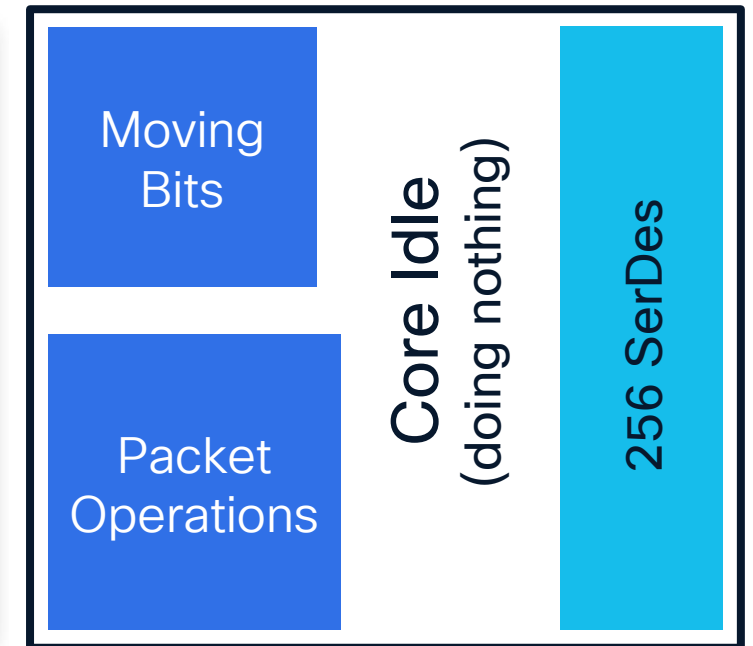
CPU / Memory

- CPU
 - Max power varies based on CPU type and load
 - Generally, good at reducing power with low loads
 - Most routers don't need server-class CPUs
- Other stuff
 - CPU DRAM
 - Flash (may drive cooling reqs)
 - Lots of small bits (console, USB)



Network Processors

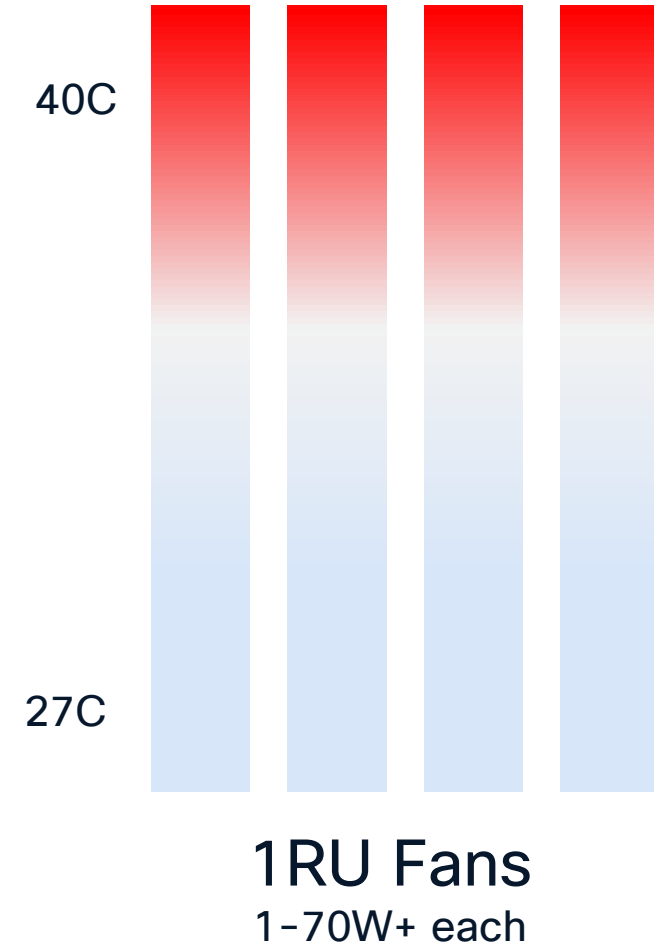
- Highly dependent on silicon process and BW/feature/scale capabilities
- Idle Power – Leakage Current
- I/O – SerDes 
 - Possible to power down
- Power from load 
- Wide range
 - Access – ~30W
 - 51.2T Leaf/Spine – 700W+



