TOSHIBA
Leading Innovation >>>

INVENTOR OF FLASH MEMORY

第179回計算機アーキテクチャ・第15回組込みシステム合同研究発表会

SSDアーキテクチャと未来

2010年 1月 28日

株式会社 東芝 セミコンダクター社
メモリ事業部　SSD応用技術部
参事
土屋 憲司

Copyright 2009, Toshiba Corporation.

Contents

1 NAND and SSD Market Over View

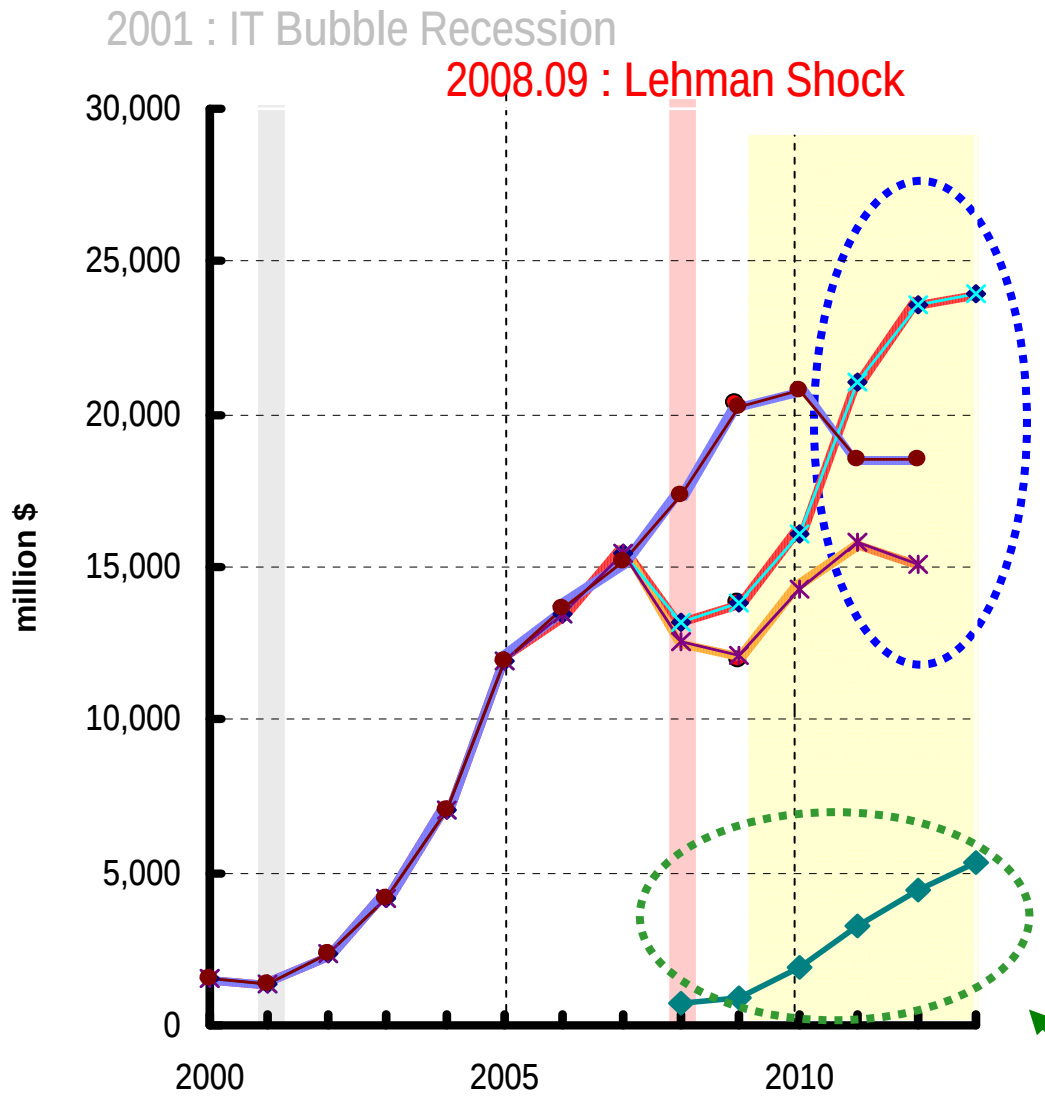
2 Reliability of SSD

3 Technologies for SSD

4 Future of SSD

1 NAND and SSD Market Overview

