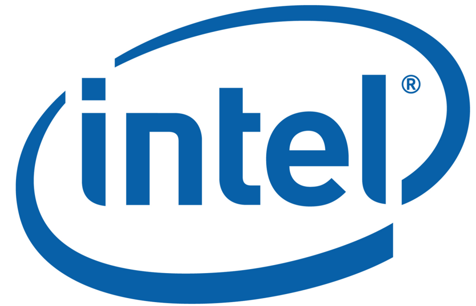
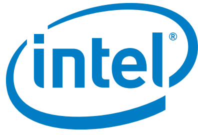


Intel® Ethernet Switch FM4224/ FM4112 24-Port 10G Ethernet L2 Switches

Design and Layout Guide

April 2013
Revision 1.2



INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

A "Mission Critical Application" is any application in which failure of the Intel Product could result, directly or indirectly, in personal injury or death. SHOULD YOU PURCHASE OR USE INTEL'S PRODUCTS FOR ANY SUCH MISSION CRITICAL APPLICATION, YOU SHALL INDEMNIFY AND HOLD INTEL AND ITS SUBSIDIARIES, SUBCONTRACTORS AND AFFILIATES, AND THE DIRECTORS, OFFICERS, AND EMPLOYEES OF EACH, HARMLESS AGAINST ALL CLAIMS COSTS, DAMAGES, AND EXPENSES AND REASONABLE ATTORNEYS' FEES ARISING OUT OF, DIRECTLY OR INDIRECTLY, ANY CLAIM OF PRODUCT LIABILITY, PERSONAL INJURY, OR DEATH ARISING IN ANY WAY OUT OF SUCH MISSION CRITICAL APPLICATION, WHETHER OR NOT INTEL OR ITS SUBCONTRACTOR WAS NEGLIGENT IN THE DESIGN, MANUFACTURE, OR WARNING OF THE INTEL PRODUCT OR ANY OF ITS PARTS.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined". Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

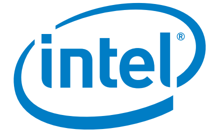
Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or go to: http://www.intel.com/design/literature.htm.

Intel and Intel logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

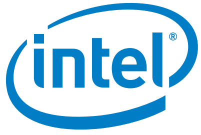
*Other names and brands may be claimed as the property of others.

Copyright © 2013, Intel Corporation. All Rights Reserved.

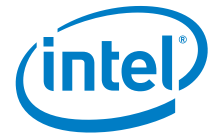


Revision History

Date	Revision	Description
April 2013	1.2	Initial Release (Intel Public).



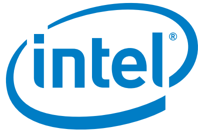
NOTE: ***This page intentionally left blank.***



1.0 General Description

1.1 Product Overview

The FM4224 and FM4112 are a fully-integrated, single-chip 24-port 10G Ethernet layer-2/3 switch chips offering wire-speed performance, extremely low-latency characteristics, and leading power efficiency. With its robust layer-2/3 switching and routing capabilities and the ubiquity of Ethernet, the FM4224 and FM4112 fit comfortably in a number of existing and emerging applications. And, with the unprecedented level of integration, these switch chips remove the cost, area, and power barrier for rapid and far-reaching 10G Ethernet deployment.

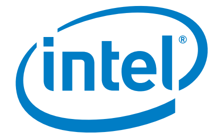


2.0 General Design Guidelines

2.1 Introduction

Meeting system performance requirements depends on good design practices. These practices minimize high-speed, digital-switching and common-mode noise, and provide shielding between internal circuits and the environment. Good design practices apply throughout the entire design, not just to the FM4224 and FM4112, and are outlined in Section 2.2 through Section 2.8.

Note: Intel recommends that these design practices be adhered to throughout the design process.

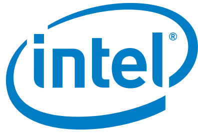


2.2 General Recommendations for Printed Circuit Board (PCB)

High speed serial interfaces such as XAUI require careful attention to trace layout in order to provide adequate signal integrity. This includes but is not limited to trace lengths, impedances, and terminations. Furthermore, the Ball Grid Array (BGA) package requires multiple PCB trace layers to accommodate routing of all signals.

Table 1 PCB Layer Stack-up Used in Evaluation Platform

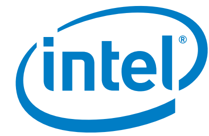
Layer #	Layer Type	Layer Name	Copper Weight (oz)	Thickness (mil)	Notes
1	Top	Top	0.5	2.1	0.7 mil Cu plus 1.4 mil plating thickness.
	Pre-preg 4450			7.8	
2	Plane	Gnd	0.5	0.7	
	Core 4350			6.6	
3	Signal	Signal-A	0.5	0.7	High speed XAUI signals Rx (1).
	Pre-preg 4450			6.6	
4	Plane	Gnd	0.5	0.7	
	Core 4350			6.6	
5	Signal	Signal-B	0.5	0.7	High speed XAUI signals Rx (2).
	Pre-preg 4450			7.8	
6	Plane	Gnd	0.5	0.7	
	Core 4350			6.6	
7	Signal	Signal-C	0.5	0.7	High speed XAUI signals Tx (all).
	Pre-preg 4450			7.8	
8	Plane	Gnd	0.5	0.7	
	Core FR4			6	
9	Signal	Signal-D	0.5	0.7	Low speed signals.



