Dense Computing with Transmeta's Crusoe™

Dean Gaudet
dgaudet@transmeta.com

TRANSMETA CORPORATION

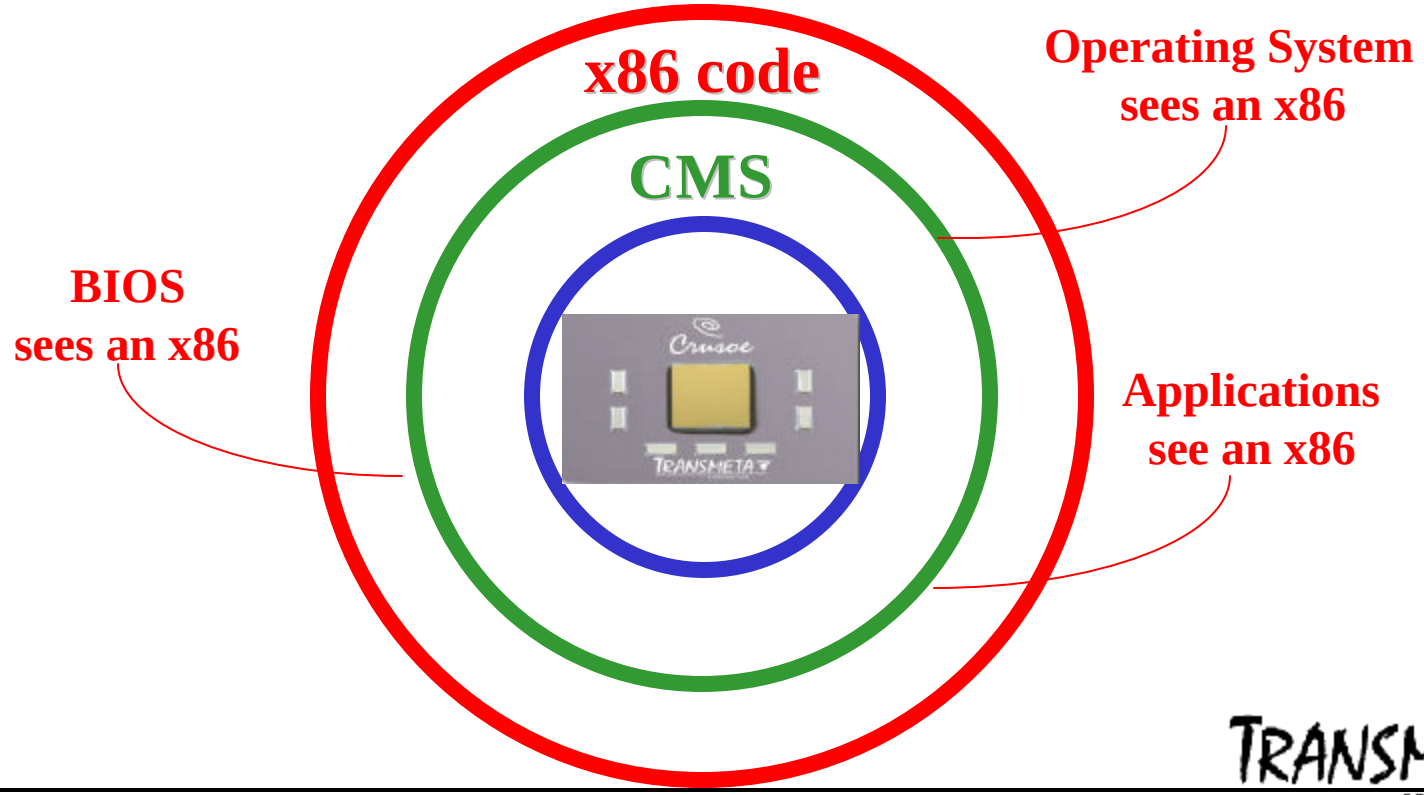
Transmeta Corp.

- http://www.transmeta.com/
- Located in Santa Clara, CA
- Transmeta develops and sells software-based microprocessors, and supporting technologies that enable low power, high performance x86 computing.
- Transmeta's family of Crusoe microprocessors is targeted at notebook, Internet appliance, and ultra-dense servers.
- Transmeta is publicly traded on NASDAQ, ticker symbol TMTA.