World Wide NAND Flash Market



2009.09 : Forecast

2008.04 : Before Lehman Shock

2009.01 : Right after Lehman Shock

NAND Market

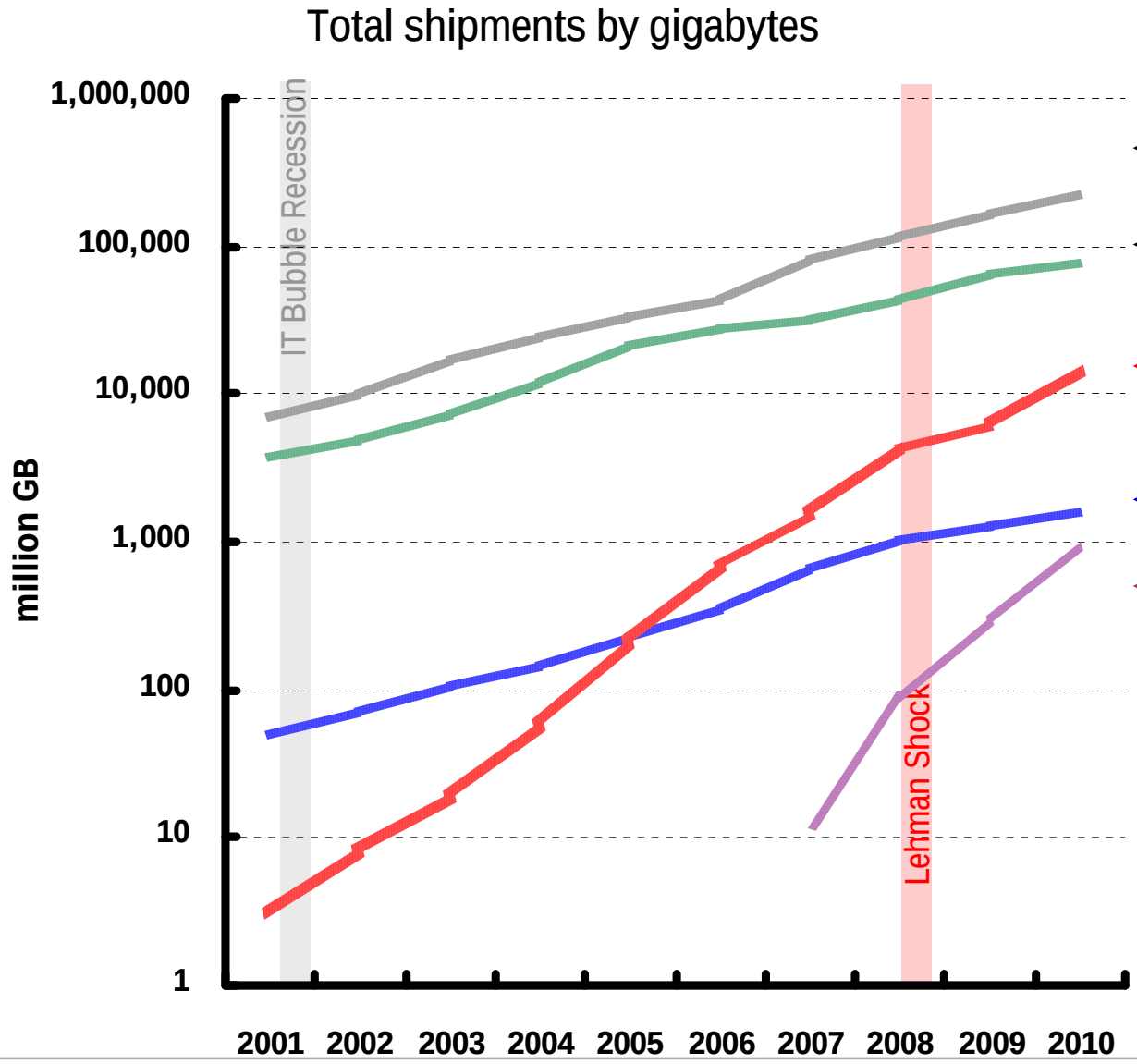
+15% CAGR can be expected for 2009 - 2013.
Unlike other semiconductor devices,
NAND can anticipate quick recovery.

Toshiba expects that NAND market will expand beyond the pre-crisis estimate.

