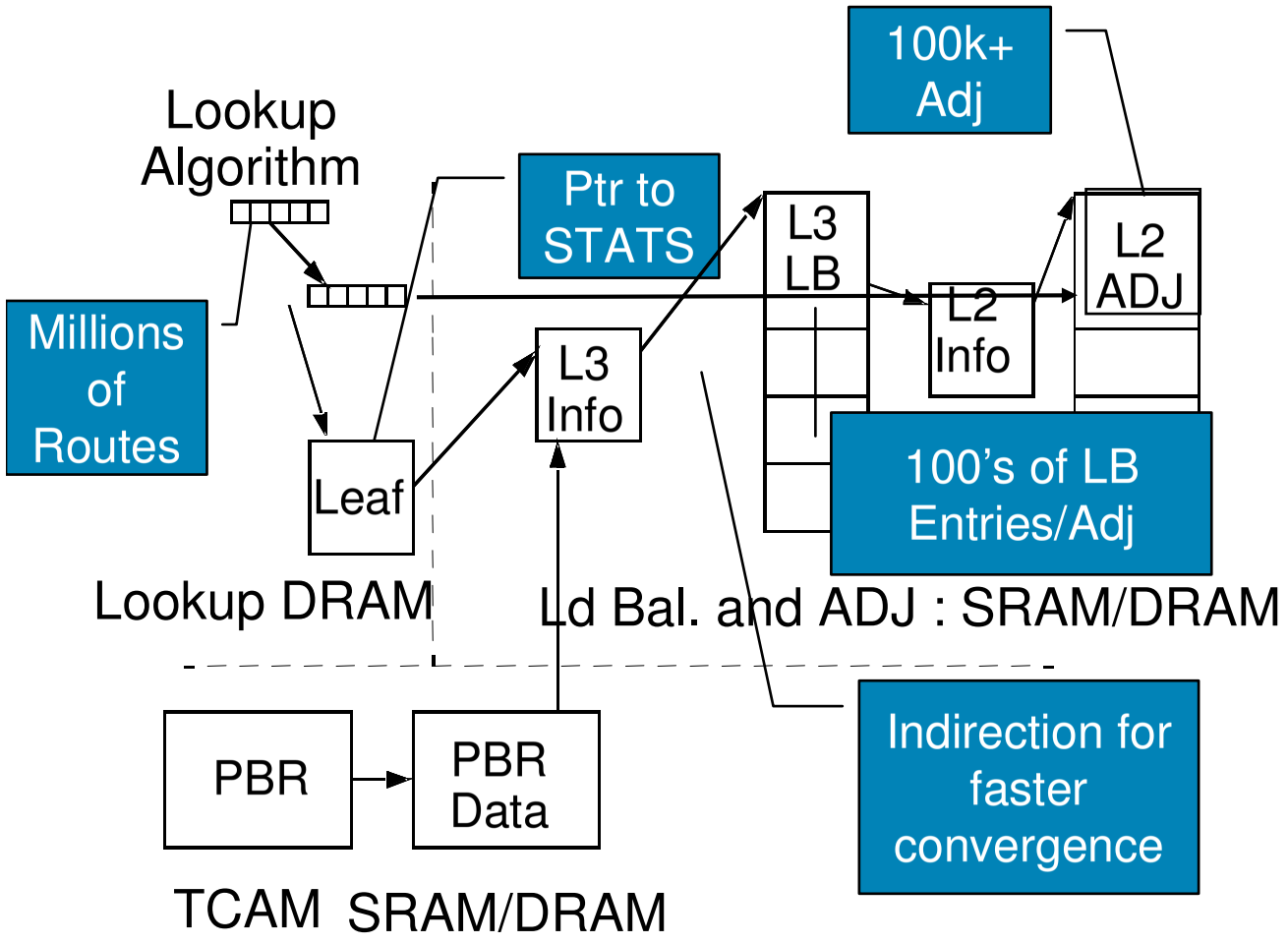
Packet Forwarding Challenges at 100Gbps

David Tsiang

Forwarding Challenges at 100 Gbps

- 2.5x everything over 40G
- More scalability
  - Global internet table growing larger.
  - Potential for address explosion as IPV4 addresses run out:
  - Address re-selling -> fragmentation of address blocks.
  - IPV6
  - Growth of VPN usage
- More flexibility as Internet deals with evolution (e.g., V4-V6 transition, LISP, pt-mpt MPLS).
- Complex recipe of flexibility, scalability, future-proofing for investment protection, and PPS performance.

