UEFI Firmware Solutions for Enterprise Servers: A Case Study in 8-way Processor Support

Kevin Li - Engineering Manager, Intel
Hu Leijun - Technology Director, Inspur
Jeff Bobzin - Senior Director Software Architecture, Insyde Software, and Secretary UEFI Board

EFIS005

Sponsors of Tomorrow.™ intel

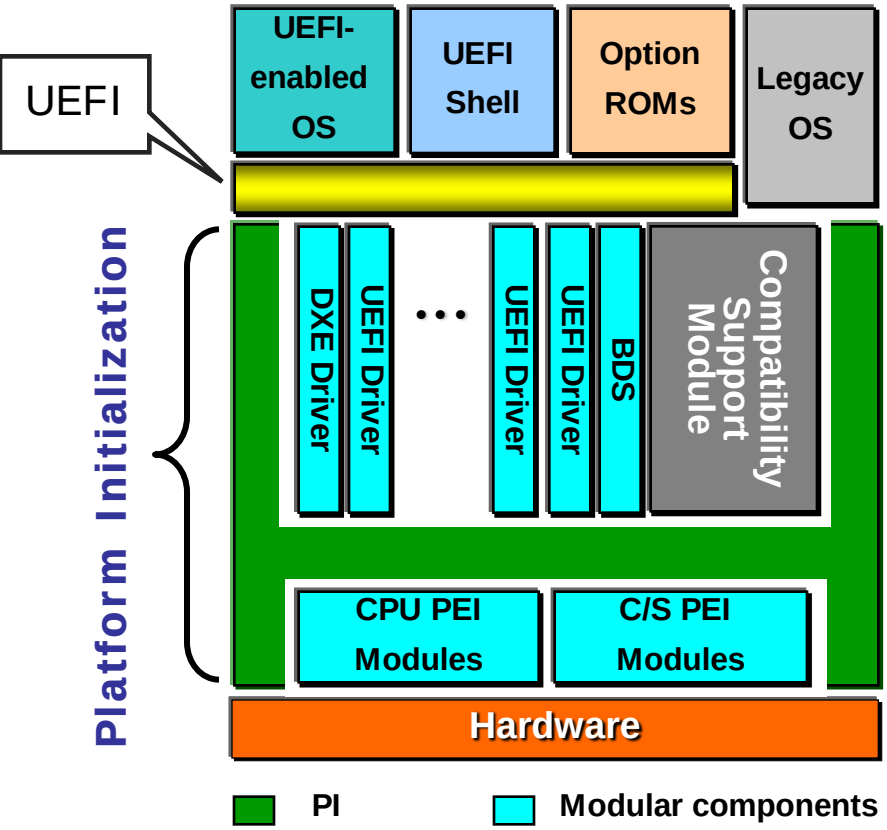
Agenda

- UEFI promotes scalability - Intel
- 8SG Server Introduction - Inspur
- Firmware for 8-way Server - Insyde

Agenda

- **UEFI promotes scalability - Intel**
- **8SG Server Introduction - Inspur**
- **Firmware for 8-way Server - Insyde**

Standard Firmware Interfaces



- UEFI: Unified Extensible Firmware Interface
  - a new model for the interface between the OS and platform firmware
- PI: Platform Initialization
  - Standardization: key to interoperability across implementations
  - Modular components like silicon drivers (e.g. PCI) and value-add drivers (security)
  - Preferred way to build UEFI

UEFI is Architected for Dynamic Modularity

IDF2010
INTEL DEVELOPER FORUM

UEFI Pre-boot Advantages

- More robust boot loader using UEFI
  - Easier to implement “Failover Boot” solution
- UEFI Shell: Full featured utility shell
- Easily reload, unload & update UEFI drivers in Pre-OS
- Pre-OS networking - Full IPv4 and IPv6 network stack, PXE boot, iSCSI
- Pre-OS applications can run in UEFI Boot Services layer
  - lot of useful DOS-like tools can be ported to UEFI applications

