SPARC M12/M10 サーバのアーキテクチャー

ホワイトペーパー

マニュアル番号：C120-E690-14
2022 年 06 月

Copyright © 2007, 2022, 富士通株式会社 All rights reserved.

本書には、オラクル社および/またはその関連会社により提供および修正された技術情報が含まれています。

オラクル社および/またはその関連会社、および富士通株式会社は、それぞれ本書に記述されている製品および技術に関する知的所有権を所有または管理しています。これらの製品、技術、および本書は、著作権法、特許権などの知的所有権に関する法律および国際条約により保護されています。

本書およびそれに付属する製品および技術は、その使用、複製、頒布および逆コンパイルを制限するライセンスのもとにおいて頒布されます。オラクル社および/またはその関連会社、および富士通株式会社およびそのライセンサーの書面による事前の許可なく、このような製品または技術および本書のいかなる部分も、いかなる方法によっても複製することが禁じられます。本書の提供は、明示的であるか黙示的であるかを問わず、本製品またはそれに付随する技術に関するいかなる権利またはライセンスを付与するものでもありません。本書は、オラクル社および富士通株式会社の一部、あるいはそのいずれかの関連会社のいかなる種類の義務を含むものでも示すものでもありません。

本書および本書に記述されている製品および技術には、ソフトウェアおよびフォント技術を含む第三者の知的財産が含まれている場合があります。これらの知的財産は、著作権法により保護されているか、または提供者からオラクル社および/またはその関連会社、および富士通株式会社へライセンスが付与されているか、あるいはその両方です。

GPL または LGPL が適用されたソースコードの複製は、GPL または LGPL の規約に従い、該当する場合に、お客様からのお申し込みに応じて入手可能です。オラクル社および/またはその関連会社、および富士通株式会社にお問い合わせください。この配布には、第三者が開発した構成要素が含まれている可能性があります。本製品の一部は、カリフォルニア大学からライセンスされている Berkeley BSD システムに由来しています。

UNIX は The Open Group の登録商標です。

Oracle と Java は Oracle Corporation およびその関連企業の登録商標です。

富士通および富士通のロゴマークは、富士通株式会社の登録商標です。

SPARC Enterprise、SPARC64、SPARC64 ロゴ、およびすべての SPARC 商標は、米国 SPARC International, Inc.のライセンスを受けて使用している、同社の米国およびその他の国における商標または登録商標です。

その他の名称は、それぞれの所有者の商標または登録商標です。

このソフトウェアまたは関連ドキュメントを、米国政府機関もしくは米国政府機関に代わってこのソフトウェアまたは関連ドキュメントをライセンスされた者に提供する場合は、次の通知が適用されます。

U.S. GOVERNMENT END USERS: Oracle programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, delivered to U.S. Government end users are "commercial computer software" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, use, duplication, disclosure, modification, and adaptation of the programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, shall be subject to license terms and license restrictions applicable to the programs. No other rights are granted to the U.S. Government.

免責条項: 本書または本書に記述されている製品や技術に関してオラクル社、富士通株式会社および/またはそのいずれかの関連会社が行う保証は、製品または技術の提供に適用されるライセンス契約で明示的に規定されている保証に限ります。

このような契約で明示的に規定された保証を除き、オラクル社、富士通株式会社および/またはそのいずれかの関連会社は、製品、技術、または本書に関して、明示、黙示を問わず、いかなる種類の保証も行いません。これらの製品、技術、または本書は、現状のまま提供され、商品性、特定目的への適合性または第三者の権利の非侵害の黙示の保証を含みそれに限定されない、明示的であるか黙示的であるかを問わない、なんらの保証も、かかる免責が法的に無効とされた場合を除き、行われないものとします。このような契約で明示的に規定されていないかぎり、オラクル社、富士通株式会社および/またはそのいずれかの関連会社は、いかなる法理論のもとの第三者に対しても、その収益の損失、有用性またはデータに関する損失、あるいは業務の中断について、あるいは間接的損害、特別損害、付随的損害、または結果的損害について、そのような損害の可能性が示唆されていた場合であっても、適用される法律が許容する範囲内で、いかなる責任も負いません。

本書は、「現状のまま」提供され、商品性、特定目的への適合性または第三者の権利の非侵害の黙示の保証を含みそれに限定されない、明示的であるか黙示的であるかを問わない、なんらの保証も、かかる免責が法的に無効とされた場合を除き、行われないものとします。

目次

はじめに……1
高性能と高い信頼性を兼ね備えたSPARC M12/M10……1
1. SPARC M12……3
製品ラインナップ……3
SPARC M12-1……5
SPARC M12-2……6
SPARC M12-2S……7
2. SPARC M10……8
製品ラインナップ……8
SPARC M10-1……10
SPARC M10-4……11
SPARC M10-4S……12
3. SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサ……14
4. システムアーキテクチャー……16
システムインターコネクト……16
1. システムバス……16
2. SPARC M12-1 インターコネクトアーキテクチャー……19
3. SPARC M12-2 インターコネクトアーキテクチャー……20
4. SPARC M12-2S インターコネクトアーキテクチャー……21
5. SPARC M10-1 インターコネクトアーキテクチャー……24
6. SPARC M10-4 インターコネクトアーキテクチャー……25
7. SPARC M10-4S インターコネクトアーキテクチャー……26
8. システムインターコネクトの信頼性機能……29
メモリ……30
1. メモリミラーリング……31
システムクロック……32
I/O バス……32
1. I/O サブシステムのアーキテクチャー……32
2. SPARC M12-1 の I/O サブシステム……33
3. SPARC M12-2 の I/O サブシステム……34
4. SPARC M12-2S の I/O サブシステム……35
5. SPARC M10-1 の I/O サブシステム……36
6. SPARC M10-4 の I/O サブシステム……37
7. SPARC M10-4S の I/O サブシステム……38
8. 内蔵ドライブと周辺装置……39

9. PCI ボックス……39
冷却……42
高効率電源……44

5. システム管理……45

信頼性、可用性、保守性……45
1. コンポーネントの冗長化と活性交換……46
2. パーティショニング機能……47
3. 高度な信頼性機能……48
4. エラー検出、診断、復旧……48
システム管理……49
1. システム監視機構……49
2. 冗長 XSCF……50
3. XSCF Control Package（XCP）……50
4. ロールベースのシステム管理……51
5. Enhanced Support Facility……51
6. 高可用性を実現するPRIMECLUSTER……51
7. FUJITSU Software Systemwalker Centric Manager……51
8. Oracle Enterprise Manager Ops Center 12c……52
エコ……52

6. Oracle Solaris 11 オペレーティングシステム……54

Oracle Solaris ZFS……54
Oracle VM Server for SPARC……58
Oracle Solaris ゾーン……59
Image Packaging System……60
Boot Environment……61

7. SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサの技術拡張……62

マイクロアーキテクチャー……62
1. チップ構成……62
2. コアのマイクロアーキテクチャー……63
3. チップ間インターフェース……69
拡張命令セットアーキテクチャー……69
1. HPC-ACE……69
2. ソフトウェア・オン・チップ（SWoC：SoftWare on Chip）……71
RAS 機能（信頼性、可用性、保守性）……72
1. エラーマーキング……74
2. 内蔵 RAM の信頼性および可用性機能……74
3. 内蔵レジスタと演算器の信頼性機能……75
4. 同期更新方式と命令リトライ……76
5. 保守性の向上……77

8. 結論……78

はじめに

高性能と高い信頼性を兼ね備えたSPARC M12/M10

ビジネス環境の急速な変化に対応するITインフラには、スピード、安定稼動、効率化と最適化が求められています。SPARC M12/M10は、高性能プロセッサ、最先端の半導体技術に加えて、富士通がメインフレームおよびSPARC Enterprise Mシリーズから継承する高い信頼性と高水準の技術力を備えた、現代のITインフラに必要な総合的プラットフォームを提供します。さらに、システムの信頼性、柔軟性、拡張性を強化し、クラウドコンピューティング環境のニーズに対応可能な環境を実現しています。

システムパフォーマンスの向上を実現するSPARC64™ XII / SPARC64™ X+ / SPARC64™ X

富士通は、SPARC64 X+（テンプラス）プロセッサおよびSPARC64 X（テン）プロセッサの高い性能と信頼性を継承する、新しいSPARC64 XII（トゥエルブ）プロセッサを開発しました。
SPARC M12は、高いコア性能を実現するプロセッサSPARC64 XIIを搭載し、お客様の基幹業務システムのデータベース／DWH処理を高速化します。
SPARC M10は、SPARC64 X+およびSPARC64 Xの2種類のプロセッサから、お客様の業務負荷の特性に応じて、適切なプロセッサを選択することができます。

- **システム・オン・チップ**
  UNIXサーバ用に開発されたSPARC64 XII/SPARC64 X+/SPARC64 Xは、周辺LSIをプロセッサに集約しています。

- **ソフトウェア・オン・チップ**
  専用の命令を追加することにより、従来ソフトウェアで処理していた様々な処理機能をプロセッサのハードウェアに構築しました。ビッグデータの分析・解析処理に役立つ同時演算実行数を拡大しています。

- **冷却技術**
  当社独自の冷却技術が、プロセッサの最大性能を引き出します（「Liquid Loop Cooling（LLC）」、および新たに導入された「Vapor and Liquid Loop Cooling（VLLC）」）。
  これらの技術により、処理の高速化（スループットの向上）とともに、システムパフォーマンスの向上を実現しています。

基幹システムやリアルタイム情報分析に最適な SPARC M12/M10

プロセッサと SPARC M12/M10 間の相互接続には、最先端の高速インターコネクト技術を採用しています。なかでも、SPARC M12-2S と SPARC M10-4S は、ビルディングブロック（BB）方式（*1）による大幅な性能・機能拡張が可能です。最大 16 サーバ（*2）を 1 つのシステムとして構成でき、SPARC M12-2S は最大 32 CPU/3072 スレッド、SPARC M10-4S は最大 64 CPU/2048 スレッドまで拡張可能な、きわめて柔軟でスケーラブルなシステムで、高い性能を発揮します。

そのため、小規模構成で導入して初期投資を軽減し、データベースシステムに情報分析システムを統合してリアルタイムな分析を行うなど、需要の増加に合わせてシステムを拡張することができます。
これらの優れた機能を備えた SPARC M12/M10 は、クラウドコンピューティング時代のデータセンターに最適なサーバといえます。

*1 ビルディングブロック方式とは、複数のCPU、メモリおよびI/O拡張スロットを含む共通のサーバ（ブロック）をまるでブロックを積み重ねるように接続する方式です。

*2 2022年4月以降に販売開始したSPARC M12-2Sは、最大4サーバまでとなります。

設置性に優れた SPARC M12/M10

SPARC M12/M10 は、優れた設置性も実現しました。周辺 LSI を SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサ内に集約し、SPARC M12/M10 は小型化を実現した高密度サーバとなりました。

SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサを搭載した SPARC M12/M10 は、多くのハードウェア技術およびソフトウェア技術を利用して、成長を続ける IT インフラストラクチャに最適なソリューションをお客様に提供します。

1.SPARC M12

製品ラインナップ

SPARC M12はデータベースやアナリティクス基盤をはじめ、幅広いニーズに最適な製品です。

- コンパクトな 1U サイズに高性能、高信頼を凝縮したエントリーサーバ「SPARC M12-1」
- データベースとアナリティクスの統合など、システム統合に最適なミッドレンジサーバ「SPARC M12-2」
- ミッドレンジからハイエンドまでの幅広いニーズに対応可能な「SPARC M12-2S」

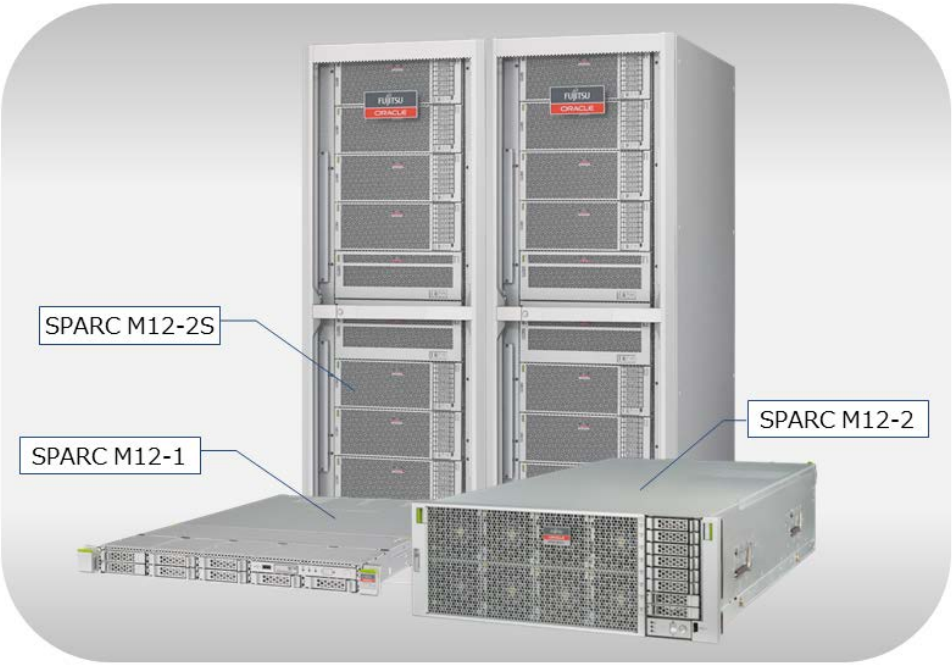

図 1-1. SPARC M12 の製品ラインナップ

表 1-1. SPARC M12 の仕様

		SPARC M12-1	SPARC M12-2	SPARC M12-2S（1BB）	SPARC M12-2S（16BB）（*1）
SPARC64 XII プロセッサ	周波数	3.2 GHz	3.9 GHz	4.25 GHz	4.25 GHz
	3次キャッシュサイズ	16 MB	32 MB	32 MB	32 MB
	CPU数	1	1または2	1または2	最大32
	コア数	6	12（1プロセッサあたり）	12（1プロセッサあたり）	最大384
	スレッド数	48	96（1プロセッサあたり）	96（1プロセッサあたり）	最大3072
メモリ	最大容量	1 TB	3 TB（1プロセッサあたり24メモリスロットの場合） 2 TB（1プロセッサあたり16メモリスロットの場合）	3 TB（1プロセッサあたり24メモリスロットの場合） 2 TB（1プロセッサあたり16メモリスロットの場合）	48 TB（1プロセッサあたり24メモリスロットの場合） 32 TB（1プロセッサあたり16メモリスロットの場合）
	最大メモリ搭載数	16	48（1プロセッサあたり24メモリスロットの場合） 32（1プロセッサあたり16メモリスロットの場合）	48（1プロセッサあたり24メモリスロットの場合） 32（1プロセッサあたり16メモリスロットの場合）	768（1プロセッサあたり24メモリスロットの場合） 512（1プロセッサあたり16メモリスロットの場合）
内蔵ストレージ	インターフェース	SAS	SAS	SAS	SAS
	ドライブ数	8	8	8	128
PCIeスロット	インターフェース	PCI Express 3.0	PCI Express 3.0	PCI Express 3.0	PCI Express 3.0
	PCIeスロット数（基本）	3	11	8	128
	PCIeスロット数（最大）	33（*2） PCIボックス3台接続時	91（*2） PCIボックス8台接続時	88（*2） PCIボックス8台接続時	1,408（*2） PCIボックス128台接続時
I/Oインターフェース	10 Gbイーサネットポート	4	4	4	64
	SASポート	1	1	1	16
	USBポート	2	2	2	32
ラックユニット		1 U	4 U	4 U	40 U x 2（XB Box含む）
仮想化機能	物理パーティション数	1	1	1（1 BBあたり1パーティション）	最大16
	最大論理ドメイン数	48	192	192	256（物理パーティションあたり）

*1 2022年4月以降に販売開始したSPARC M12-2Sは最大4BB。仕様の各項目の最大値は、16BBの1/4となります。

*2 CPU最大数搭載時、かつ、すべての制御ドメイン、ルートドメイン、I/OドメインでOracle Solaris 11を動作させる場合

SPARC M12-1

SPARC M12-1は、1ラックユニット（1 U）の省スペースな筐体に、高性能、高信頼を凝縮したエントリーサーバです。

最大6コアの高性能CPUを基本搭載し、最大1 TBメモリ、最大8台の内蔵ストレージ（HDDとSSD）を搭載可能です。内蔵I/Oスロットは、PCI Expressスロットが3スロットあり、さらにPCIボックスを増設して、最大33スロットまで拡張が可能です。これによりミッドレンジクラスまでの拡張性に対応します。また、2台の電源と7台のファンユニットが内蔵された冗長構成となっています。

最小限のCPUリソースでスタートし、必要に応じてCPUコアアクティベーション（CPUコア使用権）を購入することで、段階的にCPUリソースを拡張することができます。また、Oracle Solarisゾーンと Oracle VM Server for SPARCにより、サーバ仮想化やシステム統合も実現可能です。SPARC M12-1は、初期投資を抑え、段階的な処理能力の増強が可能なため、エントリーサーバにおいてもTCOを最適化することができます。

SPARC M12-2

SPARC M12-2 は、データベース集約、データベースとアナリティクスの統合に適した高い性能と高い信頼性のあるミッドレンジサーバです。メインフレームクラスの信頼性は、ミッションクリティカル業務に適しています。

4 ラックユニット（4 U）に、最大 2 CPU（24 コア）、最大 3 TB（1 プロセッサあたり 24 メモリスロットの場合）メモリ、最大 8 台の内蔵ストレージ（HDD と SSD）を搭載可能です。また、内蔵 I/O スロットは、PCI Express スロットが 11 スロットあります。さらなる I/O 拡張に対応するため、PCI ボックスを増設して、最大 91 スロットまで PCI Express スロットを拡張できます。4 台の電源ユニットを 2 台 1 組として冗長的に電力を供給し、冗長化された当社独自の高効率冷却技術 Vapor and Liquid Loop Cooling（VLLC）と 8 台のファンユニットで冷却します。

最小限の CPU リソースでスタートし、必要に応じて CPU コア アクティベーション（CPU コア使用権）を購入することで、段階的に CPU リソースを拡張することができます。また、Oracle Solaris ゾーンと Oracle VM Server for SPARC により、システムの変動要求に合わせた柔軟な仮想化構成変更が可能です。CPU も 1 CPU から 2 CPU に増設できますので、処理能力の増強も可能です。

SPARC M12-2S

SPARC M12-2S は、世界最高レベルの拡張性と柔軟性を備えているため、ミッドレンジからハイエンドまでの幅広いニーズに対応することができます。基幹データベースを含む大規模統合にも対応可能なモデルです。

高性能、高信頼に加えて、SPARC M12-2S はビルディングブロック方式を採用しているため、柔軟な拡張性も提供します。また、拡張接続用ラックを導入すると、業務を継続しながら、スケールアップ構成や、並列分散処理に適したスケールアウト構成を自由に組み合わせることができます。拡張接続用ラック 2 架を導入した場合は、システム全体をシャットダウンすることなく、最大 16 までのビルディングブロック接続にスケールアップも可能です。

ビルディングブロック構成の SPARC M12-2S 間は当社独自の高速インターコネクトで接続され、最大 16BB 構成までのリニアなパフォーマンス向上を実現します。スタートは必要最小限の CPU、メモリ、PCI Express スロット、およびディスクベイでシステムを構成し、必要に応じて、SPARC M12-2S を追加することで、処理能力を増強することができます。また、CPU リソースは、CPU コア アクティベーション（CPU コア使用権）を購入することで段階的に拡張できるため、CPU チップを増設するよりも細かく、着実な方法で拡張することも可能です。

2. SPARC M10

製品ラインナップ

SPARC M10 は幅広い分野に適用可能な製品を取り揃えています。

- コンパクトな 1U サイズを実現したエントリーサーバ「SPARC M10-1」
- システム統合に最適なミッドレンジサーバ「SPARC M10-4」
- ミッドレンジからハイエンドまでの幅広いニーズに対応可能な「SPARC M10-4S」

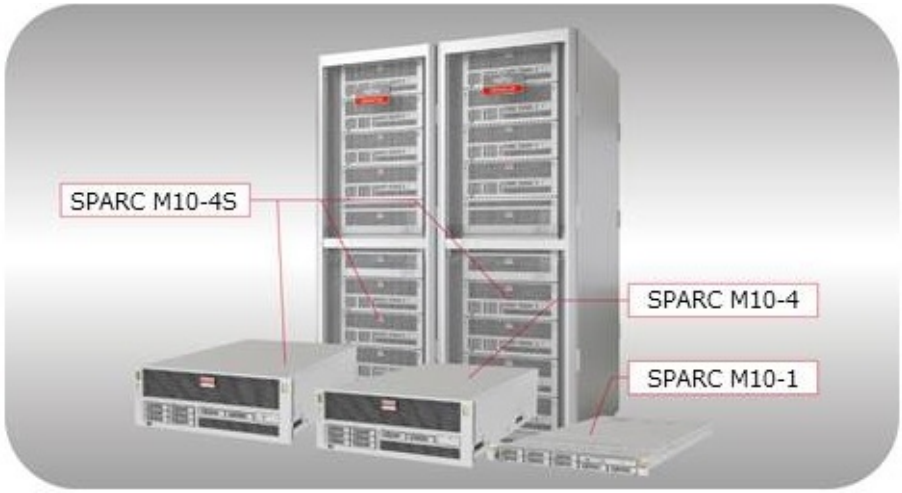

図 2-1. SPARC M10 の製品ラインナップ

表 2-1. SPARC M10 の仕様

		SPARC M10-1	SPARC M10-4	SPARC M10-4S（1BB）	SPARC M10-4S（16BB）
SPARC64 X+プロセッサ（*1）	周波数	2.8 GHz/3.2 GHz/3.7 GHz	3.4 GHz/3.7 GHz	3.7 GHz	3.7 GHz
	2次キャッシュサイズ	22 MB/22 MB/24 MB	24 MB	24 MB	24 MB
	CPU数	1	2または4	2または4	最大64
	コア数	16/16/8	16/8（1プロセッサあたり）	16（1プロセッサあたり）	最大1024
	スレッド数	32/32/16	32/16（1プロセッサあたり）	32（1プロセッサあたり）	最大2048
SPARC64 Xプロセッサ（*1）	周波数	2.8 GHz	2.8 GHz	3.0 GHz	3.0 GHz
	2次キャッシュサイズ	22 MB	24 MB	24 MB	24 MB
	CPU数	1	2または4	2または4	最大64
	コア数	16	16（1プロセッサあたり）	16（1プロセッサあたり）	最大1024
	スレッド数	32	32（1プロセッサあたり）	32（1プロセッサあたり）	最大2048
メモリ	最大容量	1 TB	4 TB	4 TB	64 TB
	最大メモリ搭載数	16	64	64	1024
内蔵ストレージ	インターフェース	SAS	SAS	SAS	SAS
	ドライブ数	8	8	8	128
PCIeスロット	インターフェース	PCI Express 3.0	PCI Express 3.0	PCI Express 3.0	PCI Express 3.0
	PCIeスロット数（基本）	3	11	8	128
	PCIeスロット数（最大）	23 PCIボックス2台接続時	71（*2） PCIボックス6台接続時	58（*2） PCIボックス5台接続時	928（*2） PCIボックス80台接続時
I/Oインターフェース	1Gbイーサネットポート	4	4	4	64
	SASポート	1	1	1	16
	USBポート	2	2	2	32
ラックユニット		1 U	4 U	4 U	40 U x 2 （XB Box含む）
仮想化機能	物理パーティション数	1	1	1（1BBあたり）	最大16
	最大論理ドメイン数	32	128	128	256（物理パーティションあたり）