Why Programmability ?
Simple Forwarding (Not so simple!)



Example FEATURES:

- **IPv4/v6**
- **MPLS–3 Labels**
- **Link Bundling**
- **Load Balancing L3**
- **Policing, QOS**
- **Marking**
- **TE/FRR**
- **Sampled Netflow**
- **WRED**
- **ACL**
- **Per prefix accounting**
- **GRE/L2TPv3 Tunneling**
- **RPF check (loose/strict)**
- **Congestion Control**

40G->100G What does it take

- 2.5 X MIPS, Memory TPS, Memory BW, FF-MHz
- Reqt to keep the same power profile

  (no forklift upgrades yet please).
- Mitigations:
  Silicon advances (110nm -> 90nm -> 65 nm).

  Lower voltages, capacitance, and leakage current.

  More efficient memory technology (DRAM, SRAM, TCAM).

  More efficient design (terminations, power supply design).

ASIC Technology

- ASIC: 110nm at 1.2v vs 65nm at 1.0v

  $P_d = V^2 * C * f$

- Capacitance goes down as does voltage.

  Clock input capacitance 26% less (factor .74)
  Data input capacitance 50% less (factor of .5)
  Combinatorial capacitance 43% less (factor of .57)
  30% FF's

- V drop gets you $1^2/1.2^2 = .77$

  23% pwr reduction due to voltage.

- Net gain ~ (.77 * ((.74/2 + .5/2)*.3 + .57*.7)) = .45

- **ASIC dynamic power gain at 100G: 2.5 * .45 = 1.13**

ASIC Technology (2)

- Static Power - Leakage current

  cu11 .096mA/Kgates
  cu65 .078mA/Kgates
  .078/.096 = .8125, 19% reduction

- Increased frequency on ASICs from 250Mhz to 400 Mhz

  Less gates per Gbps. Same dynamic power but less static (leakage) power per Gbps.

- **ASIC static power gain: 2.5 * 250/400 * .8125 = 1.26**

Memory Technologies

- DRAM

  40G: FCRAM 4Mx18 0.6W, 332 Mbps/pin, tRC 25ns
  100G: RLDRAM-II 16Mx36 : 1.6W, 800 Mbps/pin, tRC 15ns

  Power reduction is 332 * 18/.6 / (800*36/1.6) = .55
  45% reduction in DRAM power.

  DRAM power gain 2.5 * .55 = 1.38

- TCAM

  11.25W at 40G equiv Lups vs 7.5W at 100G equiv Lups.

  (7.5/100)/(11.25/40) = .27

  73% reduction in TCAM power (better TCAM cell design + better silicon).