UEFI enhances pre-OS space for server features

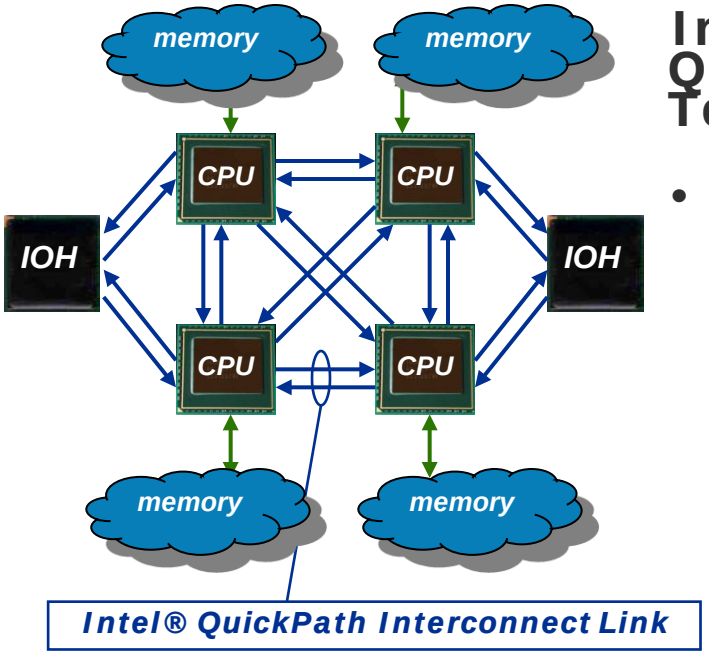
IDF2010
INTEL DEVELOPER FORUM

UEFI Advantages in Scalability

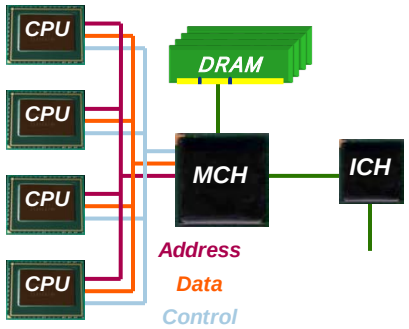
- Defines standard interfaces across different platforms & architectures
- Common code for IA32, X64, and Intel® Itanium® architecture
  - We use the same Intel® 5520 Chipset with 82801JB I/O Controller Hub (ICH10) chipset code across all the segments
- Based on protocols instead of proprietary implementations

UEFI makes scalable server support easier

Intel® QuickPath Interconnect vs. FSB



FSB



Intel® QuickPath Interconnect (Intel® QPI) drives a leap forward in Platform Technology

- **Scalable solution**
  - Much higher link bandwidth than FSB
    - Headroom for higher transfer rates
  - Vastly greater MP system bandwidth with multiple, independent memory controllers and Intel QuickPath Interconnect links
    - Scales efficiently with number of processors
  - Many system topologies with more than four processors supported
  - Common interface for Intel® Itanium® and Xeon® Processor based systems
- **Improved system robustness**
  - Additional levels of error recovery and logging for mission critical systems
  - RAS features

Highly Configurable System Interconnect

IDF2010
INTEL DEVELOPER FORUM

Agenda

- UEFI promotes scalability - Intel
- 8SG Server Introduction - Inspur
- Firmware for 8-way Server - Insyde

Inspur 8SG Server Introduction

- Inspur 8SG Server Overview
- Inspur 8SG Platform Topology
- Inspur 8SG Flexible Partitioning
- Inspur 8SG Advanced RAS Features
- Inspur 8SG Server Management

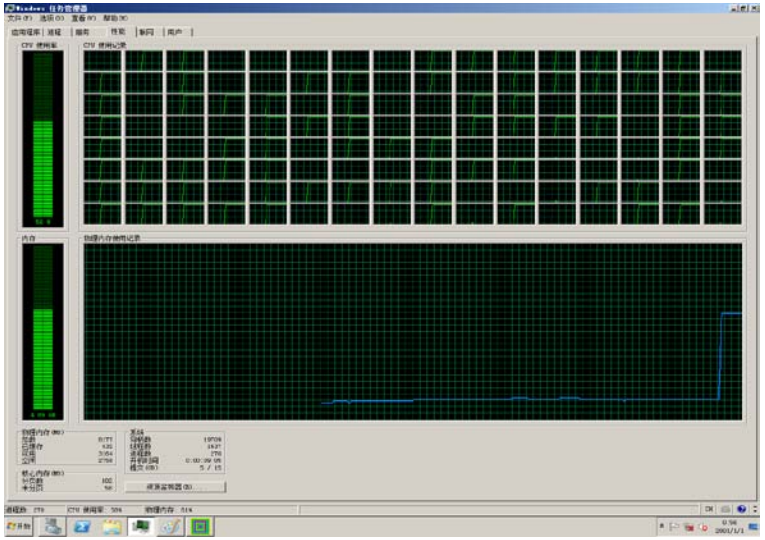
First 8-socket 8-core modular server in China

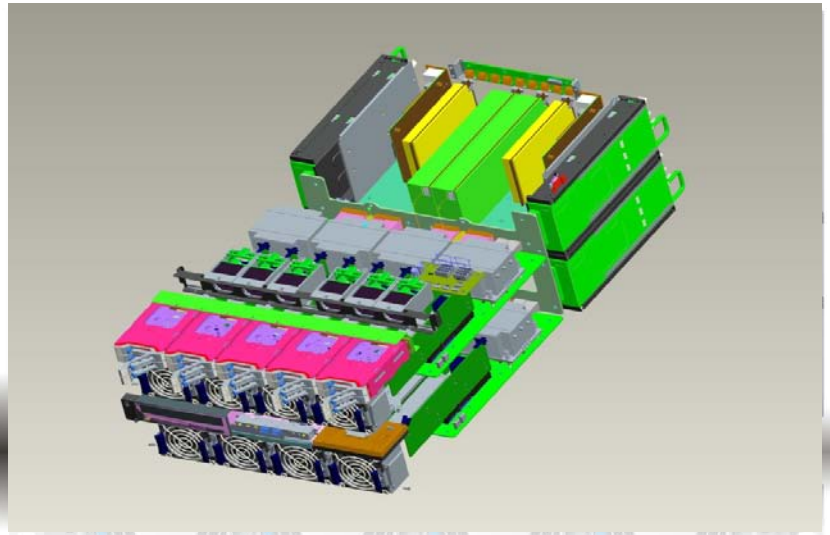
Up to 128 logical processors offering exceptional performance!

insyde®
inspur 浪潮
IDF2010
INTEL DEVELOPER FORUM

Inspur 8SG Server Overview

- 8-socket Intel® Xeon® Processor 7500 series (Nehalem-EX), with 64C/128T in 6U mainframe
- 64 DIMMs，up to 1TB size; 204.8GB（25.6GB*8）Memory bandwidth
- Exceptional I/O performance：72X PCI Express* Gen2
- Advanced RAS features: Intel® Scalable Memory Interconnect (Intel® SMI)/Intel® QPI self healing, Machine Check Architecture, Memory migration, Physical partitioning