TRANSMETA CORPORATION

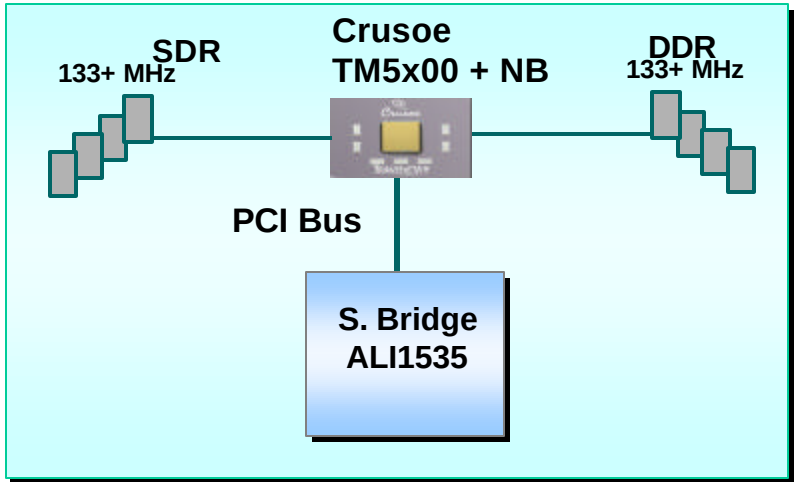
Crusoe overview

- Our product = Processor silicon + Code Morphing Software (CMS)
- x86 compatible including MMX, SMM, etc.
- CMS dynamically translates x86 code to native Very Long Instruction Word (VLIW) instructions



TRANSMETA CORPORATION

TM5x00 architectural overview



- VLIW architecture executes up to 4 operations per cycle
- Integrated northbridge
  - Two memory controllers (SDR and DDR)
- Separate memory spaces
  - x86
  - private memory for CMS (invisible to x86)
- 64KB L1 instruction cache, 64KB L1 data cache
- 256KB-512KB L2 unified cache
- 0.18µ process -> 0.13µ process

X86 Performance Challenges

- X86 performance is complicated by:
  - Complex instruction encoding
  - Too few registers
  - Branches (basic blocks too short)
  - Memory aliasing
  - Legacy 16-bit support
  - Self modifying code
- Transmeta's VLIW architecture provides many features for addressing these challenges

Key CPU features

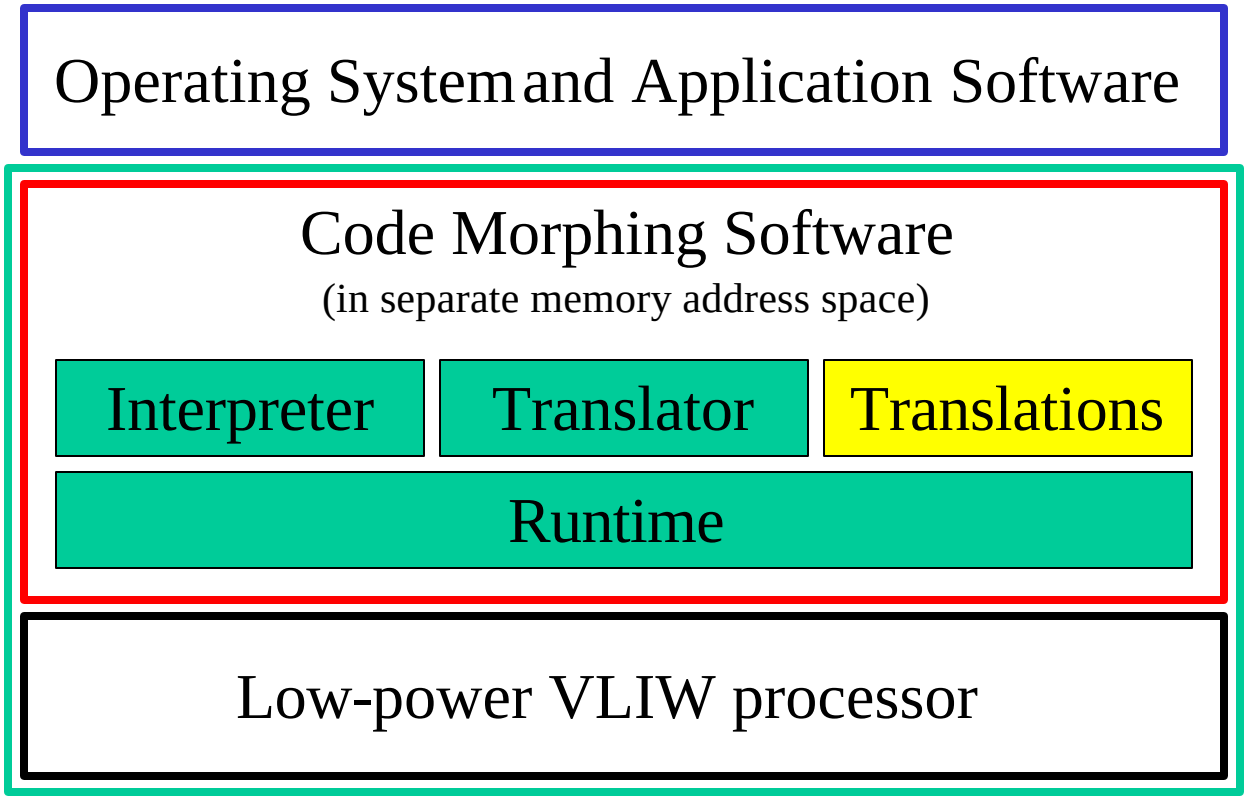
- Some x86 functionality provided by the VLIW core
  - EFLAGS, many operations, TLB entries similar to x86
- Larger number of registers (64 Int., 32 FP/MMX)
- Support for speculative execution
  - Commit/Rollback mechanisms
  - Most integer registers are shadowed
  - Stores are held in a buffer
  - Allows speculation while guaranteeing precise x86 instruction execution semantics