12.8T chip – 400W

Fans

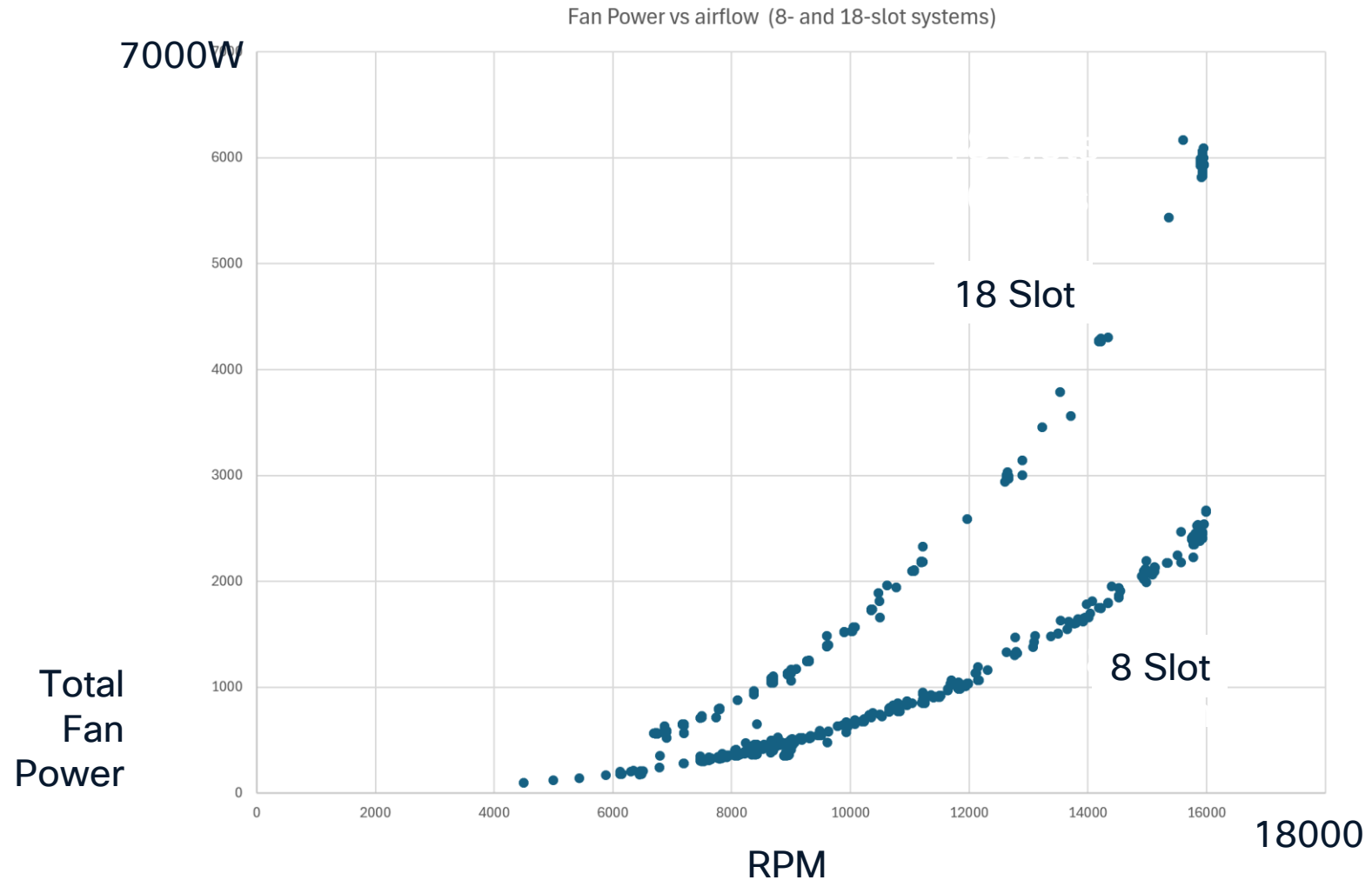
- Fan power varies greatly (cubic with airflow)
- Control may be very simple
 - 60/80/100% based on ambient
- Or extremely complex and consider
 - Line cards (for modular systems)
 - Optics types
 - Airflow direction
 - Real-time temps of <anything>



Real Fan Power Data

- Key factor in provisioning
- Rarely actually need max
- *Q: Raise your hand if*
 - Above 35C past year
 - Above 40C

High End System Airflow vs. Power



Optics Power

- Consider module-to-module variability & aging
- Module temp (DOM, not ambient)
- Traffic
 - Doesn't matter for grey optics
 - Slight variation for coherent (40+ km) based on error correction load
- Loss in voltage conversion & power distribution



Grey ($\leq 10\text{km}$)			Coherent (40+ km)	
<div>3</div>	<div>10</div>	<div>16</div>	<div>17-22</div>	<div>30</div>
100G	400G	800G	400G	800G

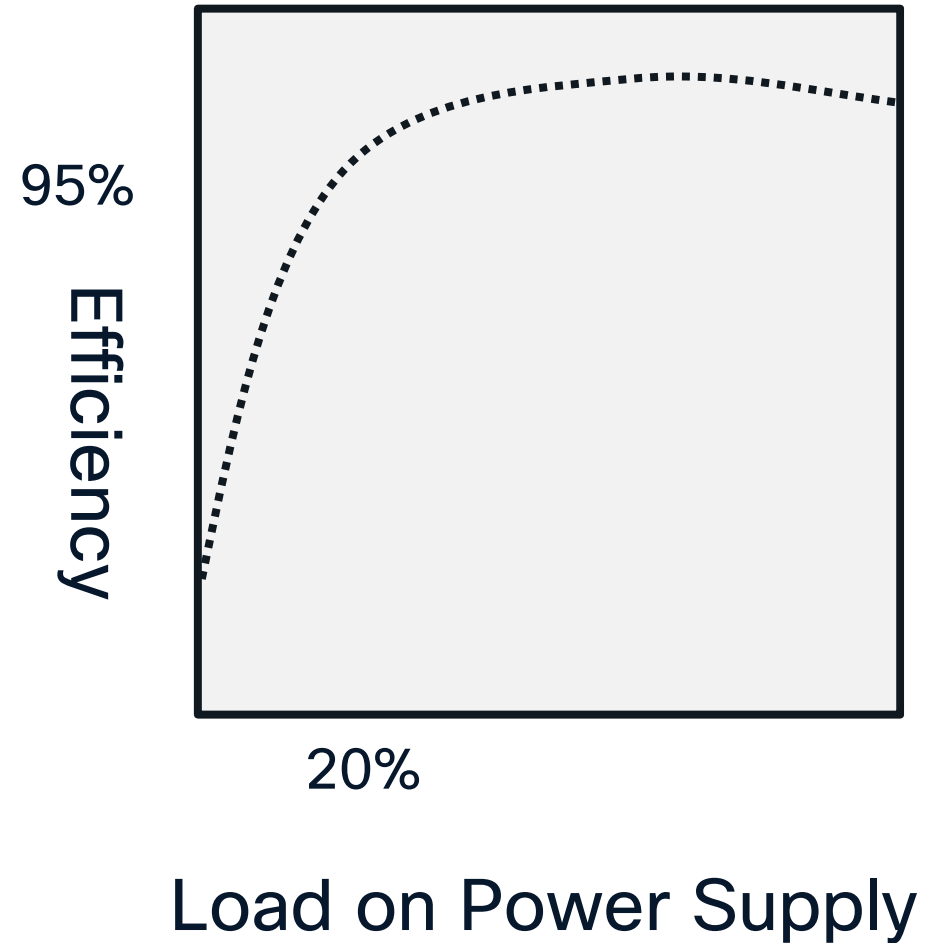
Analog – Voltage Conversion & Plane Loss

Efficiency varies with
Load, Type, Voltage, Temperature

Overprovisioning PSUs reduces efficiency

Voltage Conversion
e.g., 12V to 3.3V for optics, 0.8V for NPUs
POL – Point of Load

