

Re-Architecting Power Systems for AI Racks

Harsha Bojja
AI Platform Architect - Microsoft

Today's AI Infrastructure

Management & Fabric

Power Conversion

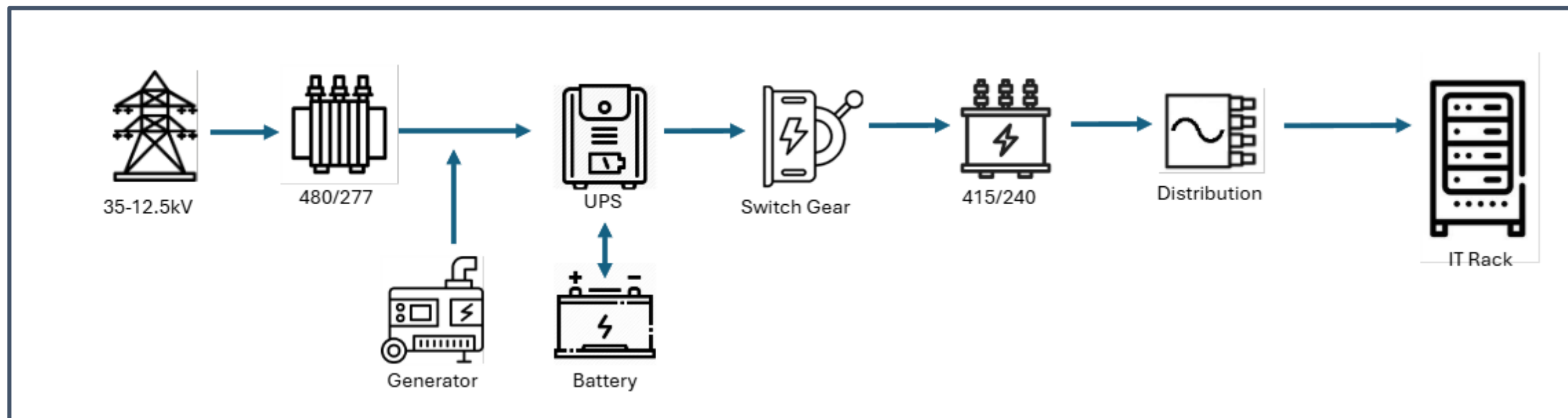
Scale Up Switches

AI Accelerators



All-in-one racks work well for today's and near future ~30-200kW solutions

Today's Power Delivery



Current power delivery path from power substation to IT Rack

Today's AI Infrastructure

Management & Fabric

Power Conversion

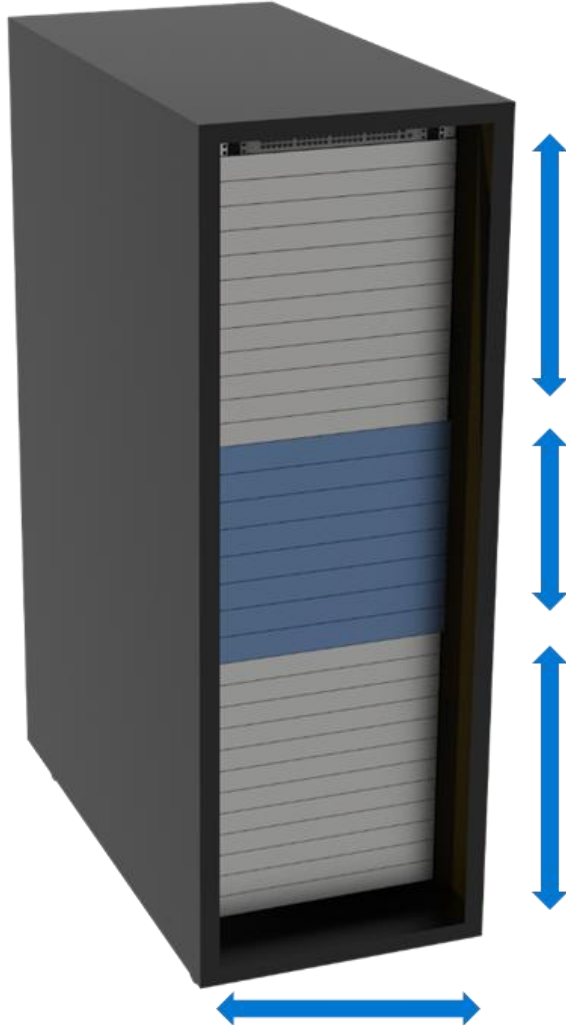
Scale Up Switches

AI Accelerators



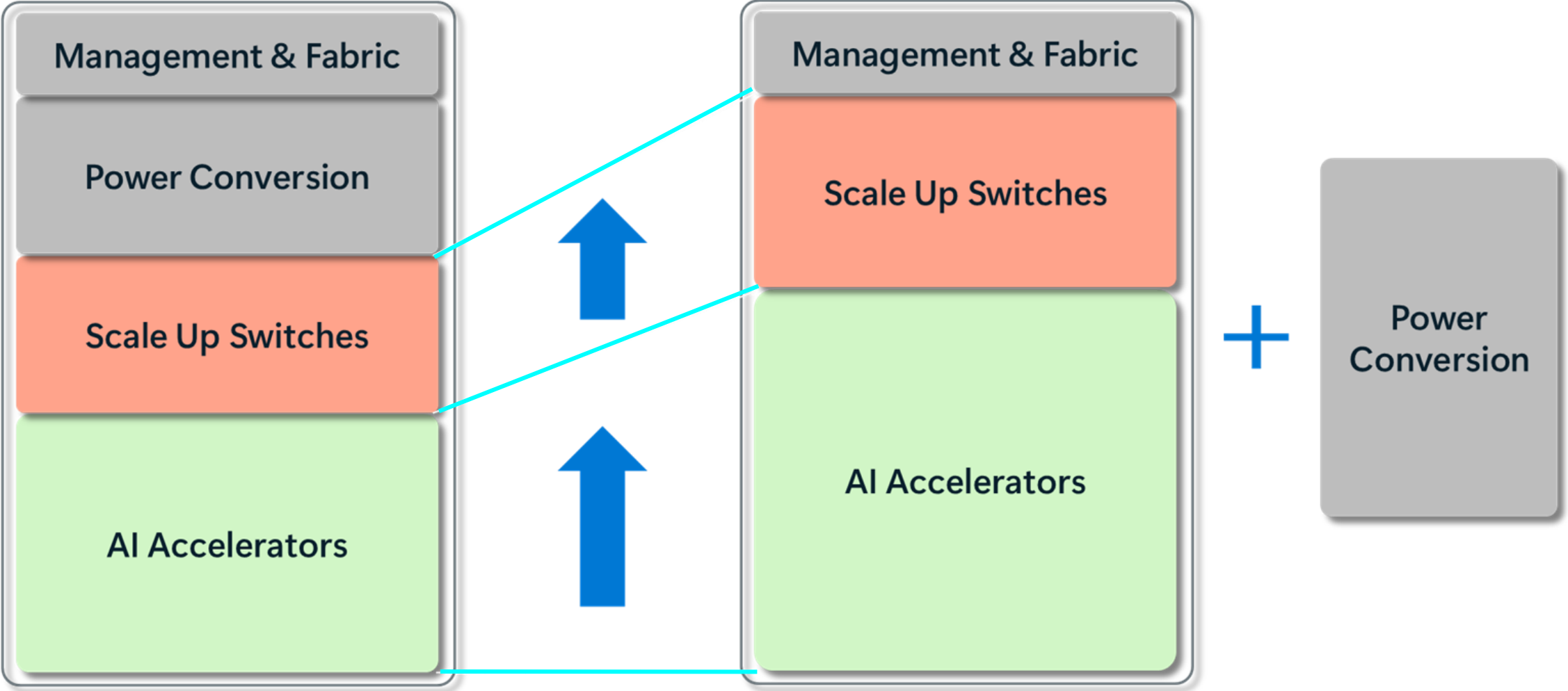
AI Rack with Liquid Cooling Sidecar

Driving Factors & Challenges



- Higher power accelerators
 - ❑ Higher liquid flow rate and/or lower temperature fluid
 - ❑ Often physically larger – drives larger trays or racks
- More Accelerators
 - ❑ Requires more and/or larger switch ASICs and trays
 - ❑ Accelerators are now further from switch ASICs, challenging SI