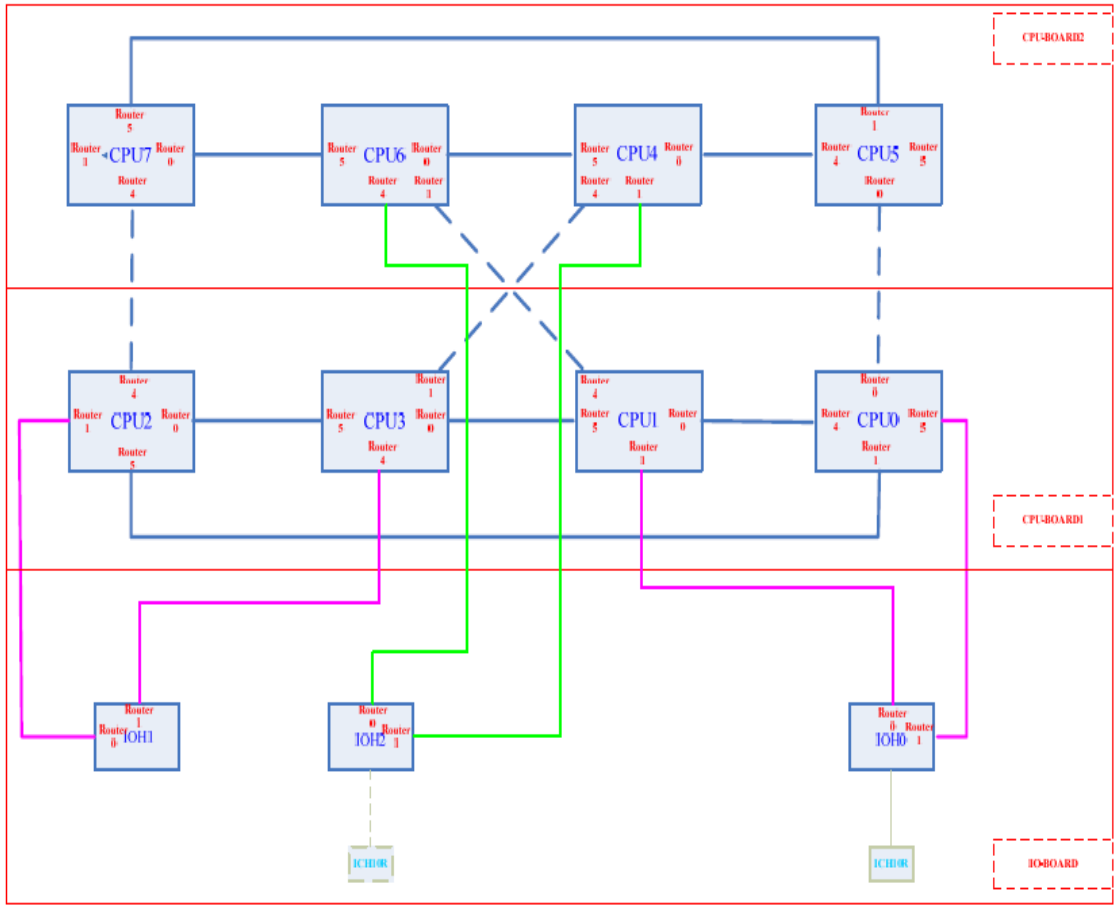


insyde®

inspur 浪潮

IDF2010
INTEL DEVELOPER FORUM

Inspur 8-Socket Glueless Platform Topology



- Built upon a 2-socket per IOH building block offering Flexibility and Modularity
- Topology selected to minimized transaction hops and latency
- Single partition includes one legacy IOH, 2 non-Legacy IOHs and one active 82801JB I/O Controller Hub (ICH10)
- 6.4 GT/s link speed is critical to 8-socket glueless performance