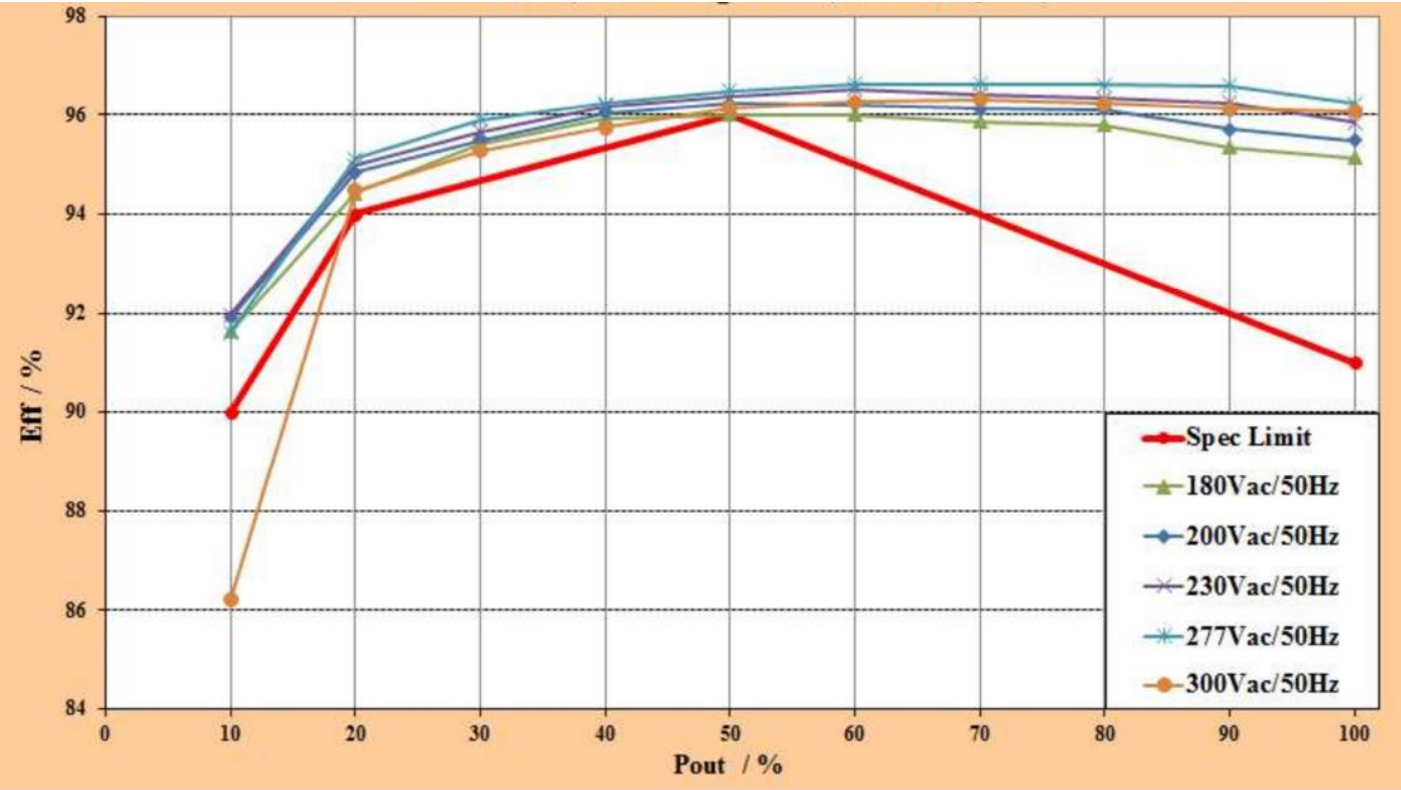
Moving current through system



80 PLUS Ratings

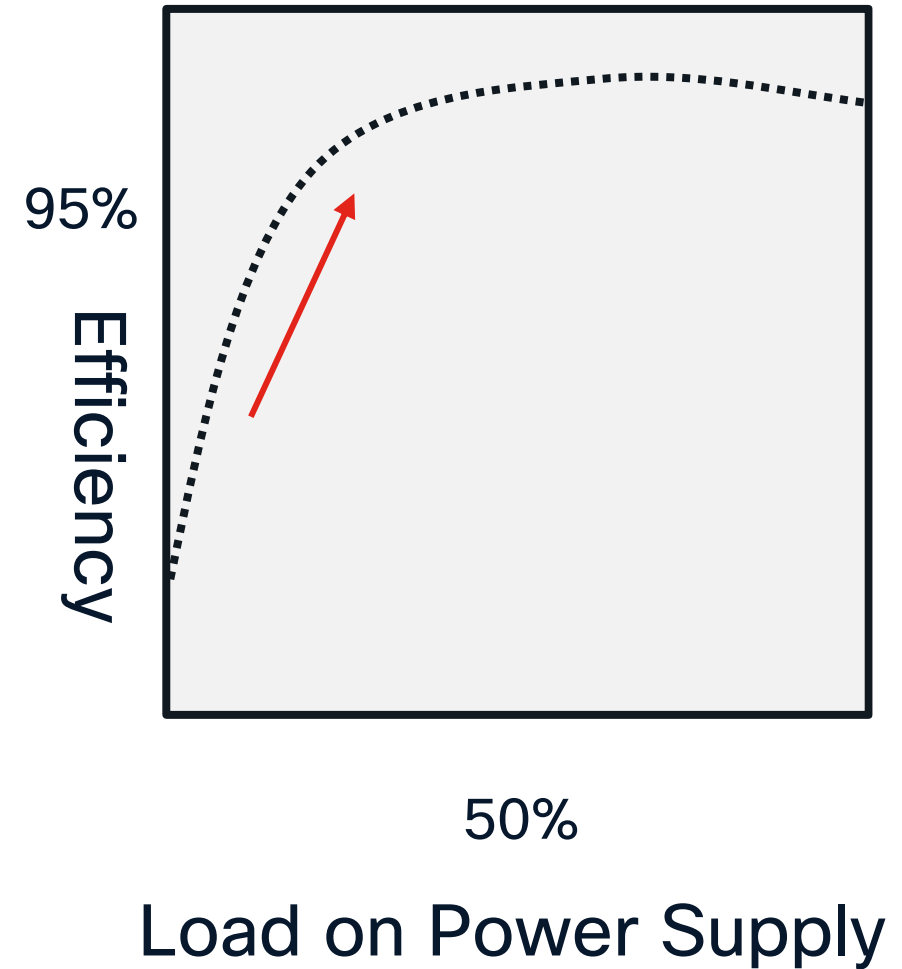


	Loading	80 Plus	Bronze	Silver	Gold	Platinum	Titanium
Efficiency	20%	80%	82%	85%	87%	90%	90%
	50%	80%	85%	88%	90%	92%	92%
	100%	80%	82%	85%	87%	89%	94%