- Accelerators need more bandwidth
 - ❑ Increase lane count (challenging physics and fit)
 - ❑ Increase lane speed (challenging SI)

Optimizing the Infrastructure



Disaggregated AI Rack and Power Delivery

Optimizing the Infrastructure

AI system + Power

AI System

Power

Server Rack



Disaggregated AI Rack and Power Delivery

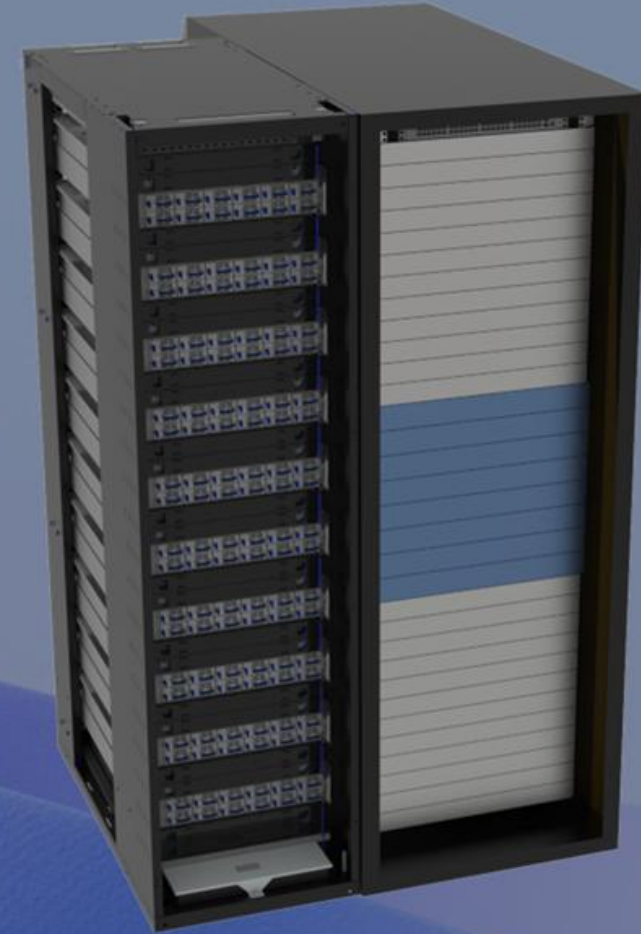
Solving Power Infrastructure

Mt. Diablo Power Rack

400 Vdc disaggregated power rack co-design provides infrastructure solution to scale for AI

Scales from 100s of kW up to 1MW

Commonality across platforms



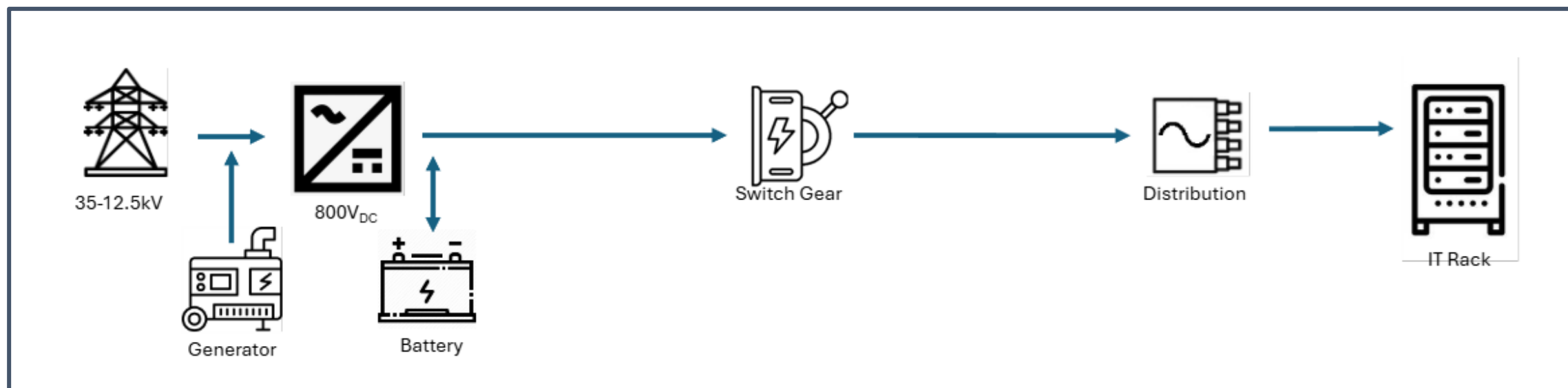
Mt Diablo Disaggregated Power Rack



- Utilizes high voltage DC (+/- 400VDC or 0-800VDC)
- Specification is a joint effort of Microsoft, Meta, and Google
- Scale the number of power shelves to support the needs of evolving AI/server racks
- Pick your availability model:
 - ☐ Feed redundancy
 - ☐ Power shelf redundancy
 - ☐ N + x power supply redundancy