Additional CPU innovations

- Alias hardware (memory aliasing)
  - Register-poor x86 code requires lots of memory accesses
  - Special load and store instructions
  - They “protect” the physical address loaded into register
  - Another load or store to the same address traps
  - Benefits:
    - Can reorder memory operations more freely
    - Can eliminate redundant memory operations
      ➔ fewer memory accesses than typical x86

- Translation protection hardware (self-modifying code)
  - Self-modifying code could lead to stale translations.
  - Extension to the TLB to keep the common case fast

Code Morphing Software (CMS)



CMS: Interpreter

- Interpret one x86 instruction at a time
- Low overhead, but slow
- Why?
  - Provides precise exceptions
  - Assists the translator
    - Collects branch probabilities
    - Collects information about type of memory accessed
  - “Filters” infrequently-executed code from being translated

CMS: Translator

- Translates frequently executed x86 code into native VLIW instructions
  - Translations can contain multiple x86 basic blocks
- Translator also optimizes and schedules the generate efficient code
  - Classical compiler optimizations like CSE, dead code elimination, etc.
  - Applies transformations to increase parallelism
- Caches the translations for re-use
- Higher overhead, but much faster than interpreter

Goal: Amortize the cost of translation

CMS: Translations

- Optimize for the common case!
- Translations aggressively use CPU features
  - Rollback/Commit mechanism allows speculation
    - Schedule aggressively for the common path
    - Control and data speculation
    - If a fault occurs, rollback and interpret
  - Alias hardware allows us to reorder memory ops
    - Translator uses static memory disambiguation
    - Alias hardware assists with the remaining cases
    - Exposes more parallelism by breaking many data dependences
    - Allows us to exploit large register set
  - T-bit HW protects us from executing stale translations

CMS: Runtime

- Manages a cache of translations
  - Garbage collection/compaction optimizes RAM cache
  - Quick look-up schemes
  - Speeds up inter-translation flow via “chaining”
- Includes out-of-line handlers for complex ops
  - Single, shared copy of big ops helps the I-cache
  - ex. Segment descriptor manipulation, IRET, etc.
- Handles interrupts, PCI, exceptions, LongRun

LongRun

- Goal: Adapt processor's power usage to the needs of the workload
- Power = $C \times V^2 \times F$
  - CPU frequency roughly linear to operating voltage
  - By running CPU at lower frequency and voltage, a cubic power reduction can be achieved
  - CMS monitors "idleness" and dynamically adapts both frequency and voltage to meet the application demands

Advantages of our approach

- Allows infrequently-used logic to be in software
  - Trades expensive logic for cheap memory
  - Smaller die size
  - Fewer logic transistors means **lower power** chip
- Easier to build and debug
  - Easily work around silicon bugs
  - Great tools for performance and correctness testing
- Combined HW and SW approach enables improvements in 4 axes!
  - process technology
  - microarchitecture and circuits
  - architecture
  - software

  Architecture and software axes are unique to Transmeta!