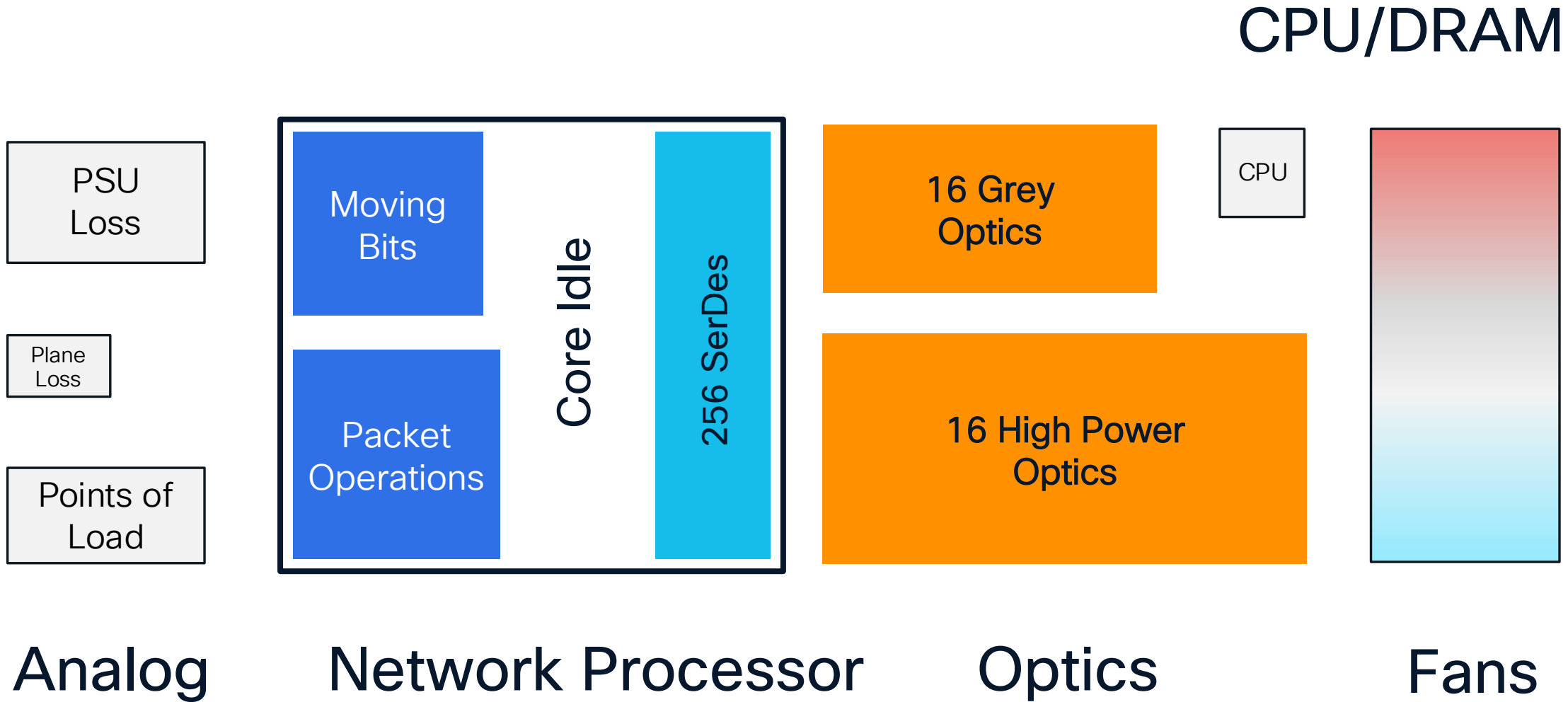
Balancing Redundancy and PSU Efficiency/Load

- More PSUs -> Lower Load
- Lower Load -> Lower Efficiency
- Warm standby
- HVAC/HVDC will improve efficiency
 - *Anyone deployed High Voltage (~400V)?*



Putting it All Together

32x 400 GbE router w/ components to scale



Power in a Large System (IOS XR Example)



RP/0/RP0/CPU0:ios#show env power

Total output power capacity (N + 1)	:	31500W +	6300W
Total output power required	:	17884W	
Total power input	:	14598W	
Total power output	:	13636W	

Power Module	Supply Type	-----Input-----				-----Output---		Status
		Volts	A/B	Amps	A/B	Volts	Amps	
0/PT2-PM0	PSU6.3KW-HV	212.0/211.7	5.7/5.6			54.2	40.7	OK
0/PT2-PM1	PSU6.3KW-HV	212.3/212.0	5.7/5.8			54.6	41.8	OK
0/PT2-PM2	PSU6.3KW-HV	212.0/212.3	5.9/5.9			54.6	43.2	OK
Total of Power Modules:		14598W/68.7A				13636W/250.5A		

Location	Card Type	Power Allocated Watts	Power Used Watts	Status
0/RP0/CPU0	8800-RP	95	81	ON
0/RP1/CPU0	8800-RP	95	78	ON
0/0/CPU0	(LC type 1)	2494	2076	ON
0/1/CPU0	(LC type 2)	1705	1197	ON
0/FC0	(Fabric 1 of 8)	417	417	ON
0/FC5	(Fabric 5 of 8)	417	407	ON
0/FT0	(Fan Tray 1 of 4)	687	601	ON
0/FT1	(Fan Tray 1 of 4)	687	591	ON