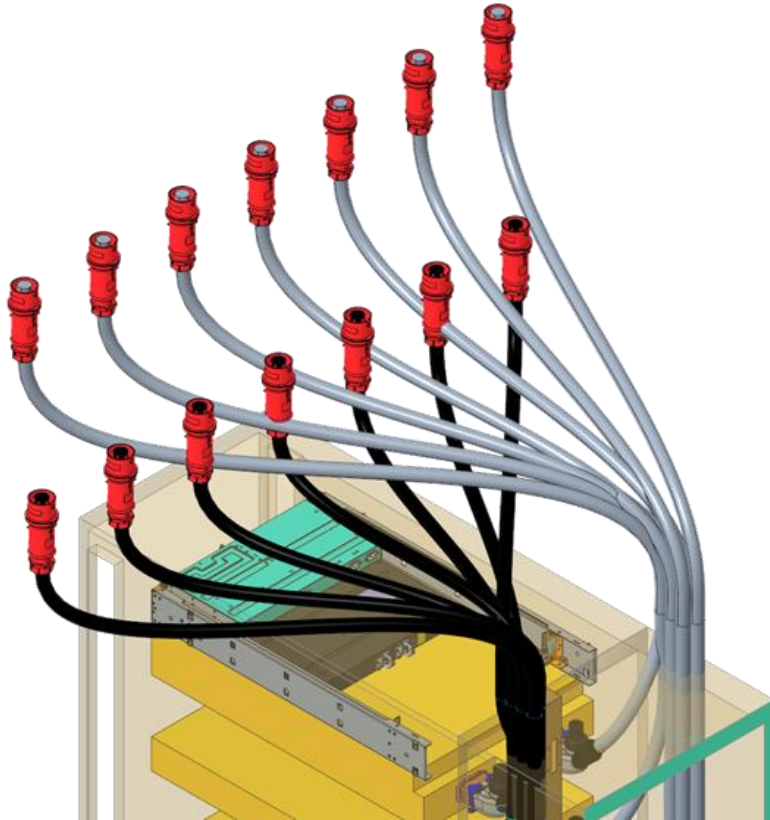
Leverage the current 48V solutions while designing the optimized +/- 400V solution

Tomorrow's Power Delivery



Future power delivery path from Substation to IT Rack

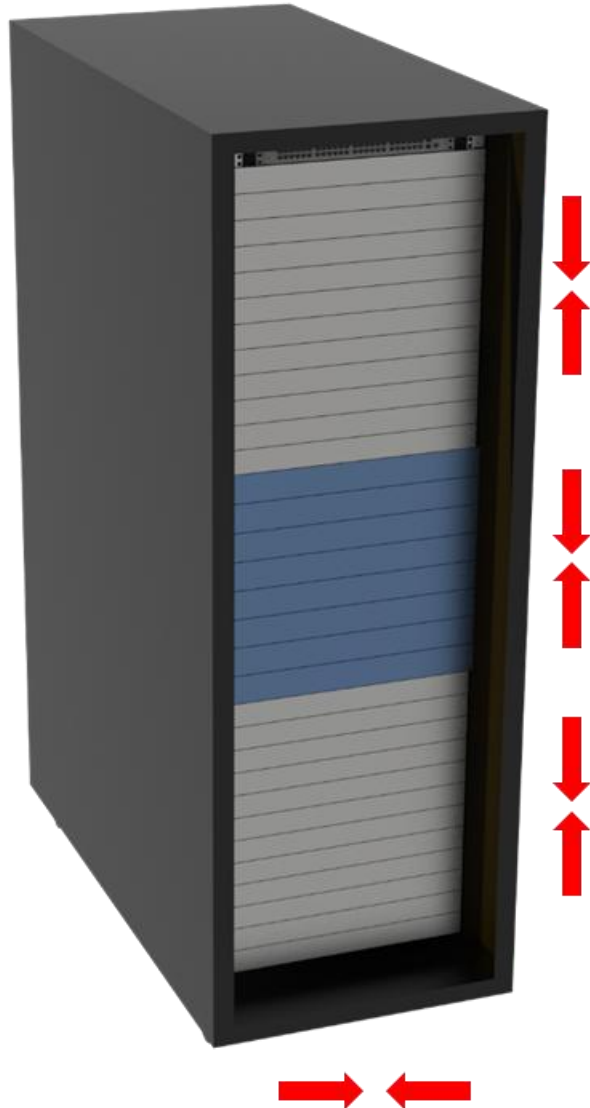
Tame the AC Whips



AC Whips



Changes & Optimizations



- Higher power accelerators
 - ☐ Larger trays
 - ☐ Little to no airflow
 - ☐ Move from ~80% LC to 99%
 - ☐ QD size increasing?
- More Accelerators & Higher Bandwidth
 - ☐ More high-speed IO between accelerators and switches
 - ☐ Move to LC to enable higher density
 - ☐ Find ways to optimize and consolidate components
 - ☐ Move to 200G, 400G, etc. interconnects
 - ☐ Higher speed Serdes reduces reach, or requires retimers
 - ☐ How far can we take copper? When do we require optics?