Layer #	Layer Type	Layer Name	Copper Weight (oz)	Thickness (mil)	Notes
	Pre-preg FR4			10	
10	Signal	Signal-E	0.5	0.7	Low speed signals.
	Core FR4			4	
11	Plane	PWR-3	1	1.4	3.3V (VDD33, VDDA33) 1.2V (VDDX).
	Pre-preg FR4			5	
12	Plane	Gnd	0.5	0.7	
	Core FR4			6	
13	Plane	PWR-2	1	1.4	1.5V (VTT).
	Pre-preg FR4			6	
14	Plane	Gnd	0.5	0.7	
	Core FR4			4	
15	Plane	PWR-1	1	1.4	1.2V (VDD).
	Pre-preg FR4			5	
16	Bottom	Bottom	0.5	2.1	0.7 mil Cu plus 1.4 mil plating thickness.

The PCB stack-up design used in the Evaluation Platform for the FM4224 uses 16 layers as listed in Table 1 and has been found to produce good results. This stack-up affords sufficient power, signal and ground planes and the proper spacing between them to provide the required power supplies and well isolated low-speed and high-speed signal routing.

This design uses high quality, low loss Rogers prepreg 4450 and core 4350 materials surrounding the high-speed signal layers instead of the more common FR4 material. This was done in order to preserve the signal integrity at the board egress CX4 connectors to the greatest extent possible and consequently drive the greatest length of CX4 cable possible. Use of the higher cost Rogers materials is not strictly required; however, FR4 material is thought to be sufficient for backplane applications in which up to 40 inches of total trace length are driven. Note that there is a trade-off to be made here. The more trace length driven in FR4 the shorter is the maximum length of CX4 cable that might subsequently be driven.



2.3 PCB Power Planes

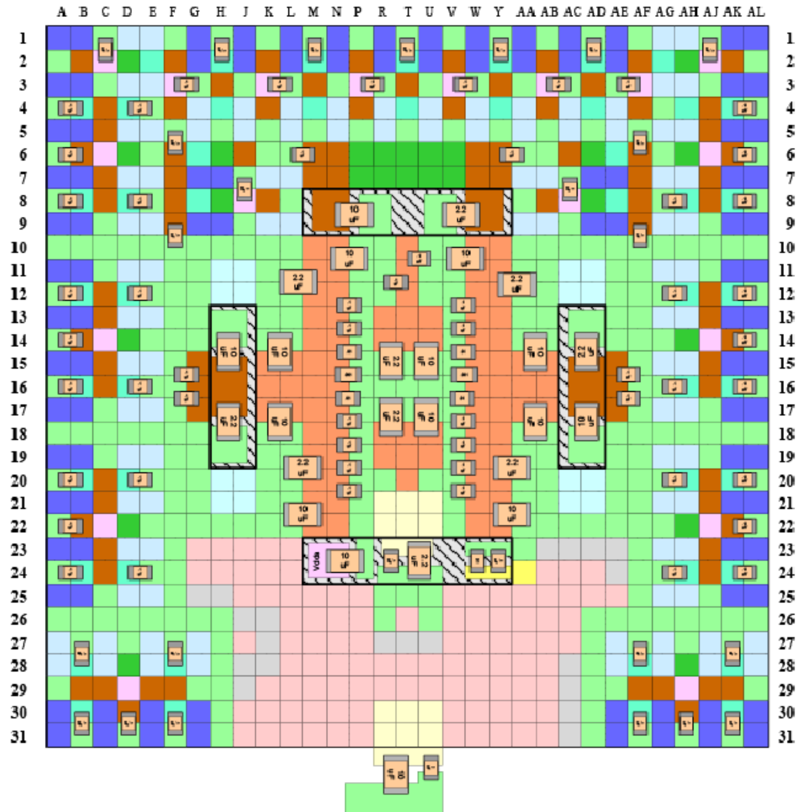
The FM4224/4112 requires 1.2 Vdc, 1.5 Vdc and 3.3 Vdc voltages. These should all be supplied through low-resistance power planes. The recommended power PCB plane thickness is a minimum 0.5 oz. copper; however, 1.0 oz. copper is preferred. The PCB ground planes should provide a low-resistance return path for high-speed switching currents. The recommended ground PCB plane thickness is a minimum 0.5 oz. copper. Prevent breaks or large voids in the ground plane by avoiding multiple ground references (split ground planes) or copper voids due to closely-spaced vias, high-density through-hole connectors, fine pitch BGA, or large board cutouts (slots). To minimize PCB layers, multiple power supplies can co-exist on the same layer, provided suitable spacing and isolation are furnished.

2.4 Power and Ground Filtering

- Follow good design practices to minimize noise from digital-switching and power-supply circuits.
- Ensure the power supply is rated for the load.
- Keep noise levels and ripple below the levels specified in Section 2.5.
- Filter the analog power circuits. Each analog power pin should have separate filter networks placed as close as possible to the respective power pin.
- Filter and shield DC-DC converters, oscillators, etc.