  TCAM power gain 2.5 * .27 = .675

Other power savers

- More efficient SI design

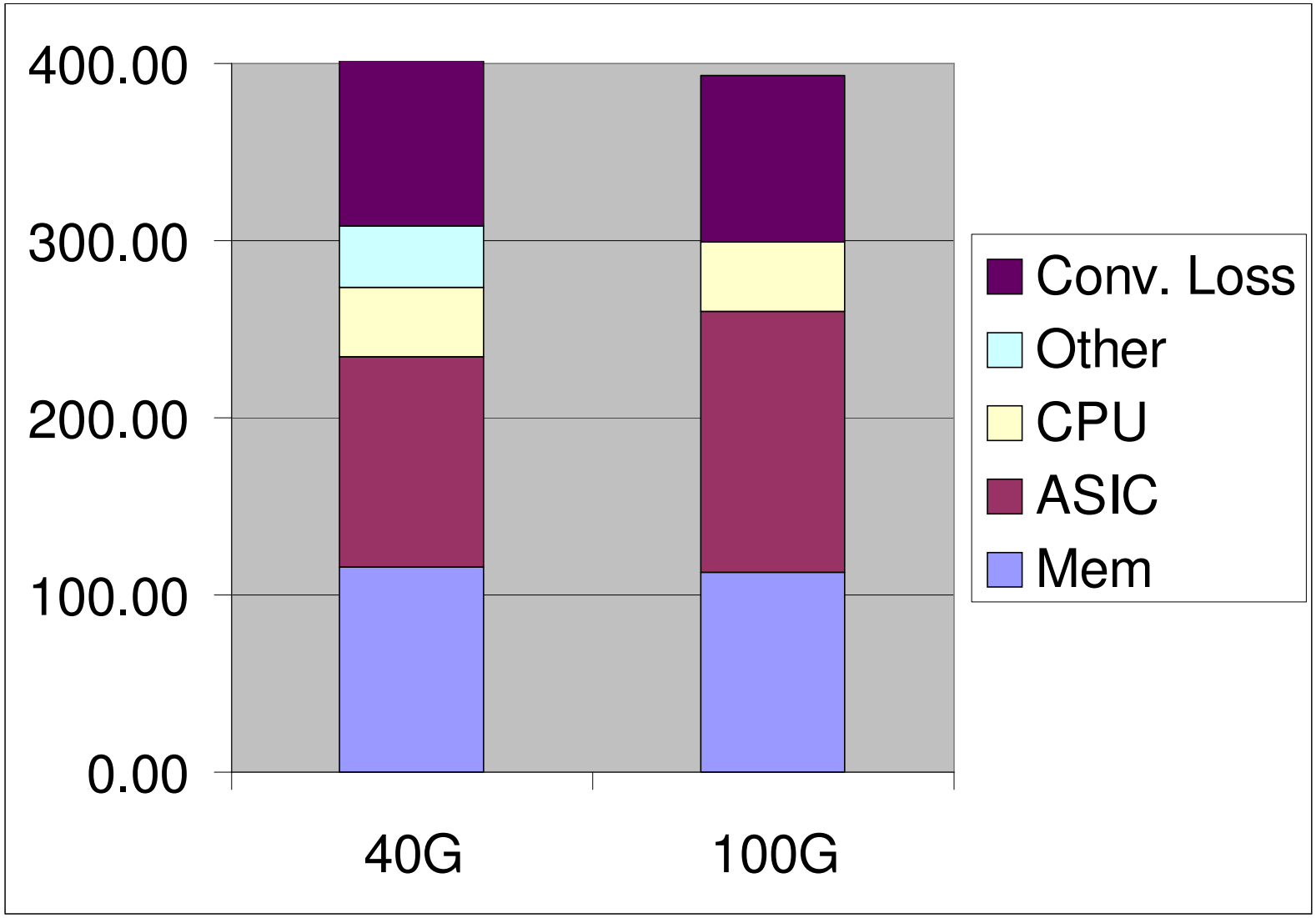
  Internal terminations vs external Thevenin terminations on memory lines.

- Integrated serdes on ASICs
- Replacement of some SRAMs with DRAMs
- Efficient power supply design

  Discrete designs optimized per load zone decreases power loss through DC-DC converter.

- Integration of Service Processor function (10W).

Power consumption 40G vs 100G



Conclusion

- Silicon advances will enable us to reach 90% of the 100G challenge.
- The remaining 10% comes from more efficient design (terminations, integration, use of DRAMs instead of SRAMs).
- However: Silicon advances lag exponential bandwidth growth.
- Hence: Aggregate power required will grow. Large SPs will need multi-chassis capability to meet exponential bandwidth growth.