High Speed Serial Links – Design Trends & SI Challenges

Jane Lim
Cisco CHG Switching SI

Dec. 5th 2014
Outline

- Introduction – IoT, data density, high-speed signal transmission
- Typical devices used on data center system boards
- Typical high-speed serial interfaces on networking products
- High-speed Serdes (HSS) electrical signaling – reach, application, power
- 3D memory integration – HSS requirements
- Chip & system levels SI/PI design challenges
  - Package design optimization: material → IL, routing → RL, pinout → XTALK
  - Board design challenges: modeling, via/antipad, IL variation, material characterization
  - Power integrity: PDN design for HSS supplies
Introduction

- The growth of IoT has resulted in a massive expansion of the number & size of data centers around the world
- Cloud computing & mega data centers demanding low power, faster compute and storage interfaces
- Cisco offers a broad range of switches and routers products which can be used in IoT solutions
- There is a constant desire for signal to run faster with less power to address the data density issues
- This trend drives more innovation and revolutionary SI design challenges
Cisco Data Center Switching Products

- **N7K/N77 series** – for highly scalable data center networks with a fabric architecture:
  - 18/10/6 slot Midplane
  - 9 slot Backplane
  - 2000+ Serdes pairs routed

- **MDS9000 series** – for storage networks of all sizes and architectures:
  - 13/9/6 slot Midplane
  - 3 slot Backplane
  - 1 RU Pizza Box
Typical Devices Used on System Boards

- **ASICS**
  - Main components, various functions (switch fabric, data forwarding, memory storage)
  - Board space and power/cooling constraint data bandwidth or ASIC counts
  - Die/pkg technology constraints port density and contents on ASIC
  - Chip-to-chip & chip-to-modules communications mainly through HSS

- **Microprocessors**
  - PowerPC cores used for packet-processing tasks
  - Configurable Ethernet interface, PCIe peripheral subsystem, DDR memory interface

- **FPGA**
  - I/O and power management, lower speed

- **EDC, Retimer**
  - Used between ASIC and optical modules & HSS interface (i.e. 4X25GE)
  - EDC/Retimer typically includes equalizations and CDR to clean up jitters
  - Extend signal reach on PCB
Typical HSS Interfaces – Backplane

- **Backplane/Midplane (long reach)**
  - Serdes interface ASIC on line card to ASIC on fabric card
  - Current generation ~10.3Gb/s; next generation ~25Gb/s, total channel length > 30”
  - Requires advanced signal equalization features like FFE, DFE & crosstalk management
  - May require eye opener devices in the channel
Typical HSS Interfaces – Front Panel

- **Front panel ports (medium to short reach)**
  - Serdes interface ASIC to optical modules with optional EDC (depend on channel length)
  - Typical speed is 10x10GE or 4X25GE, max. channel < 14” /w one connector
  - Low voltage swing, low EMI, slew rate control
Typical HSS Interfaces – Chip to Chip

- **Chip to chip & Serial memory (medium to short reach)**
  - Serdes interface ASIC to another port ASIC or to serial memory ASIC
  - Typical speed today ~12.5 Gb/s, next gen ~25.0 Gb/s, max. channel < 12”
  - Low swing low power interface, typically doesn’t require DFE in receiver
HSS General Requirements

- Full rate operation with half/quarter rate for legacy support
- Programmable output differential voltage control
- Programmable output slew rate control
- Support continuously-adapt equalization and back-channel transmitter training
- Support Forward Error Correction (FEC) for LR application
- Support DC coupled operation without external DC blocking capacitors
- Receiver with built-in eye monitoring in mission mode
- PRBS generator/checker with auto-detect P/N polarity swap feature
- Support AC JTAG
## SerDes Electrical Spec (NRZ signaling)