Processor Comparison

Processor Families	Speed	Silicon Characteristics			TDP (Max)
		Process	Die Area	# of Gates	
PIII-M (Tualatin)	**1.13 GHz**	**0.13u**	**80.4 mm2**	**44 Million**	**21.8 W**
	866 MHz	**0.13u**	**80.4 mm2**	**44 Million**	**19.5 W**
PIV	**2 GHz**	**0.18u**	**217 mm2**	**42 Million**	**75.3 W**
	1.5 GHz	**0.18u**	**217 mm2**	**42 Million**	**57.9 W**
IA-64 Itanium	**800 MHz**	**0.18u**	**440 mm2**	**320 Million**	**130 W**
UltraSparc III	**750 MHz**	**0.18u**	**246 mm2**	**29 Million**	**70 W**
PowerPC	**667 MHz**	**0.18u**	**106 mm2**	**33 Million**	**19.0 W**
	533 MHz	**0.18u**	**106 mm2**	**33 Million**	**15.2 W**
AMD Athlon	**1.4 GHz**	**0.18u**	**120 mm2**	**37 Million**	**72 W**
	1.0 GHz	**0.18u**	**120 mm2**	**37 Million**	**54 W**
TMTA Crusoe	**1.0 GHz**	**0.13u**	**55 mm2**	**28 Million**	**7.5 W**
	800 MHz	**0.13u**	**55 mm2**	**28 Million**	**6 W**

Crusoe and Clusters

- TM5800 @ 800MHz has a Thermal Design Power (TDP) of 6W, and runs SPECint2000 with less than 4W.
- Note this is at full performance – Intel LV/ULV parts have quite different power requirements on AC power, their high-performance settings.
- AMD, Intel high-end and next-generation CPUs have TDPs of 50W and (much) higher.
- The low Crusoe TDP allows computer manufacturers to achieve a much greater density of CPUs within any physical space – far less heat to remove per CPU.

IR Photographs of CPU

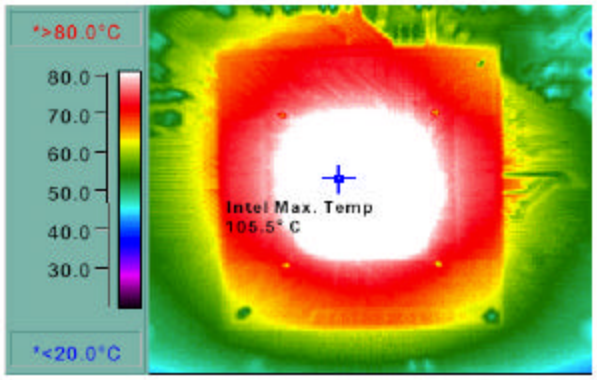


Figure 3. A Pentium III processor plays a DVD at 105° C (221° F).

Intel Pentium III CPU decoding a DVD

221 degrees

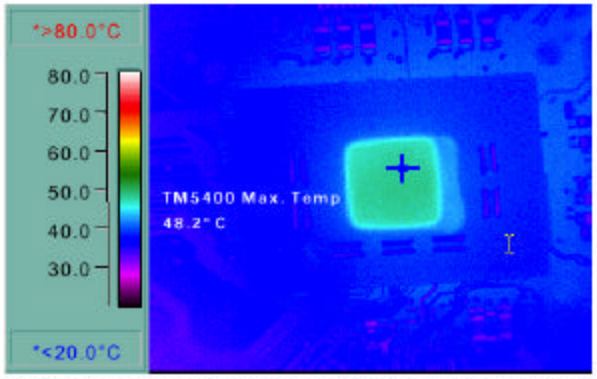


Figure 4. A Crusoe processor model TM5400 plays a DVD at 48° C (118° F).