2.5 Decoupling, Bypassing and Bulk Capacitor Guidelines

The FM4224 and FM4112 are packaged in flip-chip BGA packages. In general, this means capacitors with values from 0.001 µF to 10 µF. Figure 1 through Figure 4 show capacitor placement and values for proven designs for both part numbers. These are suggested capacitor placement schemes and modifications can be made as required by specific board designs.



0603 6.3V

0603 6.3V

0402 10V

All Capacitors are X5R Dielectric

Power Supply Legend

- VDD
- VDDX
- VTT
- VDDA
- VDD33
- VDDA33
- VSS

Signal Legend

- TX HS Signals
- RX HS Signals
- REFCLOCKS
- TTL IO
- Analog IO RREF and DIODE
- NO CONNECT

Note: Capacitor value labels are small but visible with electronically magnified view.

Figure 1 FM4112 Capacitor Placement on Underside of PCB (Bottom View)

intel

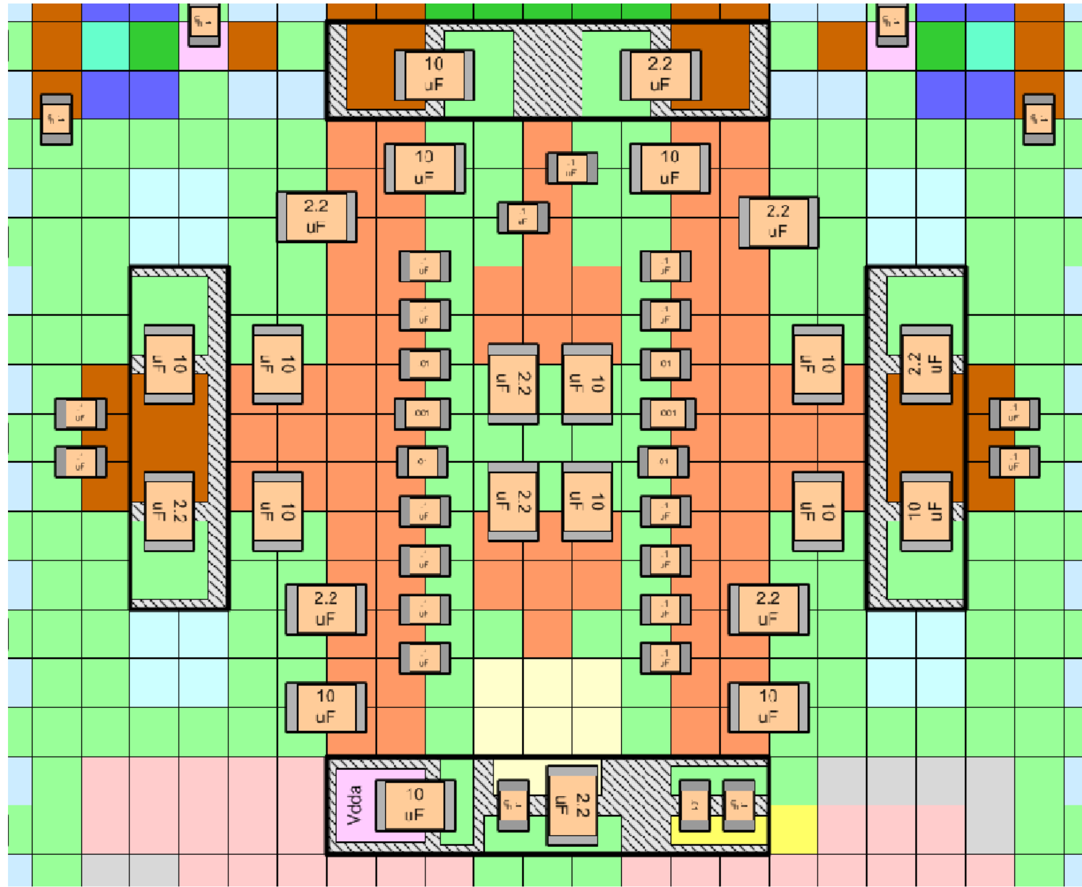


Figure 2 Magnified View of Capacitor Placement Beneath the Center Section of the FM4112