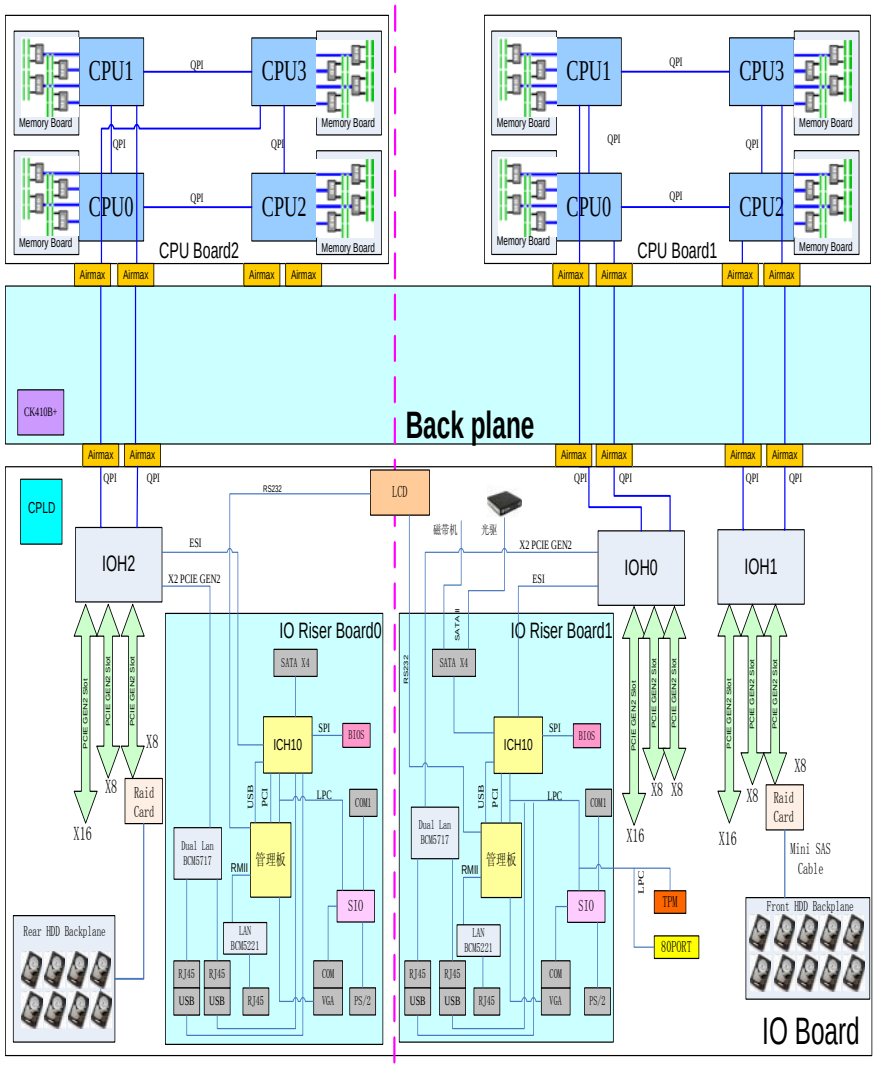
UEFI Platform Initialization Specification makes the scalable server support easier

insyde®

inspur 浪潮

IDF2010
INTEL DEVELOPER FORUM

Inspur 8SG Flexible Partitioning



- **Flexible system configuration**
  - 8SG system
    - **8SG+3 IOHs+1 ICH**
  - 4SG system
    - **4SG+2 IOHs+1 ICH**
    - **4SG+1 IOH+1 ICH**
- **Intelligent Partitioning**
  - Auto switch front-side device ports
  - Software configure without hardware alteration
  - Enhanced BIOS Design
    - **Single BIOS image used for different configurations**
  - Enhanced Partition Manage Feature
    - **Auto swap between 8S mode and dual 4S mode**

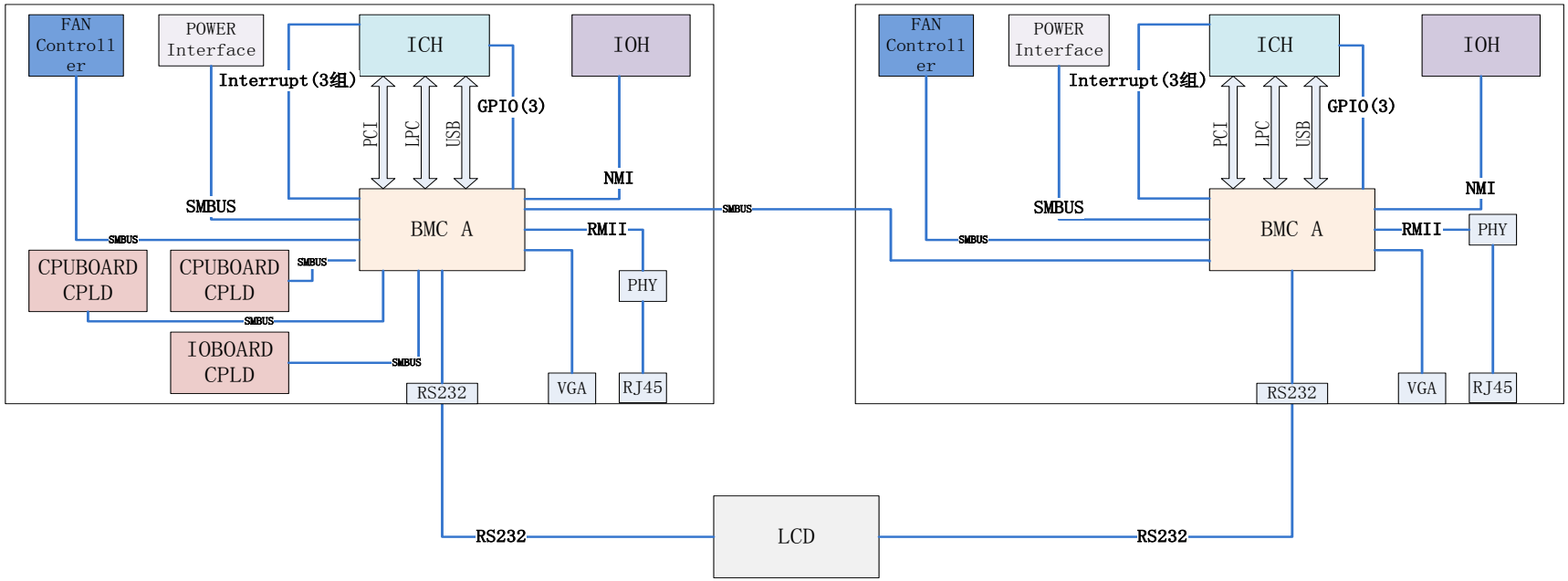
insyde®
inspur 浪潮
IDF2010
INTEL DEVELOPER FORUM