SSD Forecast

Source : WSTS + Gartner+ IDC2009

Bit Growth of NAND Flash & SSD



	CAGR 2001- 2010
HDD	45%
Opt. Media	40%
NAND	156%
DRAM	47%
SSD	334% (2008-2010)

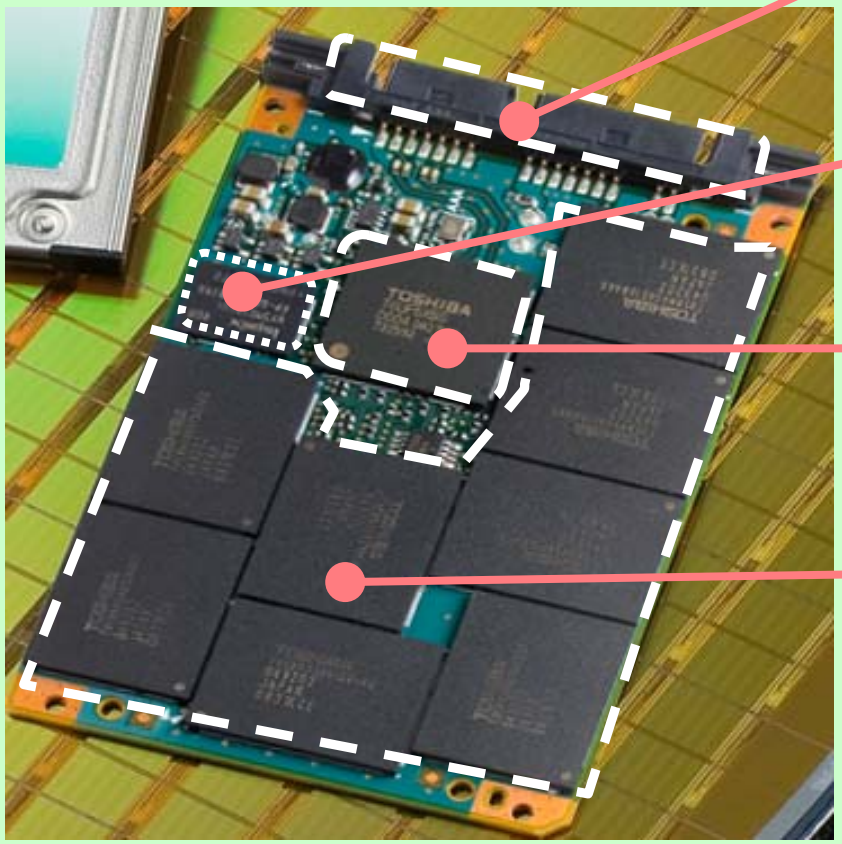
NAND bit growth rate is the highest.

Source：TSR(HDD)
iSuppli & Gartner (DRAM/NAND)
記録メディア工業会 & TSB (Opt.media)
IDC2009

2 Reliability of SSD

Basic Structure of SSD

SSDの構造 (一例)



●コネクタ
インターフェイスは、HDDと同様にSATA、micro SATAが主流。物理形状はHDDとコンパチブルなものが主流。

●DRAM
Cache機能を果たす。
(※DRAM無しの場合もあります)