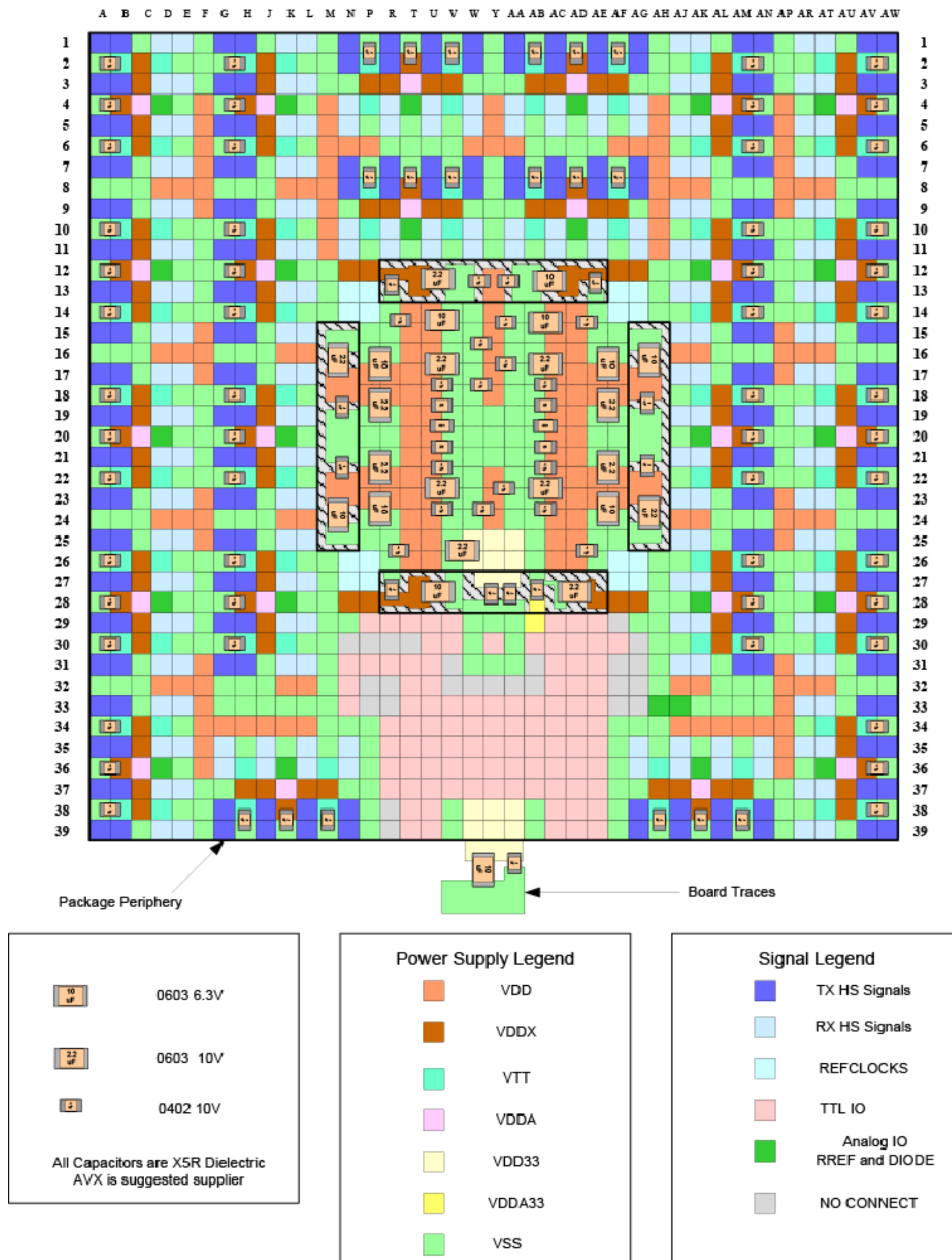


Figure 3 FM4224 Capacitor Placement on Underside of Board (Bottom View)

An exploded view of the interior portion of the chip is shown in Figure 4. Note that this capacitor layout is based upon, but does not exactly match, that of the Evaluation Platform. The layout given here is deemed to be slightly superior.

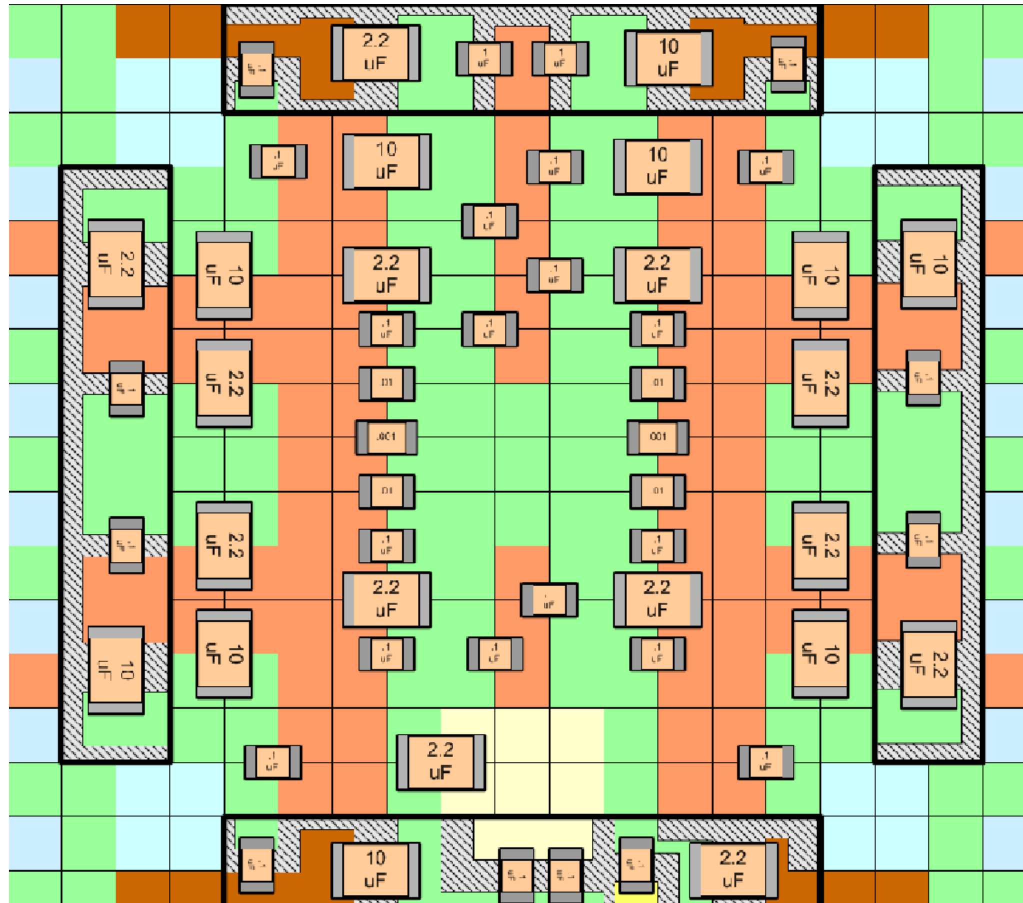
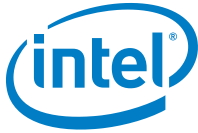