Inspur 8SG Server Advanced RAS Features

RAS Features	Types	Descriptions
Memory	Memory Mirroring	Intra socket, Inter socket
	Memory Demand and Patrol scrubbing	Demand scrubbing :write corrected data back to the memory once a correctable error is detected on a read transaction. Patrol scrubbing : proactively searches the system memory, repairing correctable errors.
	DRAM SDDC	x4 or x8 Single Device Data Correction
	…	
System Link Layer	Self-healing	Downshift link width on the error link
	Lane Failover	Adaptive routing
	CRC	Cyclic redundancy check
	Clock Failover	Redirect forwarded clocks to the clock fail-over lane
	Link Retry	Restart a cycle on the failure link
	…	
System Partitioning	Static hard partitioning	Allows a system to be divided into 2 machines, each capable of running separated OS and applications
	Partition via Virtualization	Support for Virtualization features VT-x2/VT-d2
Hot-Plug	Processor Board hot-plug	Dynamic hot-add/remove processor modules
	IOH Board hot-plug	Dynamic hot-add/remove IOH modules
	PCIe hot-plug	Native PCIe hot-plug support
	……	

insyde® inspur 浪潮 IDF2010 INTEL DEVELOPER FORUM

Inspur 8SG Server Management



- **System Management**
  - Intelligent Power Management
  - System Static Partitioning
  - IP KVM Remote Control
  - Virtual Storage Media
  - PCI-E hot-plug support
- **Human-Machine Interface**
  - Colorful LCD touch pad for end user
- **System Status Monitoring and Error Handling**
  - Chip level Error monitoring and handling
  - System thermal event monitoring and handling
  - System Boot-up Monitoring and Management
  - General states monitoring

insyde® inspur 浪潮 IDF2010 INTEL DEVELOPER FORUM

Agenda

- UEFI promotes scalability - Intel
- 8SG Server Introduction - Inspur
- Firmware for 8-way Server - Insyde

Scaling Firmware Solution to 8-way

- Insyde's challenge: scale up proven 2/4 socket firmware to support 8-way system
- Intel® QuickPath Interconnect link allows system hardware to scale up from proven 2 or 4 socket design to 8-way
- Similarly, modularity of UEFI allows firmware to scale up to meet the challenge of the larger (8-way) system

insyde® inspur 浪潮 IDF2010 INTEL DEVELOPER FORUM

Project Goals

- Tight schedule: needed to have 8-way hardware running before February Spring Festival
- Schedule goals could only be achieved if firmware project used best techniques
  1. *Analyze the Technical Challenges*
  2. *Break up required development into pieces that could be pre-tested*

insyde® inspur 浪潮 IDF2010 INTEL DEVELOPER FORUM

Agenda

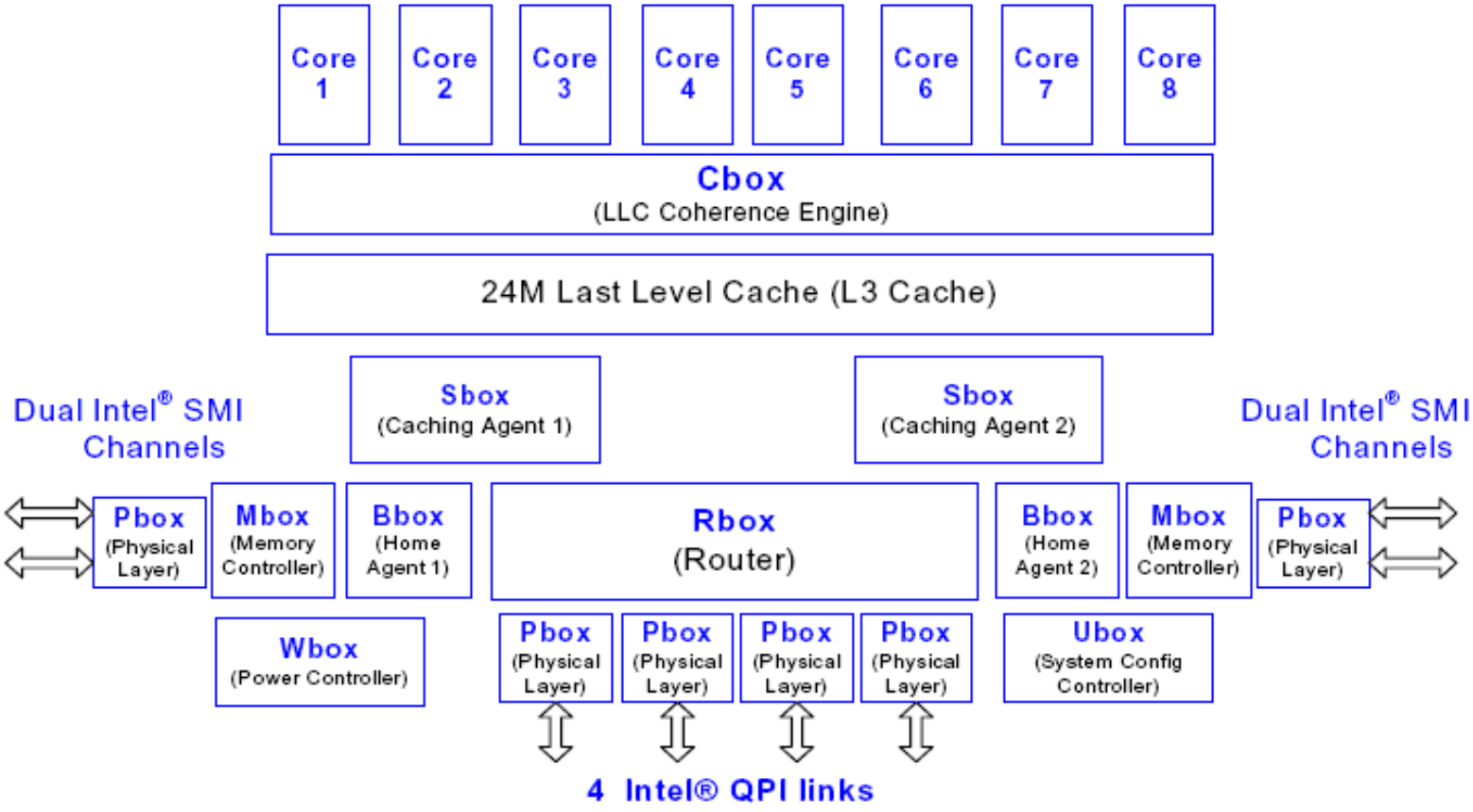
- UEFI promotes scalability - Intel
- 8SG Server Introduction - Inspur
- Firmware for 8-way Server - Insyde