37.8 kW Installed

- Max power available
 - 6x 6.3 kW PSUs
 - N+1 redundancy
 - 31500W available for allocation
- Allocated power
 - 17884W
- Available
 - 13616W for additional HW

```
RP/0/RP0/CPU0:ios#show env power
```

```
Total output power capacity (N + 1)      : 31500W + 6300W
Total output power required                : 17884W
```

Current Power Usage

- 13636W to system
- 14598W from wall
- 93.41% efficiency



```
RP/0/RP0/CPU0:ios#show env power
```

```
Total output power capacity (N + 1)
```

```
: 31500W + 6300W
```

```
Total output power required
```

```
: 17884W
```

```
Total power input
```

```
: 14598W
```

```
Total power output
```

```
: 13636W
```

PSU Details

- Volts/Amps per PSU
- 3 of 6 shown

RP/0/RP0/CPU0:ios#show env power

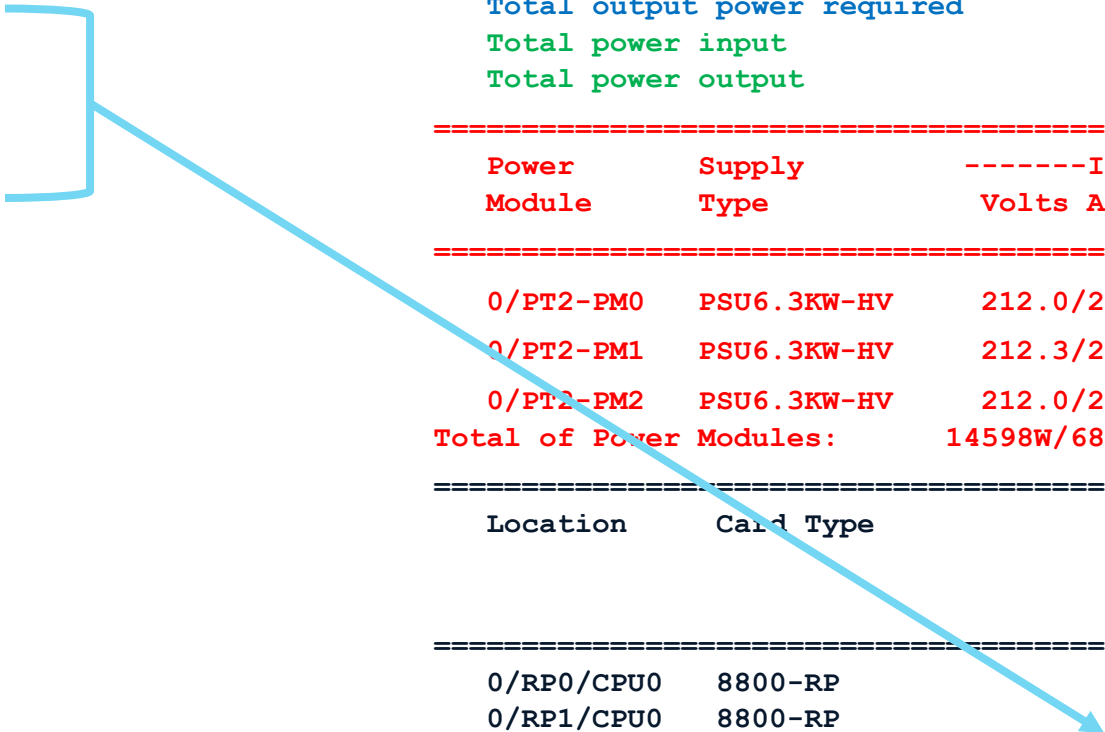
Total output power capacity (N + 1)	:	31500W +	6300W
Total output power required	:	17884W	
Total power input	:	14598W	
Total power output	:	13636W	

=====

Power Module	Supply Type	-----Input-----		-----Output---		Status
		Volts A/B	Amps A/B	Volts	Amps	
=====						
0/PT2-PM0	PSU6.3KW-HV	212.0/211.7	5.7/5.6	54.2	40.7	OK
0/PT2-PM1	PSU6.3KW-HV	212.3/212.0	5.7/5.8	54.6	41.8	OK
0/PT2-PM2	PSU6.3KW-HV	212.0/212.3	5.9/5.9	54.6	43.2	OK
Total of Power Modules:		14598W/68.7A		13636W/250.5A		

Details per PID

- PIDs
- Allocation
- Includes optics



RP/0/RP0/CPU0:ios#show env power

Total output power capacity (N + 1)

Total output power required

Total power input

Total power output

:

:

:

:

31500W +

17884W

14598W

13636W

6300W

Power Module	Supply Type	-----Input-----		-----Output----		Status
		Volts	A/B	Volts	Amps	
0/PT2-PM0	PSU6.3KW-HV	212.0/211.7	5.7/5.6	54.2	40.7	OK
0/PT2-PM1	PSU6.3KW-HV	212.3/212.0	5.7/5.8	54.6	41.8	OK
0/PT2-PM2	PSU6.3KW-HV	212.0/212.3	5.9/5.9	54.6	43.2	OK
Total of Power Modules:		14598W/68.7A		13636W/250.5A		

Location	Card Type	Power Allocated Watts	Power Used Watts	Status
0/RP0/CPU0	8800-RP	95		
0/RP1/CPU0	8800-RP	95		
0/0/CPU0	(LC type 1)	2494		
0/1/CPU0	(LC type 2)	1705		
0/FC0	(Fabric 1 of 8)	417		
0/FC5	(Fabric 5 of 8)	417		
0/FT0	(Fan Tray 1 of 4)	687		
0/FT1	(Fan Tray 1 of 4)	687		

Details per PID

- Current power draw
- Varies with load

RP/0/RP0/CPU0:ios#show env power

```
Total output power capacity (N + 1)      : 31500W + 6300W
Total output power required                : 17884W
Total power input                          : 14598W
Total power output                         : 13636W
```

Power Module	Supply Type	-----Input----- Volts A/B	Amps A/B	-----Output----- Volts	Amps	Status
0/PT2-PM0	PSU6.3KW-HV	212.0/211.7	5.7/5.6	54.2	40.7	OK
0/PT2-PM1	PSU6.3KW-HV	212.3/212.0	5.7/5.8	54.6	41.8	OK
0/PT2-PM2	PSU6.3KW-HV	212.0/212.3	5.9/5.9	54.6	43.2	OK
Total of Power Modules:		14598W/68.7A		13636W/250.5A		