*1 SPARC64 X+プロセッサ、SPARC64 Xプロセッサのどちらかを搭載できます。SPARC64 Xプロセッサを搭載したSPARC M10-4SとSPARC64 X+を搭載したSPARC M10-4Sを、ビルディングブロック構成で接続できます。

*2 CPU最大数搭載時

SPARC M10-1

SPARC M10-1は、1ラックユニット（1 U）の省スペースな筐体で、データセンターの統合や仮想化に対応可能な高い性能と高い信頼性に優れたエントリーサーバです。

最大16コアの高性能CPUを1台基本搭載し、最大1 TBメモリ、最大8台の内蔵ストレージ（HDDとSSD）を搭載可能です。また、内蔵I/Oスロットは、PCI Expressスロットが3スロットあります。さらなるI/O拡張に対応するため、PCIボックスを増設して、最大23スロットまでPCI Expressスロットを拡張できます。また、2台の電源と7台のファンユニットが内蔵されており、冗長的に電源供給と冷却が行われます。

最小限のCPUリソースでスタートし、必要に応じてCPUコアアクティベーション（CPUコア使用権）を購入することで、段階的にCPUリソースを拡張することができます。また、Oracle SolarisゾーンとOracle VM Server for SPARCにより、サーバ仮想化やシステム統合も実現可能です。SPARC M10-1は、初期投資を抑え、段階的な処理能力の増強が可能なため、エントリーサーバにおいてもTCOを最適化することができます。

SPARC M10-4

SPARC M10-4 は、SPARC M10-1 よりも多くのプロセッサ、メモリおよび I/O 容量を必要とするデータセンターの統合および仮想化に適した高い性能と高い信頼性を持つミッドレンジサーバです。お客様のミッションクリティカル業務のサポートに必要な柔軟性、拡張性、信頼性を備えています。

4 ラックユニット（4 U）に、最大 4 CPU（64 コア）、最大 4 TB メモリ、最大 8 台の内蔵ストレージ（HDD と SSD）を搭載可能です。また、内蔵 I/O スロットは 11 スロットの PCI Express を採用しています。さらなる I/O 拡張に対応するため、PCI ボックスを増設して、最大 71 スロットまで PCI Express スロットを拡張できます。2 台の電源ユニットで冗長的に電力を供給し、冷却技術 Liquid Loop Cooling（LLC）と 5 台のファンユニットで冷却します。

最小限の CPU リソースでスタートし、必要に応じて CPU コア アクティベーション（CPU コア使用権）を購入することで、段階的に CPU リソースを拡張することができます。また、Oracle Solaris ゾーンと Oracle VM Server for SPARC により、サーバ仮想化やシステム統合も実現可能です。もちろん、SPARC M10-1 同様、初期投資の削減と段階的な処理能力の増強も可能です。CPU も 2 CPU から最大 4 CPU まで増設できるので、メモリと I/O を最大限に活用できます。

SPARC M10-4S

SPARC M10-4S は、世界最高レベルの拡張性と柔軟性を備えているため、ミッドレンジからハイエンドまでの幅広いコンピューティングニーズに対応することができます。そして、現代の大規模システム、あるいは今後の急成長が見込まれるクラウドコンピューティング、ビッグデータ処理などのインフラにも最適なモデルです。

SPARC 10 モデルに共通する高性能、高信頼に加えて、SPARC M10-4S はビルディングブロック方式を採用しているため、柔軟な拡張性も提供します。また拡張接続用ラックを導入すると、稼動中のパーティションを停止することなく、スケールアップ構成や、並列分散処理に適したスケールアウト構成を自由に組み合わせることができます。拡張接続用ラック 2架を導入した場合は、システム全体をシャットダウンすることなく、最大 16 までのビルディングブロック接続にスケールアップも可能です。

ビルディングブロック構成の SPARC M10-4S 間は当社独自のインターコネクト技術で接続され、最大 16BB 構成までのリニアなパフォーマンス向上を実現します。スタートは必要最小限の CPU、メモリ、PCI Express スロット、およびディスクベイでシステムを構成し、必要に応じて、SPARC M10-4S の追加により処理能力を増強することができます。また、CPU リソースは CPU コア アクティベーション（CPU コア使用権）を購入することで段階的に CPU リソースを拡張できるため、CPU チップを増設するよりも

細かく、着実な方法で拡張することも可能です。

さらに、SPARC64 X プロセッサを搭載した SPARC M10-4S は、SPARC64 X プロセッサを搭載した SPARC M10-4S はもちろん、SPARC64 X+プロセッサを搭載した SPARC M10-4S にもビルディングブロック接続を行うことができます。

SPARC64 X プロセッサを搭載した SPARC M10-4S の代わりに、SPARC64 X+プロセッサを搭載した SPARC M10-4S を使用することにより、すでに使用している SPARC64 X 搭載の SPARC M10-4S を無駄にすることなく、さらに高性能なシステムにアップグレードすることも可能です。

3. SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサ

UNIX サーバ向けに富士通が開発した SPARC64 XII は、高性能、高信頼を兼ね備えた最新プロセッサです。高いコア性能、高周波数、マルチコア/マルチスレッド機能、高いメモリスループット性能を備えているのが特長です。さらに、単一システムに CPU チップを SPARC64 XII は最大 32 個まで、SPARC64 X+/SPARC64 X は最大 64 個まで結合できるため、きわめて高い性能と拡張性を提供することができます。そして、プロセッサおよびその搭載システムの随所に、富士通がメインフレーム開発で培ってきた信頼性技術が活かされています。

SPARC64 XII/SPARC64 X+/SPARC64 X は、従来の SPARC64 プロセッサから信頼性の高い技術を継承しつつ、様々な機能拡張を行っています。マルチコア化、周辺 LSI 機能のプロセッサ内への集約といったマイクロアーキテクチャーの強化に加え、レジスタ拡張、SIMD 命令を活用した暗号処理やデータベース処理のパフォーマンス向上を行う大幅な命令拡張を行っています。

プロセッサ内に周辺 LSI の機能を集約することで、処理の高速化、消費電力あたりのパフォーマンスの向上を実現しています。具体的には、SPARC64 X は SPARC64 VII+と比較して、CPU コアあたりのパフォーマンスは約 2 倍、プロセッサあたりのパフォーマンスは約 7.5 倍に向上しています。SPARC64 X+では、さらに向上しており、そのパフォーマンスは CPU コアあたりで 2.4 倍、プロセッサあたりで 9.5 倍になります。さらに、SPARC64 XII では、SPARC64 X+と比較して、CPU コアあたりで約 2.5 倍、プロセッサあたりで約 2 倍と飛躍的に性能が向上しています（SPECint_rate で測定）。

SPARC64 XII/SPARC64 X+/SPARC64 X の技術拡張の詳細については、「7. SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサの技術拡張」を参照してください。

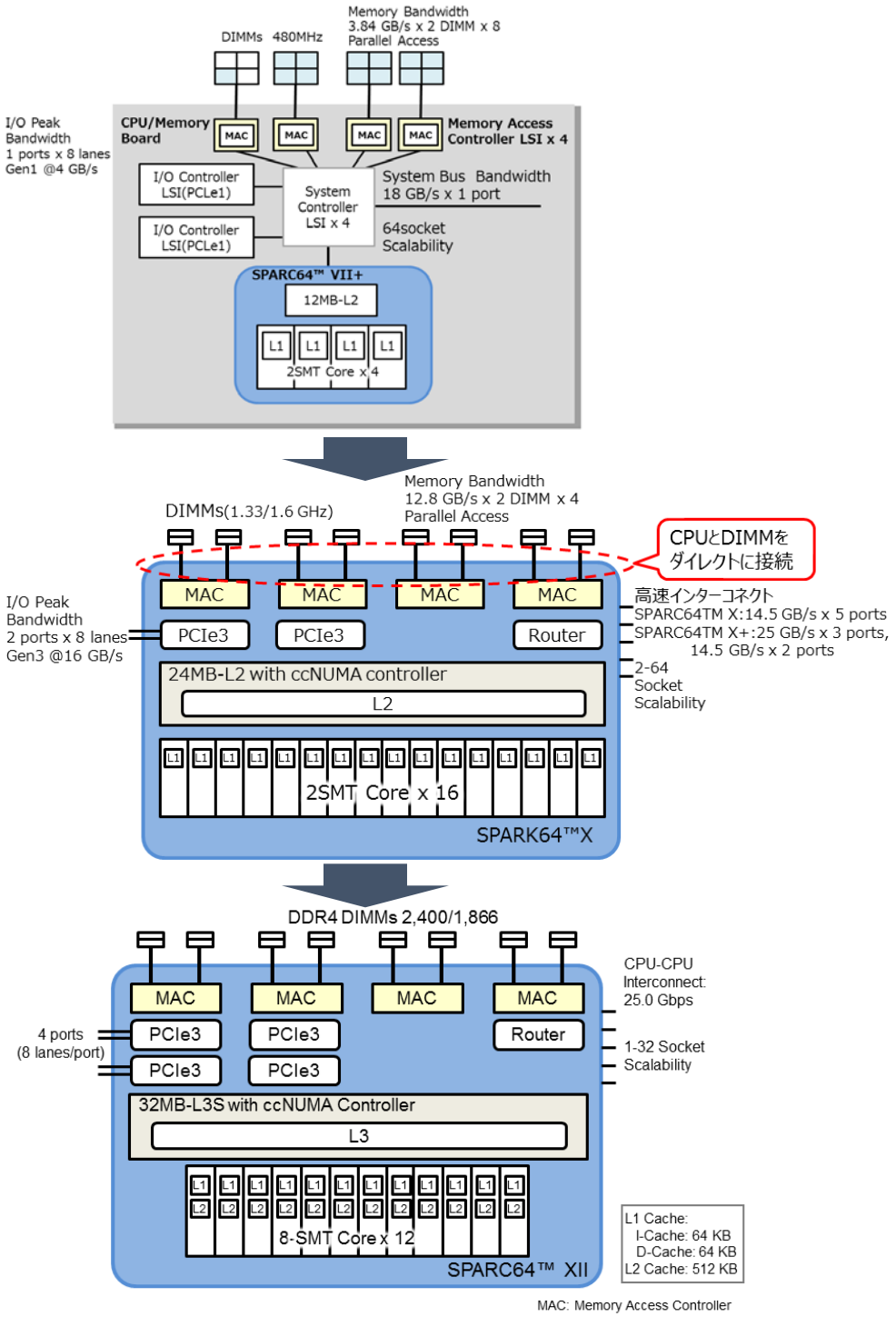

図 3-1. SPARC64 XII/SPARC64 X/SPARC64 VII+の機能拡張

4. システムアーキテクチャー

SPARC M12/M10では、プロセッサ、システムインターコネクト、メモリサブシステム、I/Oサブシステムが相互に連携しています。これにより、企業の汎用コンピューティングの統合から、大量データの高速かつ安定稼動が不可欠となるデータベース処理まで、幅広いワークロードのニーズに対応します。つまり、拡張性と高い性能を兼ね備えた最適なプラットフォームを実現します。

SPARC M12/M10 は、ccNUMA アーキテクチャーのメモリローカリティのメリットを最大限に活かした高性能と高信頼性に重点をおいて設計されています。このこだわりは、SPARC M12/M10 内のサブシステムすべての特性や機能においても同じことがいえます。例えば、SPARC M12/M10 内の高帯域幅のシステムバス、強力な SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサ、高密度メモリサポート、高速 PCI Express などは、エンタープライズアプリケーション向けに高い信頼性、拡張性はもちろん、最高水準の稼動時間や処理能力をもたらします。

システムインターコネクト

システムインターコネクトは、SPARC M12/M10の高性能、拡張性と信頼性を支えています。例えば、システム内の複数のシステムコントローラーとクロスバーユニットを経由して、CPU、メモリ、I/Oサブシステム間をダイレクトに接続します。また、コンポーネント間の複数のバスルートによって、障害発生時でもシステムオペレーションを中断することなく継続できる設計になっています。

1. システムバス

ハイエンドサーバには膨大な数の CPU が搭載されていますが、アプリケーションのパフォーマンス向上のためにすべてのプロセッサを活用できなければ、リニアなスケーラビリティを実現することができません。リニアなスケーラビリティを実現し、多様なアプリケーションに高速なパフォーマンスを確実に提供するためには、システムバスの能力が重要な役割を担っています。そのため、SPARC M12/M10では、コンポーネント間に莫大な帯域幅と低いレイテンシを維持するように設計されたシステムインターコネクトを採用しています。このようなシステムバスは、バランスのとれた確実なパフォーマンスをアプリケーションに提供します。これにより、お客様の IT 業務に大きく寄与することができます。

SPARC M12/M10の性能は、先に述べたインターコネクトの設計によって最大限に活か

すことができます。システムバスは、ダイレクト接続とパケット交換技術の活用により複数のデータストリームを転送することで、応答時間を短縮しています。パケット交換はバス上の「デッド」サイクルを排除し、これによってシステム全体で高い処理性能が実現されます。また、システムバスは単方向バスで構成されており、競合を起こすことなくデータストリームを転送することができます。そして、データストリーム内のアドレス、データ、制御データは多重化されて転送されます。

サーバのシステムコントローラーは、各サーバ内の CPU、メモリ、I/O サブシステム、インターコネクトパス間のトラフィックの転送先制御を行っています。

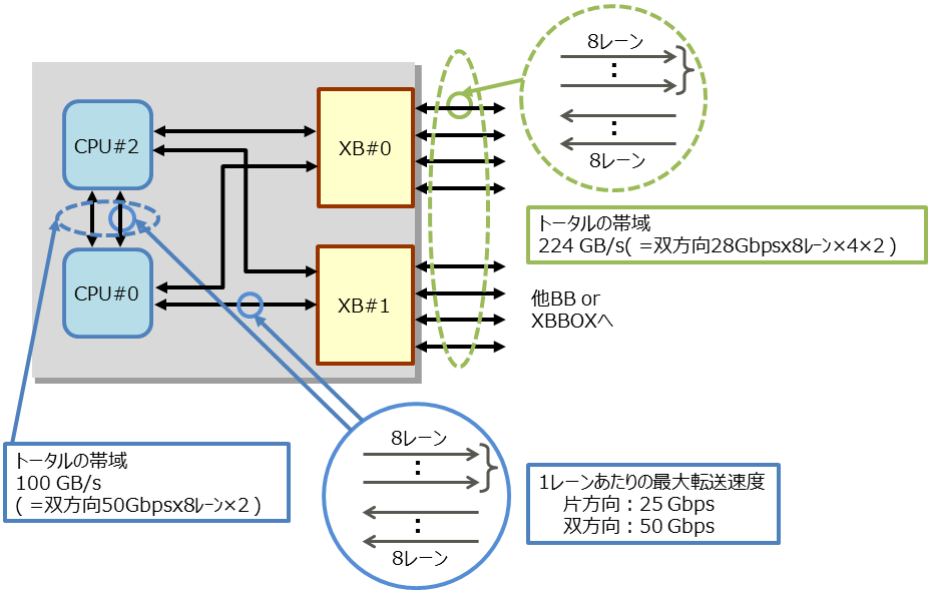

図 4-1. SPARC M12 の帯域幅とデータ転送速度

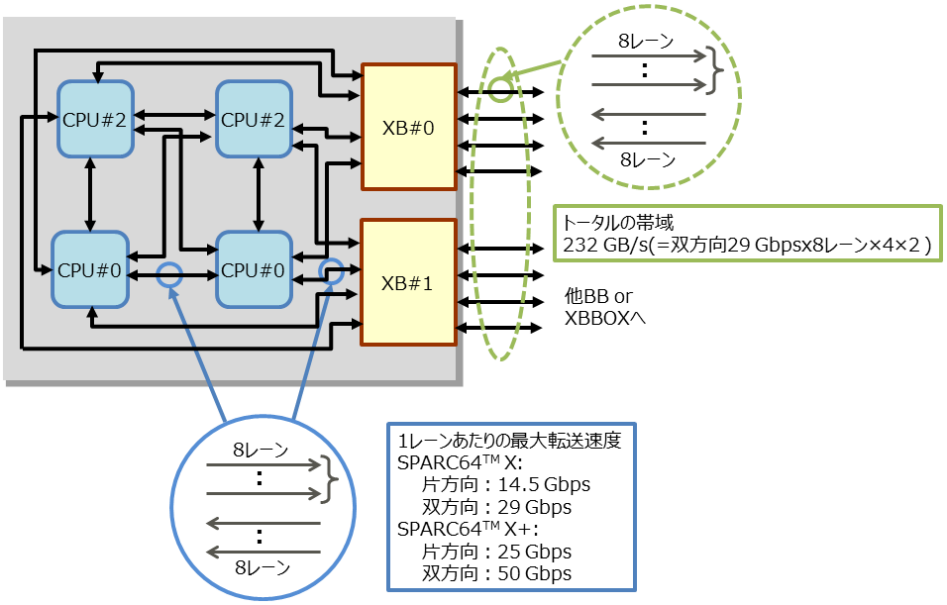

図 4-2.　SPARC M10 の帯域幅とデータ転送速度

2. SPARC M12-1 インターコネクトアーキテクチャー

SPARC M12-1 では、SPARC64 XII プロセッサを 1 個搭載した 1 つのシステムボードが実装されています。プロセッサにメモリアクセスコントローラー、I/O コントローラー、システムコントローラーを内蔵しています。プロセッサは、DIMM、PCI Express スイッチに直接接続しています。

SPARC M12-1 のインターコネクトアーキテクチャーを図 4-3 に示します。

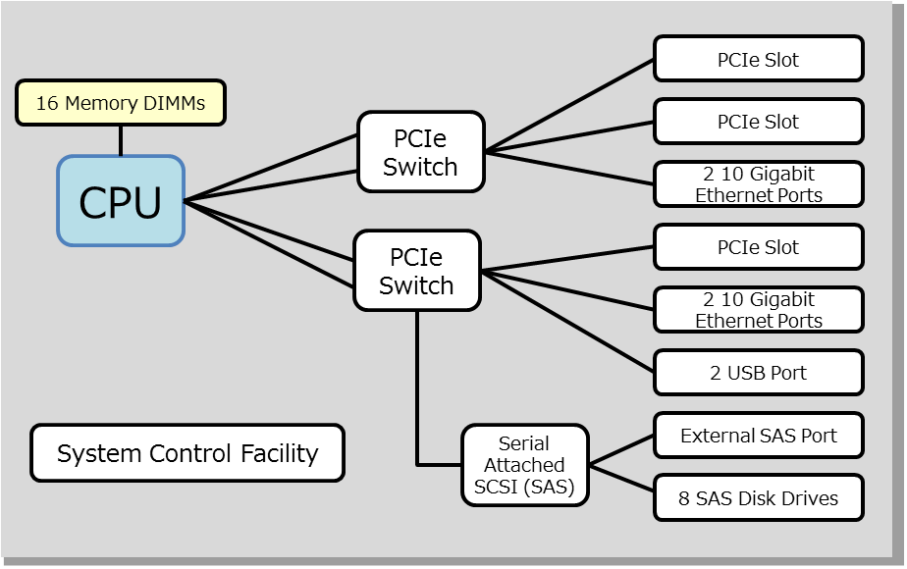

図 4-3. SPARC M12-1 インターコネクトアーキテクチャー

3. SPARC M12-2 インターコネクトアーキテクチャー

1 台の SPARC M12-2 に、SPARC64 XII プロセッサを標準で 1 個搭載、最大 2 個搭載可能です。プロセッサに、メモリアクセスコントローラー、I/O コントローラー、システムコントローラーが内蔵されています。各プロセッサは DIMM および PCI Express スイッチと接続しており、2 個のプロセッサ間はシステムバスにより直接接続されます。I/O バスの帯域幅を広げ、プロセッサ障害の影響を防ぐために、PCI Express スイッチはそれぞれ 2 個のプロセッサに接続しています（図 4-4）。

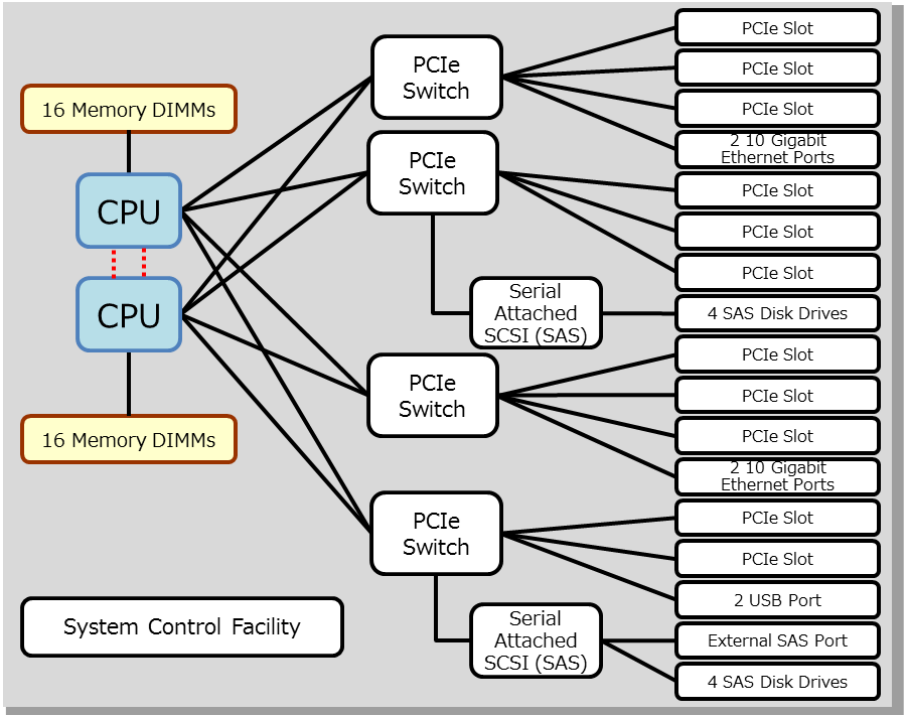

図 4-4. SPARC M12-2 インターコネクトアーキテクチャー

4. SPARC M12-2S インターコネクトアーキテクチャー

SPARC M12-2S は、ビルディングブロック方式で接続するために、サーバとクロスバーボックスにクロスバーユニットを実装し、複数 SPARC M12-2S 間を接続しています。SPARC M12-2S 1 台あたり、SPARC64 XII プロセッサを標準で 1 個搭載、最大 2 個搭載可能です。プロセッサに、メモリアクセスコントローラー、I/O コントローラー、システムコントローラーが内蔵されています。各プロセッサは DIMM および PCI Express スイッチと接続しており、2 個のプロセッサ間はシステムバスにより直接接続されます。I/O バスの帯域幅を広げ、プロセッサ障害の影響を防ぐために、PCI Express スイッチはそれぞれ 2 個のプロセッサに接続しています（図 4-5）。