DC Compatibility

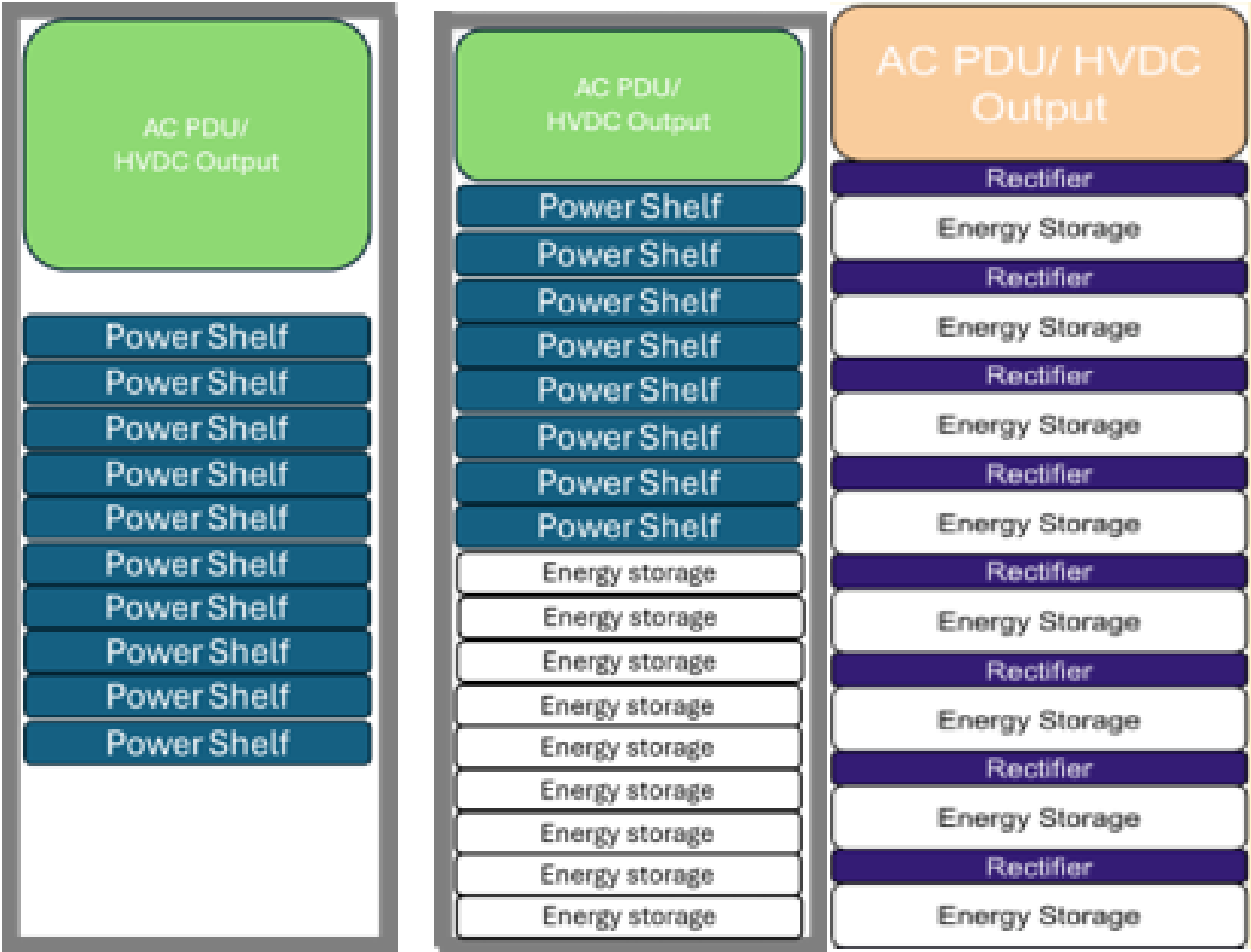


- Fully liquid cooled racks
- Power density **5-10x** higher than DCs were built for
- Disaggregated Power Racks
- Service from both sides
- Heavy – likely cement floors only vs raised floor
- Casters vs forklift movement
- More complex service

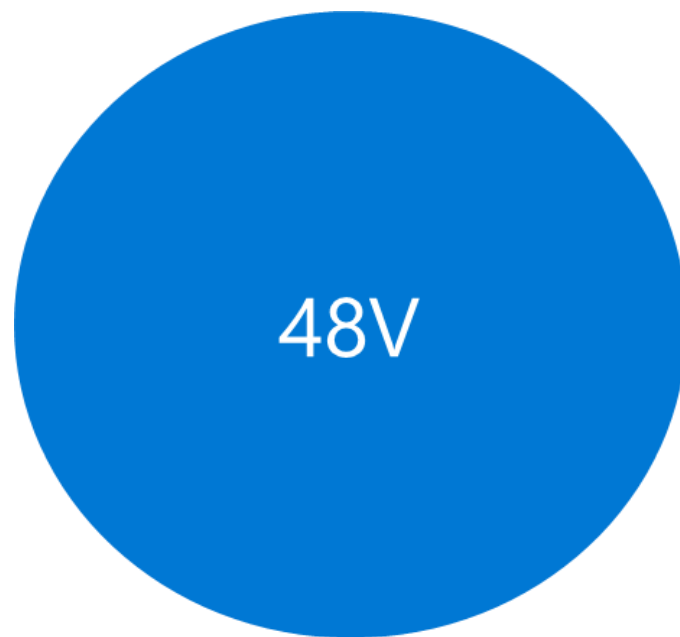
Disaggregated Power Rack Specifications

- Rack Frame
 - *Support various standard, non-standard rack specifications*
- Power Shelves
 - *Blind mate power and center justified busbar*
- Provision for intrusion detection, lugs/quick connect attach
- 3-phase AC input (3P+E) -> AC Power Distribution -> AC/DC Power Supplies
- AC Feed Configuration Options (**480-415-400** Vrms nominal)
 - *3-phase AC or Dual AC*
 - *WYE 5-Wire*
 - *Delta 4-Wire*
- AC loss holdup and ride through support via **bulk storage capacitors**
- Line overvoltage, Line sag, Holdup sans battery backup
 - *ITIC upto 20 msec at 100% loading*
- Support battery backup, capacitor bank units and dynamic, pulse loading
- Active management and control via **CAN, Modbus, Ethernet** etc. protocols
- Overall target rack efficiency at 480Vac - > **97%** at **100%** loading
- Upto **12x 200A** AC whips
- PE, High resistance mid-point grounding for **+/- 400V**