●コントローラ
SSDの心臓部となる。このコントローラでの制御により、高速化/書換寿命の長期化/高信頼性化を実現

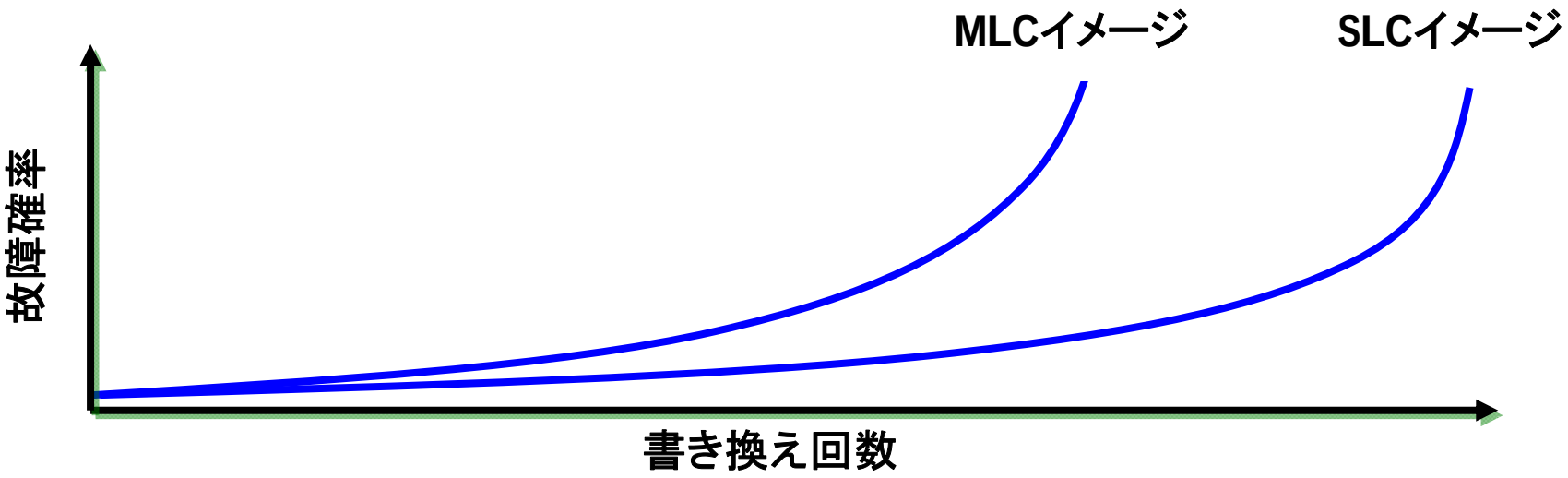
●NANDフラッシュメモリ
データを蓄積するNANDフラッシュメモリ。多値NAND(MLC)技術により低価格/大容量化を実現。
(※2値NAND(SLC)が使用される場合もある)

●形状
・2.5inch case (コネクタ:標準SATA)
・1.8inch case (コネクタ:Micro SATA)
・1.8inch case less <筐体なし> (コネクタ:Micro SATA)
など

SSDの寿命に対する誤解

- MLC NANDは、SLC NANDに比べ 書き換え耐性が低い　**Truth**
- NANDの書き換え回数が規定値を超えると、SSDはすぐ壊れてしまう　**Myth**