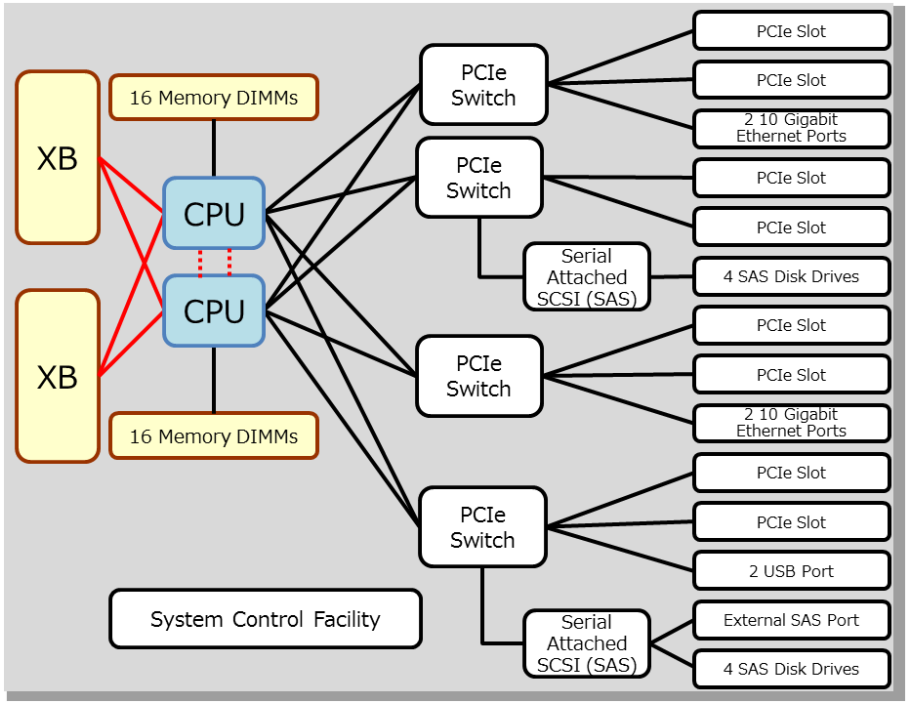

図 4-5. SPARC M12-2S インターコネクトアーキテクチャー

SPARC M12-2S では、ビルディングブロック方式を採用しており、SPARC M12-2S 間のシステムバスを接続するためにクロスバーユニットが実装されています。このシステムバスは、低レイテンシでかつ高いスループットのデータ転送を提供します。

また、ビルディングブロック構成時のメモリの物理アドレスは、システム性能を最適化するために、すべての SPARC M12-2S のシステムコントローラーへ均等に分散するように設計されています。

SPARC64 XII プロセッサがクロスバーユニットに直接接続されるため、複数の SPARC

M12-2S にまたがったデータ転送のレイテンシが最小化されます。また、それぞれの SPARC M12-2S に内蔵されたクロスバーユニットを Point to Point で直接接続することにより、最大 4 台の SPARC M12-2S に接続して 8 CPU チップまで拡張できます（図 4-6）。最大 16 台の SPARC M12-2S を接続するシステム構成の場合は、各 SPARC M12-2S はクロスバーボックスを経由してほかのすべての SPARC M12-2S に接続されます（図 4-7）。SPARC M12-2S のシステムインターコネクトは、ピーク時に 5,453 GB/s という帯域幅を実現し、リニアなスケーラビリティを提供します。

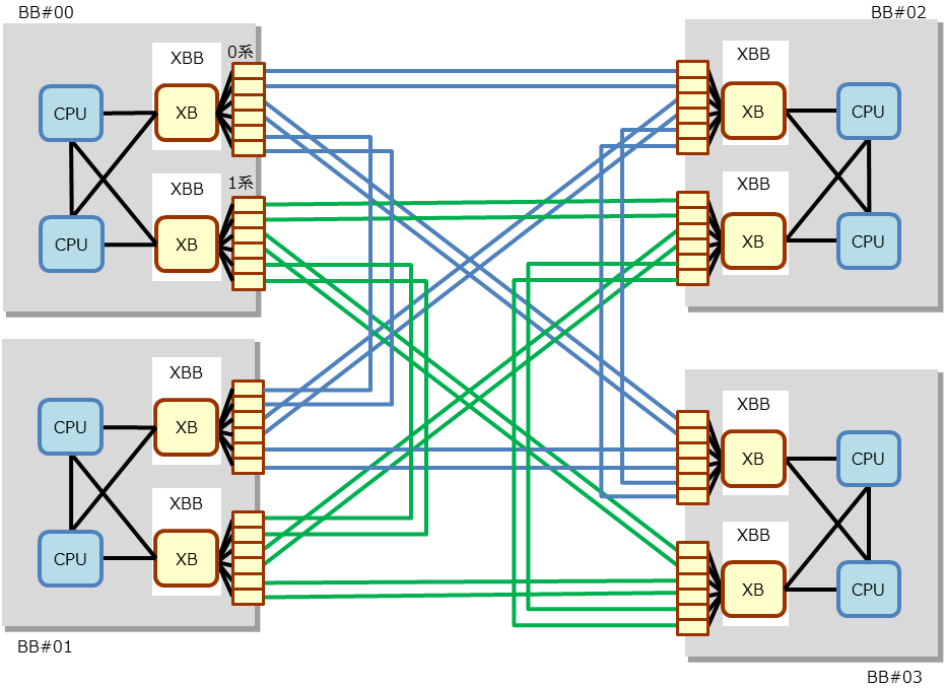

図 4-6. サーバどうしを直接接続するシステム構成（SPARC M12-2S 4BB 構成）

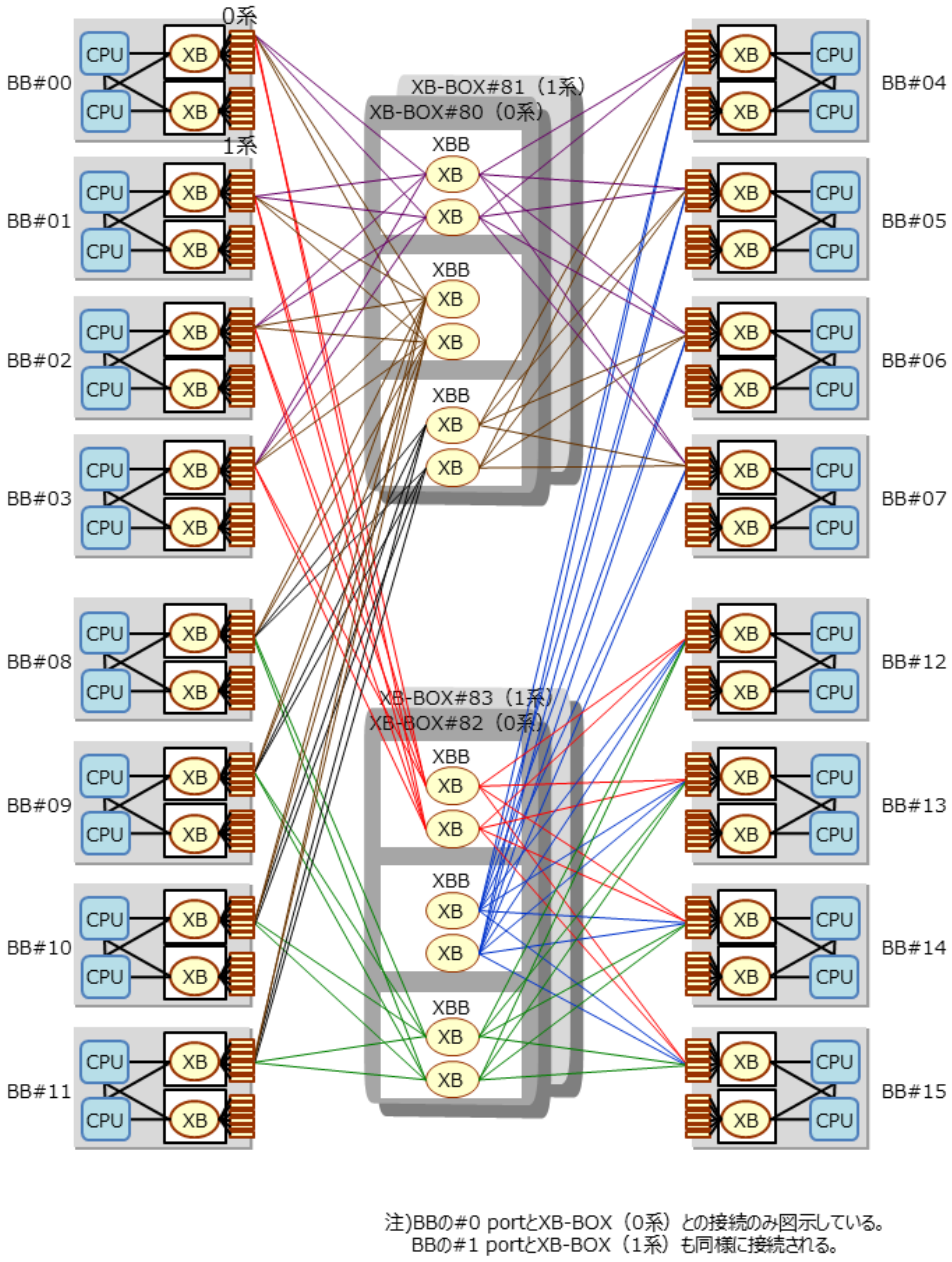

図 4-7. クロスバーボックスを経由して接続するシステム構成（SPARC M12-2S 16BB 構成）

5. SPARC M10-1 インターコネクトアーキテクチャー

SPARC M10-1 では、1 つのシステムボードが実装されています。SPARC64 X+または SPARC64 X プロセッサを 1 個搭載し、プロセッサにメモリアクセスコントローラー、I/O コントローラー、システムコントローラーを内蔵しています。プロセッサは、DIMM、PCI Express スイッチに接続しています。

SPARC M10-1 のインターコネクトアーキテクチャーを図 4-8 に示します。

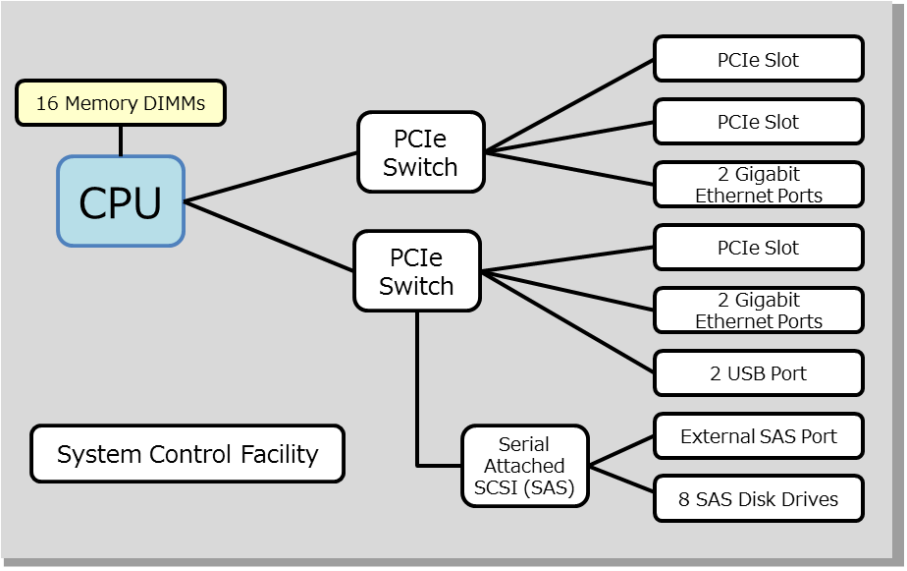

図 4-8. SPARC M10-1 インターコネクトアーキテクチャー

6. SPARC M10-4 インターコネクトアーキテクチャー

SPARC M10-4 では、2 つのシステムボードが実装されていますが、1 つの論理システムボードとして機能します。SPARC64 X+プロセッサまたは SPARC64 X プロセッサを最大 4 個まで搭載可能です。SPARC M10-1 と同様、CPU に、メモリアクセスコントローラー、I/O コントローラー、システムコントローラーを内蔵しています。各コントローラーは DIMM、PCI Express スイッチと接続し、4 個のプロセッサ間はすべて直接接続されています。I/O バスの帯域幅を広げ、プロセッサ障害の影響を防ぐために、PCI Express スイッチはそれぞれ 2 個のプロセッサに接続しています（図 4-9）。

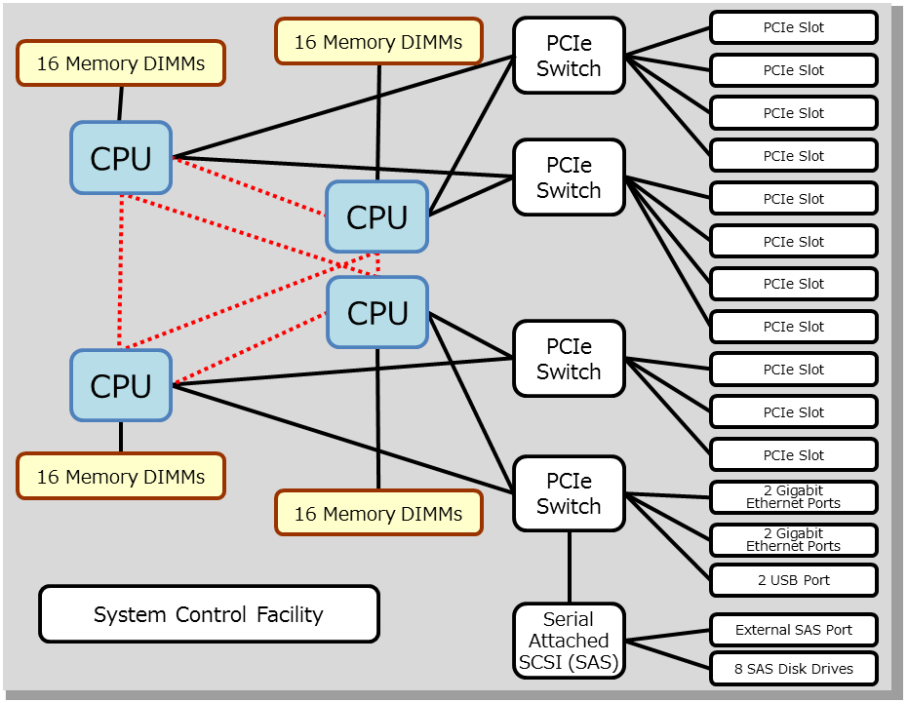

図 4-9. SPARC M10-4 インターコネクトアーキテクチャー

7. SPARC M10-4S インターコネクトアーキテクチャー

SPARC M10-4S は、ビルディングブロック方式で接続するために、サーバとクロスバーボックスにクロスバーユニットを実装し、複数の SPARC M10-4S 間を接続しています。SPARC M10-4 と同様、各サーバでは 2 つのシステムボードが実装されていますが、1 つの論理システムボードとして機能します。SPARC64 X+プロセッサまたは SPARC64 X プロセッサを最大 4 個まで搭載可能です。SPARC M10-1、SPARC M10-4 と同様、プロセッサに、メモリアクセスコントローラー、I/O コントローラー、システムコントローラーを内蔵しています。各コントローラーは DIMM、PCI Express スイッチと接続し、4 個のプロセッサ間はすべて直接接続されています。I/O バスの帯域幅を広げ、プロセッサ障害の影響を防ぐために、PCI Express スイッチはそれぞれ 2 個のプロセッサに接続しています（図 4-10）。

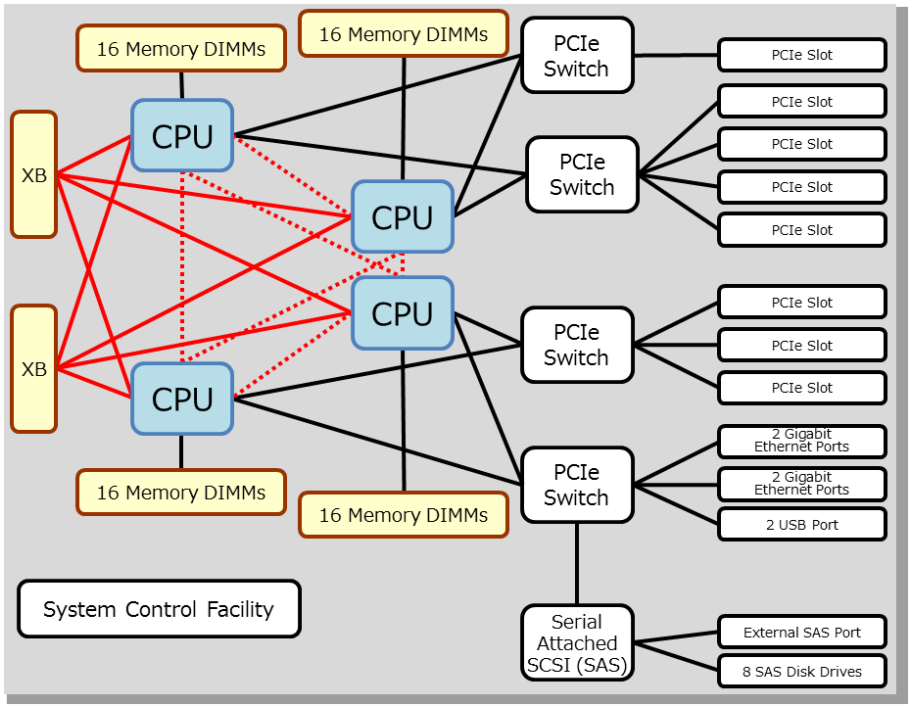

図 4-10. SPARC M10-4S インターコネクトアーキテクチャー

SPARC M10-4S では、ビルディングブロック方式を採用しており、SPARC M10-4S 間のシステムバスを接続するためにクロスバーユニットが実装されています。このシステムバスは、低レイテンシでかつ高いスループットのデータ転送を提供します。

また、ビルディングブロック構成時のメモリの物理アドレスは、システム性能を最適化するために、すべての SPARC M10-4S のシステムコントローラーへ均等に分散するよ

うに設計されています。

SPARC64 X+/SPARC64 X プロセッサがクロスバーユニットに直接接続されるため、複数の SPARC M10-4S をまたがったデータ転送のレイテンシが最小化されます。また、それぞれの SPARC M10-4S に内蔵されたクロスバーユニットを Point to Point で直接接続することにより、最大 4 台の SPARC M10-4S に接続して 16CPU チップまで拡張できます(図 4-11)。最大 16 台の SPARC M10-4S を接続するシステム構成の場合は、各 SPARC M10-4S はクロスバーボックスを経由してほかのすべての SPARC M10-4S に接続されます（図 4-12)。SPARC M10-4S のシステムインターコネクトは、ピーク時に 6,553 GB/s という帯域幅を実現し、富士通の前世代ハイエンドサーバのおよそ 9 倍もの処理能力を発揮します。

SPARC64 X プロセッサを搭載した SPARC M10-4S は、ビルディングブロックを構成するサーバとして SPARC64 X プロセッサを搭載した SPARC M10-4S、および SPARC64 X+ プロセッサを搭載した SPARC M10-4S、両方に接続することが可能です。

よって、すでに稼動している SPARC64 X プロセッサを搭載した SPARC M10-4S を無駄にすることなく、さらに高性能なシステムにアップグレードできます。

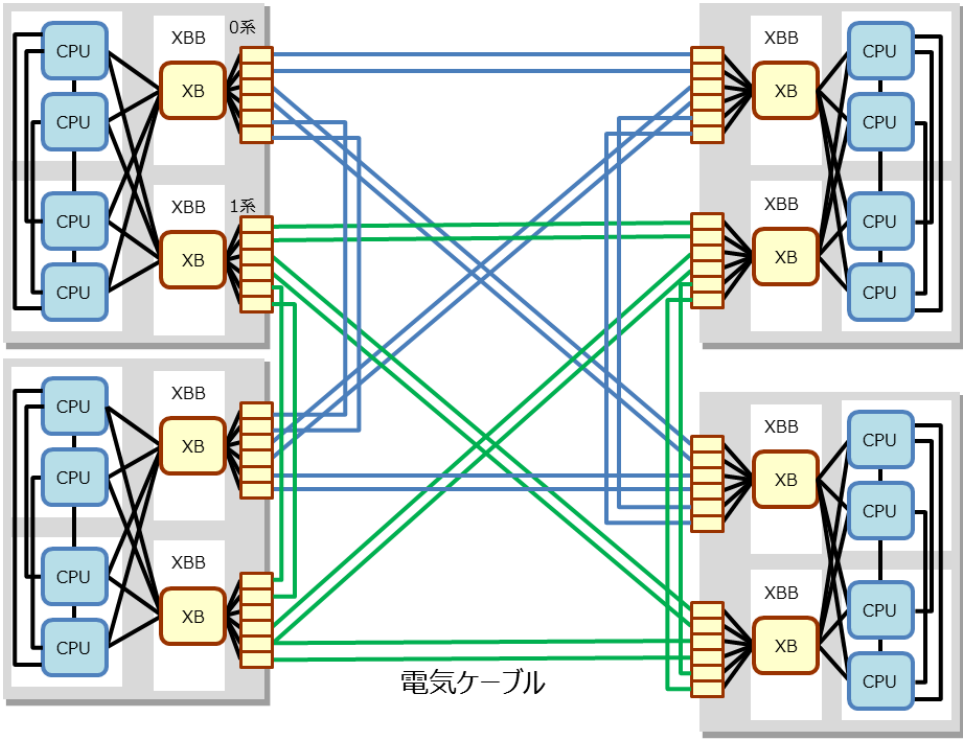

図 4-11. サーバどうしを直接接続するシステム構成（SPARC M10-4S 4BB 構成）

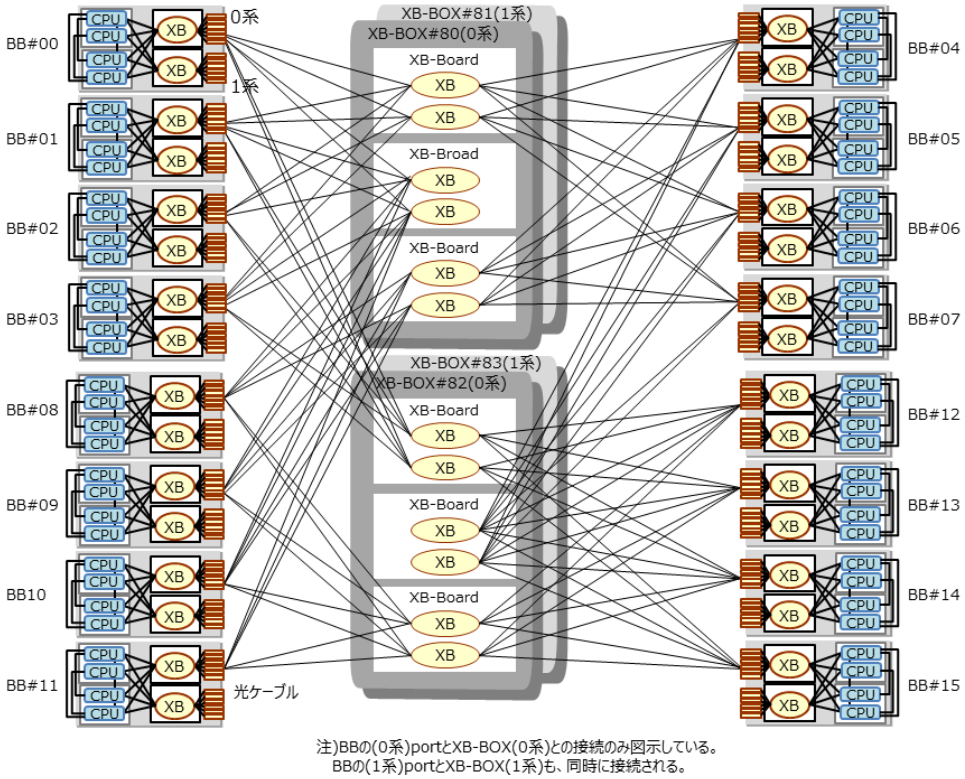

図 4-12. クロスバーボックスを経由して接続するシステム構成
（SPARC M10-4S 16BB 構成）

8. システムインターコネクトの信頼性機能

SPARC M12/M10 に組み込まれているシステムインターコネクトの冗長性と信頼性の機能が、サーバの安定性をさらに高めます。このインターコネクトは、トランザクションアドレス、データともに、ECC あるいは CRC で保護されています。これにより、すべてのシステムバス上で損失や破損に対して保護されています。例えば、CPU、メモリアクセスコントローラー、I/O コントローラー内で 1 ビットデータエラーが検出されると、ハードウェアはエラーのデータを訂正して転送を続けます。さらに、富士通の高速インターコネクトで接続されるシステムバス（クロスバーとクロスバー間、クロスバーとプロセッサ間、プロセッサとプロセッサ間）内で CRC がマルチビットデータエラーを検出した場合は、ハードウェアが自動的にそのデータを再送します。万が一、データ再送によるエラー復旧が見込めない場合は、そのレーンを縮退します。ハードウェア障害がシステムインターコネクト内で発生した場合には、問題のあるレーンを切り離し、残る正常なバスルートを利用して業務を再開させます。