Disaggregated Power Rack Configurations

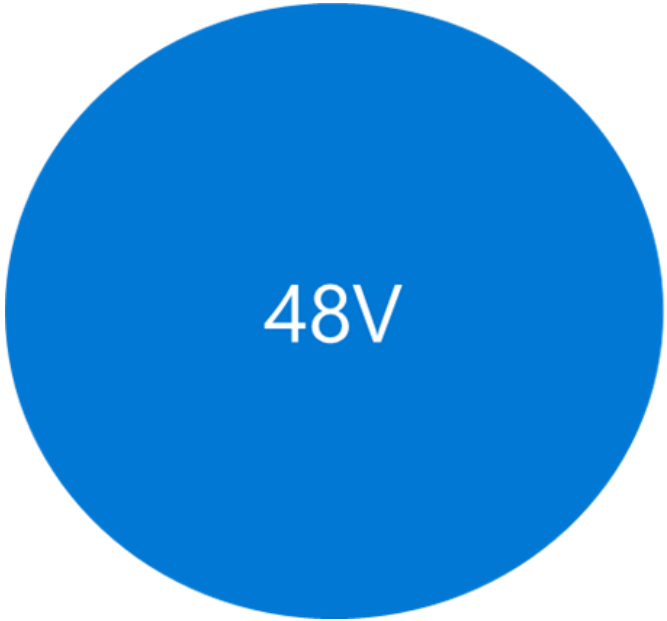


Building a 500kW AI System



Copper diameter = 56 mm

Higher Voltage Options



48V

Diameter = 56 mm



400V

Diameter = 20 mm

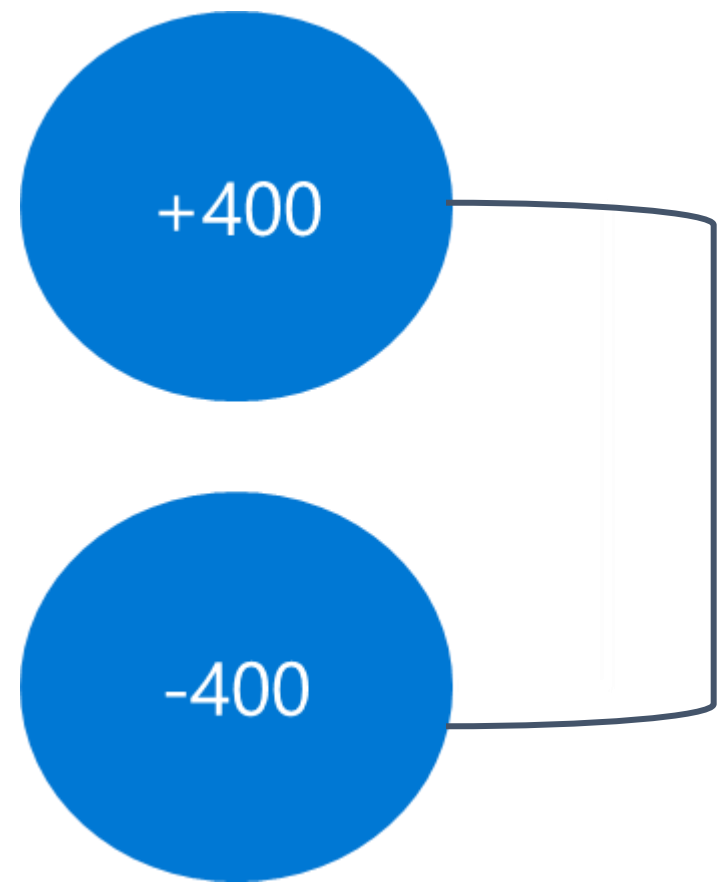


+/-
400V

Diameter = 14 mm

Effective Voltage - 800V

Busbar Dimension Comparison



800V Effective Voltage



Weight

- 14mm busbar 3 kg / 6.5 lbs
- 56mm busbar 47 kg / 103 lbs

Safety Considerations

- Follow existing IEC/UL 62368-1 baseline guidelines
- Short Circuit Fault Current Rating - 10kA to 40kA
- Cable Assemblies and Power Cables - UL certified 60 Vdc or 42.4 Vac at min temperature rating of 90C
- No service access to HVDC components exceeding PS2 limits
- Compliant creepage and clearance per UL guidelines
- Protective earthing and bonding per IEC 60417
- Leakage current - < 3.5mA for AC and < 10 mA for DC
- Energy Storage Systems - Compliant to UL 9450 (including fire propagation testing)
- High Resistance Grounding
- Solid Grounding

High Voltage DC

- 48V doesn't scale to the power distribution required for the next wave on AI systems
- Auto industry is already moving to 800V architectures
- +/- 400V provides for the same 800V effective voltage
- The industry is ready with the downstream converters