TRANSMETA™

TM 5400 CPU decoding a DVD

118 degrees

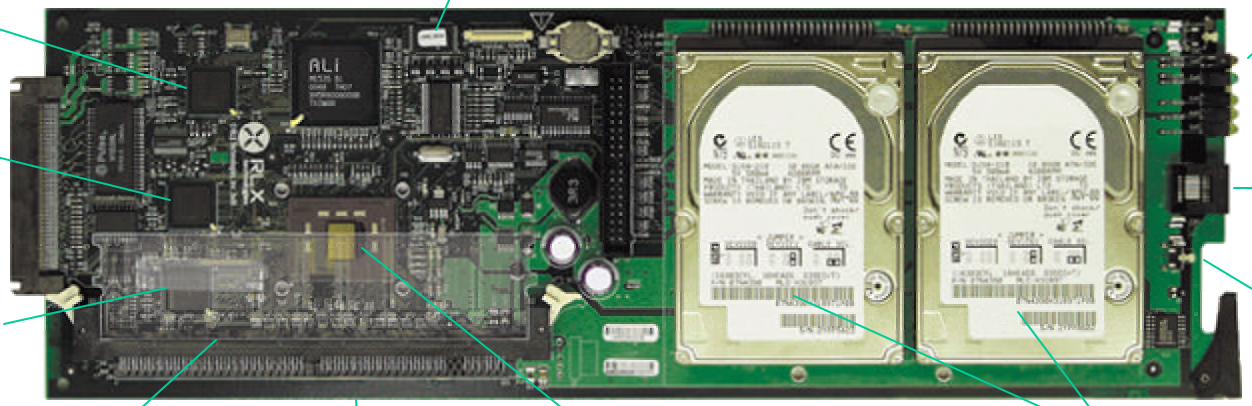
TRANSMETA CORPORATION

RLX System 324™ Overview

- RLXtechnologies located in Houston, TX
- http://www.rlxtechnologies.com/
- RLX ServerBlade™
  - **Transmeta™ Crusoe™ CPU, TM5600**
  - **Single DIMM, up to 512MB**
  - **0, 1 or 2 hard drives, up to 30GB each**
  - **3 NICs for public, private and management nets**
- RLX Chassis
  - **3U, houses 24 ServerBlades**
  - **3 network cards, 1 management hub**
  - **2 power supplies for redundancy**
  - **Front door fans and status LEDs**
- Management Software
  - **RLX Control Tower™ for web-based provisioning and monitoring**
  - **RLX Remote Console for console access via network**