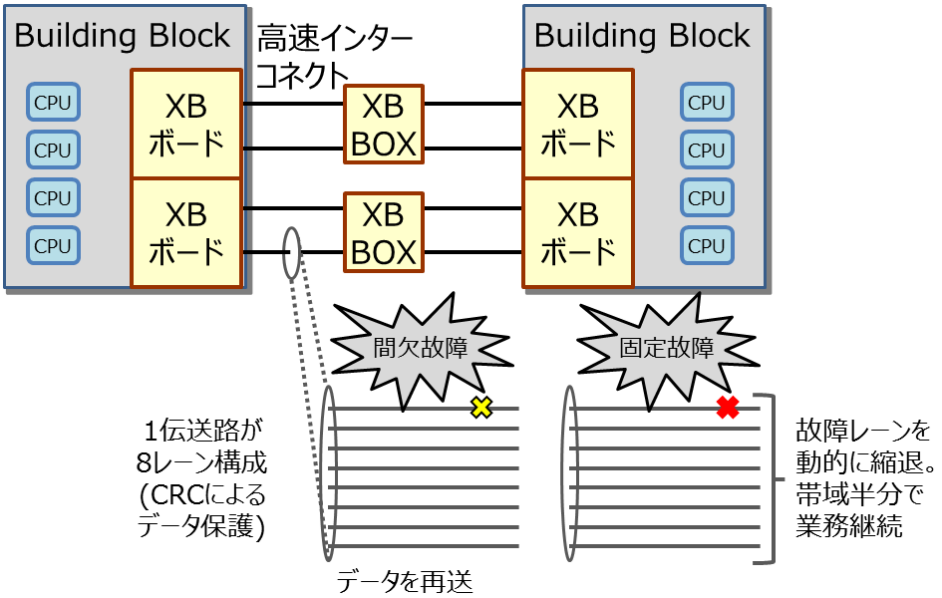

図 4-13. システムインターコネクトの信頼性機能

メモリ

SPARC M12/M10 のメモリサブシステムは、システムのスケーラビリティとスループットの向上に貢献しています。16BB 構成で、SPARC M12 は最大 3 TB（1 プロセッサあたり 24 メモリスロットの場合）または 2 TB（1 プロセッサあたり 16 メモリスロットの場合）まで、SPARC M10 は最大 64 TB までのメモリが搭載可能です。また SPARC M12 は DDR4 DIMM、SPARC M10 は、DDR3 DIMM をサポートしており、システム性能を強化するために、どちらも最大 4way メモリインターリーブ構成をとっています。使用可能な DIMM サイズは、8 GB、16 GB、32 GB、64 GB です。表 4-1、表 4-2 にメモリサブシステムの詳細を示します。

表 4-1. SPARC M12 のメモリサブシステム仕様

	SPARC M12-1	SPARC M12-2	SPARC M12-2S（1BB）	SPARC M12-2S（16BB）（*1）
		DDR4-2400/1866（*2）		
最大メモリ容量	1 TB	3 TB（1プロセッサあたり24メモリスロットの場合） 2 TB（1プロセッサあたり16メモリスロットの場合）	3 TB（1プロセッサあたり24メモリスロットの場合） 2 TB（1プロセッサあたり16メモリスロットの場合）	48 TB（1プロセッサあたり24メモリスロットの場合） 32 TB（1プロセッサあたり16メモリスロットの場合）
最大DIMMスロット数	16	48（1プロセッサあたり24メモリスロットの場合） 32（1プロセッサあたり16メモリスロットの場合）	48（1プロセッサあたり24メモリスロットの場合） 32（1プロセッサあたり16メモリスロットの場合）	768（1プロセッサあたり24メモリスロットの場合） 512（1プロセッサあたり16メモリスロットの場合）
メモリインターリーブ単位	4または8 DIMM	8 DIMM	8 DIMM	8 DIMM
最大インターリーブ数	4	4	4	64

*1　2022 年 4 月以降に販売開始した M12-2S は最大 4BB。仕様の各項目の最大値は、16BB の 1/4 となります。

*2　1 プロセッサあたり 24 メモリスロットで、24 DIMM 搭載した場合

表 4-2. SPARC M10 のメモリサブシステム仕様

	SPARC M10-1	SPARC M10-4	SPARC M10-4S（1BB）	SPARC M10-4S（16BB）
		DDR3-1600/1333		
最大メモリ容量	1 TB	4 TB	4 TB	64 TB
最大DIMMスロット数	16	64	64	1024
メモリインターリーブ単位	4 または 8 DIMM	8 DIMM	8 DIMM	8 DIMM
最大インターリーブ数	4	8	8	128

SPARC M12/M10 のメモリサブシステムは、性能はもちろん、信頼性にも配慮されています。例えば、メインメモリに格納されたすべてのデータは ECC で保護されています。さらに、以下に示す高度な機能により、早期の診断を可能とするとともに障害分離が強化されているため、システムの整合性が保障され、アプリケーションの可用性が向上します。

- メモリパトロール
  メモリパトロールは、エラーをチェックするためにメモリを定期的にスキャンする機能です。システムやアプリケーションのエラー発生前にメモリの障害領域を使用しないようにするため、システムの信頼性が守られます。
- メモリ拡張 ECC
  メモリ拡張 ECC は、1 ビットエラーの訂正を行います。そのため、読み込みエラーなどのメモリデバイス障害が起こったときでも、中断することなく処理を継続することができます。この機能は、IBM のチップキル技術に似ています。

1. メモリミラーリング

SPARC M12/M10 は、メモリミラーリング機能を採用しています。メモリミラーリング機能は、高可用性を実現するための機能であり、可用性要件が厳しいアプリケーションの実行に適しています。例えば、SPARC M12/M10 上でメモリミラーリングモードが有効になっている場合は、メモリサブシステムは書き込み時にデータを複製します。そして、読み出し時にはデータをメモリミラーの両サイドのデータと比較します。このほか、バスレベルまたは DIMM レベルでエラーが発生した場合は、もう一方のメモリバスと代替 DIMM セットを活用して通常のデータ処理を継続させることができます。さらに SPARC M12/M10 では、SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサに内蔵された

メモリアクセスコントローラー（MAC）を使用して、メモリモジュール間でメモリをミラーリングします（図 4-14）。

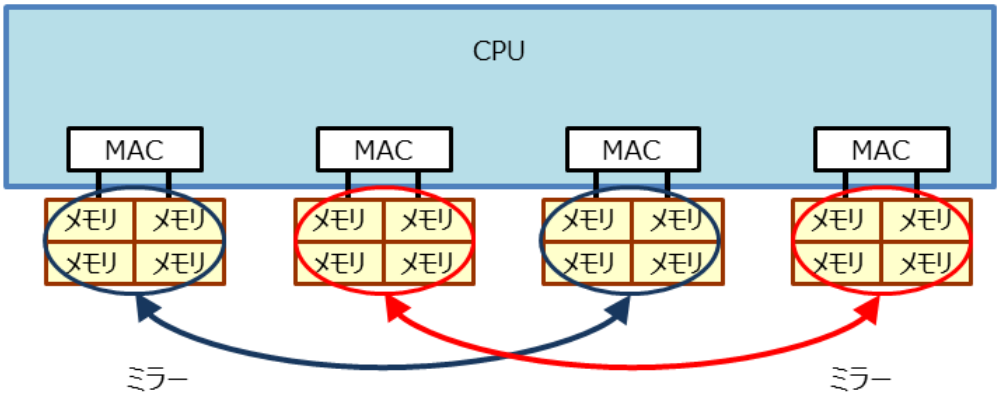

図 4-14.　SPARC M12/M10 のメモリミラーリングのアーキテクチャー

システムクロック

SPARC M12/M10 のシステムクロックは、CPU ごとに独立して実装されています。このため、万が一システムクロックに障害が発生した場合でも、障害が発生したシステムクロックを使用しているプロセッサのみを縮退させてシステムを再起動することができます。

I/O バス

SPARC M12/M10 は、PCI Express バスによる I/O サブシステム内の高速データ転送を実現しています。様々な PCI Express カード（Fibre Channel、InfiniBand、Gigabit Ethernet、Flash PCI Express など）が最適な I/O 性能を発揮できるように、SPARC M12/M10 では各プロセッサ内に PCI Express 3.0 プロトコルを実装しています。PCI Express 3.0 のデータ転送速度はピーク時で 8 GB/s に達します。

1.　I/O サブシステムのアーキテクチャー

SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサと、PCI Express スロット、オンボードデバイスや内蔵ドライブは、PCI Express スイッチにより接続されます。なお、PCI Express のバスを PCI ボックスにより拡張することができ、PCI Express スロットを増設可能です。このように SPARC M12/M10 の I/O アーキテクチャーは、サーバの統合や仮想化に必要な拡張性と高い性能（最新の PCI Express 3.0 の高い処理能力）を提供できるように設計されています。

PCI Express カードのホットプラグを容易にするため、SPARC M12-2/M12-2S/M10-4/M10-4S や PCI ボックスには、PCI Express カセットを採用しています。PCI ホットプラグをサポートしている PCI Express カードを PCI Express カセットに搭載することで、稼動中のサーバに内蔵された PCI Express スロットまたは PCI ボックスのスロットに挿入できます。

2. SPARC M12-1 の I/O サブシステム

図 4-15 に、SPARC M12-1 の I/O サブシステムのアーキテクチャーを示します。図には、SPARC M12-1 のシステムボードに実装された 2 つの PCI Express のスイッチがあります。このスイッチがすべての I/O コンポーネントとプロセッサ内の I/O コントローラーに接続します。I/O サブシステムは、3 つの PCI Express スロット、1 つの外部 SAS ポート、2 つの USB ポート（Front:USB 2.0、Rear:USB 3.0）を実装しています。外部 SAS ポートは SAS テープデバイスまたはストレージデバイスとの接続に、USB ポートはサポート対象の DVD デバイスとの接続に使用します。さらに、PCI ボックスを接続することにより、PCI Express スロットを増やすこともできます。

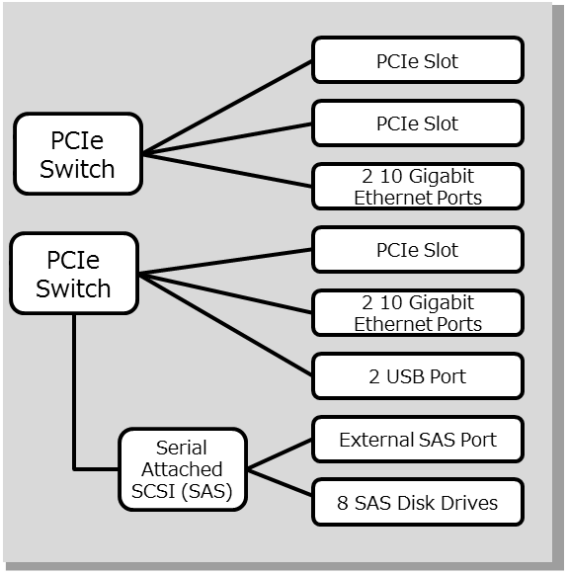

図 4-15. SPARC M12-1 I/O サブシステムのアーキテクチャー

3. SPARC M12-2 の I/O サブシステム

SPARC M12-2 のシステムボードには、4 つの PCI Express のスイッチが実装されています。このスイッチがすべての I/O コンポーネントとプロセッサ内の I/O コントローラーを接続しています。また、PCI Express スイッチと CPU 内蔵の I/O コントローラーは、1 対 1 あるいは 1 対 2 のように CPU チップの搭載数に応じて接続することができます。I/O サブシステムは、内部 SAS コントローラー 2 つと 10GbE コントローラー 2 つによる冗長構成が可能です。11 の PCI Express スロット、1 つの外部 SAS ポート、2 つの USB ポート（Front:USB 2.0、Rear:USB 3.0）を実装しており、外部 SAS ポートは SAS テープデバイスまたはストレージデバイスとの接続に、USB ポートはサポート対象の DVD デバイスとの接続に使用します。さらに、PCI ボックスを接続することにより、PCI Express スロットを増やすこともできます。

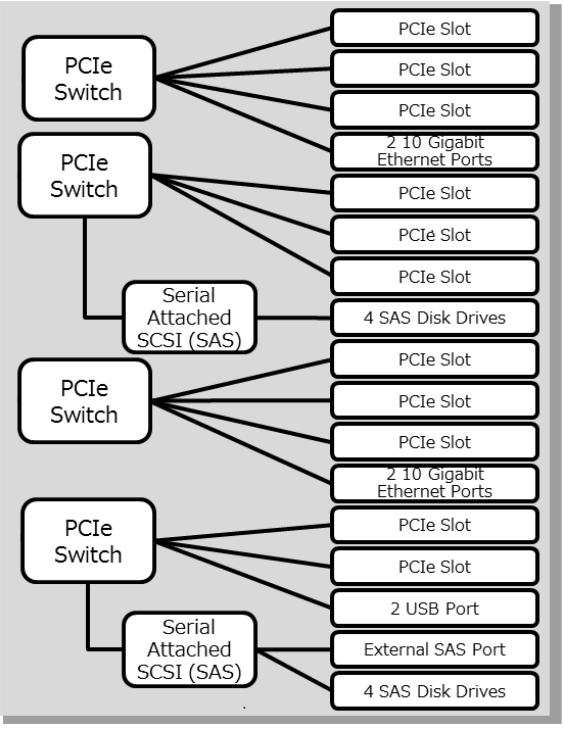

図 4-16. SPARC M12-2 I/O サブシステムのアーキテクチャー

4. SPARC M12-2S の I/O サブシステム

SPARC M12-2S のシステムボードには、4 つの PCI Express のスイッチが実装されています。このスイッチがすべての I/O コンポーネントとプロセッサ内の I/O コントローラーを接続しています。また、PCI Express スイッチと CPU 内蔵の I/O コントローラーは、1 対 1 あるいは 1 対 2 のように CPU チップの搭載数に応じて接続することができます。I/O サブシステムは、内部 SAS コントローラー 2 つと 10GbE コントローラー 2 つによる冗長構成が可能です。8 つの PCI Express スロット、1 つの外部 SAS ポート、2 つの USB ポート（Front:USB 2.0、Rear:USB 3.0）を実装しており、外部 SAS ポートは SAS テープデバイスまたはストレージデバイスとの接続に、USB ポートはサポート対象の DVD デバイスとの接続に使用します。さらに、PCI ボックスを接続することにより、PCI Express スロットを増やすこともできます。

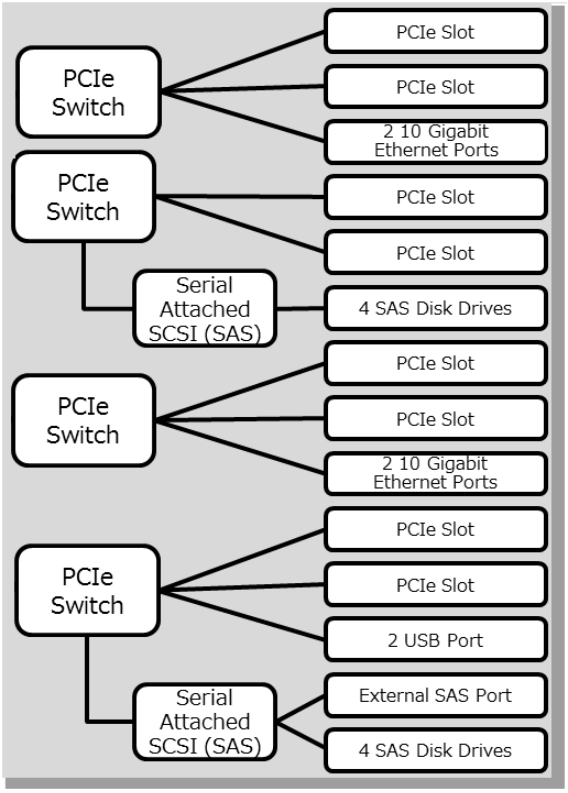

図 4-17. SPARC M12-2S I/O サブシステムのアーキテクチャー

5. SPARC M10-1 の I/O サブシステム

SPARC M10-1 のシステムボードには、2 つの PCI Express のスイッチが実装されています。このスイッチがすべての I/O コンポーネントと SPARC64 X+/SPARC64 X プロセッサ内の I/O コントローラーを接続しています。I/O サブシステムは、3 つの PCI Express スロット、1 つの外部 SAS ポート、2 つの USB 2.0 ポートを実装しており、外部 SAS ポートは SAS テープデバイスまたはストレージデバイスとの接続に、USB 2.0 ポートはサポート対象の DVD デバイスとの接続に使用します。さらに、PCI ボックスを接続することにより、PCI Express スロットを増やすこともできます。

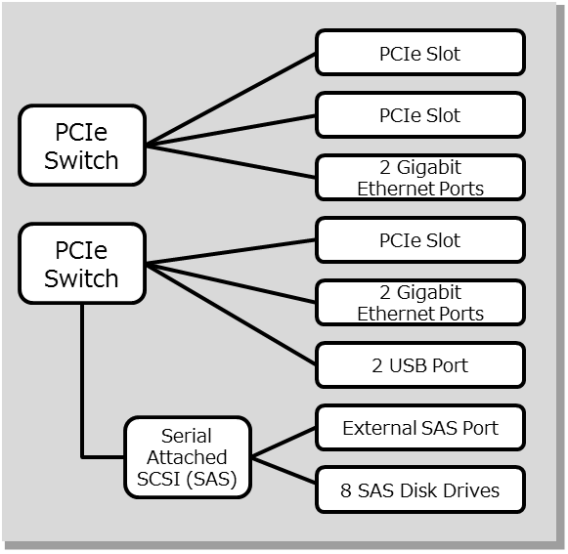

図 4-18. SPARC M10-1 I/O サブシステムのアーキテクチャー

6. SPARC M10-4 の I/O サブシステム

SPARC M10-4 のシステムボードには、4 つの PCI Express のスイッチが実装されています。このスイッチがすべての I/O コンポーネントと SPARC64 X+/SPARC64 X プロセッサ内の I/O コントローラーを接続しています。また、PCI Express スイッチと CPU 内蔵の I/O コントローラーは、1 対 1 あるいは 1 対 2 のように CPU チップの搭載数に応じて接続することができます。I/O サブシステムは、11 の PCI Express スロット、1 つの外部 SAS ポート、2 つの USB 2.0 ポートを実装しており、外部 SAS ポートは SAS テープデバイスまたはストレージデバイスとの接続に、USB 2.0 ポートはサポート対象の DVD デバイスとの接続に使用します。さらに、PCI ボックスを接続することにより、PCI Express スロットを増やすこともできます。

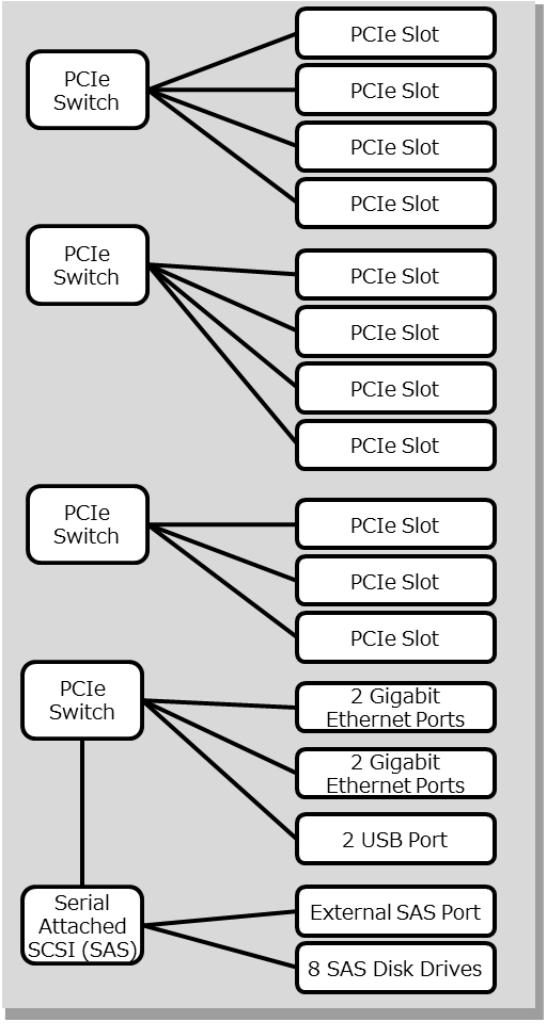

図 4-19. SPARC M10-4 I/O サブシステムのアーキテクチャー

7. SPARC M10-4S の I/O サブシステム

SPARC M10-4S のシステムボードには、4 つの PCI Express のスイッチが実装されています。このスイッチがすべての I/O コンポーネントと SPARC64 X+/SPARC64 X プロセッサ内の I/O コントローラーを接続しています。また、PCI Express スイッチと CPU 内蔵の I/O コントローラーは、1 対 1 あるいは 1 対 2 のように、各 SPARC M10-4S の CPU チップの搭載数に応じて接続することができます。I/O サブシステムは、SPARC M10-4S あたり 8 つの PCI Express スロット、1 つの外部 SAS ポート、2 つの USB 2.0 ポートを実装しており、外部 SAS ポートは SAS テープデバイスまたはストレージデバイスとの接続に、USB 2.0 ポートはサポート対象の DVD デバイスとの接続に使用します。さらに、PCI ボックスを接続することにより、PCI Express スロットを増やすこともできます。