NANDは書き換え回数が増えるほど、"故障確率" が高くなるデバイスである



NANDの故障モードは幾つかあり、SSDとして予防・救済出来るものもある

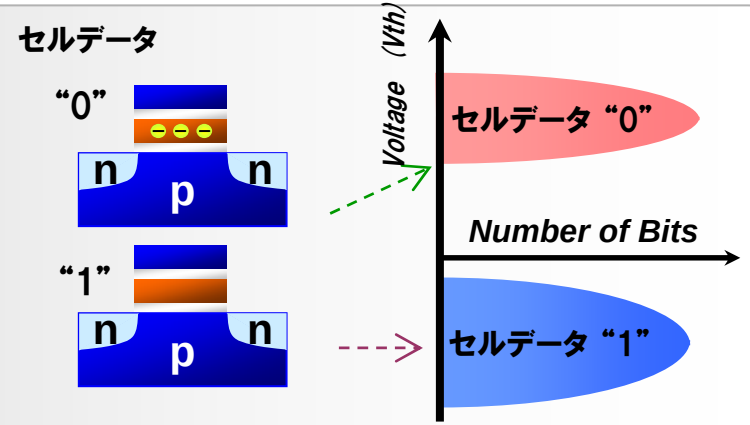
NANDフラッシュとは？

SLC（Single Level Cell）：2値NAND技術

・1セルに1ビット分のデータの格納が可能（1 bit/Cell）

●1セルに格納されているデータ

1 OR 0



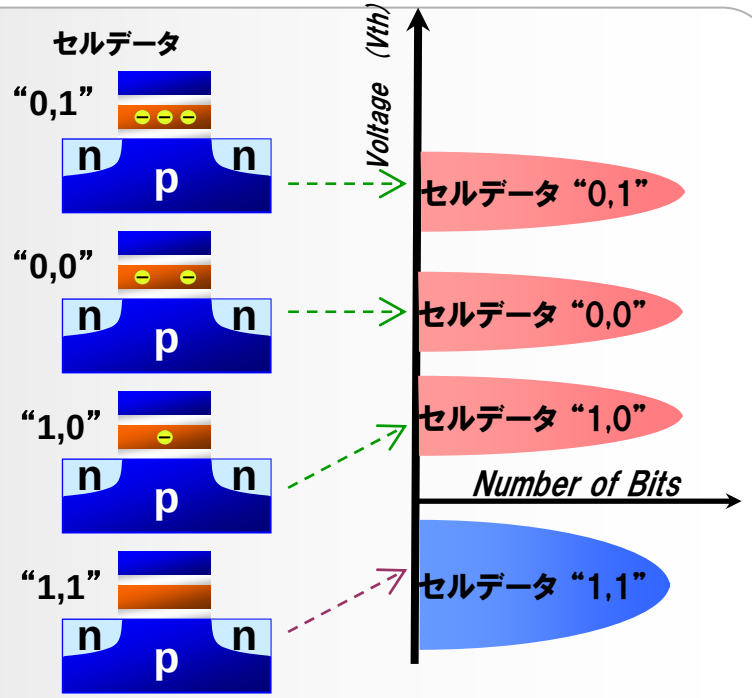
MLC（Multi Level Cell）：多値NAND技術

・1セルに2ビット分のデータの格納が可能（2bit/Cell）
これにより、1セルあたりに格納できるデータ量が
SLCの2倍になる
・1ビットセルあたり、3ビットを格納できるMLCも製品化予定