Figure 4 Magnified View of Capacitor Placement Beneath the Center Section of the FM4224

2.5.1 V_{DD} - 1.2V Core Voltage Supply

V_{DD} is the supply with the largest current transients. This supply needs the highest quality and largest values of decoupling. The supply voltage tolerance is ±5% and the min/max ripple should be kept to -100/+100 mV (10 MHz - 2.13 MHz).

In some cases it is not desired or possible to attach capacitors directly to the ball locations on the PCB. In those cases, it is critical that the layer guidelines be followed closely to ensure that there is as much high-quality distributed capacitance as possible lying within the PCB power planes.



2.5.2 V_{DDA} - 1.2V SerDes Bias Generator Supply

±5% tolerance. Filtering with a ferrite and capacitor is suggested to keep this supply quiet.

2.5.3 V_{DDX} - 1.2V SerDes PLL, CDR and Digital Supply

2.5.4 V_{DD33} - 3.3V LVTTL Supply

The V_{DD33} supply is an LVTTL supply with a ±10% voltage tolerance. As such, it might be fairly noisy but the LVTTL I/O are not very noise sensitive. Frequencies are relatively low. See the layout in Figure 1 through Figure 4 and note that several values are used. 0.1 µF and 2.2 µF close to the die with added 10 µF and 0.1 µF external to the package.

2.5.5 V_{DDA33} - 3.3V Analog Supply

This should be a very quiet supply. In the Evaluation Platform PCB stack-up shown, V_{DDA33} is derived from V_{DDA33} and should be filtered to keep ripple better than -100/ +100 mV. Reference clock frequencies are in the range of 100 to 400 MHz (312.5 MHz for XAUI). The SerDes PLL transfer function with respect to power supply noise is a high-pass function. Power supply noise within the PLL loop bandwidth of 31.2 MHz is transferred to the XAUI outputs.

2.5.6 V_{TT} - 1.2V to 1.8V Transmitter Termination Voltage

The driver termination voltage, V_{TT}, can be set to values between V_{DD} (1.2V) and 1.8V. V_{TT} in conjunction with V_{SW}, the single-ended voltage swing, controls the Transmitter Common Mode Voltage, V_{TCM} according to the following equation:

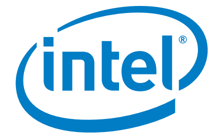
$$V_{TCM} = V_{TT} - V_{SW}$$

where V_{SW} is the product of driver current, I_{DR}, and the termination impedance, 25Ω. The drive current, in turn, is determined by setting the Nominal Drive Current bits (HiDrv and LoDrv) in the SERDES_CNTL_2 register and the Dtx bits in the SERDES_CNTL_1 register.

The single ended V_{OH} and V_{OL} values can then be computed from V_{TT} and V_{SW}:

$$V_{OH} = V_{TT} - 0.5 * V_{SW}$$

$$V_{OL} = V_{TT} - 1.5 * V_{SW}$$



Example:

1. Assume HiDrv = LoDrv = 0, implying a nominal drive current of 20 mA.
2. Assume the Dtx bits for the output in question are set to 0111b (0x7), corresponding to a drive current multiplier of 1.35. Total I_{DR} is then 20 x 1.35 = 27 mA.
3. The single ended voltage swing, V_{SW} is then 27 mA x 25Ω = 0.675V ($V_{SW(diff)}$ = 1.35V).
4. Assume V_{TT} is set to its maximum value of 1.8V.
5. V_{TCM} is computed as V_{TT} - V_{SW}, or 1.8 - 0.675 = 1.125V.
6. V_{OH} and V_{OL} are computed as 1.4625V and 0.7875V, respectively. The resulting output waveforms would look like those shown in Figure 5.