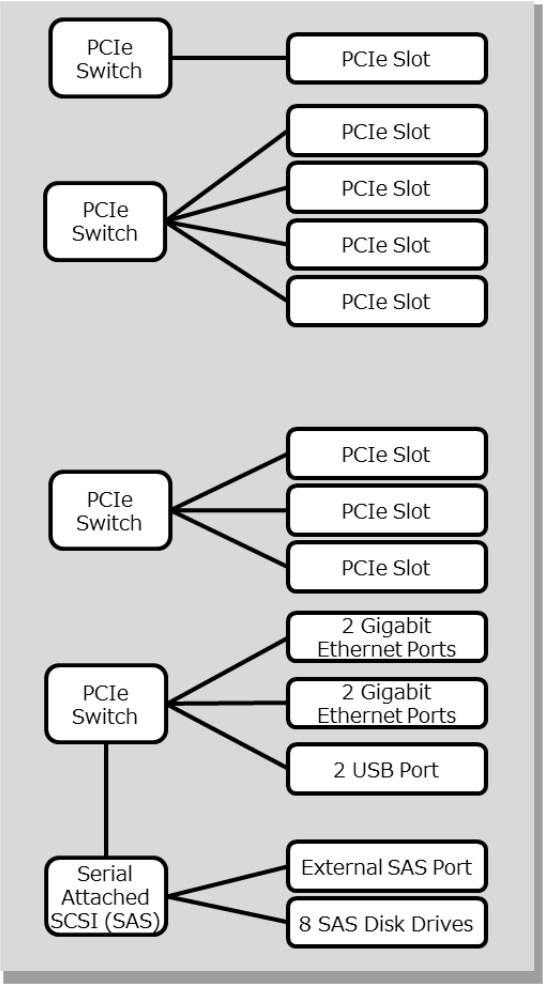

図 4-20. SPARC M10-4S I/O サブシステムのアーキテクチャー

8. 内蔵ドライブと周辺装置

SPARC M12/M10は、活性交換可能な 8台の内蔵 Serial Attached SCSI（SAS）2.5インチハードディスクドライブ、および2.5インチ SSDを搭載できます。オンボード SAS コントローラーは、SPARC M12では RAID 0、RAID 1、RAID 1E、RAID 10ボリュームおよびホットスワップドライブを、SPARC M10では RAID 0、RAID 1、RAID 1Eボリュームおよびホットスワップドライブをサポートしています。加えて、SAS テープデバイスやストレージデバイスを接続する外部SASポートを1つ実装しています。このSASポートは 4レーンで構成されており、合計で最大 24 Gbps の帯域幅を実現します。さらに SPARC M12/M10は、DVD ドライブを接続する USBポートをそれぞれ 2つ（SPARC M12は USB 3.0ポートを 1、USB 2.0ポートを 1、SPARC M10は USB 2.0ポートを 2）を実装しています。

9. PCI ボックス

SPARC M12/M10では、I/O接続追加用としてオプションの PCIボックスを提供しています。この PCI ボックスは 2 Uのラックに搭載可能で、PCI Express スロットを 11スロット内蔵しています。PCI ボックスは、PCI Express カセットを使用することで、PCI Express カードの活性交換をサポートします。

PCI ボックスは、SPARC M12/M10の内蔵 PCI Express スロットに実装されたリンクカードに接続され、サイドバンド信号で制御されます。リンクカードはロープロファイルカードで、帯域幅 8 GB/s の 8レーン PCI Express バス 1本を使用します。SPARC M12/M10の PCIボックスのアーキテクチャーは高い処理能力の I/Oを提供し、現在はもちろん将来的にも、様々な PCI Express カード（Fibre Channel、InfiniBand、Gigabit Ethernet、Flash PCI Express）の最大データ速度をサポートします。

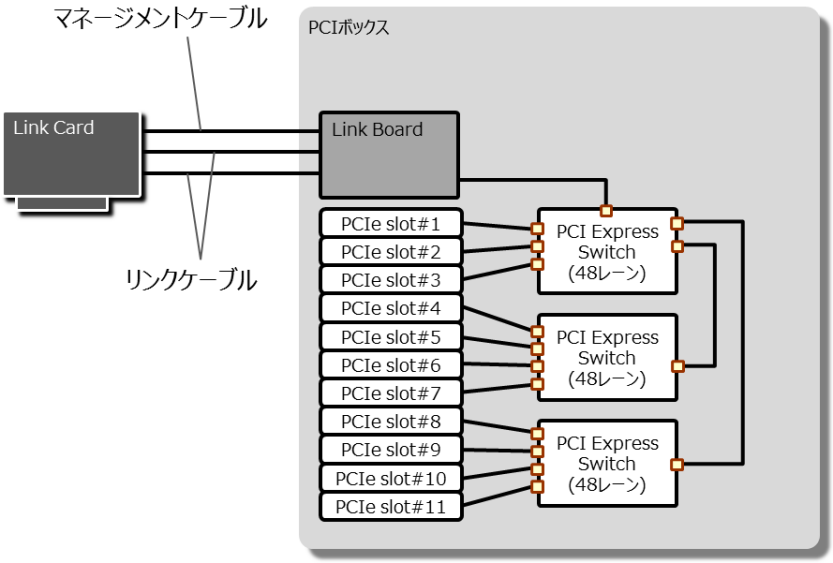

図 4-21. PCI ボックスのアーキテクチャー

PCI ボックスは、サーバの PCI Express スロットに搭載されたリンクカードと、PCI ボックス側に搭載されたリンクボードをケーブルで接続します。このケーブルは、電気ケーブルキットあるいは光ケーブルキットを選択可能です。SPARC M12/M10 では、表 4-3、表 4-4 に示すように、複数の PCI ボックスの接続をサポートしています。

表 4-3. SPARC M12 の PCI ボックス接続可能台数

	CPU構成	OS版数	PCIボックス接続可能台数（最大）
SPARC M12-1	1 CPU	Oracle Solaris 11（*1）	3台
		Oracle Solaris 10（*2）	2台
SPARC M12-2	1 CPU	Oracle Solaris 11（*1）	4台（*3）または 3台（*4）
		Oracle Solaris 10（*2）	2台
	2 CPU	Oracle Solaris 11（*1）	8台（*3）または 6台（*4）
		Oracle Solaris 10（*2）	6台
SPARC M12-2S（1サーバあたり）	1 CPU	Oracle Solaris 11（*1）	4台（*3）または 3台（*4）
		Oracle Solaris 10（*2）	2台
	2 CPU	Oracle Solaris 11（*1）	8台（*3）または 5台（*4）
		Oracle Solaris 10（*2）	5台

*1 Oracle Solaris 11使用時：制御ドメイン、ルートドメイン、I/OドメインのOS版数がすべてOracle Solaris 11の場合。

*2 Oracle Solaris 10使用時：制御ドメイン、ルートドメイン、I/OドメインのOS版数がすべてOracle Solaris 10、または、Oracle Solaris 11とOracle Solaris 10 が混在している場合。

*3 ファームウェア版数がXCP 3040以降の場合

*4 ファームウェア版数がXCP 3030以前の場合

表 4-4. SPARC M10 の PCI ボックス接続可能台数

	CPU構成	PCIボックス接続可能台数（最大）
SPARC M10-1	1 CPU	2台
SPARC M10-4	2 CPU	3台
	4 CPU	6台
SPARC M10-4S（1サーバあたり）	2 CPU	3台
	4 CPU	5台

冷却

SPARC M12-1 と SPARC M10-1 はファンによる空気冷却を行っています。SPARC M12-2/M12-2S、SPARC M10-4/M10-4S は、高性能化するプロセッサの発熱を効率的に冷却する技術を採用しています。

Liquid Loop Cooling（LLC）

SPARC M10-4 と SPARC M10-4S では、Liquid Loop Cooling（LLC）を採用しています。LLC は、高い冷却性能を持つ液体冷却技術と導入・運用が容易な空気冷却技術を組み合わせた冷却方法で、以下の 3 つの機能ブロックで構成されています。

1. 送熱部：配管、ポンプ
2. 受熱部：クーリングプレート
3. 放熱部：ラジエータ

冷媒は送熱部のポンプによって LLC システム内を循環します。そして、受熱部へ送られた冷媒は、CPU や電源部品の発熱をクーリングプレートにより受け取ります。冷媒は、送熱部の配管を経由してラジエータに送られ、筐体前面のファンの冷却風によって冷却されます。そして、配管、ポンプを経由してクーリングプレートに戻ります。

冷媒を筐体内部で循環させることで、外部からの液体供給や液体冷却のためのメンテナンスが不要になり、サーバ導入の容易性や保守性を向上させました。

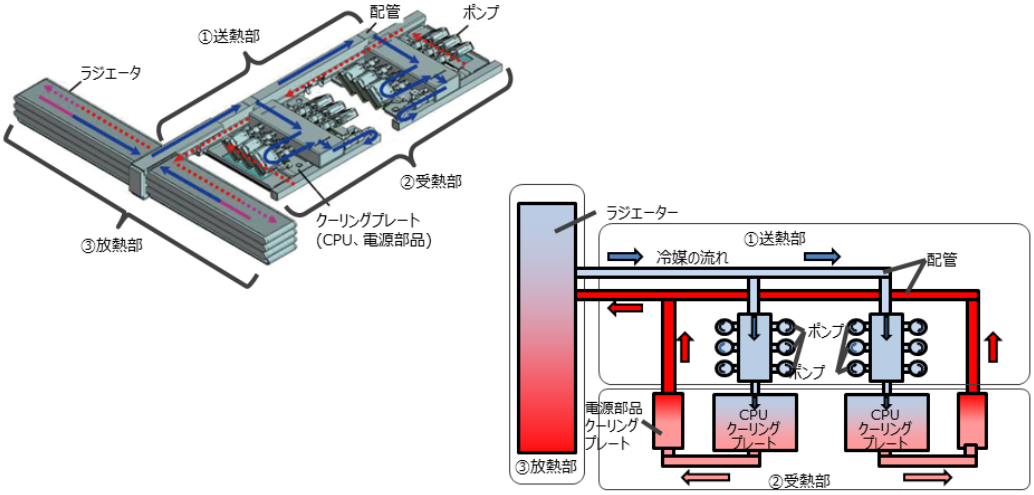

図 4-22. Liquid Loop Cooling の仕組み

Vapor and Liquid Loop Cooling（VLLC）

SPARC M12-2 と SPARC M12-2S では、Vapor and Liquid Loop Cooling（VLLC）を採用しています。VLLC は、LLC に気化冷却を使用する新しい冷却技術です。

LLC と同様、VLLC も 3 つの機能ブロックで構成されています。

1. 送熱部：配管、ポンプ
2. 受熱部：クーリングプレート
3. 放熱部：ラジエータ

LLC と同様、冷媒は送熱部のポンプによって VLLC システム内を循環し、クーリングプレートで受けた熱をラジエータが空気中に放出します。

VLLC は、LLC の空冷技術と水冷技術を組み合わせた冷却方法に加えて、液体が気体に変化する際に熱を奪う気化冷却も利用しています。

本技術により、LLC のメンテナンスフリー、サーバ導入の容易性・保守性を維持しつつ、LLC の約 2 倍の冷却性能を実現しました。

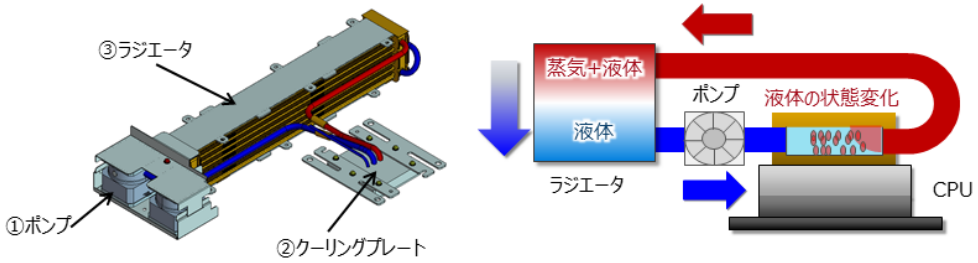

図 4-23. Vapor and Liquid Loop Cooling の仕組み

高効率電源

SPARC M12/M10 は、コンピュータ用電源ユニットの電力変換効率の高さを示す「80PLUS（*1）」認証の電源ユニットを採用しています。「80PLUS PLATINUM」の認証を取得しています。

SPARC M12-2/M12-2S と SPARC M10-4/M10-4S の電源ユニットは、次の技術によって高変換効率を実現しています。

- **低損失部品：SiC ダイオード、Super Junction FET**
  スイッチング電源回路に、シリコンを材料とする従来の電力半導体に代えて、SiC ダイオードや Super Junction FET といった低損失電力半導体を採用し、電力損失を抑えています。

- **高効率回路：ブリッジレス整流回路、スタンバイレス給電回路**
  整流回路と後段の PFC 回路（力率改善回路）を一体制御するブリッジレス整流回路を採用し、電力変換回路の段数を減らすことで高い変換効率を実現しています。また、主電源だけで給電する富士通独自のスタンバイレス給電方式によりスタンバイ電源回路を削減することで、高効率化を実現しています。

- **二次側低損失配線：二次側出力部配線バスバー構造**
  電源ユニットと二次側の電力部品間の接続に、従来のプリント基板に代えて、銅を材料とした直流抵抗を低減するバスバー構造の配線を行うことで、電力損失を抑えています。

*1 80PLUSとは、80PLUS®プログラムが推進する電気機器の省電力化プログラムで、電源ユニットが交流から直流に変換する変換効率が80%以上ということを表しています。
効率ランクは、STANDARD、BRONZE、SILVER、GOLD、PLATINUM、TITANIUMの順に高くなります。
URL : https://plugloadsolutions.com/80pluspowersupplies.aspx

5. システム管理

信頼性、可用性、保守性

ダウンタイムを短縮して、重要なサービスを提供し続けるためには、耐故障性や迅速な復旧が必要です。さらに、短時間で拡張できるような仕組みが盛り込まれている必要があります。その点、SPARC M12/M10は、複雑なネットワークコンピューティングソリューションと、厳しい高可用性要件を満たすことに最も重点をおいて設計されました。特に、システムコンポーネントの冗長化および活性交換対応、システム全体にわたる診断機能とエラー復旧機能、標準装備されているリモート管理機能がサーバの信頼性を支えています。これらの信頼性のあるサーバの高度なアーキテクチャーが、アプリケーションの可用性やハードウェア障害からの迅速な復旧を可能にします。これらは、システム運用を単純化し、企業のコスト削減を実現します。

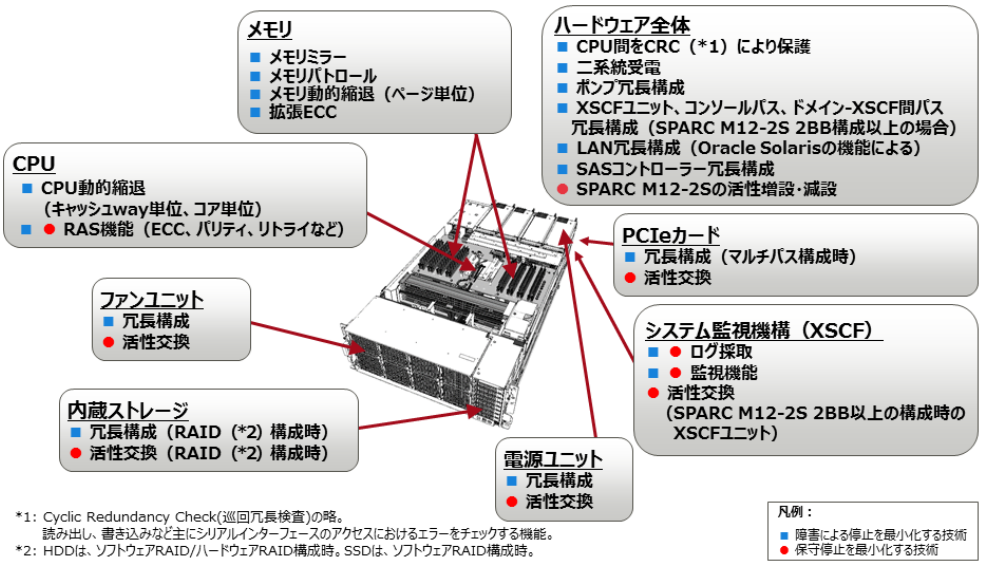

図 5-1. SPARC M12 の信頼性、可用性、保守性

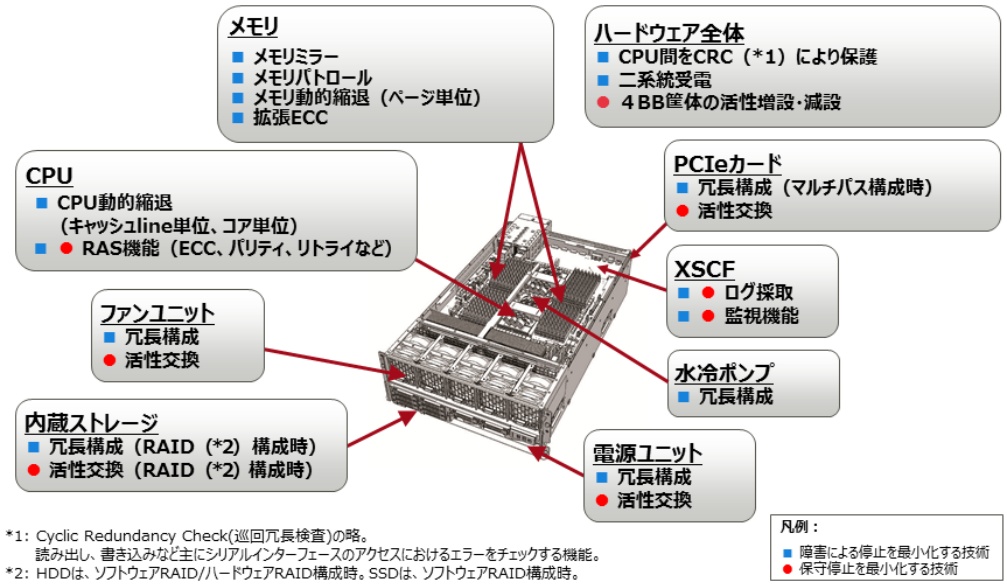

図 5-2.　SPARC M10 の信頼性、可用性、保守性

1.　コンポーネントの冗長化と活性交換

今日の IT 企業には、ノンストップで動き続けるビジネスオペレーションへの対応が求められます。そのため、計画的ダウンタイムを短縮し、場合によっては完全になくすことが求められています。このような要求に応え、システム構成の変更時や、万一のトラブル発生時にもシステムの稼動を継続するために、SPARC M12/M10 ではハードウェアを冗長化し、活性交換を可能にしています。ほとんどの場合、ユーザーやシステム機能に影響することなく、ハードウェア障害から復旧することが可能です。

SPARC M12/M10 は、電源ユニットとファンユニットの冗長構成と活性交換を採用しています。また、SPARC M12-2/M12-2S と SPARC M10-4/M10-4S では PCI Express カードの活性交換も可能です。管理者は、SPARC M12/M10 の活性交換可能なディスクドライブと、ハードウェア RAID もしくはディスクミラーリングソフトウェアを組み合わせて、冗長な内蔵ストレージを構成することができます。SPARC M12-2S または SPARC M10-4S では、さらに、ビルディングブロック構成中の SPARC M12-2S または SPARC M10-4S の活性保守により、ほかの物理パーティションの動作中に CPU やメモリの構成を変更できます。複数の SPARC M12-2S または複数の SPARC M10-4S からなるビルディングブロック構成のシステムでは、SPARC M12-2S または SPARC M10-4S のサービスプロセッサは冗長化されます。さらに、SPARC M12-2S と SPARC M10-4S には縮退可能なクロスバーユニットも用意されています。そのため、障害が発生しても、二重化されているこれらのコンポーネントが運用を継続させます。エラーが生じた場所と状況によ

ってては、システムは縮退モードで動作を継続したり、あるいは障害を自動的に診断し、コンポーネントをシステムから自動的に切り離してリブートしたりすることもあります。さらに、SPARC M12/M10内の活性交換可能なハードウェアはサービスを迅速化します。そして、ほかの物理パーティションを停止せずに、コンポーネントの交換や増設が簡単に行えるようになっています。

2. パーティショニング機能

コストを削減し、管理上の負担を軽減するために、多くの企業でサーバ統合が進行しています。これに伴い、複数のアプリケーションを1台のサーバ上に集約する際のセキュリティと有効性を高めることができるツールが必要とされています。SPARC M12-2SやSPARC M10-4S（以降、1筐体を1ビルディングブロック（BB）と呼びます）では、ビルディングブロック方式により複数のビルディングブロックを高速接続し、ビルディングブロック単位で物理パーティションを作成できます。例えば、複数のビルディングブロックで物理パーティションを構成したときは、物理パーティションを構成しているすべてのビルディングブロック内のハードウェアリソースが使用できます。物理パーティショニング機能によって、最大 16BB 構成で、16 の独立した物理パーティションを構築することも、単一の物理パーティションを構築してすべてのビルディングブロックにまたがる大規模システムを構築することも、その間の構成にすることもできます。Oracle Solarisは、それぞれの物理パーティションで独立して動作します。また、Oracle VM Server for SPARCにより、物理パーティション上に複数の独立した論理ドメインを構築できます。この場合、Oracle Solarisはそれぞれの論理ドメインで独立して動作します。そのため、1つのパーティションやドメイン内のソフトウェア障害は分離され、ほかのパーティションやドメインのオペレーションに影響を与えることはありません。

さらに、SPARC M12-2S と SPARC M10-4S では動的再構成（DR）機能もサポートしています。DR機能とは、OSを停止せずにビルディングブロックを追加することで、CPUやメモリなどのハードウェアリソースを拡張できる機能です。さらに、ハードウェアリソースをパーティションから別のパーティションに移動することもできます。SPARC M12-2S と SPARC M10-4S は、DR機能を使用して、高い柔軟性と可用性を達成しています。DR機能は、次のような場合に利用できます。

- 作業負荷が増えたときに、稼動環境を停止することなくハードウェアリソースを拡張する
- 計画に基づき、稼動環境を停止することなくハードウェアリソースを割り当てし直す（例：四半期の終わりごとに、販売管理用のデータベースソフトウェアがあるパーティションに、別のパーティションから CPU やメモリを移動する）
- 故障した CPU／メモリを修理するために、稼動環境を停止することなく、故障したビルディングブロックを取り外し、交換して接続する

3. 高度な信頼性機能

SPARC M12/M10のコンポーネントには高度な信頼性機能が用意され、プラットフォーム全体の安定性を高めています。周辺ASIC機能をプロセッサ内に集約したことによって、サーバアーキテクチャー内のコンポーネントの数と複雑さを削減させ、信頼性を高めています。さらに、CRCとECC保護メカニズムによりプロセッサのデータパスの整合性も保証されます。そして、自律的なエラー復旧が行われるため、修正措置開始までの時間を短縮し、稼動時間を拡大することができます。また、大容量データパスにおいても、富士通の高速インターコネクト技術によって高度なエラー検出と訂正機能を実現しています。さらにSPARC M12-2SやSPARC M10-4S内の複数の縮退可能なクロスバースイッチは、システムバスを冗長化しているため、可用性の高いシステムを構築することができます。

また、Oracle Solarisの予測的セルフヒーリングソフトウェアが、SPARC M12/M10の信頼性をさらに高めます。このソフトウェアは、CPUやメモリを定期的に監視します。エラーが発生したときには、エラーの状況に応じて、まずプロセッサ内の1つのスレッド、1つのコアあるいはコアすべてを自動的にオフラインの状態にします。そして、繰り返し発生するCPUのソフトエラーを解決します。また、Memory Page Retirement機能では、特定メモリ（DIMM）で訂正可能なエラーが複数発生した場合に、メモリページをオフラインの状態にできます。

4. エラー検出、診断、復旧

SPARC M12/M10は、障害を早期に訂正し、周辺コンポーネントが繰り返しダウンすることを防止するための重要な技術を採用しています。以下に示す、サーバのハードウェアサブシステム内のエラー検出機能と復旧機能は、本質的に信頼性を向上させるアーキテクチャーです。