"Yesterday" and "today" AC voltage distribution to rack

**MVAC to LVAC
transformer**

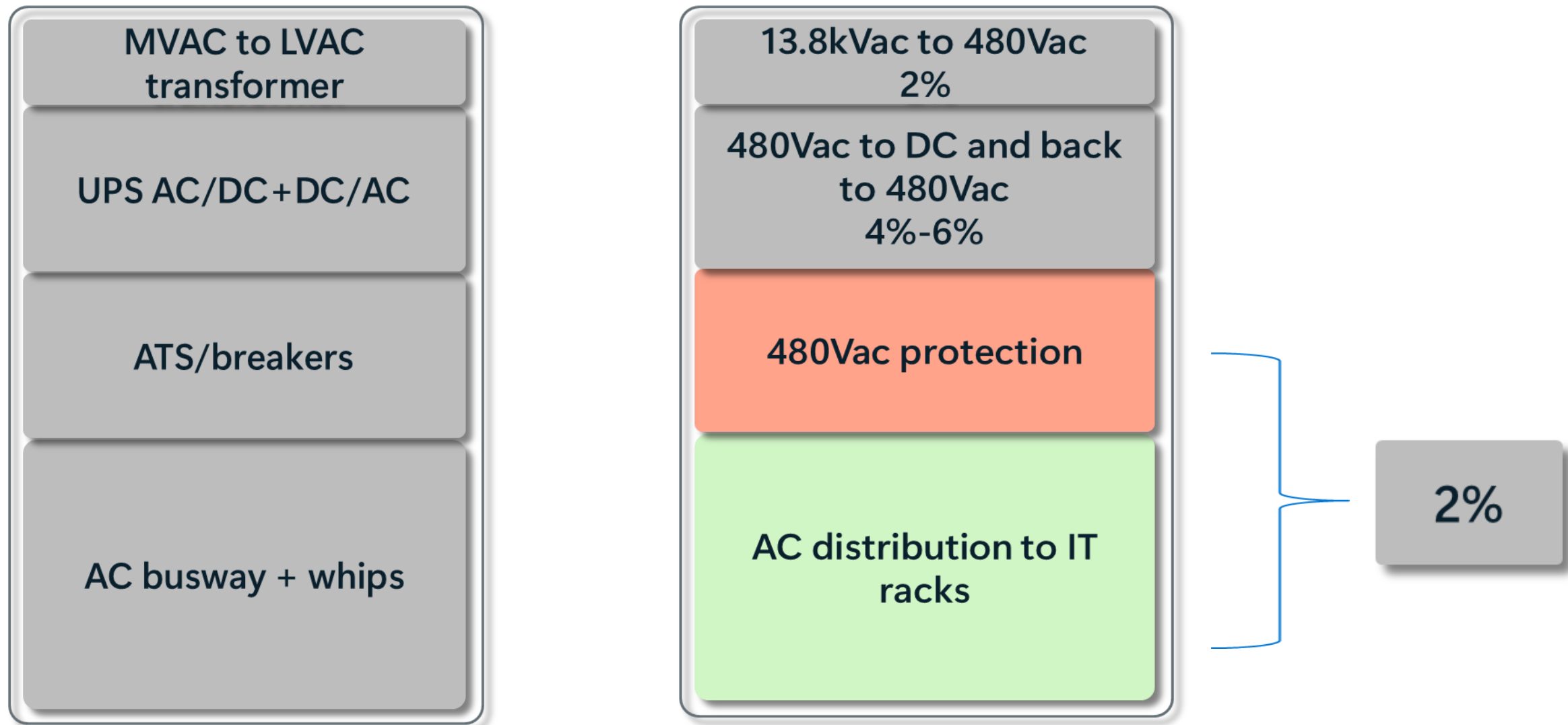
UPS AC/DC+DC/AC

ATS/breakers

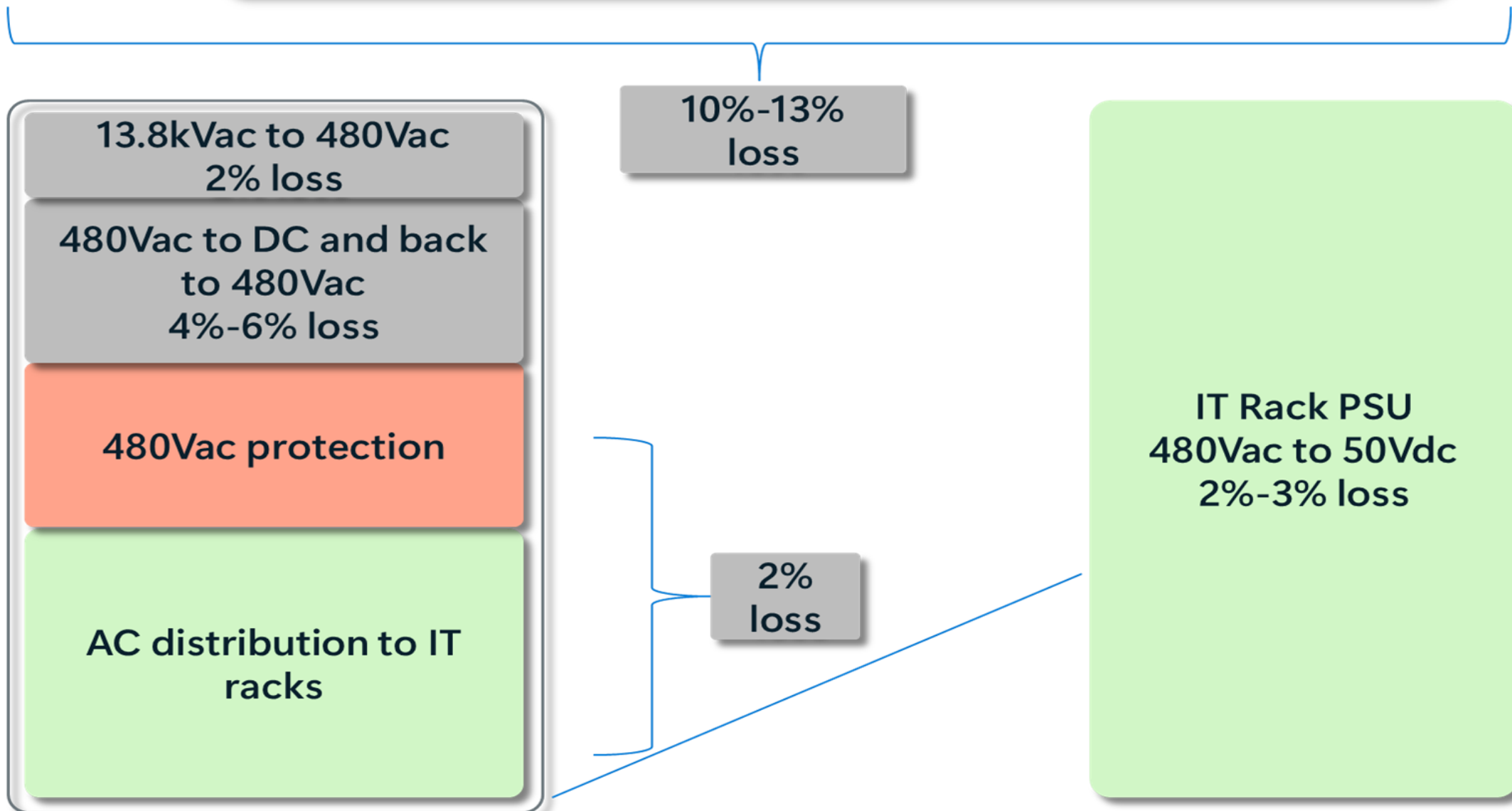
AC busway + whips

MVAC - Medium Voltage Alternating Current
LVAC - Low Voltage Alternating Current
UPS - Uninterrupted Power Supply
ATS - Automatic Transfer Switch