TRANSMETA CORPORATION

RLX ServerBlade™



- **Public NIC**
- **33 MHz PCI**

Private NIC
33 MHz PCI

Management NIC
33 MHz PCI

CMS 1 MB

Status LEDs

Serial RJ-45
debug port

Reset Switch

128MB, 256MB, 512MB
DIMM SDRAM
PC-133

512KB
Flash ROM

Transmeta™
TM5600 633 MHz

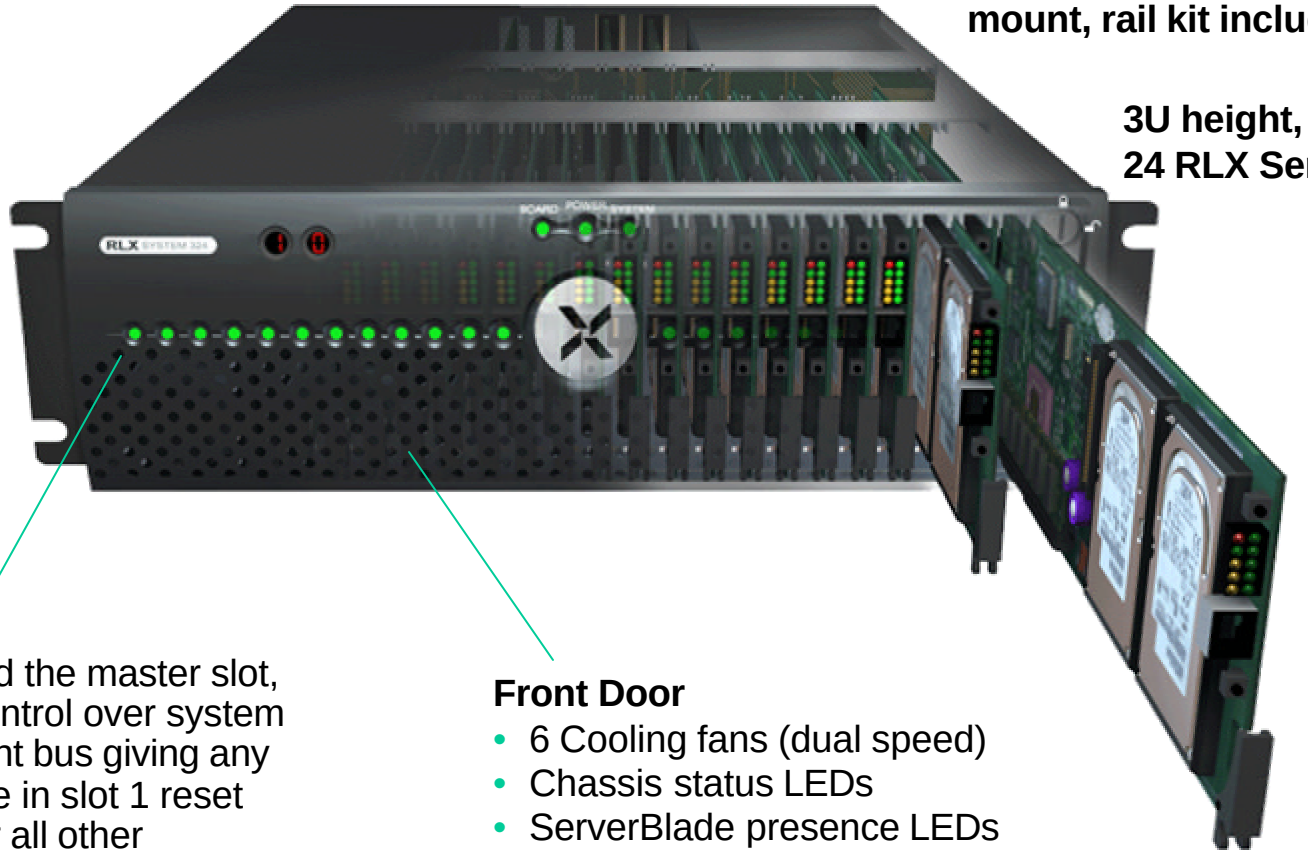
ATA 66
0, 1 or 2 - 2.5” HDD
10 or 30 GB each

128KB L1 cache, 512KB L2 cache
LongRun, Southbridge, X86 compatible

RLX System 324™ Chassis

Standard 19” 4-post rack mount, rail kit included

3U height, holds up to 24 RLX ServerBlades



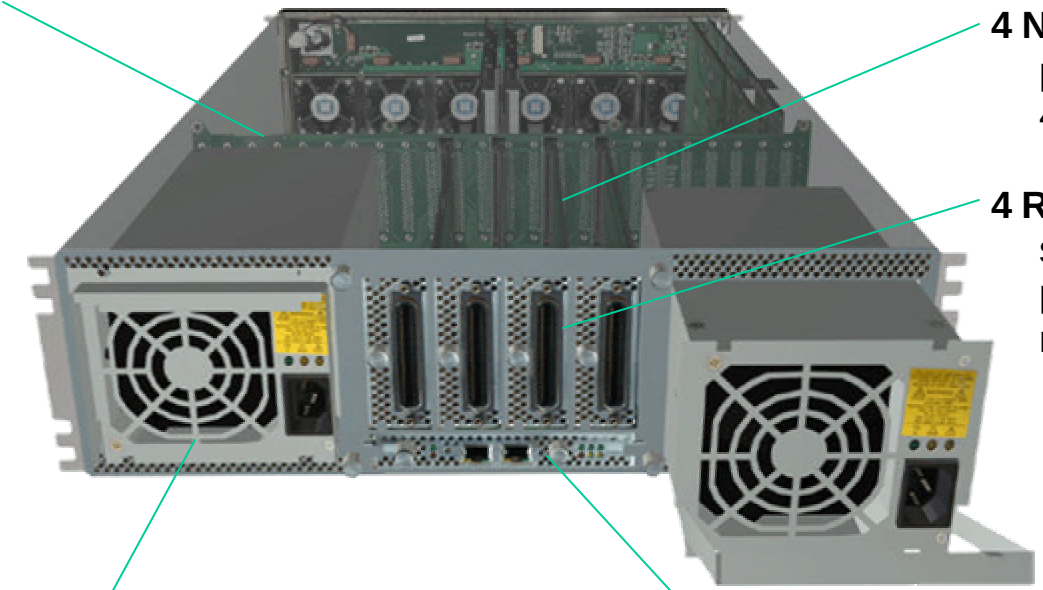
Slot 1 called the master slot, has hard control over system management bus giving any ServerBlade in slot 1 reset control over all other ServerBlades in chassis