Location	Card Type	Power Allocated Watts	Power Used Watts	Status
0/RP0/CPU0	8800-RP	95	81	ON
0/RP1/CPU0	8800-RP	95	78	ON
0/0/CPU0	(LC type 1)	2494	2076	ON
0/1/CPU0	(LC type 2)	1705	1197	ON
0/FC0	(Fabric 1 of 8)	417	417	ON
0/FC5	(Fabric 5 of 8)	417	407	ON
0/FT0	(Fan Tray 1 of 4)	687	601	ON
0/FT1	(Fan Tray 1 of 4)	687	591	ON

Case Study

Platform Evaluation & Modeling

Example from Silicon One
Ports up/down
Traffic load
Fan Power

Case Study Overview

- Evaluated systems with 1 to 86 ASICs
- Silicon One in 7nm TSMC process node
- 400 GbE w/ DAC, grey, and 40 km ZR optics
- Range of ambient temps
- ~10K “snapshots” w/ 1000+ data points each

Test Approach

- Start at typical ambient: 22C (impact of higher temps also investigated)
- Shut down all ports – collect data
- Power on ports at each supported speed (e.g., 40G, 100G, 400G) w/o load
- Test mix of packet sizes and bandwidth

	0	25%	50%	75%	100%
256B					
512B					
1500B					
9000B					

Parser .xls output (one of many formats)

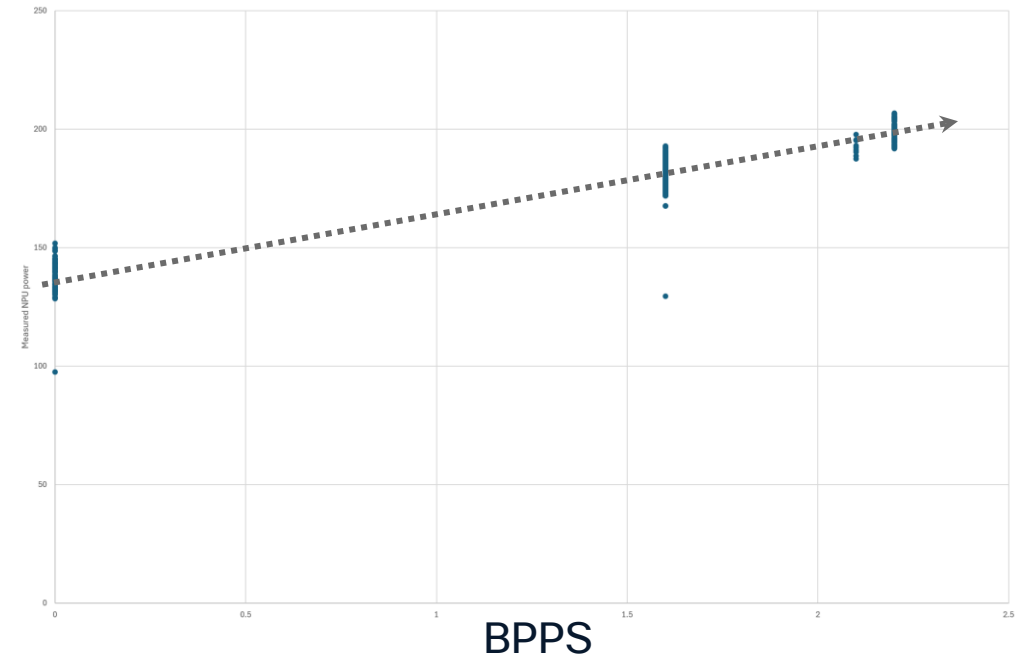
- Pick what fields we want
- Build XLS from those
- Do math!
 - Separate optics power
 - Idle power
 - Watts / Tbps
 - Watts / Bpps
 - Etc.
- Regression to build model

	B	C	D	E	F	I	J	K	L	M
1	timestamp	rack_pid	annotation	cli_bw	cli_paksi	pwrIn	pwrOut	total pwrE	Inlet Ter	avg_fans
2	08/15 21:07	8711-32FH-M	SteadyState	0	0	1237	1138	92.0%	22	16425
3	08/15 21:08	8711-32FH-M	SteadyState	0	0	1237	1135	91.8%	22	16425
4	08/15 21:08	8711-32FH-M	SteadyState	0	0	1239	1138	91.8%	22	16440
5	08/15 21:09	8711-32FH-M	SteadyState	0	0	1239	1135	91.6%	22	16425
6	08/15 21:10	8711-32FH-M	SteadyState	0	0	1237	1138	92.0%	22	16440
12	08/15 21:30	8711-32FH-M	768B_25LR	25	768	1260	1171	93.0%	22	16440
13	08/15 21:30	8711-32FH-M	768B_25LR	25	768	1260	1171	93.0%	22	16425
14	08/15 21:31	8711-32FH-M	768B_25LR	25	768	1260	1171	93.0%	22	16440
15	08/15 21:32	8711-32FH-M	768B_25LR	25	768	1262	1171	92.8%	22	16440
16	08/15 21:33	8711-32FH-M	768B_25LR	25	768	1262	1171	92.8%	22	16440
17	08/15 21:37	8711-32FH-M	768B_50LR	50	768	1305	1205	92.3%	22	16455
18	08/15 21:38	8711-32FH-M	768B_50LR	50	768	1305	1205	92.3%	22	16440
19	08/15 21:39	8711-32FH-M	768B_50LR	50	768	1305	1205	92.3%	22	16440
20	08/15 21:40	8711-32FH-M	768B_50LR	50	768	1305	1206	92.4%	22	16425
21	08/15 21:41	8711-32FH-M	768B_50LR	50	768	1305	1205	92.3%	22	16425
22	08/15 21:43	8711-32FH-M	768B_75LR	75	768	1328	1234	92.9%	22	16425
23	08/15 21:44	8711-32FH-M	768B_75LR	75	768	1328	1235	93.0%	22	16305
24	08/15 21:45	8711-32FH-M	768B_75LR	75	768	1328	1235	93.0%	22	16425
25	08/15 21:46	8711-32FH-M	768B_75LR	75	768	1328	1235	93.0%	22	16425
26	08/15 21:46	8711-32FH-M	768B_75LR	75	768	1328	1236	93.1%	23	16425
27	08/15 22:05	8711-32FH-M	768B_100LR	100	768	1374	1260	91.7%	23	16425
28	08/15 22:06	8711-32FH-M	768B_100LR	100	768	1351	1260	93.3%	23	16425
29	08/15 22:07	8711-32FH-M	768B_100LR	100	768	1351	1261	93.4%	23	16425