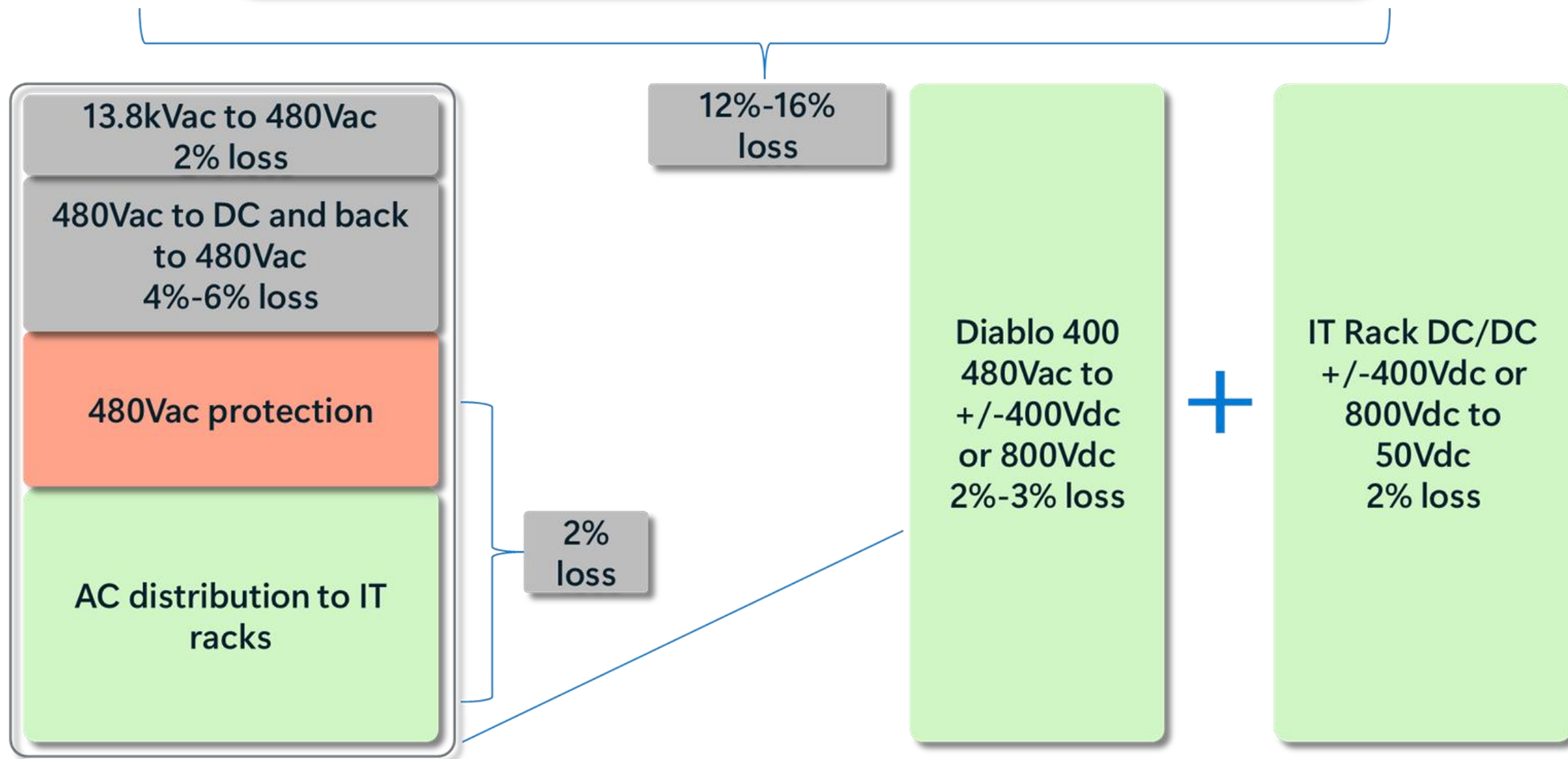
Breaking down the “yesterday” and “today” losses



"Yesterday" IT rack power distribution effect



"Today" IT rack power distribution effect



"Tomorrow" DC voltage distribution to rack

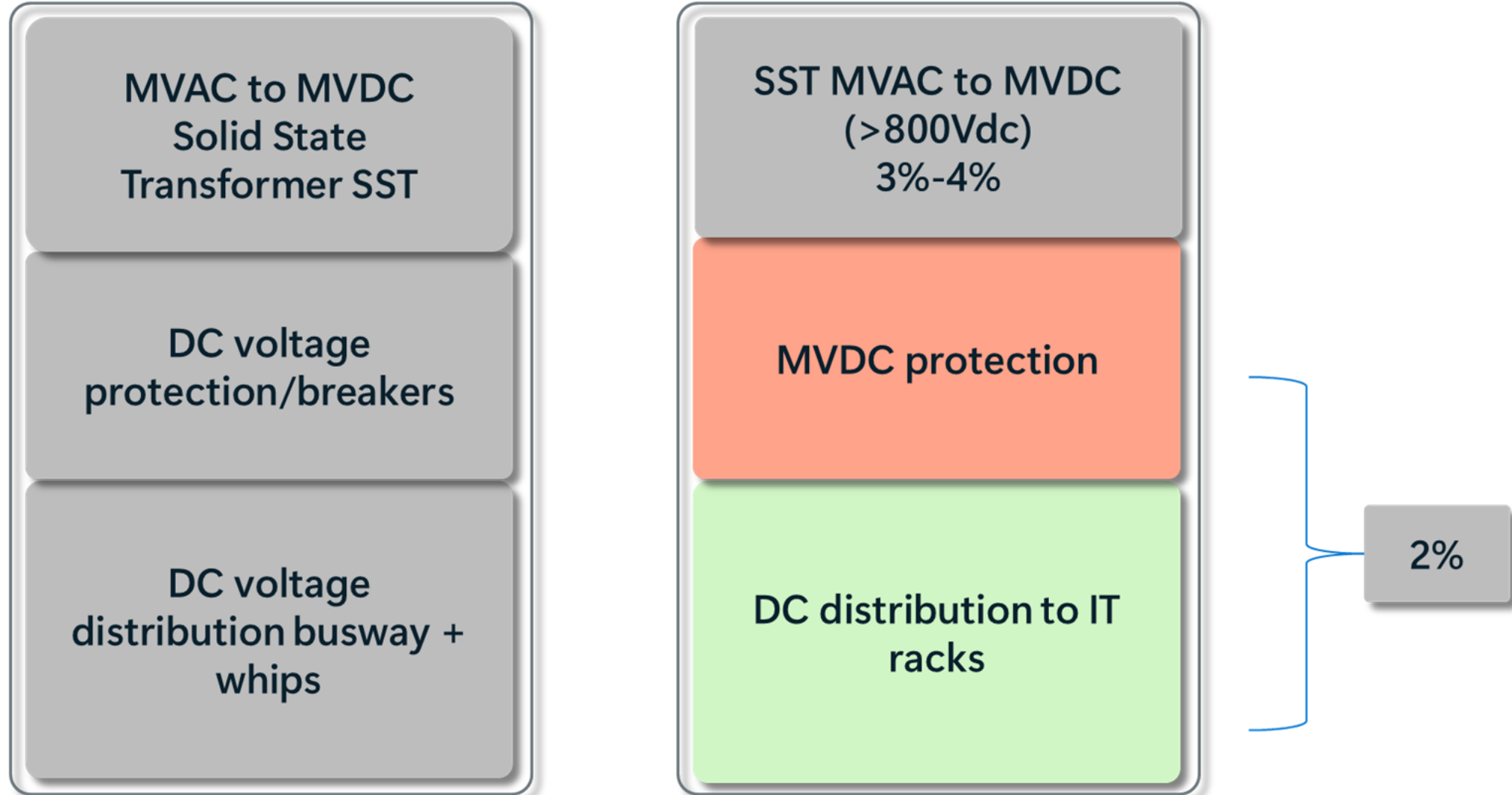
MVAC to MVDC
Solid State
Transformer SST