1. ***Analyze the Technical Challenges***
2. *Break up development into pieces that could be pre-tested*

insyde® inspur 浪潮 IDF2010 INTEL DEVELOPER FORUM

Challenge: Dissect the CPU

The Intel® Xeon® Processor 7500 series has a complex Internal Structure



insyde®
inspur 浪潮
IDF2010
INTEL DEVELOPER FORUM

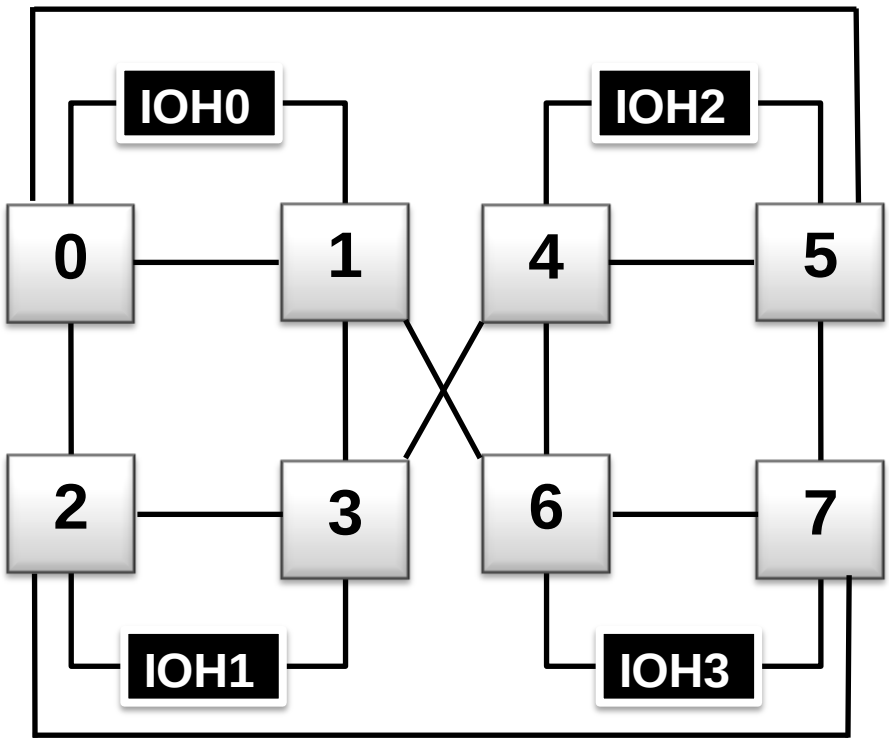
Challenge: Processor Initialization

Boot Mode?	Designer chose to use Intel® QPI boot mode – cores will boot from ICH Flash
Core Count?	Expand tables for 8 (sockets) x 8 (cores) x 2 (threads) == 128
APIC ID?	3 bits socket ID 5 bits thread ID == 256 IDs Work -- Expand tables
x2APIC?	Available, must be supported, but APIC still available too. (this is largest server that fits in APIC)

insyde® inspur 浪潮 IDF2010 INTEL DEVELOPER FORUM

Challenge: 8 Socket Routing

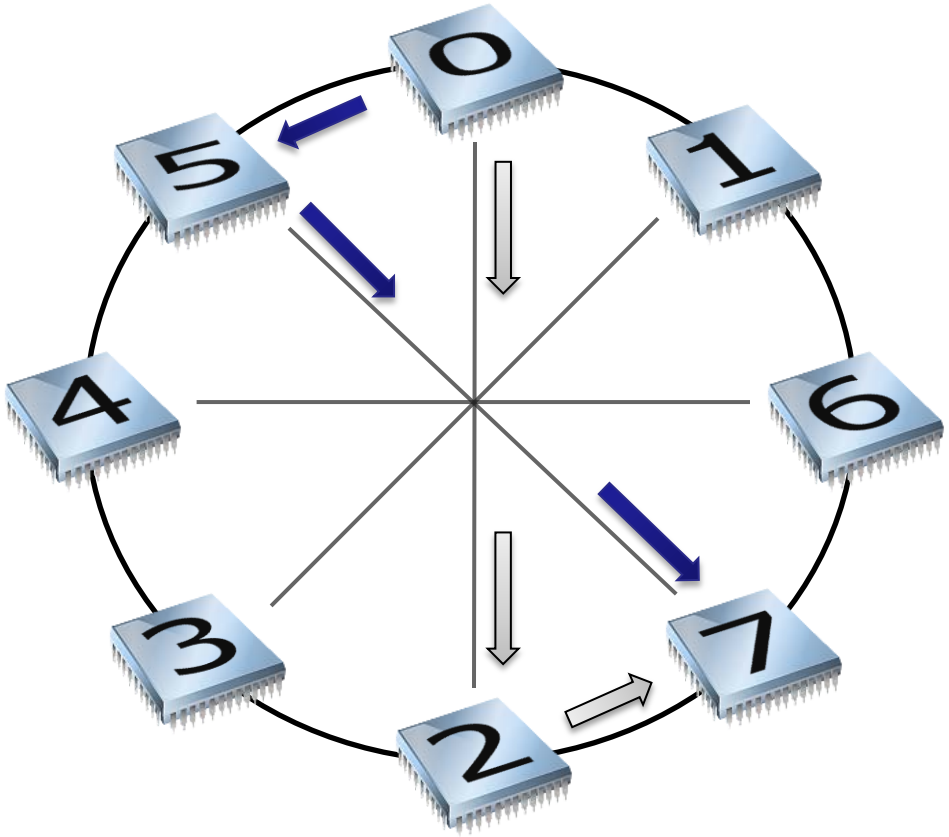
- This design is 'Glueless' → No Node Controller
- Intel® QPI fabric expanded to include 8 Processor sockets
- Processor-to-processor links are 1-hop or 2-hop
- Maximum of 2-hop for performance