<table>
<thead>
<tr>
<th>Standard</th>
<th>SerDes Architecture</th>
<th>Insertion Loss Target</th>
<th>Application</th>
<th>WC power target (mW/ch)</th>
</tr>
</thead>
<tbody>
<tr>
<td>USR</td>
<td>Tx: FFE, Rx: CTLE</td>
<td>1.5dB@14GHz 3dB@28GHz</td>
<td>Bump-to-bump Inside MCM or Interposer</td>
<td>&lt; 20 (28G)</td>
</tr>
<tr>
<td>XSR</td>
<td>Tx: FFE, Rx: CTLE</td>
<td>4dB@14GHz 8dB@28GHz</td>
<td>Ball-to-ball Across PCB</td>
<td>&lt; 50 (28G)</td>
</tr>
<tr>
<td>VSR</td>
<td>Tx: FFE, Rx: CTLE</td>
<td>10dB@14GHz 20dB@28GHz</td>
<td>Ball-to-ball Across PCB</td>
<td>&lt;100 (28G)</td>
</tr>
<tr>
<td>MR</td>
<td>Tx: FFE, Rx: CTLE, DFE</td>
<td>20dB@14GHz 40dB@28GHz</td>
<td>Ball-to-ball Across PCB /w 1 conn</td>
<td>&lt; 200 (28G)</td>
</tr>
<tr>
<td>LR</td>
<td>Tx: FFE, Rx: CTLE, DFE</td>
<td>35dB@14GHz</td>
<td>Ball-to-ball Across PCB /w 2 conn</td>
<td>&lt; 300 (28G)</td>
</tr>
</tbody>
</table>
After 28Gb/s, PAM4 vs. NRZ

**Choices for 50Gb/s Modulation Strategy**

<table>
<thead>
<tr>
<th>SERDES Core Type</th>
<th>IL</th>
<th>Source: IEEE802.3bs 400GE / November 2014, San Antonio, TX</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>USR</td>
<td>1.5dB@14GHz 3dB@28GHz Bump-to-bump Inside MCM or 3D Stack</td>
</tr>
<tr>
<td></td>
<td>XSR</td>
<td>4dB@14GHz 8dB@28GHz Ball-to-ball Across PCB</td>
</tr>
<tr>
<td></td>
<td>VSR</td>
<td>10dB@14GHz 20dB@28GHz Ball-to-ball</td>
</tr>
<tr>
<td></td>
<td>MR</td>
<td>20dB@14GHz 40dB@28GHz Ball-to-ball</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>35dB@14GHz Ball-to-ball</td>
</tr>
<tr>
<td></td>
<td>C2M</td>
<td>4dB@14GHz 8dB@28GHz Ball-to-ball</td>
</tr>
<tr>
<td></td>
<td>C2C</td>
<td>10dB@14GHz 20dB@28GHz Ball-to-ball</td>
</tr>
<tr>
<td></td>
<td>C2F</td>
<td>35dB@14GHz Ball-to-ball</td>
</tr>
</tbody>
</table>

C2C Reach is a Worry

Safe Bet Across Technologies

© 2014 Cisco. All rights reserved
3D Memory Integration – HSS Requirement

- Data growth will drive the scaling of the interconnects
- Use 3D integration to enhance HSS to >12Gb/s of high density DRAM data
- Industry is settling on two main approaches:

<table>
<thead>
<tr>
<th>High Bandwidth Memory (HBM)</th>
<th>Hybrid Memory Cube</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrate with ASIC on Interposer</td>
<td>Packaged Component</td>
</tr>
</tbody>
</table>

- Wide I/O Low Power
- Serial I/O Higher Power
Chip & System Levels SI/PI Challenges

- Serial channel data rate for chip to chip communication is being pushed to ~25Gb/s range and the next gen. is ~50Gb/s and higher
- Require special packaging/PCB technologies to tackle signal integrity (SI) and power integrity (PI) issues
- Every discontinuity on the HSS channel has to be well taken care of including die/package/board transition, trace fanout, via stub, etc
- Impacts of dielectric loss, conductor loss, surface roughness and glass weave effects need to be well understood
- Holistic channel modeling methodology is needed
- Accurate measurement and verification is important
Package Design
ASIC Package Design Optimization