Front Door

- 6 Cooling fans (dual speed)
- Chassis status LEDs
- ServerBlade presence LEDs
- Numeric ID LEDs

RLX System 324™ Chassis

Passive Midplane
- Passes signal and power
- Consolidates network cabling
- Enables front door function and reporting



4 Network Connect Cards
provide pass-through for 48 (2x24) NICs

4 RJ-21 Connectors
switched or routed to public or private networks

2 Power Supplies
- Load balanced
- Hot pluggable
- Provide 450 watts each

1 Management Hub Card
supports 24 NICs,
1 RJ-45 connector out to Management Network

Crusoe vs. P-III relative performance

- Relative performance of similarly configured Crusoe and P-III systems
- Results scaled by MHz
- Note that no attempt has been made to factor in power costs!

Benchmark	Relative Performance	Notes
SPECweb99	100%	2x 100baseT, 512MB memory, 4200rpm 2.5” IDE, Linux/TUX
SPECint2000	71%	Intel 5.0 compiler on Win2K
BLAST	85%	
RSA1024	80%	Openssl-0.9.6 “openssl speed rsa1024”
Misc crypto.	80 to 110%	Various other “openssl speed” results
FinSim	~70%	Based on a verilog sample from Crusoe design

Power Measurements

- Direct power comparisons of the CPUs Crusoe vs. P-III are difficult/misleading because:
  - Crusoe has integrated northbridge, P-III has separate northbridge.
  - No access to P-III "development systems" – no power measurement pins.
  - CPU power is only one component of overall system power.
- Direct power comparisons of overall system power requires us to choose a P-III system vendor to compare with.
- P-III systems can vary from 70W to more than 100W for full system power under a load such as SPECweb99.
- RLX ServerBlade consumes only 15W at peak load, idles at 7W.

More Web Pages Served

- Compare Pages/Sec/Rack at 5,000 max watts per 42U rack*

	RLX System 324	Cobalt 4i	Cobalt 4R XTR	Compaq DL320
Number of Servers	254	42	42	42
Watts per Server	15.7	33.9	72.0	76.4
Pages/Sec per Server ZD WebBench (peak request load)	1466 **TM 5600 633MHz, 512MB SDRAM,IDE 10GB HD, Red Hat 6.2, 2.4.2, Zeus**	208 **AMD k6-2 450MHz, 128MB SDRAM, IDE 10GB HD, Cobalt SW**	916 **PIII, 933 MHz, 512MB, SDRAM, IDE 3x30 GB, RAID, Cobalt SW**	1706 **PIII 800MHz 512MB SDRAM, 18GB Ultra SCSI3HD, Red Hat 6.2, 2.4.2, Zeus**
Pages/Sec per Rack	372,364 Baseline	8,736 2.3%	38,472 10.3%	71,652 19.2%
RLX Advantage		43x	10x	5x

*Calculations based on 80% of max watts = 4000 watts

Conclusion

- Transmeta's unique technology has enabled a new class of low-power, high-performance computing.
- The industry is now awakening to the issue of power consumption.
- RLXtechnologies has packaged the Crusoe processor in a very dense, highly scalable platform suitable for cluster applications.
- RLXtechnologies System 324, fully loaded costs less than US$30,000, and consumes only 400W at peak power with 24 nodes.
- In addition to the power/cooling benefits within a machine room, the System 324 solution can easily fit "under your desk", enabling researchers to build clusters without special IT/MIS assistance.