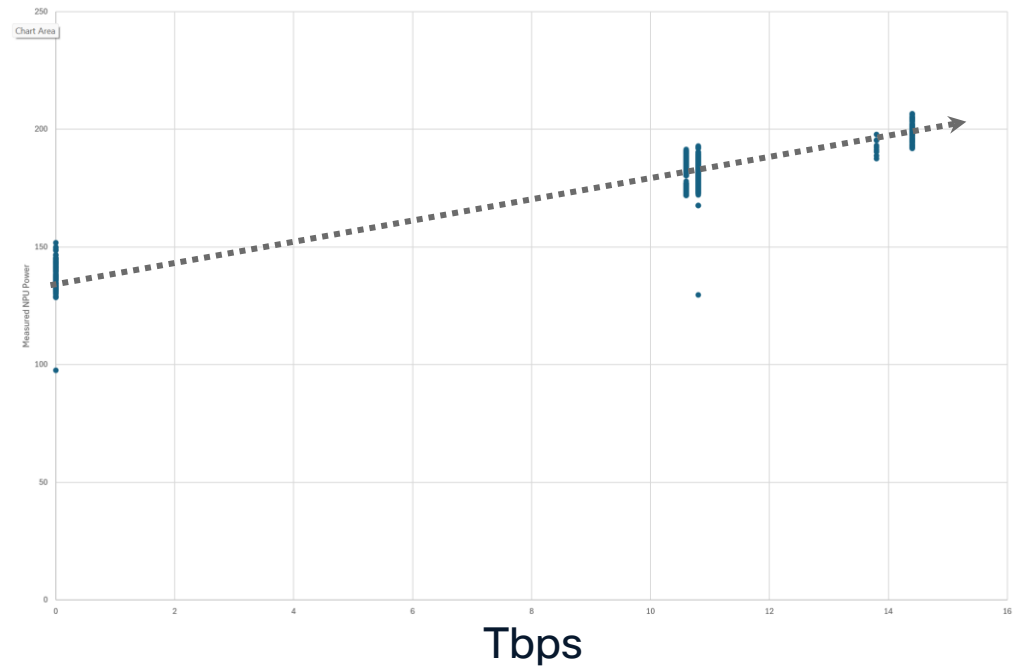
NPU Power

- NPU power is a function of:
 - Static power (always on)
 - Number/Speed of SerDes (port & fabric)
 - Traffic rate
 - Bandwidth rate

Watts



Watts



Multi-variate Analysis

- Lots of data
- Isolate the variables you care about
- Run regression (excel, python, etc.)
- Sanity check coefficients
- R-square values

Idle Power (SerDes on)

Watts per

Watts per

Does line match data set?

R Square

0.976775

Intercept

676.3208

Bpps

14.14087

Tbps

12.14574

Creating a Model

- Idle power for all hardware
- Adjust ports/cards down
- Interpolate dynamic power from traffic
- Fan power driven by most demanding component
 - Ambient
 - Type of boards in modular systems
 - Optics – longer reach modules require more airflow
- Power supply efficiency based on PSU type and load

How Can You Save Energy?

Network design

Silicon process nodes

Partial device shutdown

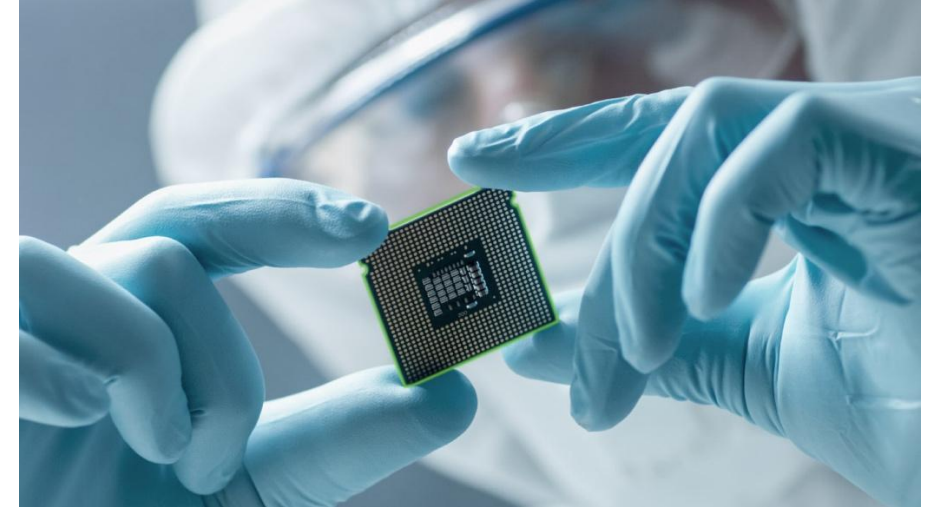
Network Design

- Number of devices
- Type of router/switch
 - “Core” vs. full feature/scale
 - Single vs. multi-chip
- Flatten leaf/spine when possible
- Integrate optical network
- Eliminate redundant redundancies