- Package design optimization is required due to the impacts of
  - Dielectric loss, Cu surface roughness → IL
  - trace/via routing density → RL
  - pin-out → XTALK

- Package substrate material selection (GX13, GZ41, GY11) is very important to constraint the package loss, especially for backplane interface

- Package loss exceeds -2dB for a 30mm trace @14GHz with GY11, the conductor loss is dominant

- Solution: consider smaller ball pitch, i.e. 0.8mm to reduce package body size and trace length for pin count limited designs

<table>
<thead>
<tr>
<th></th>
<th>GX13</th>
<th>GZ41</th>
<th>GY11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diel. Constant</td>
<td>3.1</td>
<td>3.3</td>
<td>3.2</td>
</tr>
<tr>
<td>Loss Tangent</td>
<td>0.019</td>
<td>0.0074</td>
<td>0.005</td>
</tr>
<tr>
<td>Roughness (nm)</td>
<td>~600 (CZ)</td>
<td>~300 (ICZ)</td>
<td>~100 (FB)</td>
</tr>
</tbody>
</table>
Impact of Build-Up Material to IL

- Two candidates of BU materials (GX13 vs. GZ41) for ASIC packages
- At 7.5GHz, the insertion loss (IL) improvement from GZ41 to GX13 is 0.2dB over a 20mm trace
- Dielectric loss is only 20% for GZ41 case @ 7.5GHz
Impact of Trace Routing to RL

- Stackup is typically available in 2 core thicknesses: 400um & 800um
- HSSdiff pairs are routed above or below the laminate core
- Lower layer routing will bring worse return loss (RL) – not recommended!
Package Pin-out – Crosstalk Mitigation

- Package pin-out design is very important to mitigate crosstalk which is dominant by PCB via

- Various TX/TX and TX/RX pin-out are designed for >25Gbps interface

- Both frequency and time domains analysis are required to investigate channel crosstalk performance

- The overall effect includes, via-via, via-trace and trace-via coupling needs to be considered along with the PCB routing layer selection

- The optimal pin-out should be selected based on bit rate, channel margin and PCB routing constraint
Package BGA Pinout - FEXT

FEXT pattern 1  FEXT pattern 2  FEXT pattern 3

Tx-Tx or Rx-Rx Pinfield FEXT

Frequency (GHz)
0 15 30 45
-100
-80
-60
-40
-20
0
Crosstalk (dB)

FEXT pattern 1
FEXT pattern 2
FEXT pattern 3

TABLE I.  TX-TX OR RX-RX PINFIELD FEXT COMPARISON

<table>
<thead>
<tr>
<th></th>
<th>S:G Ratio</th>
<th>Total Far End Crosstalk (dB)</th>
<th>Peak-Peak Noise (mV)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>14GHz</td>
<td>28GHz</td>
</tr>
<tr>
<td>FEXT Pattern 1</td>
<td>1:1</td>
<td>-25.49</td>
<td>-15.89</td>
</tr>
<tr>
<td>FEXT Pattern 3</td>
<td>1:2</td>
<td>-39.65</td>
<td>-33.96</td>
</tr>
</tbody>
</table>
Package BGA Pinout - NEXT

![Diagrams of NEXT patterns](image)

**TABLE II. Tx-Rx NEXT**

<table>
<thead>
<tr>
<th>NEXT Pattern</th>
<th>Total Near End Crosstalk (dB)</th>
<th>Peak-Peak Noise (mV)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14GHz 28GHz 42GHz</td>
<td></td>
</tr>
<tr>
<td>NEXT Pattern 1</td>
<td>-43.25 -28.28 -33.53</td>
<td>3.14</td>
</tr>
<tr>
<td>NEXT Pattern 2</td>
<td>-48.43 -30.52 -33.65</td>
<td>2.26</td>
</tr>
<tr>
<td>NEXT Pattern 3</td>
<td>-54.11 -28.98 -24.33</td>
<td>1.71</td>
</tr>
</tbody>
</table>
Model to Hardware Correlation

- Diff pairs on test vehicles were measured on probe station
- Good agreement btw measurement & simulation data is achieved upto 20GHz

IL (GX13, 400 um core)  IL (GZ41, 400 um core)
Board Design
The transition from ASIC package to board is very important for HSS links.