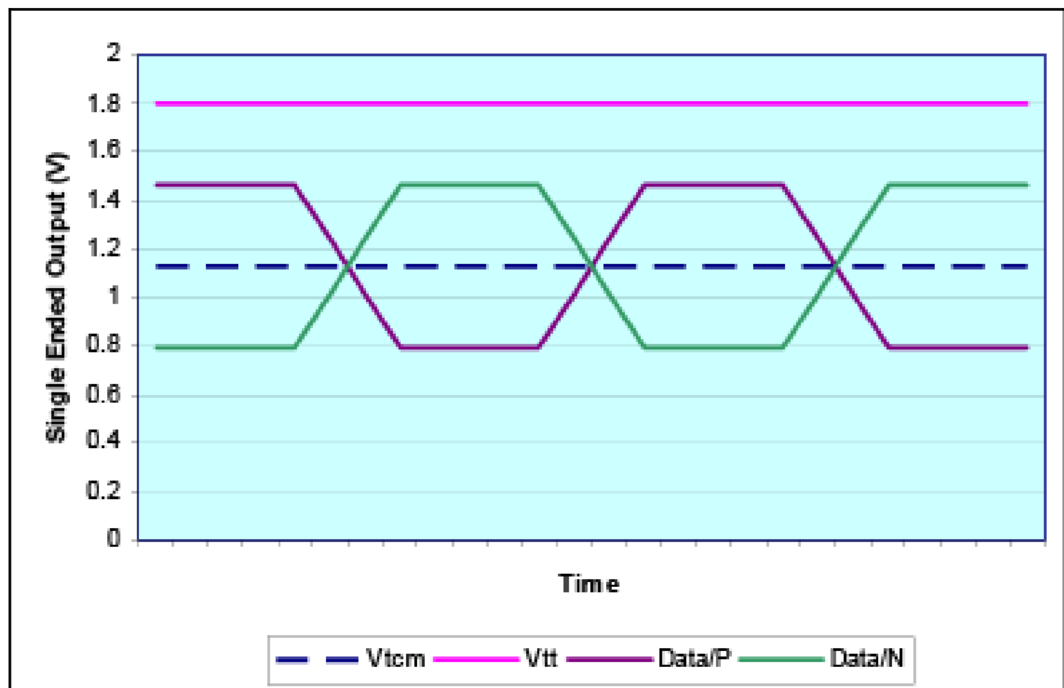


Figure 5 Output Waveforms Resulting from the Example Settings

These settings can be used to drive a long distance over a backplane or CX4 cable. In that case, setting the DEQ bits in the SERDES_CNTL_1 register to a high level of pre-emphasis might also be desirable.

2.6 Power and Ground Planes

- Provide ample power and ground planes.
- Avoid breaks in the ground plane, especially in areas where the plane shields high-frequency signals.
- Route high-speed signals adjacent to a continuous, unbroken ground plane.
- Stagger vias to avoid creating moats caused by anti-pad voids in the planes.
- When possible, fill in unused areas of the signal planes with solid copper, and attach them with vias to a V_{DD} or ground plane that is not adjacent to the signal layer. This technique is referred to as signal layer filling and can improve capacitive coupling of the power planes.

2.7 Differential Signal Layout

- Route differential pairs close together, the same distance, and away from other signals.
- Keep each differential pair on the same plane and groups of four XAUI lanes on the same plane. To minimize crosstalk at the receiver, it is best to route groups of four Rx XAUI lanes that are physically adjacent at the switch chip on alternate layers (see . This is shown in Figure 6.

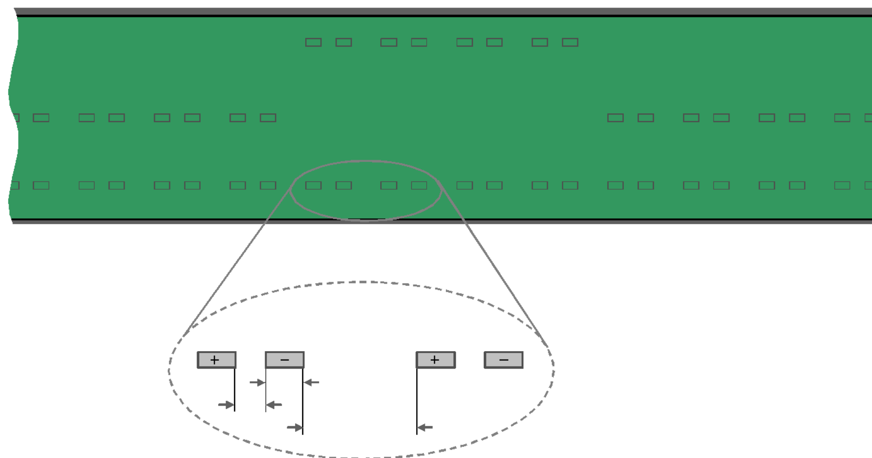
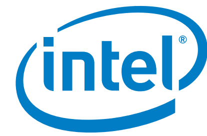


Figure 6 Cross Sectional View of XAUI Tx and Rx Signal Routing

- Minimize vias and layer changes.
- Transmit XAUI lanes can be interleaved: P01_TA - P02_TA - P01_TB - P02_TB, etc. This is shown in Figure 6, and the reason for doing so is because the Tx signal traces are most effectively routed away from the switch chip in this manner. An example of the interleaved Tx signal routing at the switch chip is also shown in Figure 6. This is allowed electrically because the crosstalk is small (about 10 mV). Since all the Tx



lines have relatively large signal amplitudes the crosstalk does not interfere with the data.