insyde®
inspur 浪潮
IDF2010
INTEL DEVELOPER FORUM

Processor Topology is a Ring

- Key to Maximum Performance is traffic balance
- For example, there are 2 shortest routes for traffic between 0 and 7
- Always choose the route least used by previous selections
- Dynamic algorithm gives good result – but can be further tuned



Traffic Balance is a Key to Performance

insyde®
inspur 浪潮
IDF2010
INTEL DEVELOPER FORUM

Route Table Entries - RTA

- Key Feature of Rbox
- Initialized by Firmware
- Route Table Entries (RTA) direct messages
  - From Internal Box to Correct External Port
  - Incoming from one External Port to another External Port (pass-through)
- Each RTA Entry Includes choice of VN0/VN1
  - VN0 and VN1 are alternate transaction trackers
  - Prevent deadlocks that might occur if only VNA was used

insyde® inspur 浪潮 IDF2010 INTEL DEVELOPER FORUM

Agenda

- UEFI promotes scalability - Intel
- 8SG Server Introduction - Inspur
- Firmware for 8-way Server - Insyde

1. *Analyze the Technical Challenges*
2. ***Break up development into testable pieces***

insyde® inspur 浪潮 IDF2010 INTEL DEVELOPER FORUM

Steps to be Ready for Power-On

- Project unit-test criteria were created
- 8-Way features were tested on 4-way

STAGE	REQUIREMENT
SEC - (pre-PEI CPU Startup)	Expand check-in table
PEI – QPI Init	Discovery topology, Build Routing,
PEI – Memory Init	Verify 8-way Compatibility
DXE – MPCPU Protocol	Expand data tables
DXE – ACPIPLATFORM	Build expanded ACPI Tables
SMI – SMI Handler	Expand TSEG allocation
RAS – RAS Drivers	Identify 8-way impact on RAS

insyde® inspur 浪潮 IDF2010 INTEL DEVELOPER FORUM

Supporting 8-way Routing

- 8-socket is not a standard supported feature of Intel® QPI Init Reference Code

Steps of Routing Code

1. **SBSP Selected**
2. **ID neighbors. ID neighbor's neighbors.**
3. **Walk all links building node list (tree)**
4. **Assign Virtual Network VN0, VN1 to links**
5. **Program RTIDs for directed Messages**
6. **Program RTIDs for Broadcast**

insyde® inspur 浪潮 IDF2010 INTEL DEVELOPER FORUM

Unit Test the Router

- Insyde needed to write complex code to build 8-way router entries but CRB was only 4-way
- Solution: QPISIM software simulator executes router code on desktop
- Insyde and Inspur engineers studied output and compared against desired
- Simulator can also test routing with non-standard configuration
  1. Fewer than 8 Sockets populated
  2. Intel® QPI Link failures – does the system fallback?

insyde® inspur 浪潮 IDF2010 INTEL DEVELOPER FORUM

Agenda

- UEFI promotes scalability - Intel
- 8SG Server Introduction - Inspur
- Firmware for 8-way Server - Insyde

Conclusions

insyde® inspur 浪潮 IDF2010 INTEL DEVELOPER FORUM

Smooth Scaling

- Power-on completed ahead of schedule with all sockets working
- Intel® QPI allows system to scale from 2 to 4 to 8 socket while maintaining consistent architectural elements
- UEFI Firmware also scales using:
  - Modularity
  - Defined Interfaces
  - Architected Flow of Control

Challenges Met With UEFI Firmware

insyde®

inspur 浪潮

Next Steps

- Key Understandings:
  - ***Intel® QPI Architecture scales smoothly from 1 to 8 Sockets***
  - ***UEFI Firmware is mature and ready for the toughest challenges***
  - ***For increased schedules confidence break your project into testable pieces***

insyde® inspur 浪潮 IDF2010 INTEL DEVELOPER FORUM

Additional resources on UEFI:

- Other UEFI Sessions – Next slide
- More web based info:
  - Specifications and Implementation sites: www.tianocore.org, www.uefi.org, www.intel.com/technology/efi