Challenges coming with the BGA package:
- Big BGA solder ball
- Excess capacitance, which can’t be eliminated
- Impedance discontinuity in BGA-to-PTH
- Impact on signal quality

3D models port settings are important if the cascading method is used in channel analysis.

BGA dimension needs to be adjusted within package model, as the BGA ball shape may change once the package is mounted on the PCB.
Board Level Via/Antipad Design

- Inevitable impedance discontinuities at the transition PTH via to fan out from traces
  - Impact return loss (RL)
  - Distort insertion loss (IL)
  - Decrease eye opening
  - Data misinterpretation at receiver

- The transition structure from PTH to trace inevitably creates both capacitive and inductive discontinuities.

- The low impedance “transition traces” method is employed to compensate the extra inductance resulting from increasing anti-pad size
Based on “counter discontinuity” idea, a short segment of low impedance traces is inserted in fan-out region, so that the extra inductance can be compensated.
IBIS-AMI Channel Analysis

Chip-to-chip, with Tx/Rx package, 4” traces

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>22mil Anti-pad; 3.5/3.5/3.5 mil transition traces</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eye height (before Rx)</td>
<td>0.172 V</td>
<td>0.212 V</td>
</tr>
<tr>
<td>Eye width (before Rx)</td>
<td>19.1 ps</td>
<td>20.7 ps</td>
</tr>
</tbody>
</table>

BER = 1E-15 @ the input of Rx
PCB IL Variation Over Temperature

- Some PCB materials are sensitive to temperature
- @14GHz, IL variation over temp (25C–75C) is up to -6dB for a 30" trace
- Require a continuously adapt CTLE in the Rx
- @28GHz, needs better material for LR applications
Power Integrity
PDN Design for HSS Supplies

- As the data rate goes up and the voltage level decreases, power distribution networks (PDN) is increasingly challenging.

- To ensure HSS works reliably in the large system board, needs accurate modeling in both pre-layout and post-layout.

- Pre-layout is to verify self-induced noise can meet design specs
  - Cisco Generic PCB power model (CGPM)
  - Extracted 3D package model
  - Accurate die model

- Post-layout is to verify adequate filter design for clean analog supplies
  - Extracted actual board plane model
  - Accurate ferrite bead and capacitor model
Serdes Analog VDD PDN
Pre-Layout: AC Effects of Board

|Z| Profile of PDN for SerDes

Effects of PCB

Impedance (ohm)

Frequency (1kHz ~ 10GHz)
Pre-Layout: Transient Effects of Board

![Graph showing voltage (V) over time (0 ~ 3µs) for different configurations: Die + PKG, VDD Logic Core, Die + PKG + MB (CS), and VDD Logic Core.](image)
Post-Layout: Power Filtering Response

Ideal Filter Response (pre-layout)

Sentinel PSI (post-layout)
-5.81dB @ 187MHz

SIwave (post-layout)
-8.29dB @ 214MHz
From Siwave simulation: Mode 1 of 221MHz agrees with the spike @ 214MHz in the filter response.
Post-Layout: Power Filtering Response (cont.)

Power Filter w/o 1uF PFC: -5.8dB @ 187MHz
Power Filter including PFC: -17.7dB @ 290MHz
PFC + 4 caps added to shift resonance: -21.9dB @378MHz
Conclusions

- HSS links dominate high-end switches → power wired internet backbone
- HSS interfaces are challenging system designs → reach and power trade-off
- Modulation strategy (NRZ vs. PAM4) beyond 28Gb/s under discussion
- Packaging material selection and substrate design are critical for HSS
- Optimizing signal transition structures is a must to achieve a lower BER
- High-end low loss PCB laminate is needed for 25+ Gb/s applications
- Holistic channel modeling is essential to predict HSS performance
- HSS AVDD requires detailed pre-layout & post-layout PDN analysis
Thank you.