- エンドツーエンドのデータ保護機能がシステム全体でエラーを検出して訂正するため、データの整合性が保証されます。
- SPARC M12/M10は、最先端の障害隔離機能を使用して、まずコンポーネント境界内のエラーを切り離します。次に、コンポーネント全体ではなく、関連デバイス部分のみをオフライン状態にします。また、チップもしくはチップ内のブロック単位でエラーを切り離して安定性を改善し、最大コンピュータパワーの可用性も維持します。これらの機能は、CPU、メモリ、CPU内のメモリアクセスコントローラー、クロスバーASIC、CPU内のPCI Express I/Oコントローラーに適用されます。
- 常に動作状況を監視し、適切な状態とエラー状態の履歴ログを提供します。
- ホストのウォッチドッグ機能は、OSとアプリケーションが正常に稼動していることを監視するために、プロセッサの動作を定期的にチェックします。ウォッチドッグ

にタイムアウトが発生した場合は、自動的に復旧機能が起動されるようになっています。

- 定期的なコンポーネントステータスチェックを実行します。その結果、多くのシステムデバイスの状態を判断し、障害の兆候を検出します。また、復旧メカニズムを起動して、システムおよびアプリケーションの障害も防ぎます。
- エラーロギング、多段階アラート、電子的な FRU（Field Replaceable Unit）識別情報、システム障害の LED インジケーターは、迅速な問題解決に貢献します。

システム管理

サーバシステムを直接操作してローカルなシステム管理を行うことは、もはや現実的ではありません。24 時間体制のシステムオペレーション、ディザスタリカバリ、地理的に分散しているシステムのリモート管理が主流となっています。富士通サーバにおける数ある利点の 1 つは、完全に自動化されたデータセンターのサポートです。サポートスタッフがネットワークにアクセスし、リモートでのサーバ管理を可能にします。さらに SPARC M12/M10 では、強力なシステム監視機構（XSCF）と XSCF Control Package（XCP）により、ハードウェアへの構成変更を伴わないほぼすべてのタスクを、システム管理者がリモート機能で実行・制御できるようにしています。これらのリモート機能によって、管理上の負荷を軽減し、時間を節約し、運営費を削減します。

1. システム監視機構

システム監視機構（XSCF：eXtended System Control Facility）は、SPARC M12/M10 のリモート監視機能および管理機能の中核ともいえます。XSCF は、サーバシステムに依存せず、XCP を実行する専用プロセッサで構成されています。

XSCF は、定期的にシステム全体の環境センサーを監視しています。その結果、潜在的なエラーの状態を事前に警告し、必要に応じて予防的システム保守手続きを実行します。例えば、物理的なシステム損傷の可能性がある温度状態に応じて、XSCF がサーバのシャットダウンを開始します。また、サービスプロセッサ上で作動する XCP を利用して、管理者がプラットフォームと同様にパーティションとドメインをリモートで制御し、監視することもできます。

XSCF へのネットワーク接続またはシリアル接続を使用すると、管理者はネットワーク上のどこからでもサーバを効果的に管理することができます。また、サービスプロセッサへのリモート接続は、オペレーティングシステムから切り離されて動作します。そして、システムコンソールを完全に制御し、管理権限を付与します。

2. 冗長 XSCF

ビルディングブロック方式で接続された複数の SPARC M12-2S または SPARC M10-4S では、1 つの XSCF はアクティブ用、そしてもう 1 つの XSCF はスタンバイ用として構成されています。この 2 つのサービスプロセッサ間の XSCF ネットワークでは、システム管理情報を容易に交換できます。XSCF に障害が発生した場合、すでにサービスプロセッサが同期されているので、すぐにフェイルオーバーし、役割を変更できるようになっています。さらに、SPARC M12-2S では、故障した XSCF を、もう 1 つの XSCF の正常動作中に活性交換することができます。

3. XSCF Control Package（XCP）

XCP は、SPARC M12/M10 と個々のパーティショニング機能を迅速かつ効果的に制御し、監視することができます。また XCP は、コマンドラインインターフェース（CLI）とウェブブラウザユーザーインターフェースを提供し、管理者とオペレーターがシステムコントローラー機能にアクセスできるようにしています。そして、固有の管理能力のある安全なアカウントは、パーティションやプライマリ論理ドメインコンソールにもセキュリティをもたらします。さらに、XSCF とドメイン間の通信は、Secure Shell（SSH）と Secure Socket Layer（SSL）を基盤とする暗号化された接続をサポートしています。そのため、XCP を使用して入力されたコマンドを安全にリモート実行できます。

XCP は、次の主要サーバ機能に対するインターフェースとなります。

- SPARC M12-2S/M10-4S を物理パーティションに割り当てるなど、物理パーティション管理
- XSCF とパーティションおよびプライマリ論理ドメイン間におけるインタラクションの記録などの監査管理
- サーバ内コンポーネントへの電力の監視および制御
  提示されたハードウェア情報の解釈、および温度上昇あるいは電源トラブルなどの緊急問題の通知
- Oracle Solaris の障害管理アーキテクチャーとの組み合わせによる、正確な障害診断と予測的障害分析を通じた可用性の向上
- OpenBoot PROM（OBP）や power-on self test（POST）など、診断プログラムの実行および監視
- CPU コア アクティベーション操作の実行、ステージング機能の提供による追加された処理リソースのアクティブ化
- 冗長 XSCF 構成における障害の監視、必要に応じた自動フェイルオーバーの実行

4. ロールベースのシステム管理

XCP を活用すると、複数のシステム管理者とオペレーターが、複数の物理パーティションを個別に管理できます。そして、単一の SPARC M12/M10 として連携させることもできます。この管理ソフトウェアは、グループに編成された複数のユーザーアカウントをサポートします。そして、編成された各グループには異なる権限が割り当てられます。その権限は、特定のハードウェアセット（物理コンポーネント、物理パーティション、または物理パーティション内の物理コンポーネントなど）に関する特定の行動セットを実行することができます。1 人のユーザーが、多数の物理パーティション上で、複数の異なる権限を所有することもできます。

5. Enhanced Support Facility

Enhanced Support Facility は、XSCF と連携して SPARC M12/M10 の運用管理性や保守性を向上させるソフトウェアです。

このソフトウェアでは、本体装置の構成、状態の表示や故障情報を表示することができます。また、OS 上からディスクや電源、PCI カードなどの状態を監視し、トラブルが発生すると検出された情報を XSCF に通知します。

また、システム情報の一括採取、/etc/systems ファイルの設定チェック、サーバの電源投入・切断のスケジュール運用、ディスクの活性保守手順などの表示することも可能です。

6. 高可用性を実現する PRIMECLUSTER

システム監視機構（XSCF）と連携したサーバ強制停止機能により、確実なサーバの停止と高速な切り替えによる高品質なクラスタシステムを実現しています。

Oracle VM Server for SPARC による仮想環境においても、ゲストドメイン単位での切り替えと、制御ドメインを含めたサーバ単位での切り替えが可能です。

PRIMECLUSTER は、Symfoware や Oracle RAC などの並列データベースと組み合わせて利用できます。そして、切り替え（FailOver）型クラスタシステム、シングルノードクラスタシステムだけでなく、非クラスタのディスクミラー運用、ネットワーク二重化運用にも対応しています。

7. FUJITSU Software Systemwalker Centric Manager

FUJITSU Software Systemwalker Centric Manager は、システム運用のライフサイクル（導入／設定、監視、復旧、評価）に従い、ソフトウェア資源の配付、システムやネットワークの集中監視、リモートからのトラブル復旧などの優れた機能で運用管理作業を軽減します。

また、このライフサイクル管理によりマルチプラットフォーム環境やインターネット環境など、最新のビジネス環境におけるシステムの統合管理、運用プロセスの標準化（ITIL）、運用セキュリティの統制を支援します。

8. Oracle Enterprise Manager Ops Center 12c

Oracle Enterprise Manager Ops Center 12c は、多数の SPARC M12/M10 の物理サーバと仮想サーバを一元管理するシステム管理ソフトウェアです。システム管理のコストを削減します。Ops Center を使用することで、CPU、メモリなどハードウェアリソースの割り当て、物理サーバと仮想サーバの構築、システムファームウェアのアップデート、Oracle Solaris 11 OS のパッチ適用が簡単にできます。

エコ

SPARC M12/M10 は、未使用のハードウェアや使用率の低いハードウェアの無駄な消費電力をカットする、省電力機能を備えています。初期導入時は、節電よりも性能を優先するように、省電力機能は無効に設定されていますが、性能よりも節電を優先する場合は省電力モード（Elastic Mode）に設定することができます。省電力モードは、物理パーティションごとに設定できます。

さらに SPARC M12 では、省電力モードに加えて、性能と電力の両者をバランスさせたパフォーマンスモード（Performance Mode）に設定することができます。パフォーマンスモードは、物理パーティションごとに設定できます。

SPARC M12/M10 は次のような省電力機能を備えています。

- **ハードウェア部品の低消費電力化**
  SPARC M12/M10 は、低消費電力を考慮したハードウェア部品を選定し、設計しています。

- **未使用ハードウェア部品の消費電力削減**
  システムに実装されるハードウェアには、バージョン、物理パーティション構成、論理ドメイン構成によっては、使用しないハードウェアがあります。例えば、CPU がパーティションやドメインに割り当てられていない場合、その CPU は自動的に省電力状態になります。同様に、割り当てられていないメモリも自動的に省電力状態になります。

- **センサー監視機能**
  センサー監視機能を活用して、消費電力とエアーフローの監視と記録を行います。

実際の電力消費データを収集することにより、データセンターの電力容量の設計を最適化することが可能です。同様に、エアーフローデータを収集することにより、データセンターの冷却設備の最適化も可能です。

- power capping 機能

  power capping 機能を活用すると、システムの消費電力の上限値を設定することができます。設定した上限値を超えないように CPU の周波数が自動的に制御されます。この機能によって、データセンターの施設に合ったシステムの消費電力を制御できるようにもなります。

6. Oracle Solaris 11 オペレーティングシステム

Oracle Solaris は、パフォーマンス、スケーラビリティ、信頼性、セキュリティに優れた、業界標準の UNIX オペレーティングシステムです。

Oracle Solaris は、高信頼性システムに適した設計になっています。カーネルをコンパクトに構成しているため、オペレーティングシステムの障害や、障害によるプラットフォームのダウンタイムの発生を抑えることができます。また Oracle Solaris は、カーネルや共有ライブラリ、アプリケーションを明確に区別することができます。これは、アプリケーションの障害による影響を限定的にするためです。さらに、Oracle Solaris は、逐次更新されるソフトウェアを、システムを再起動せずにインストールすることもできます。そのため、企業や団体のシステムのアップタイムを増やし、メンテナンスによる負荷を軽減します。

Oracle Solaris 11 は、IPS（Image Packaging System）、BE（Boot Environment）、強化されたセキュリティ、ネットワークの仮想化、Oracle Solaris 10 ゾーンなどの最先端の機能を提供します。これらの機能が、システムの導入や運用効率の向上、安全で柔軟な仮想環境の構築、投資保護の強化を実現し、ミッションクリティカルなシステムにおけるクラウドコンピューティングのための強固な技術基盤を提供します。

SPARC M12/M10 は、ベアメタルの Oracle Solaris 10 と Oracle Solaris 11 環境をサポートしていますので、お客様資産の投資対効果を高め、投資リスクを最小にします。

Oracle Solaris ZFS

ストレージ仮想化機能として、Oracle Solaris ZFS が標準装備されています。Oracle Solaris ZFS では、複数の物理ストレージデバイスをストレージプールで管理します。ストレージプールから必要な領域を割り当てることで、仮想化されたボリュームを作成することができます。

Oracle Solaris ZFS は、より効率的で最適化されたストレージデバイスの利用を実現し、同時に、信頼性と拡張性を飛躍的に向上させます。そして、物理ストレージは、ストレージプールに対して動的に追加、取り外しが可能です。その際、サービスを中断する必要はありません。そのため、今までにないレベルの柔軟性、可用性、パフォーマンスを提供します。

Oracle Solaris ZFS は、256 ビットのチェックサムによって全データを保護します。つまり、99.999999999999999%のエラー検出と訂正を行います。また、冗長化したストレ

ージプールでエラーを検出した場合は、自動的に破損データを修復します。そのため、費用も時間も消費するハードウェア障害やソフトウェア障害によるデータの損失からデータを保護します。そして、ファイルシステムに関わる作業中に管理者がミスを起こす可能性を減らすことによって、高い可用性の実現に貢献します。

Oracle Solaris ZFS は、128 ビットのファイルシステムですが、実用上は無限大の容量を管理することができます。Oracle Solaris ZFS を管理するためのメタデータは、必要に応じて動的に割り当てることができ、ファイルシステム数やファイル数の制限がありません。従来のファイルシステムでは、ファイルシステムのサイズが物理デバイスのサイズに制限されていました。ところが、Oracle Solaris ZFS では、ストレージプールで物理デバイスを隠ぺいするため、サイズは特定の物理デバイスに制限されません。Oracle Solaris ZFS は、初期化せずにファイルシステムの階層を簡単に作成できます。そして、ZFS ストレージプールに割り当てられたディスク容量の範囲で自動的に拡張します。

Oracle Solaris ZFS ストレージプールは、物理ディスクを集約して管理する仕組みです。非冗長、ミラー、RAID-Z（シングルパリティ）、RAID-Z2（ダブルパリティ）、RAID-Z3（トリプルパリティ）から冗長構成を選択することができます。ストレージプールに書き込まれたデータはすべて、使用可能なデバイス上に動的にストライプ化されます。また、冗長構成のストレージプールでは、不正なデータブロックを検出すると、自己修復するために、別の冗長コピーから正しいデータを取得します。

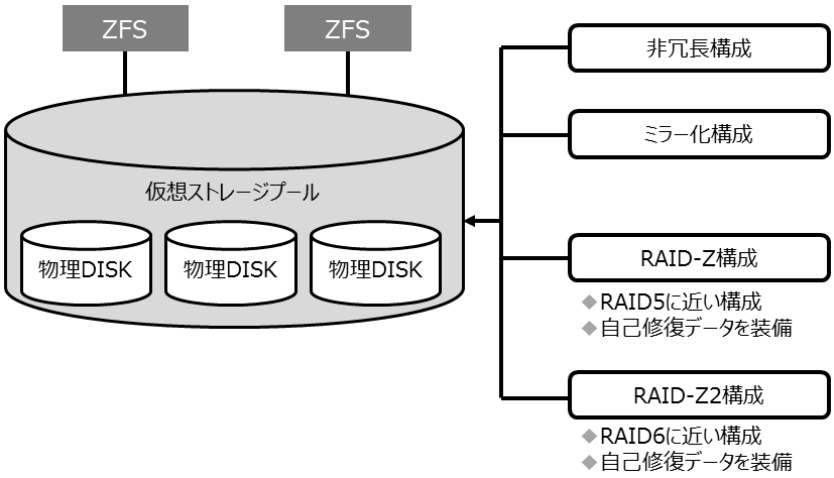

図 6-1. ZFS ストレージプール

ZFS の重複排除機能によって、ZFS ストレージプール内で重複するデータブロックを共有することで、格納するデータ量を削減してストレージをより効率的に使用できます。また、この重複排除機能は、圧縮機能や暗号化機能と組み合わせて使用することもでき

ます。

ストレージプールは、オンラインの状態でディスクの増設や交換が可能ですので、初期導入時に将来使用分のディスク容量を確保する必要がありません。

また、性能向上と消費電力削減を目的として開発されたのが、ZFS ハイブリッドストレージプールです。これは、1 つのストレージプール内にメモリ、SSD、ディスクを組み合わせたものです。ZFS インテントログ（ZIL : ZFS Intent Log）は、通常ストレージプール内の領域を使用します。そして、SSD などの高速デバイスへの割り当てによって、同期書き込みの性能が向上する仕組みになっています。また、メモリとディスクの間に SSD などの高速デバイスをキャッシュとして追加すると、ZFS キャッシュデバイス（L2ARC : Level 2 Adaptive Replacement Cache）は、静的データのランダムな読み込みの性能を向上することができます。

Oracle Solaris ZFS は、トランザクションファイルシステムです。このシステムでのデータ書き込みは、上書きせずに、元データをコピーして更新（コピーオンライト）します。そして、一連のデータ更新処理の完了後に新旧データへのポインタを切り替えます。これらによって、ファイルシステムの整合性を維持することができます。また、突然の電源切断のような場合でも、ファイルシステムが破壊されることはありません。

図 6-2 の(1)～(3)の処理により、突然のシステム中断が発生した場合でもデータが失われることなく、古いデータまたは変更後のデータに常にアクセスできるようになります。

(1) Data を変更する場合、上書きを行わず、新しい領域にコピーして変更を実施する。
(2) Data の変更後、管理情報も Data と同様にコピーして変更を実施する。
(3) uberblock を新しい管理情報に書き換え、変更後のデータにアクセスを行う。

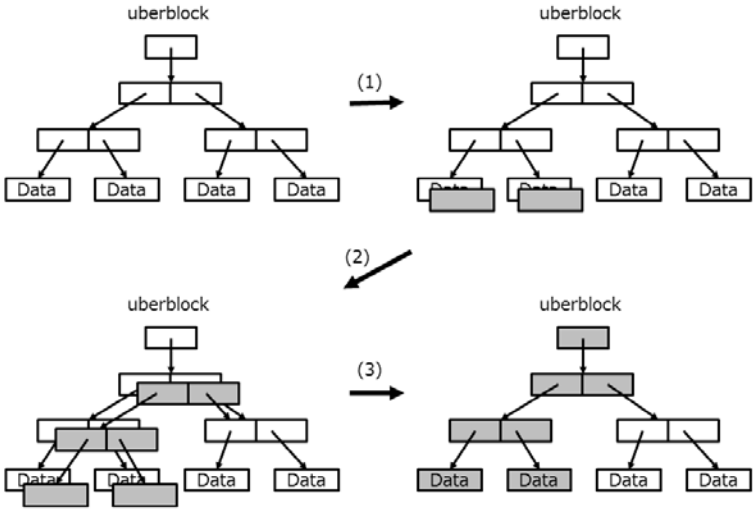

図 6-2. コピーオンライト

ファイルシステムを作成すると、Oracle Solaris ZFS は自動的にマウントポイントを生成します。つまり、サーバ起動時に自動的にマウントされるので、マウントの管理が必要ありません。また、ZFS ボリュームは、ブロックデバイスとしてボリュームを作成するときに初期サイズの領域が予約され、使用量に応じて自動拡張されます。

Oracle Solaris ZFS は、すべてのユーザーデータおよび管理情報のメタデータに、エンドツーエンドのチェックサムを持っています。例えば、データブロックのチェックサムは親ブロックに保持されています。そして、トップの管理ブロック（uberblock）まで順に続くので、データツリー全体の自己検証が可能です。また、異常を検出したときは、冗長コピーからデータを復旧します。そして、信頼性を高めるために、Oracle Solaris ZFS のメタデータは異なるディスクにまたがって何度か自動的に保存されます（ditto ブロック）。

さらに、Oracle Solaris ZFS では、ユーザーデータのコピーを複数保存することもできます。もし、複数ディスクで冗長化できない場合でも、ディスクブロックの読み取り障害から復旧することができます。

ZFS スナップショットは、ファイルシステムまたはボリュームの読み取り専用コピーです。コピーする場合は、ディスク領域を消費せずに瞬時に作成することができます。スナップショットは直接参照できません。しかし、クローンやバックアップなどは可能です。例えば、ZFS クローンは、ファイルシステムまたはボリュームの書き込み可能なコピーです。これは、スナップショットから作成することができます。そしてスナップショットと同様、ディスク領域を消費せず瞬時に作成されます。クローンはデータを変更した分だけディスク領域を消費します。

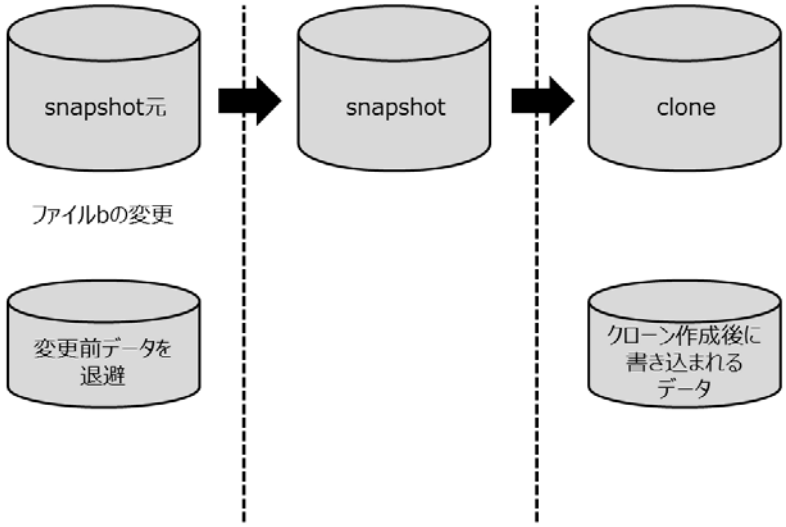

図 6-3. ZFS スナップショット

さらに、物理ストレージの盗難や外部からの侵入者による攻撃から保護するために、ZFS データセットごとに暗号化することもできます。暗号化により、ストレージ内のセキュアな削除を実現できます。

Oracle VM Server for SPARC

Oracle VM Server for SPARC は、独立した Oracle Solaris 環境が稼動する論理ドメインを構成するために、物理サーバをファームウェア層のハイパーバイザにより仮想サーバに分割します。Oracle Solaris 11 または 10 のいずれかで実行可能な Domain Manager により、CPU、メモリ、I/O デバイスは柔軟に割り当てられます。

論理ドメインは、CPU、メモリおよび I/O デバイスを論理的にグループ化した仮想サーバです。論理ドメインごとに、独立した Solaris OS が稼動します。各論理ドメインは、個々に起動、停止できます。論理ドメインは、物理パーティションあたり最大 256 まで作成することができます。そして、ハイパーバイザの論理ドメインチャネル（LDC : Logical Domain Channel）経由で相互に通信します。また、ディスクやネットワークなどの仮想デバイスは、LDC で仮想サービスと通信して、物理デバイスにアクセスします。

仮想 CPU、メモリ、仮想 I/O デバイスなどのリソースは、論理ドメインが稼動中であっても動的に再構成することができます。そして、論理ドメインの仮想 CPU 数は、動的なリソース管理ポリシーに従って自動的に増減させることができます。この動的なリソース管理ポリシーは、使用可能なリソース数、使用率、上限、下限、期間などを組み

合わせて作成することができます。さらに、CPU 電源管理では、業務負荷の変化により使用されていない CPU の電源を切断して、消費電力を削減します。

業務負荷の変動やサーバ増設などに伴い、ゲストドメインを別の物理サーバへ移行させることが可能です。ライブマイグレーションを利用して、ゲストドメインの動作を停止せずに、メモリ内容を圧縮して高速転送できます。

また、サーバを統合する場合は、Physical-to-Virtual（P2V）移行ツールで物理サーバを論理ドメインに移行できます。このツールは、物理サーバの構成情報を収集し、ファイルシステムイメージを作成します。次に、この収集した構成情報から論理ドメインを作成し、仮想ディスクにファイルシステムイメージを復元します。

Oracle Solaris ゾーン

Oracle Solaris ゾーンは、1 つの OS 空間を仮想的に分割します。その分割は、まるで複数の OS 空間が稼動しているように見えます。このほか、CPU やメモリなどのハードウェアリソースを柔軟に割り当てる Oracle Solaris Resource Manager の機能も備えています。