- Keep transmit and receive pairs away from each other; run orthogonally, or separate with a ground plane layer. The Tx and Rx signal must not run together. This is because the XAUI specification allows for long cable or trace runs. At the end of a trace or cable run the resulting signal is quite small. Adjacent line crosstalk between differential pairs is seen as single ended noise. With a strong Tx signal adjacent to an Rx signal returning through a long cable or trace length the resulting strong aggressor to weak victim crosstalk can easily cause data corruption.
- Within each XAUI data path all the differential pairs should be matched in electrical length as closely as possible.
- Beneath the package, route signal traces between rows of package balls as shown in Figure 7. Signal traces are 4 mils wide between package balls and expand to 5.5 mils outside the package perimeter.
- Below each package ball there should be an anti-pad in the ground plane. An anti-pad is a cutout in the ground plane roughly the size of the ball to avoid a large capacitive discontinuity at the package board interface.
- Careful attention must be paid to the electrical length of each signal. Signals routed on the surface of the board as microstrip has a different propagation velocity than signals routed as stripline within the PCB.
- All traces should be run above or between a fixed, solid-ground plane(s) that is continuous across the length of the traces.
- All traces should be implemented as 100 Ω differential pairs. These can be either micro-strip or stripline-type differential pairs.
- Each trace is routed continuously on one layer with a maximum of two vias located at the driver BGA and receiver component.

- All XAUI PCB traces should be designed to the specifications listed in Table 2.

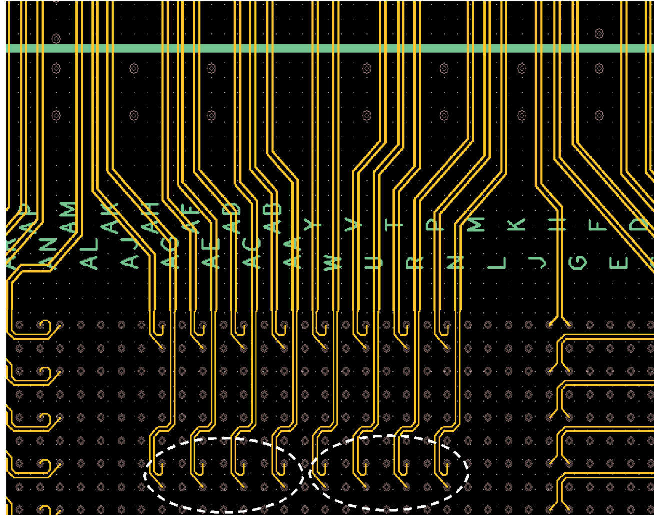


Figure 7 Trace Routing on Tx Interface Showing Interleaving of XAUI lanes (Four XAUI Tx Interfaces Shown)

XAUI interfaces are AC coupled at the receive end of the link, so provision must be made for the coupling capacitors as shown in Figure 8. Signal trace information is listed in Table 2.

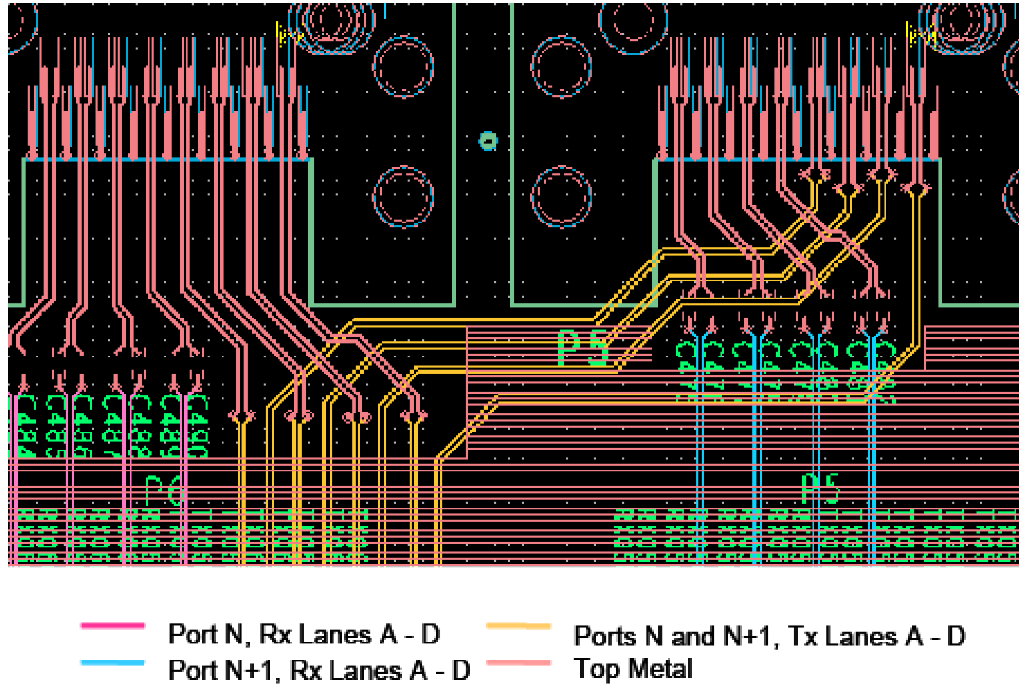
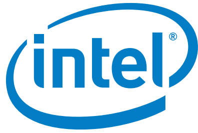


Figure 8 XAUI Tx Lane De-interleaving (Accommodation for Rx Coupling Capacitors Also Shown)

Table 2 XAUI PCB Trace Specifications

	Transmit	Receive
Trace Width (w)	4/5.5 mil	-
Differential Trace Separation (ds)	4/5.5 mil	-
Inter-pair Separation (is)	3 x ds	3 x ds
Trace Impedance	100 Ω diff	-
Trace Length	-	10 inches
Trace Difference	-	50 mil



2.8 Clock Requirements

2.8.1 Frame Handler

A reference clock for the frame handler operation is input on the FH_PLL_REFCLK pin. This clock input has minimum/maximum values of 1.2 MHz and 70 MHz, respectively and is a single-ended input.