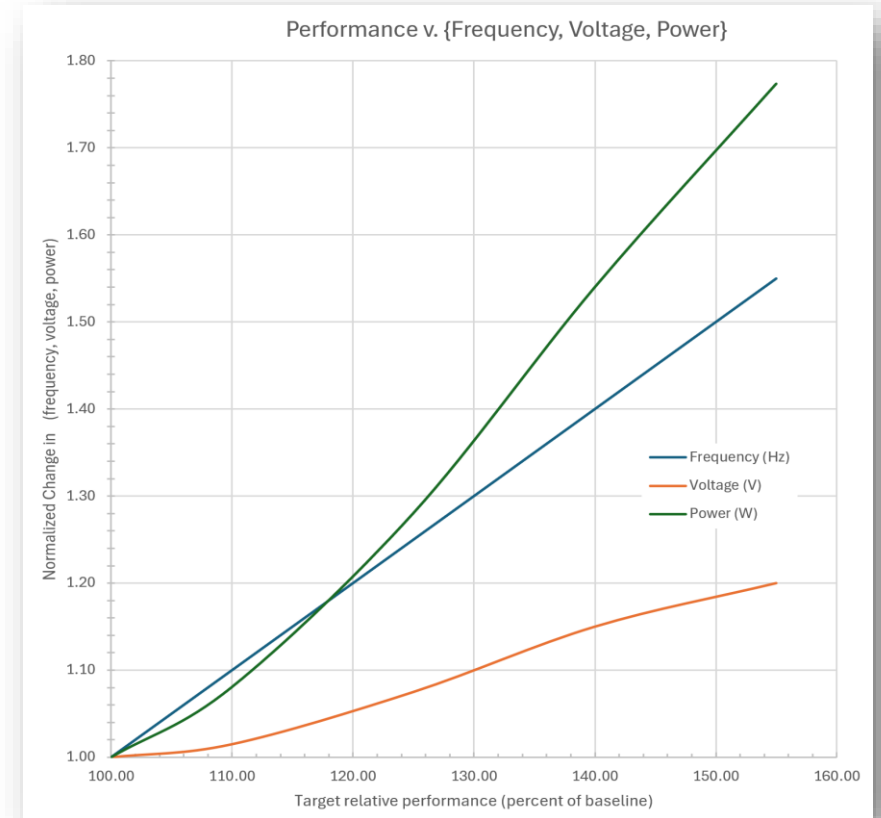
Silicon Process Nodes

- 28/16/7/5nm widely deployed (up to 51.2T devices)
- Investment protection vs. deploying latest tech
 - Far greater if they allow chip consolidation
- Different chip blocks have different power profiles
 - Memory vs. I/O vs. MACs vs. Engines vs. Crypto

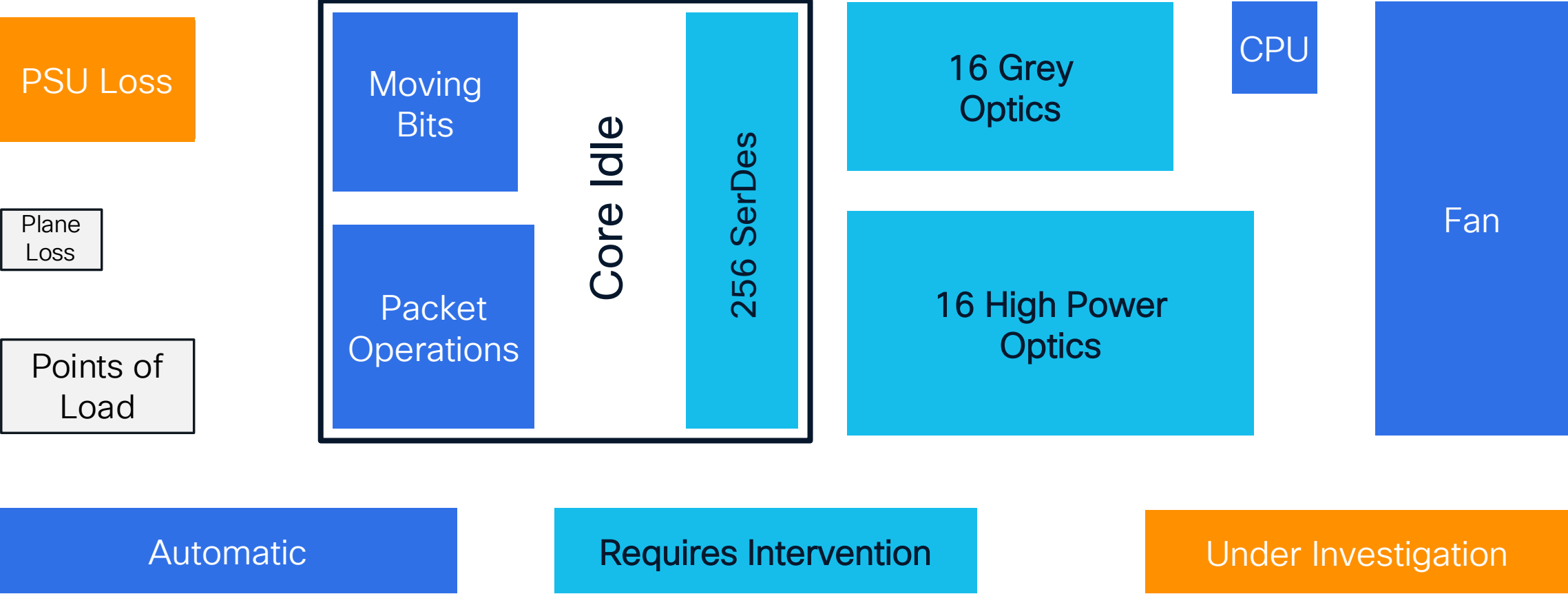


How can ASICs Dynamically Save Energy

- Customizing voltage and frequency to application
- Powering down unused I/O
 - Ports or fabric connections
- Clock gating
- Combination of automatic + config



Automatic Savings vs. Manual Intervention



In Summary

Thanks for Coming

**We'll be available after the session for Q&A
See slides additional Sustainability Sessions**

Complete your session evaluations



Complete a minimum of 4 session surveys and the Overall Event Survey to claim a Cisco Live T-Shirt.



Earn up to 800 points by completing all surveys and climb the Cisco Live Challenge leaderboard.



Level up and earn exclusive prizes!



Complete your surveys in the Cisco Live Events app.

Continue your education



Visit the Cisco Stand for related demos



Book your one-on-one Meet the Expert meeting



Attend the interactive education with Capture the Flag, and Walk-in Labs



Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

Contact me at: Insert preferred comms method

Thank you

CISCO Live !