Oracle Solaris ゾーンは、仮想化された OS 環境で、アプリケーションの実行に適した安全で隔離された環境を実現します。また、ゾーンごとに実行されるプロセスは分離されるため、ほかのゾーンへの影響はありません。

グローバルゾーンは、Oracle Solaris システム内に存在する唯一のゾーンです。このゾーンがシステム全体を管理しています。そして、ノングローバルゾーンの作成や管理、物理 I/O デバイスの割り当てなども、グローバルゾーンでしかできません。

ノングローバルゾーンは、仮想 Solaris 環境のソフトウェアパーティションです。ほかのゾーンに影響することなくアプリケーションを実行します。また、最大 8,191 までのノングローバルゾーンを作成できます。各ゾーンは許可されたファイルシステム、許可された物理 I/O デバイスしか使用できません。

ノングローバルゾーンを構成するシステムファイルでゾーンを作成する場合は、必要なパッケージを選択してインストールすることができます。グローバルゾーンでパッケージが更新された場合は、これらのファイルも更新され、また、すべてのノングローバルゾーンのファイルも同期して更新されます。

Oracle Solaris Resource Manager は、リソースの使用状況を定期的に監視します。そして、ゾーンを停止せずに設定されたリソース量の範囲内で自動的にリソースを割り当てます。ゾーンに割り当てる CPU リソースは、リソースプールで管理されます。メモリは、資源上限デーモンで制御されます。さらに、リソースプールは、CPU をグループ化するプロセッサセットと CPU の配分を制御するスケジューリングクラス（タイムシェア

リング、公平配分）で構成されます。

既存のゾーンをコピーして、新しいゾーンを容易に作成することもできます。ゾーンのコピーは、ZFS クローンを使用することで瞬時にでき、使用するディスク領域も節約できます。

物理サーバで CPU やメモリなどのリソースが不足する場合、あるいはゾーンの組み合わせや使用方法を変更する場合は、ゾーンを別サーバに移行することができます。移行元サーバでゾーンを切り離し（detach）、移行先サーバに組み込みます（attach）。もし、両方のサーバでパッケージ構成などの環境が異なっていたとしても、組み込み時には、ゾーンを構成するシステムファイルが移行先サーバのグローバルゾーンと同期されます。

Oracle Solaris 11 では、Oracle Solaris 10 ゾーンを標準装備しています。

P2V（Physical-To-Virtual）機能や V2V（Virtual-To-Virtual）機能を活用すると、既存の Oracle Solaris 10 環境を Oracle Solaris 11 のゾーン上に、そのまま移行することができます。つまり、1 台のサーバに Oracle Solaris 10 と Oracle Solaris 11 の環境を集約できます。これにより、既存の ICT 資産の投資保護や TCO 削減に貢献します。

仮想ネットワーク機能を活用すると、NIC（Network Interface Controller）を仮想化することができます。そして、複数の Oracle Solaris ゾーンに VNIC（Virtual Network Interface Controller）を割り当てることができます。こうすることで、複数の Oracle Solaris ゾーンによる独立したネットワーク環境が構築でき、使用するサーバや NIC の数を削減することができます。さらに、仮想スイッチ（etherstub）が作成されるので、ハードウェアに依存しない仮想ネットワーク環境も実現でき、スイッチ数が削減できます。加えて、リソース管理機能を活用すると、VNIC の帯域幅制限を設定することもできます。

Oracle Solaris 11.3 では、Oracle Solaris カーネルゾーンをサポートしています。Oracle Solaris カーネルゾーン機能は、ゾーン内の完全なカーネルおよびユーザー環境を提供し、さらにグローバルゾーン（カーネルゾーンホスト）とほかのゾーン間でカーネルの分離を強化します。

Image Packaging System

IPS（Image Packaging System）は、OS 環境をパッケージ単位で管理、インストール、更新、削除できるフレームワークです。

IPS は、基本的なサーバ運用に最適な OS 環境をメディアからインストールした後、業務に必要な追加パッケージをネットワーク経由でインストールします。管理者は、独自にネットワークインストールサーバを用意する必要はなく、リポジトリサーバを利用できます。また IPS は、パッケージの複雑な依存関係を自動的に判断します。そして、従

来のパッチ適用もパッケージの入れ替えでできるので、依存関係を意識することなく更新し、適用漏れも防ぎます。

このように、IPS は導入および運用管理の効率化を図っており、直接お客様のコスト削減につながります。

Boot Environment

BE（Boot Environment）は、複数のブート環境を管理します。また、オンラインアップグレードを容易に行うことができる機能です。

Oracle Solaris 11 OS のルートファイルシステムは Oracle Solaris ZFS です。パッケージの追加や更新を行う場合は、ZFS スナップショットや ZFS クローンで、短時間にブート環境を複製することができます。また ZFS ツールは、追加や更新に必要なデータブロックだけをコピーするため、ディスク容量の増加も抑えます。そして、複製されたブート環境からのリブートはもちろんのこと、パッケージの更新で不具合が発生した場合には、従来のブート環境に戻すことが可能です。

BE の活用がブート環境の世代管理を可能にします。さらに、保守作業のための業務停止時間も大幅に短縮します。

7. SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサの技術拡張

マイクロアーキテクチャー

1. チップ構成

SPARC64 X+/SPARC64 X の半導体は、28 nm CMOS プロセスを採用、1 プロセッサあたり最大 16 コアを搭載し、最大 24 MB の 24way の共有 2 次キャッシュ（SPARC M10-1 は 22way 共有 2 次キャッシュ）を備えています。SPARC64 X+の動作周波数は最大 3.7 GHz、SPARC64 X では最大 3 GHz です。

一方、SPARC64 XII の半導体は、20 nm CMOS プロセスを採用、1 プロセッサあたり最大 12 コアと、最大 32MB の 16way の共有 3 次キャッシュを搭載しています。SPARC64 XII の動作周波数は最大 4.25 GHz です（表 7-1 参照）。

表 7-1. SPARC64 XII/SPARC64 X+/SPARC64 X の仕様

	SPARC64 XII	SPARC64 X+	SPARC64 X
コア数	最大12	最大16	
ラストレベルキャッシュ	最大32 MB（3次キャッシュ）	最大24 MB（2次キャッシュ）	
動作周波数	最大4.25 GHz	最大3.7 GHz	最大3 GHz
プロセステクノロジー	20 nm CMOS	28 nm CMOS	
ダイサイズ	25.8 mm x 30.8 mm	24.0 mm x 25.0 mm	23.5 mm x 25.0 mm
トランジスタ数	54億5000万	29億9000万	29億5000万
メモリ帯域幅（理論ピーク値）	153 GB/s	102 GB/s	

SPARC64 XII を搭載した SPARC M12-2S では、CPU の高速モード（*1）を有効にすることで、CPU 動作周波数が 4.25 GHz よりも高くなります（最大 4.35 GHz）。

*1 XSCF コマンド（sethsmode）により設定可能です。高速モードに設定した場合のCPU動作周波数は、最大の4.35 GHzを保証するものではありません。

SPARC64 XII/SPARC64 X+/SPARC64 X にはメモリアクセスコントローラー（MAC）が内蔵されています。これにより、メモリアクセスの低レイテンシと高い処理能力を実現

し、メモリの帯域幅は理論ピーク値で 153 GB/s（SPARC64 XII）または 102 GB/s（SPARC64 X+/SPARC64 X）です。

さらに、SPARC64 XII/SPARC64 X+/SPARC64 X には CPU 間をつなぐためのインターフェースが内蔵されています。そのため、SPARC M12-2/M12-2S/M10-4/M10-4S 内にある複数の CPU はインターコネクト接続しています。この CPU 間のインターコネクトには高速シリアルインターフェースを採用し、複数の CPU にまたがる処理でも高い処理能力を発揮します。

SPARC64 XII/SPARC64 X+/SPARC64 X のシステム・オン・チップ機能を最大限に発揮させるために、シリコンには I/O コントローラーも組み込まれています。SPARC64 X+/SPARC64 X ではプロセッサごとに 8 レーンの PCI Express の 8 GB/s ポート 2 個が実装されており、SPARC64 XII ではプロセッサごとに 8 レーンの PCI Express の 8 GB/s ポート 4 個が実装されています。

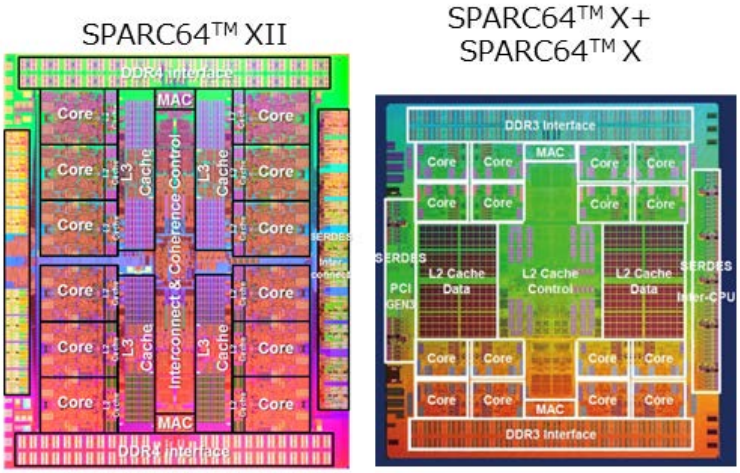

図 7-1. SPARC64 XII/SPARC64 X+/SPARC64 X のフロアプラン

2. コアのマイクロアーキテクチャー

SPARC64 XII/SPARC64 X+/SPARC64 X のコア

SPARC64 X+/SPARC64 X のコアは、命令フェッチ部と命令実行部をそれぞれ 1 つ持っています（図 7-2）。命令フェッチ部には命令 1 次キャッシュがあり、命令実行部にはオペランド用のデータ 1 次キャッシュ、演算器、レジスタなどがあります。

一方、SPARC64 XII のコアは、2 次キャッシュを内蔵しています。さらに命令フェッチ部 1 つに対して、命令実行部（以下、命令パイプライン）を 2 つ持っています（図 7-3）。

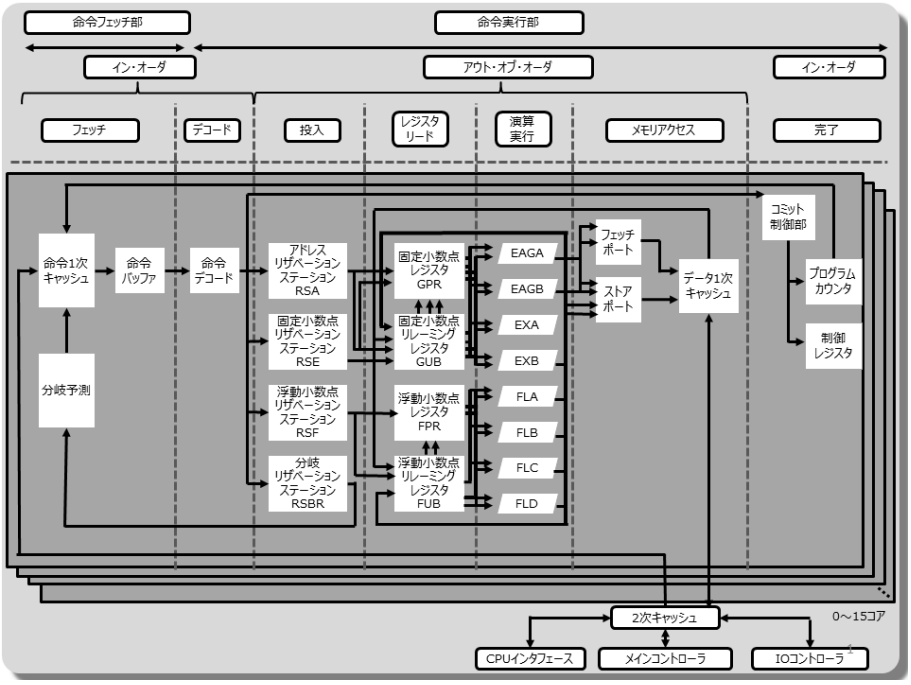

図 7-2. SPARC64 X+/SPARC64 Xのパイプライン

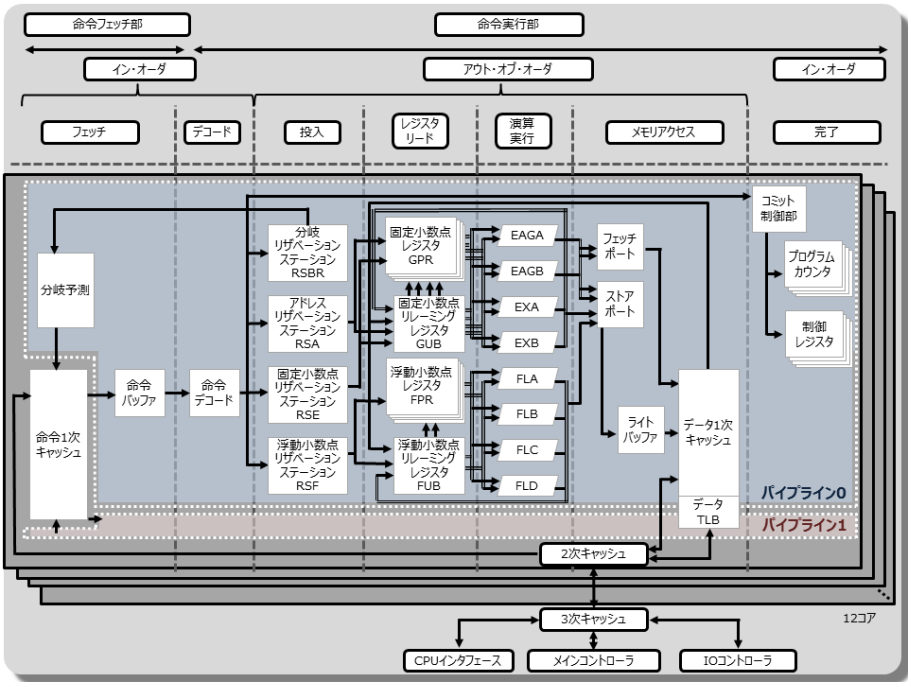

図 7-3. SPARC64 XII のパイプライン

同時マルチスレッディング（SMT：Simultaneous Multi-Threading）

SPARC64 XII/SPARC64 X+/SPARC64 X は SPARC64 VII+で初めて採用された同時マルチスレッディング（SMT）テクノロジーを継承しています。この技術によって、1 つのコアで複数のスレッドを同時に処理することができます。各スレッドは、ソフトウェアから見ると、独立した CPU として認識されます。そして、ハードウェアは命令バッファ、リザベーションステーション、パイプライン、キャッシュのような実行リソースをスレッド間で共有します。例えば、1 つのスレッドがデータ待ちでストールしている場合は、別のスレッドがパイプラインで処理を続行させます。したがって、コアとしての処理能力も向上します。また、1 つのスレッドがアイドル状態の場合は、別のスレッドが実行リソースを占有して処理を実行することも可能です。

SPARC64 X+/SPARC64 X では、1 つのコアあたり最大 2 個のスレッドを同時に実行可能です。SPARC64 XII は命令パイプラインあたり最大 4 個、1 つのコアあたり最大 8 個のスレッドを同時に実行可能です。

次の各項では、命令フェッチ部と命令実行部について説明します。

命令フェッチ

命令フェッチ部は、分岐予測に従って一連の命令をメモリまたは命令 1 次キャッシュから読み出します。命令フェッチが一度に読み出せる命令は、SPARC64 X+/SPARC64 X では 8 命令、SPARC64 XII では命令パイプラインごとに 8 命令、コアごとに 16 命令です。読み出された命令列は命令バッファに一度に取り込まれます。SPARC64 XII では命令バッファは命令パイプラインごとに分割されています。命令パイプラインで複数スレッドが稼動している場合は、命令バッファをスレッドごとに均等に分割して使用します。また、命令バッファがフル状態になった場合、または命令 1 次キャッシュでミスが発生した場合には、ハードウェア命令プリフェッチに切り替えると処理を継続することができます。このハードウェア命令プリフェッチは、2 次キャッシュまたはメモリから 1 次キャッシュに後続の命令をロードします。

命令フェッチは命令実行部とは関係なく稼動します。したがって、命令実行が滞った場合でも、利用可能な命令バッファがあるかぎり命令フェッチは継続されます。そのため、キャッシュミスなどが発生した場合でも、命令バッファ内に命令があるかぎり命令バッファから命令実行部への命令供給も継続されることになります。

命令実行の処理能力は、命令パイプラインごとに 1 サイクルあたり 4 命令（後述する前置命令 SXAR を含む場合は最大 6 命令）です。一方、命令フェッチは命令パイプラインごとに 1 サイクルあたり 8 命令です。これにより、命令 1 次キャッシュのアクセスレイテンシを隠ぺいし、システム性能を向上します。

SMT 使用時は、1 つのスレッドの命令フェッチは同じサイクルに行われます。その際、サイクルごとにスレッドを切り替えます。

また、SPARC64 XII/SPARC64 X+/SPARC64 X では分岐予測機構を強化しました。自命令の分岐履歴だけではなく、直前の命令列の複数の分岐命令が分岐したか否かの履歴パターンにも十分考慮した分岐予測を行うため、予測精度も向上します。

命令実行

SPARC64 X+/SPARC64 X では、命令フェッチ部で命令バッファに格納された命令は、1 サイクルあたり 6 命令（うち 2 命令は後述する前置命令 SXAR）までが命令実行部に供給されます。供給された命令は、命令実行部内で、デコード、発行、実行、コミットの順序で処理されます。

SPARC64 XII では命令パイプライン（命令実行部）をコアあたり 2 つ実装しており、命令フェッチ部で命令バッファに格納された命令は、命令パイプラインそれぞれに 1 サイクルあたり 6 命令までが供給されます。供給された命令は、命令パイプラインごとに独立してデコード、発行、実行、コミットの順序で処理されます。これ以降の命令実行の説明では 1 つの命令パイプラインの動作に着目して解説します。

(1) 命令デコードおよび発行

命令デコードおよび発行ステージでは、前置命令 SXAR を含む場合、6 命令（うち 2 命令は SXAR）までを同時にデコードします。前置命令 SXAR は、直後の 2 命令分の拡張情報を持つ新規命令です。命令デコードステージにて後続する 2 つまでの命令に拡張を行います。さらに、後のパイプラインステージでは、実行リソースを消費しないように、直後の命令と「結合処理」も行っています。次の処理が実行されると、SXAR により拡張された命令は「有効命令」として扱われます。

- 命令拡張の SXAR 前置命令に続く命令は、「結合処理」を実行します。
- 上記以降の命令も、SXAR で拡張された場合は「結合処理」を実行します。

命令デコードステージでは、4 つまでの有効命令を同時にデコードします。

SMT 使用時は、SPARC64 X+/SPARC64 X の 2 つ、SPARC64 XII の 4 つのスレッドのうち、いずれか 1 つのスレッドの命令デコードが同じサイクルで行われます。その際、サイクルごとにスレッドを切り替えます。

命令デコードステージでは、各種リザベーションステーション、フェッチポート、ストアポート、レジスタ更新バッファなどの実行に必要なリソースを決定します。SPARC64 XII/SPARC64 X+/SPARC64 X は、命令処理を高速化するために、リザベーションステーションなどの実行に必要なリソースを、前機種の SPARC64 VII+プロセッサよりも増強しています。

同時に命令デコードを実行する場合でも、命令種類の組み合わせに制限はありません。つまり、リソースの空きさえあれば命令の発行は可能です。また、4命令分も空きがない場合でも、プログラム上の命令順にできるだけ多くの命令を発行します。このような命令発行における停滞の排除によって、どのようなバイナリコードでも常に高い並列度が保証されます。

(2) 命令実行

デコードされた命令はリザベーションステーションに登録されます。そのうち、固定小数点命令の加減算、論理演算はRSE（Reservation Station for Execution）あるいはRSA（Reservation Station for Address generation）に登録可能です。その他の固定小数点命令はRSEのみの登録となります。また、ロード命令、ストア命令のアドレス計算はRSAに登録されます。つまり、RSA は固定小数点命令の加減算、論理演算とロード命令、ストア命令のアドレス計算で共用されています。このほか、浮動小数点演算命令はRSF（Reservation Station for Floating-point）、分岐命令はRSBR（Reservation Station for BRanch）に登録されます。

RSEは演算パイプラインEXA と EXBを起動します。RSAは EAGA と EAGBのパイプラインを起動します。RSFは FLA、FLB、FLC、FLDのパイプラインを起動します。リザベーションステーションに格納された命令は、アウトオブオーダで対応する演算器に投入されます。このとき、入力オペランドが用意できたものの中からプログラム順に古いものが優先されます。

SMT使用時は、複数のスレッドでパイプラインを同時に使用することができます。

演算部は、2つの固定小数点演算パイプライン（EXA/B）とロード、ストアのアドレス計算、固定小数点の加減算・論理演算を行う 2 つの演算パイプライン（EAGA/B）、そして、4つの浮動小数点演算パイプライン（FLA/B/C/D）を装備しています。

特に、浮動小数点演算パイプラインは前機種に比べても大幅に拡張されています。例えば、SPARC64 VIIIfx スーパーコンピュータプロセッサで新規に導入した SIMD（Single Instruction Multiple Data）機能の採用です。この機能は、1つの命令で 2つのパイプライン（FLA と FLC、あるいはFLBと FLD）を用いて並列処理を行います。各パイプラインには、従来と同様に、浮動小数点積和（FMA：Floating-point Multiply and Add）演算器を備えています。1つの FMA 演算器で、浮動小数点の乗算と加算を1サイクルで処理できます。さらに、各コアでは、8個の倍精度浮動小数点演算を1サイクルで行うことができます。

後述するように、ソフトウェア・オン・チップ（SWoC：SoftWare on Chip）では、暗号化処理や、IEEE754 標準 10進浮動小数点数 DPD（Densely Packed Decimal）形式および Oracle NUMBER 形式の演算をサポートします。また命令実行部では、演算器をコアに追加拡張しています。ビジネスアプリケーション（例えばデータベース）向けの演算を

並列処理することによって、高速化を図っています。

SPARC64 XII/SPARC64 X+/SPARC64 X では、浮動小数点レジスタ FPR（Floating-Point Register）を 4 倍まで増加し、ソフトウェアによる命令スケジューリング（ループアンローリングやソフトウェアパイプライニング）が可能になりました。

演算種類を強化した浮動小数点演算パイプラインと、4 倍まで増強したレジスタを活用することで、多種多様なアプリケーションでの処理性能も大幅に向上しました。

1 次データキャッシュでは、ロード命令、ストア命令を処理します。そのデータキャッシュは、命令列の後半にあるストア命令のアドレス計算を待たずに、後続のロード命令にデータを提供することもできます。また、1 次キャッシュは 2 つのロード命令が同時にアクセスできるデュアルポート構成となっています。2 つの 16 バイトの SIMD ロード命令、あるいは 1 つの 16 バイトの SIMD ストア命令を同時に実行することができます。バンクが競合しないかぎり、ロード命令による読み出しとストア命令による書き込みを同時に処理します。そうすることよって、キャッシュ処理能力を向上させています。

(3) 命令コミット

アウトオブオーダで実行したデータは、固定小数点データならば GPR Update Buffer（GUB）に格納されます。そして、浮動小数点データは FPR Update Buffer（FUB）、ストアデータはストアポートに格納されます。それぞれ格納された後は、プログラム順序に従って命令コミットが行われます。さらに、汎用レジスタ GPR（General Purpose Register）や FPR などのレジスタ、メモリの更新を行います。