2.8.2 SerDes Clocks

The FM4224/4112 requires reference clock inputs for each XAUI port, whether functioning in quad SerDes mode (10G) or in single SerDes mode (2.5G/1G). The clocks are used for the internal PLLs that generate the clocks used for data serialization and transmission in the transmit (serialize) direction and for clock and data recovery in the receive (de-serialize) direction.

The reference clocks are input on pins RCK[1:4](A/B) and require a source between 100 MHz and 400 MHz. This source should be generated by a high-quality, low-jitter crystal oscillator such as the Epson* EG-2101CA, -2121CA or -2102CA families of devices.

This clock input is compatible with CML, LVPECL or LVDS drivers and must be enabled at all times unless the port is powered down. Peak-to-peak jitter should be less than 30 pS.

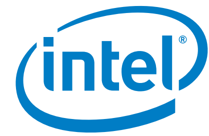
High-quality differential signal reference clock inputs are required for each SerDes. Up to four different reference clocks can be used, each supporting a group of six SerDes grouped as follows:

- Ports 2, 4, 6, 8, 10, 12
- Ports 14, 16, 18, 20, 22, 24
- Ports 13, 15, 17, 19, 21, 23
- Ports 1, 3, 5, 7, 9, 11

Note that one, two, three or four clock sources can be routed to the four clock inputs. The reference clock input should have less than 0.1 UI of jitter, peak to peak. The PLL loop bandwidth is set to $1/10^{th}$ of the reference clock frequency, so for XAUI signals, jitter of frequencies less than about 31.2 MHz are transferred from the reference clock to the XAUI outputs (ref clock freq = 312.5 MHz).

2.8.3 Clock Layout Guidelines

- Keep the clock traces as short as possible.
- Route the clock traces adjacent to an unbroken ground plane.
- Route the clock traces as 100 Ω differential traces.
- Minimize vias and layer changes.



2.8.4 CPU Interface

The FM4224 requires an LVTTL input clock on the CPU_CLK pin for the CPU interface. This clock must not exceed 100 MHZ and must have a maximum duty cycle variance of 45/55.

2.9 Unused Inputs/Outputs

2.9.1 XAUI Interface

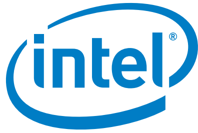
Unused XAUI inputs and outputs as well as unused reference clock inputs can be left floating.

2.9.2 SPI

If the SPI interface is not used, SPI_SI should be tied high, while SPI_SO, SPI_CS_N and SPI_SCK can be left floating.

2.9.3 JTAG

If the JTAG interface is not used, drive TRST_N low. TDI, TCK and TMS should be tied high; TDO can be left floating.



3.0 Thermal Design

3.1 Heat Sinking

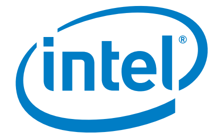
Under most operating conditions, a heat sink is required in order to keep the case temperature below the specified case temperature. The maximum case temperature is as follows:

- Samples, pre-production: 70 ºC.
- Production: TBD, pending thermal characterization.

Actual heat sinking requirements should be calculated based on air flow, air inlet temperature and anticipated power dissipation and must take into account the parallel thermal impedance of 3 ºC/Watt between the switch chip and the PCB. A sample calculation follows and is based on the following parameters:

Parameter	Definition	Value	Unit
P_D	FM4224 power dissipation	36	W
T_A	Ambient air inlet temperature	40	ºC
T_{cmax}	Maximum case temperature	70	ºC

With 40 ºC ambient temperature and 70 ºC maximum allowable case temperature, a 30 ºC case to ambient temperature rise is allowed even when the FM4224 is running at its maximum of 36 W power dissipation. The two thermal paths from the case - one thermal impedance through the solder balls to the PCB (θ_{PCB}) and the other thermal impedance through the externally provided heat sink (θ_{HS}) - when considered in parallel must limit the temperature rise to 30 ºC. The two thermal resistances in parallel equal a total thermal resistance that must be at or below approximately 0.75 ºC/W:



$$\theta_{TOT} = \frac{\theta_{HS} \times \theta_{PCB}}{\theta_{HS} + \theta_{PCB}}$$

Or solving for R_{HS}:

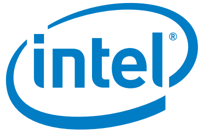
$$\theta_{HS} = \frac{-\theta_{TOT} \times \theta_{PCB}}{\theta_{TOT} - \theta_{PCB}}$$

With R_{TOT} constrained to be 0.75 ºC/W or lower and R_{PCB} equal to 3 ºC/W, R_{HS} must be 1.0 ºC/W or lower.

Intel has successfully used copper round pin BGA heatsinks from Radian Heatsinks* (www.radianheatsinks.com), specifically model# INM40002-25PCU/2.6Y+T710. This heatsink results in the following air-flow dependent thermal resistances:

Air flow (LFM	100	200	400	600
Theta (ºC/W)	2.1	1.4	0.9	0.7

Under these conditions, an air-flow of 400 LFM or more results in thermal resistances lower than 1 ºC/W.



NOTE: This page intentionally left blank.