- Technical book from Intel Press: "Beyond BIOS: Implementing the Unified Extensible Firmware Interface with Intel's Framework" www.intel.com/intelpress
- UEFI Plugfest Event at Intel in Dupont Washington, June 22-25, 2010 www.uefi.org or email: laurie.jarlstrom@intel.com

IDF 2010 UEFI Spring Sessions
April 14

EFI#	Company	Description	Time	RM
S001 ✓	Intel, IBM, HP	Using the Latest EFI Development Kit (EDK II) for UEFI Advanced Development and Innovation	11:10	302AB
S002 ✓	Intel, HP, Byosoft	Notebook Advancements for Unified Extensible Firmware Interface (UEFI) for Pre-boot Productivity	13:00	302AB
S003 ✓	Intel, Byosoft	Unified Extensible Firmware Interface (UEFI): Best Platform Security Practices	14:00	302AB
S004 ✓	Intel, Microsoft, Insyde	UEFI Fast Boot for Microsoft* Windows* 7 : Fast Boot Without Compromising your BIOS	15:00	302AB
S005 ✓	Intel, Inspur, Insyde	UEFI Firmware Solutions for Enterprise Servers: A Case Study in 8-way Processor Support	16:00	302AB

✓ DONE

Session Presentations - PDFs

The PDF for this Session presentation is available from our IDF Content Catalog at the end of the day at:

intel.com/go/idfsessionsBJ

URL is on top of Session Agenda Pages in Pocket Guide

Please Fill out the Session Evaluation Form

Give the completed form to the room monitors as you exit!

Thank You for your input, we use it to improve future Intel Developer Forum events

Q&A

Legal Disclaimer

- 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. INTEL PRODUCTS ARE NOT INTENDED FOR USE IN MEDICAL, LIFE SAVING, OR LIFE SUSTAINING APPLICATIONS.
- Intel may make changes to specifications and product descriptions at any time, without notice.
- All products, dates, and figures specified are preliminary based on current expectations, and are subject to change without notice.
- Intel, processors, chipsets, and desktop boards 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.
- Nehalem and other code names featured are used internally within Intel to identify products that are in development and not yet publicly announced for release. Customers, licensees and other third parties are not authorized by Intel to use code names in advertising, promotion or marketing of any product or services and any such use of Intel's internal code names is at the sole risk of the user
- Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance.
- Intel, Xeon and the Intel logo are trademarks of Intel Corporation in the United States and other countries.
- *Other names and brands may be claimed as the property of others.
- Copyright © 2010 Intel Corporation.

Risk Factors

The above statements and any others in this document that refer to plans and expectations for the first quarter, the year and the future are forward-looking statements that involve a number of risks and uncertainties. Many factors could affect Intel's actual results, and variances from Intel's current expectations regarding such factors could cause actual results to differ materially from those expressed in these forward-looking statements. Intel presently considers the following to be the important factors that could cause actual results to differ materially from the corporation's expectations. Demand could be different from Intel's expectations due to factors including changes in business and economic conditions; customer acceptance of Intel's and competitors' products; changes in customer order patterns including order cancellations; and changes in the level of inventory at customers. Intel operates in intensely competitive industries that are characterized by a high percentage of costs that are fixed or difficult to reduce in the short term and product demand that is highly variable and difficult to forecast. Additionally, Intel is in the process of transitioning to its next generation of products on 32nm process technology, and there could be execution issues associated with these changes, including product defects and errata along with lower than anticipated manufacturing yields. Revenue and the gross margin percentage are affected by the timing of new Intel product introductions and the demand for and market acceptance of Intel's products; actions taken by Intel's competitors, including product offerings and introductions, marketing programs and pricing pressures and Intel's response to such actions; defects or disruptions in the supply of materials or resources; and Intel's ability to respond quickly to technological developments and to incorporate new features into its products. The gross margin percentage could vary significantly from expectations based on changes in revenue levels; product mix and pricing; start-up costs, including costs associated with the new 32nm process technology; variations in inventory valuation, including variations related to the timing of qualifying products for sale; excess or obsolete inventory; manufacturing yields; changes in unit costs; impairments of long-lived assets, including manufacturing, assembly/test and intangible assets; the timing and execution of the manufacturing ramp and associated costs; and capacity utilization;. Expenses, particularly certain marketing and compensation expenses, as well as restructuring and asset impairment charges, vary depending on the level of demand for Intel's products and the level of revenue and profits. The majority of our non-marketable equity investment portfolio balance is concentrated in companies in the flash memory market segment, and declines in this market segment or changes in management's plans with respect to our investments in this market segment could result in significant impairment charges, impacting restructuring charges as well as gains/losses on equity investments and interest and other. Intel's results could be impacted by adverse economic, social, political and physical/infrastructure conditions in countries where Intel, its customers or its suppliers operate, including military conflict and other security risks, natural disasters, infrastructure disruptions, health concerns and fluctuations in currency exchange rates. Intel's results could be affected by the timing of closing of acquisitions and divestitures. Intel's results could be affected by adverse effects associated with product defects and errata (deviations from published specifications), and by litigation or regulatory matters involving intellectual property, stockholder, consumer, antitrust and other issues, such as the litigation and regulatory matters described in Intel's SEC reports. An unfavorable ruling could include monetary damages or an injunction prohibiting us from manufacturing or selling one or more products, precluding particular business practices, impacting our ability to design our products, or requiring other remedies such as compulsory licensing of intellectual property. A detailed discussion of these and other risk factors that could affect Intel's results is included in Intel's SEC filings, including the report on Form 10-Q.

Rev. 1/14/10