同時にコミットできる有効命令は 4 つまでです。したがって、SPARC64 XII/SPARC64 X+/SPARC64 X では、固定小数点演算命令あるいは固定小数点ロード命令によって GPR に書き込まれる際、同時に 4 つまで書き込みます。さらに、4 つの固定小数点演算パイプラインとロードパイプラインによって処理能力が向上し、システム性能が最大限に発揮されます。

SMT 使用時は、同じサイクルに、いずれかのスレッドの命令をコミットします。そして、サイクルごとにスレッドを切り替えます。

命令をコミットするまで、実行結果はソフトウェアからアクセスできない状態になっています。前述したように、GPR や FPR、PC（Program Counter）などの制御レジスタおよびメモリは、コミットステージでプログラム命令順に一括更新されます。この同期更新方式によって正確な割り込みが保証されています。また、実行処理もいつでも取り消すことができます。さらに、この方式によって後述する命令リトライが可能となり、信頼性の向上に役立っています。

3. チップ間インターフェース

SPARC64 XII/SPARC64 X+/SPARC64 X は、CPU 間接続のための CPU インターフェース機能を内蔵しています。2 CPU（SPARC64 XII）または 4 CPU（SPARC64 X+/SPARC64 X）までは直接接続です。これより大きな構成では、CPU はクロスバー（XB）チップを経由して接続し、SPARC64 XII では 32 CPU まで、SPARC64 X+/SPARC64 X では 64 CPU まで拡張できます。CPU 単体の性能向上に合わせて、CPU 間接続には高い処理能力を持つ高速シリアル転送プロトコルを採用しています。

拡張命令セットアーキテクチャー

SPARC64 XII/SPARC64 X+/SPARC64 X は、Oracle 社の SPARC サーバがサポートしている sun4v アーキテクチャーと完全互換の仮想マシンアーキテクチャーを採用しています。

また、SPARC64 XII/SPARC64 X+/SPARC64 X では、HPC-ACE（High Performance Computing-Arithmetic Computational Extensions）を導入しました。これは SPARC-V9 アーキテクチャー命令セットの拡張で、SPARC64 VIIIfx スーパーコンピュータプロセッサで初めて導入されたものです。HPC-ACE では、演算の並列処理を可能にする SIMD（Single Instruction Multiple Data）機能が拡張され、浮動小数点レジスタ（FPR）の数も増強されます。

さらに、SPARC64 XII/SPARC64 X+/SPARC64 X の新たな機能としてソフトウェア・オン・チップ（SWoC : SoftWare on Chip）も追加されました。そして、従来汎用命令の組み合わせで処理を行っていた暗号化処理機能の専用ハードウェアの活用によって、処理が著しく高速になりました。また拡張機能として、IEEE754 標準 10 進浮動小数点数 DPD 形式および Oracle NUMBER 形式の演算を実行できる機能も追加されています。

データベース処理などのビジネスアプリケーションの傾向にも対応できるように、データの並列処理を管理する命令拡張も行っています。

次の各項では、HPC-ACE と SWoC、および並列データ処理によるビジネスアプリケーションの高速化について説明します。

1. HPC-ACE

浮動小数点レジスタ（FPR）の拡張

SPARC-V9 内の浮動小数点演算レジスタ（FPR）は 32 本です。これは、多種多様なアプリケーションの性能を最大限に引き出すためには十分な本数ではありません。しかし、レジスタ数を増やすためには 32 ビット固定長の SPARC アーキテクチャーでは命令長が不足し、これを変更するのは困難でした。この問題を解決するために、HPC-ACE で

は SXAR（Set eXtended Arithmetic Register）という前置命令により、直後の 2 つまでの命令に対しレジスタのアドレッシングの拡張を行います。また SXAR は、レジスタアドレスを 2 ビット拡張しました。さらに、浮動小数点レジスタ（FPR）の本数を SPARC-V9 の 4 倍、128 まで指定可能にしました（図 7-4 参照）。コンパイラはこの大容量レジスタを活用して、ソフトウェアパイプライニングなどの最適化を行い、アプリケーションが持つ命令レベルの並列性を最大限に引き出します。

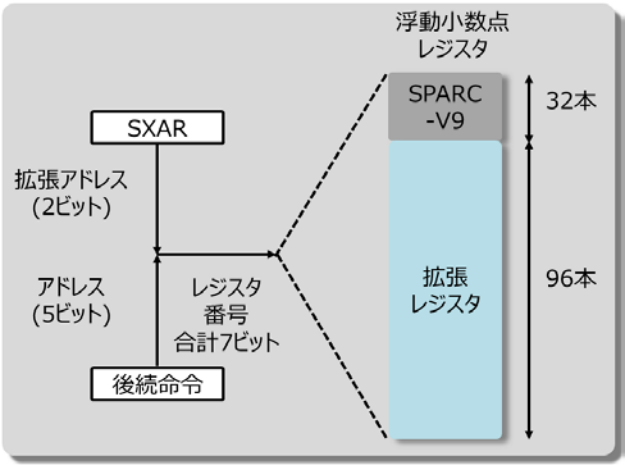

図 7-4. SXAR 命令によるレジスタアドレス拡張

SIMD（Single Instruction Multiple Data）

SIMD とは、1 つの命令で複数のパイプラインを使ってデータを並列に処理する機能です。HPC-ACE は、この SIMD の採用によって、1 つの演算命令で 2 つの浮動小数点積和演算（FMA：Floating-point Multiply and Add）を実行することを可能としました。

従来のスカラ演算では、1 つのデータに対して 1 命令を処理します。したがって、4 個の処理結果を出すには 4 回の加算命令を実行していました。それに対して、SPARC64 XII/SPARC64 X+/SPARC64 X では命令パイプラインごとに 2 個の SIMD 命令を同時に実行することにより、4 つの 64bit データを同時に積和演算処理することが可能です。また、新たな命令を追加することにより、SPARC64 XII/SPARC64 X+では、命令パイプラインごとに 8 ビットデータを 32 個まで同時に比較できます。さらに、SPARC64 XII では、4 ビットデータを 64 個まで同時に比較できます。

SPARC64 XII/SPARC64 X+/SPARC64 X では、こうした SIMD 命令を活用し、暗号化処理やデータベースなどのビジネスアプリケーションの高速化を可能にしています。

さらに、ロード命令とストア命令にも SIMD 命令を活用することができます。特にロード命令は、キャッシュラインをまたぐ場合でもペナルティなく処理を行うことができ

ます。

大量データの検索やデータの圧縮/解凍（暗号化/復号）をする際に、この機能を活用すると、大量データ処理やメモリ内データベースの高速化にも効果を発揮します。

2. ソフトウェア・オン・チップ（SWoC：SoftWare on Chip）

従来複数の命令を組み合わせて行っていた処理を、SPARC64 XII/SPARC64 X+/SPARC64 X では専用ハードウェアを使用できるようにすることで高速化を図っています。これはソフトウェア・オン・チップと呼ばれます。具体的な高速化処理を以下に説明します。

- 暗号化処理

SPARC64 XII/SPARC64 X+/SPARC64 X には、暗号化演算ユニットが実装されています。実装されている暗号化演算ユニットは、SPARC64 X+/SPARC64 X はコアあたり 2 つ、SPARC64 XII には命令パイプラインあたり 2 つ、コアあたり 4 つです。

専用の暗号命令により、アドオンアダプタを使用せずに高速の暗号化処理（暗号化/復号）を実行します。暗号形式は AES、DES、3DES、DH、DSA、ECC（*1）、RSA、SHA です。

暗号化処理をハードウェアで実行するので、追加費用や性能の低下が生じることはありません。また、データベースのフル暗号化によって、安全な環境を構築することができます。

OpenSSL や暗号化/復号コマンドは、すでに SPARC64 XII/SPARC64 X+/SPARC64 X の暗号化演算処理機能を使用できるように対応済みです。libpkcs11 などの標準ライブラリも、暗号化演算処理高速化の恩恵を受けることができます。

*1 SPARC64 XIIのみ

- IEEE754 標準 10 進浮動小数点数 DPD 形式の演算

SPARC64 XII/SPARC64 X+/SPARC64 X はメインフレームの高速演算技術を継承しています。

10 進浮動小数点演算ユニットをプロセッサに実装したことにより、10 進浮動小数点の演算処理はハードウェアで直接高速に実行されます（従来、10 進浮動小数点の演算処理はソフトウェアで処理されていました）。

従来の 10 進浮動小数点の演算処理は、ソフトウェアで処理されていました。ソフトウェアで処理を行う場合、10 進数データをソフトウェア変換で 2 進数データに変換し、変換後にハードウェアで演算処理を行います。その後、ソフトウェアで、演算結果の 2 進数データを 10 進数データに変換します。

これに対して、SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサでは、ソフトウェアでの変換をせずに、10 進数データのままハードウェアで演算処理ができます。したがって、売上計算、原価計算、リベート計算、複利計算などの一般的な処理タスクにおいて飛躍的な高速化を実現しました。その結果、流通／製造／金融分野など

多岐にわたる業界で演算の高速化が可能になりました。
この SPARC64 XII/SPARC64 X+/SPARC64 X 演算機能は、10 進浮動小数演算の標準規格である IEEE754-2008 に対応しています。

- **比較操作とコピー操作**
  SPARC64 XII/SPARC64 X+/SPARC64 X ではメモリアクセスバス幅が拡大されています。また、SIMD 拡張命令も実装され、複数ブロックデータのメモリからのロード、比較、およびメモリへの格納を可能としています。拡張 SIMD 命令によりメモリバス容量は 100％利用できるようになりました。その結果、SPARC64 XII/SPARC64 X+/SPARC64 X は最大の演算能力を発揮することができます。このようなエンハンスにより、標準 libc ライブラリの memcpy（3C）などのメモリ処理関数も高速化の恩恵を受けることができます。

従来、アプリケーションソフトウェアで実行されていた処理は、SPARC64 XII/SPARC64 X+/SPARC64 X ハードウェアで処理できるようになりました。これは、アプリケーションの性能向上にも直結します。また、Oracle Solaris 11 は、SPARC64 XII/SPARC64 X+/SPARC64 X のソフトウェア・オン・チップ機能における利点を最大限に活用すべく最適化されています。これは、アプリケーションへの変更を必要とせずに、性能を向上できることを意味しています。

RAS 機能（信頼性、可用性、保守性）

SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサでは、先進的なエラー検出機能およびエラー訂正機能を提供することによって、システムの信頼性向上を図っています。実際に、このエラー検出機能またはデータ訂正機構によって、SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサ回路の 99%が保護されています。そして、RAM ユニットは ECC によって保護あるいは複製されます。また、1 ビットエラーはすべて訂正することができます。さらに、ラッチおよび実行ユニットはすべてパリティ保護されます。例えば、プロセッサが 1 ビットエラーを検出すると、命令をリトライします。また、キャッシュは way 単位または CPU コア単位での動的縮退（エラーの種類によってはリブート後に切り離し）を行います。SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサのその他の信頼性機能として、エラーのマーキング、ハードウェアによる命令リトライ（再実行）、予防保守のサポートがあります。これらについては、後述します。

前述のとおり、SPARC64 XII/SPARC64 X+/SPARC64 X の RAS 機能は当社のメインフレームに匹敵します。これらの RAS 機能では、確実なエラーの検出、影響範囲の特定、復旧処理を実行します。さらに、エラーの記録とソフトウェアへのエラー通知なども行っています。このように、RAS 機能の実装により、SPARC64 XII/SPARC64 X+/SPARC64 X はミッションクリティカルな UNIX サーバの理想的なプロセッサとして、高い信頼性、

可用性、保守性、データ整合性を提供しています。

表 7-2. エラー検出機構

ユニット	エラー検出、訂正方法
キャッシュ（タグ部）	ECC 二重化＋パリティ
キャッシュ（データ部）	ECC パリティ
レジスタ	ECC（整数/浮動小数点） パリティ（その他）
演算器	パリティ／剰余チェック
キャッシュの動的縮退	あり
ハードウェア命令リトライ	あり
プロセッサ履歴ロギング	あり

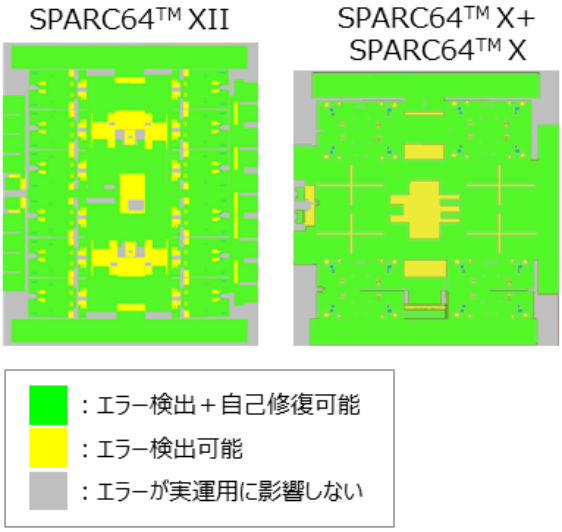

図 7-5. RAS によるエラー検出および修復領域

ユニット	エラー検出／訂正	エラー処理		記録
		訂正	縮退	
1次キャッシュメモリ（タグ部）	二重化+パリティ、ECC	リトライ、ECC	動的 way 縮退（*1）	ヒストリー
2次キャッシュメモリ（データ部）	ECC	ECC	動的 way 縮退（*1）	
3次キャッシュメモリ（データ部）（*2）	ECC	ECC	動的 way 縮退（*1）	
演算器	パリティ(*3)、ECC	ECC、ハードウェア命令リトライ	コア縮退	
レジスタ	パリティ(*3)、ECC	ECC、ハードウェア命令リトライ	コア縮退	

*1 wayはキャッシュメモリを構成する単位です。
*2 SPARC64 XIIプロセッサのみ。
*3 パリティエラー検出時は、ハードウェア命令リトライ機能によりエラーを訂正します。

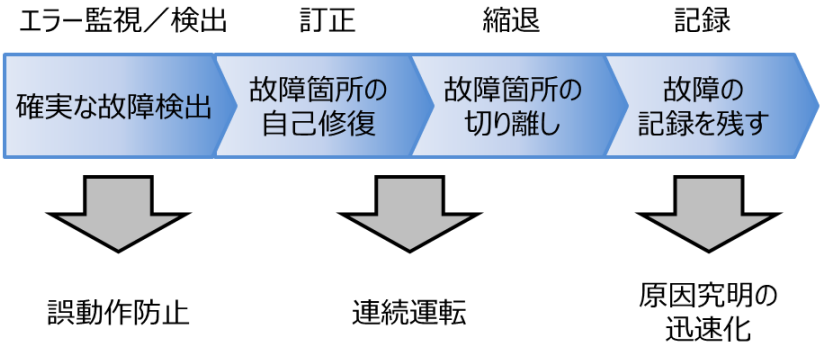

図 7-6. エラー検出と復旧のスキーム

1. エラーマーキング

メインメモリやキャッシュからの読み出しデータがマルチビットエラーになっているときは、エラーマーキング機能がエラーの発生箇所を識別するための特別の値をデータに書き込みます。まず ECC シンドロームレポートと照らし合わせて、障害の原因となったエラー箇所を特定します。それが正確な縮退と部品交換への支援につながります。

2. 内蔵 RAM の信頼性および可用性機能

SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサは、高度なデータ整合性をサポートする信頼性と可用性の機能を提供します。表 7-3 に、SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサのエラー検出／データ保護方法、エラー訂正方法を示します。

表 7-3. SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサにおける内蔵 RAM のエラー検出およびデータ保護

RAM種別	エラー検出／データ保護方法	エラー訂正方法
命令1次キャッシュデータ	パリティ	無効化および再読み込み
命令1次キャッシュタグ	パリティ+二重化	二重化データの再読み込み
1次データキャッシュデータ	SECDED（*1）ECC	ECCによる1ビットエラー訂正
1次データキャッシュタグ	パリティ+二重化	二重化データの再読み込み
2次キャッシュデータ	SECDED（*1）ECC	ECCによる1ビットエラー訂正
2次キャッシュタグ	SECDED（*1）ECC	ECCによる1ビットエラー訂正
3次キャッシュデータ（*2）	SECDED（*1）ECC	ECCによる1ビットエラー訂正
3次キャッシュタグ（*2）	SECDED（*1）ECC	ECCによる1ビットエラー訂正
命令TLB	パリティ	無効化
データTLB	パリティ	無効化
分岐履歴	パリティ	分岐予測失敗からの復旧

*1 SECDED：Single Error Detect Double Error Detect

*2 SPARC64 XIIプロセッサのみ。

1 次キャッシュ、2 次キャッシュ、3 次キャッシュ、TLB は、way 単位での動的縮退が可能です。SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサは、1 次キャッシュ、2 次キャッシュ、3 次キャッシュ、および TLB を way 単位に分割するセットアソシアティブ方式を採用しています。エラーの発生回数は way 単位でカウントされます。エラー発生回数が上限値を超える場合は、まず該当 way を縮退します。そして、それ以降はその way を使用しないようにします。way の動的縮退はハードウェアが担当します。その際、ソフトウェアが受ける影響は way の縮退によるわずかな処理速度の低下のみで、稼動への影響はありません。

3. 内蔵レジスタと演算器の信頼性機能

信頼性をさらに高めるために、SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサではレジスタや演算器もエラー保護しています。汎用レジスタ（GPR）と浮動小数点レジスタ（FPR）の大容量化に伴い、ECC による保護を採用しています。エラーが発生すると、ECC 回路によってエラーを訂正します。ほかのレジスタはパリティにより保護されています。さらに、パリティ予測回路が演算器内に実装されています。乗除算器には剰余チェック回路が実装され、出力結果のエラー検出が行われます。万一エラーを検出した

場合は、次に述べるように、ハードウェアが自動的に命令を再実行して復旧処理を行います。

4. 同期更新方式と命令リトライ

SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサは、コミット時の同期更新方式を採用しています。エラーを検出せずにコミットした命令だけが、プログラム可能なリソース（GPR、FPR、その他レジスタ、メモリなど）にある実行結果を更新します。エラーが検出された場合は、コミット前の処理はすべて取り消します。このようにすることによって、データ整合性に問題が発生するのを防止します。そして、処理を一旦取り消した後に、ハードウェアが命令を再実行します。これを「命令リトライ」と呼びます。

図 7-7 に示すように、エラーが発生した場合、命令リトライを自動的に起動します。リトライは 1 命令ずつ実行されます。それは実行が成功する可能性を高くするためです。実行が成功すると、自動的に正常な実行状態に戻ります。この処理に関して、ソフトウェアの介在は不要です。また、命令リトライが成功すれば、ソフトウェアにはエラーの影響もありません。命令リトライは、リトライ試行回数の閾値に達するまで繰り返されます。閾値を超えた場合は、プロセッサはエラーの内容を記録します。さらに異常の発生をオペレーティングシステムに通知します。そして、以降の処理はオペレーティングシステムが実行します。

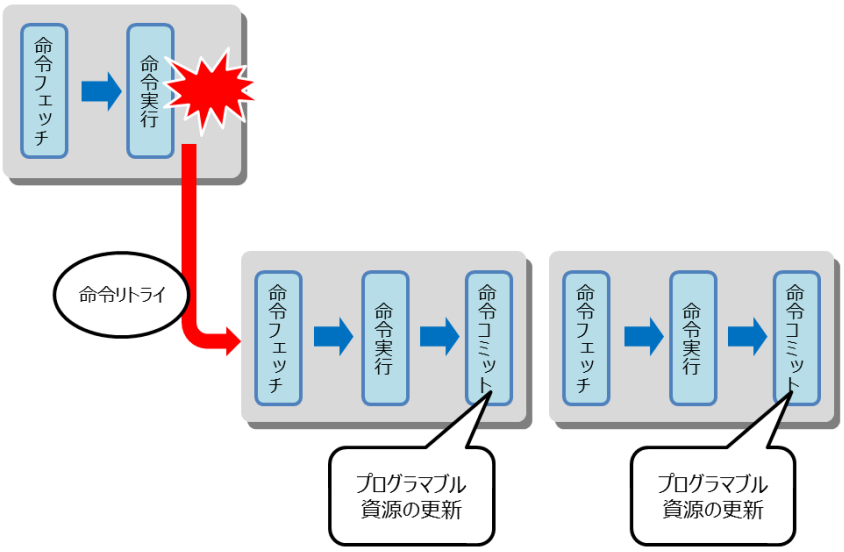

図 7-7. 命令リトライ

5. 保守性の向上

SPARC64 XII/SPARC64 X+/SPARC64 X プロセッサは前述したとおり、様々なエラーのチェック機構を装備しています。エラーがないかを監視し、エラー発生時にはその情報を XSCF（eXtended System Control Facility）に通知します。通知された XSCF ファームウェアではエラーログを収集し解析します。このように、プロセッサの豊富なエラー機構を活用することによって、SPARC M12/M10 は運用を中断せずに故障箇所や故障種別を、迅速かつ正確に特定することができます。その結果、システムは予防保守に有用な情報を提供するとともに、保守性をも向上させています。

8. 結論

ビッグデータとクラウドの時代といわれる今日、データはますます大容量化、多様化しています。このような時代にあって、多岐にわたるデータ処理をかかえるお客様が求めるものは何か。それは、これまでにない拡張性と柔軟性、信頼性と管理性です。そして何より高い性能を求めています。これら時代のあらゆるニーズに応えられるのが、SPARC M12 と SPARC M10 です。

富士通 12 世代目となる SPARC64 XII プロセッサと、富士通 10 世代目となる SPARC64 X+/SPARC64 X プロセッサは、高性能スーパーコンピューティングとミッションクリティカルな処理において長年培った技術の集積となる製品です。優れた性能はもちろん、ソフトウェア・オン・チップなどのビジネスコンピューティングの画期的な概念を提供します。このアーキテクチャーの大胆な革新によって、すべての鍵となるデータベース機能の加速化を図ることができ、かつハードウェアとソフトウェアの将来的な姿を想像させました。

ユニークで新しいビルディングブロック方式の採用。Oracle VM Server for SPARC の完全サポート。Oracle Solaris ゾーンによるさらなる柔軟性の確保。これらすべてを備えているのが、SPARC M12 と SPARC M10 です。そして、これら技術がコアごとのリソースの有効化と相まって、SPARC M12 と SPARC M10 はお客様とともにこれまでにないレベルへと着実に進化していくことでしょう。当社が誇るメインフレーム級の RAS 機能を随所に採用することで、今までにない保護レベルへと拡張します。さらに、エントリーサーバの価格でありながら、ミッドレンジクラスの性能を備えた SPARC M10-1 や SPARC M12-1 の単一 CPU モデルから、SPARC M10-4S 16BB 構成（64 CPU）または SPARC M12-2S 16BB 構成（32 CPU）のハイエンドクラスに至るまで、XSCF による一貫した管理が可能です。

LLC や VLLC などの先進技術、信頼性という伝統、世界最先端のシステムパフォーマンス、さらに、信頼されたパートナーシップによるハードウェアとソフトウェアの連携。SPARC M12 と SPARC M10 はこれらの融合によって、お客様とパートナー会社様の進歩し続けるニーズや成長へのベストアンサーとなることでしょう。

お問い合わせ先
富士通株式会社
〒105-7123
東京都港区東新橋 1-5-2 汐留シティセンター
Web サイト：https://www.fujitsu.com/jp/products/computing/servers/unix/sparc/