A New Redundancy Strategy for High-Availability Power Systems
Outline

- RAS Philosophy
- Review
  - Power 6 – Traditional approach
  - Power 7 – Hybrid approach
- Load lines and Redundancy
- New architecture
RAS Philosophy

- **Goal: 100% Availability**
  - High-end server should never go down

- **Conventional implementation**
  - Full redundancy
  - No single points of failure due to active parts
  - Full diagnostics with 100% Error Detection & FRU ID
  - Concurrent replacement of faulty hardware
Power 6 Processor Book Power

- **DCA (Distributed Converter Assembly)**
  - 15 voltage levels
  - **1600** amps total current
  - Technology: Multi-phase synchronous buck (60 phases)

- **RAS**
  - 1+1 redundant → replace after 1st failure
Lessons learned from Power 6

- Designing for 100% concurrent replacement results in poor efficiency
- DCAs must be closer to load to reduce distribution losses!
Solution to efficiency problem

Power 6

350V → Bus Converter* → 11V → VRM → 1V → Load

DCA too far away from load!

High copper losses!

Power 7

350V → Bus Converter* → 44V → PRM* → 40V → VTM* → 1V → Load

OK for DCA to be far away from load

Negligible losses

Reduced losses

Current multiplication

High current produced close to load
Power 795 Processor Book Power

- **DCA**
  - 16 voltage levels
  - 200 amps total current
  - Technology: Vicor Factorized Power

- **Processor board**
  - 8 voltage levels
  - **2475 amps** of logic current
  - Technology: Vicor Factorized Power

- **RAS**
  - DCA: 1+1 → replace on 1st failure
  - VTM: N+2 → replace on 2nd failure

Power distribution board

- **VTM's (37x)**
- **ORing FET's (222x)**
Power 775 Supercomputing Drawer

- Factorized Power is ideally suited for IBM Power 775 Supercomputer, where high power density was essential
- Contains compute power of 8 Power 795 processor books

Specifications
- 256 processor cores
- 24 TB DDR3 memory
- Load power: 17,500 watts
- Load current: **15,000 amps**!
- 100% water-cooled

*Extremely dense package → no space for l-f decoupling caps!*
Power 775 processor board (bottom view)

VTM assembly with ORing FETs (192x)
Lessons learned from Power 7

- **N+2 fail-in-place works!**  
  - Delivered higher currents redundantly - *without penalizing efficiency*

- Mandatory for very high power systems like P775  
  - Large N reduces “redundancy overhead”

- Would like more granularity (for smaller systems)  
  - N+2= 4 has same overhead as N+1= 2

- **Goal is utilize more fail-in-place for future systems**
Challenges of Concurrent Maintenance

- Every level is “disturbed” during DCA replacement
- DCA’s with faults on different levels
  - Results in deferred maintenance
  - Probability increases with # of levels
- Connector misplugs (rare)
- Negative customer perception

![Number of voltage levels by System](image)

- G7
- Power 4
- Power 5
- Power 6
- Power 7
- Projected
Goals of proposed architecture

- N+2 redundancy with no single points of failure
- Power Conversion *and Controls* at point of load
- Smaller power trains for improved granularity

*But how will independent controllers share current?*
Redundant supplies must share current

- Why force equal current sharing?
  - Optimum reliability and transient response
  - Simplifies diagnostics

- Current sharing approach
  - "Droop" sharing method
  - Voltage droops as current increases

- Advantage of "droop" sharing
  - Current sharing is automatic
  - No interconnections to fail
  - Well suited for redundant systems

- Disadvantages
  - Requires factory calibration
  - Firmware balancing to correct for drift

\[ V_1 = V_C + \Delta I \cdot R_O \]
\[ V_2 = V_C - \Delta I \cdot R_O \]
\[ V_{O1} = V_1 - I_1 \cdot R_O \]
\[ V_{O2} = V_2 - I_2 \cdot R_O \]
Load-line implementation with 1+1 power supplies

- Must use shallow load line (~1%)
- Module + Socket R provide load-line regulation

Steep load line results in large voltage drop after supply failure!

Chip load line = \( \frac{LL_{SUPPLY} + R_{MOD}}{2} \)
Power 6: 1+1 VRM’s

- N phases per DCA
- 1 Master per level
- 1 Protection controller per level
- Precision DAC
- Current sharing
  - Initial factory calibration
  - Firmware balance routine to compensate for drift
Load-line regulation with N+2 power supplies

- Sense inside module at chip
- Module & Socket R is now inside feedback loop
- Small voltage drop after supply failure

**Chip load line** = \( \frac{LL_{SUPPLY}}{N+2} \)

Steep supply load lines ensure current sharing without precision DAC or firmware balancing.
Current sharing with N+2 redundant regulators

- Current sharing depends on setpoint tolerance and load-line slope

- Setpoint tolerance: $V_{REF} \sim 0.5\%$

- Load line slopes:
  - Steep: $\sim 10\%$ when sensing on chip ($V_{core}$)
  - Shallow: $\sim 2\%$ when sensing on board (memory, I/O, etc.)

- Current sharing error for 35A Slave
  - 25% with 2% LL $\Rightarrow \pm 8.75A$ per phase
  - 5% with 10% LL $\Rightarrow \pm 1.75A$ per phase

Comparison with N+1=2:

$\Delta I = 50\%$ of rated load

(1% load line)
N+2 Master-per-Slave POL converter

- N+2 phases per POL card
- N+2 Masters
- N+2 Protection controllers
  - Originally developed for Phase Redundant VRMs in IBM Power 780 Server
- Current sharing
  - Uses internal DAC
  - No calibration
  - No firmware balancing
Each DCA communicates over independent bus.

- Tolerates Master or I2C faults
- Replace after 2\textsuperscript{nd} failure
  - MUX is not critical for power conversion
Summary of proposed architecture

- **Disadvantages**
  - Higher voltage ripple (no clock sync)
  - Needs a load line for current sharing
  - Multiple masters complicate SVID communication

- **Advantages**
  - Fault tolerant power conversion, controls and telemetry
  - Fail-in-place RAS (replace only after 2\textsuperscript{nd} failure)
  - Works with any VRM chipset (w/serial interface)
  - No firmware current balancing
  - Improved DCA reliability (12V bulk + low-current levels only)
  - Low redundancy overhead
Thank you!