DC voltage
protection/breakers

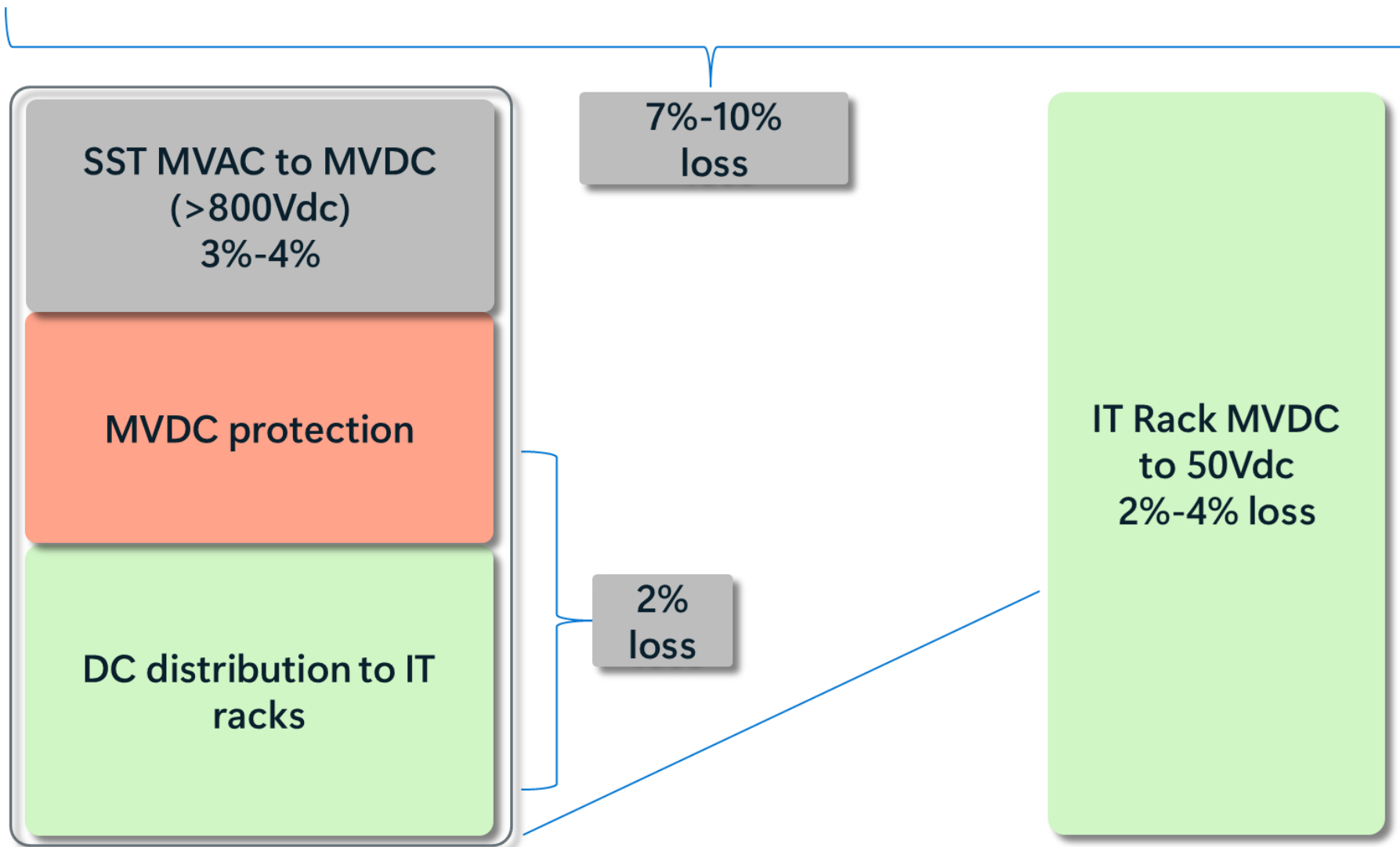
DC voltage
distribution busway +
whips

SST - Solid State Transformer

Breaking down the "tomorrow" losses



"Tomorrow" IT rack power distribution effect



Energy storage in “today” and “tomorrow” power delivery

- Energy storage (UPS function or peak shaving):
- “today” Diablo 400 power distribution:
 - Same as “yesterday” UPS
 - In-rack BBU/CBU
- “tomorrow” power distribution:
 - In-SST BBU/CBU
 - In-row BBU/CBU

UPS - Uninterrupted Power Supply
BBU - Battery Backup Unit
CBU - Capacitor Bank Unit