SLCの2倍のデータ量

●1セルに格納されているデータ

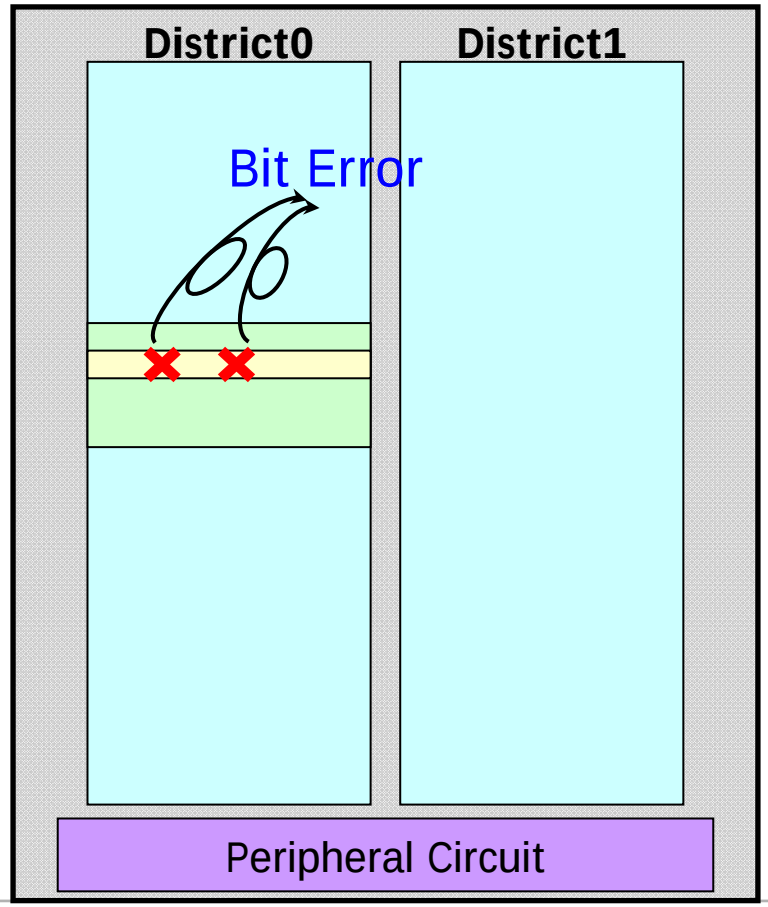
11 OR 10 OR 00 OR 01



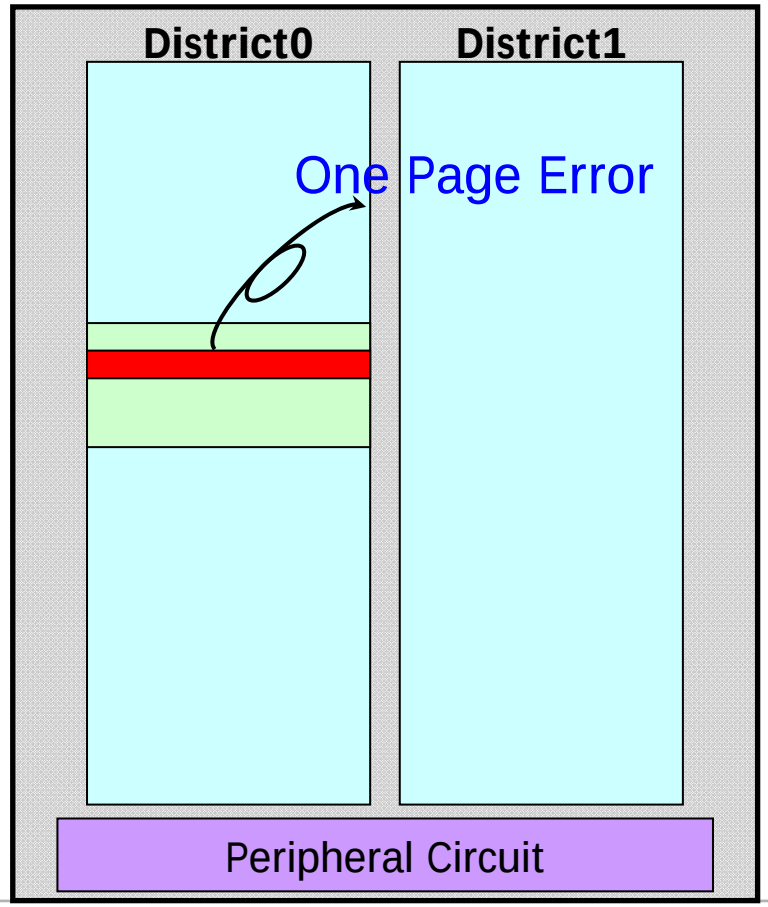
What type of error is recoverable?

Possible NAND failure modes						
No Failure	Random Bit Error	Single Page Error	Multiple Page Error	Single Block Error	District Error	Single Chip Error

Random Bit Error



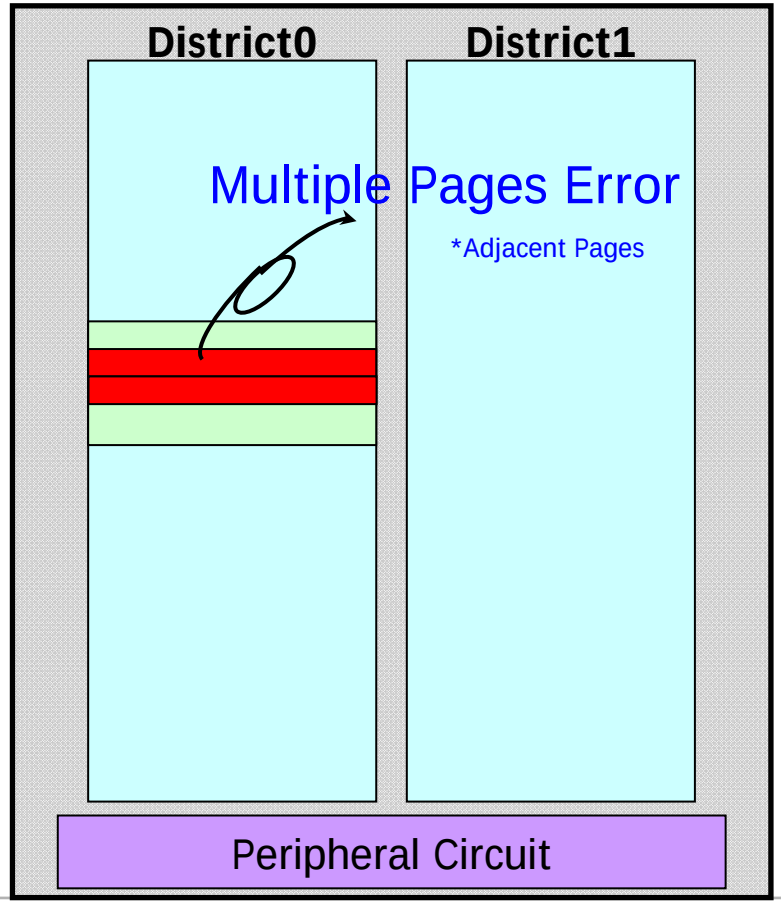
Single Page Error



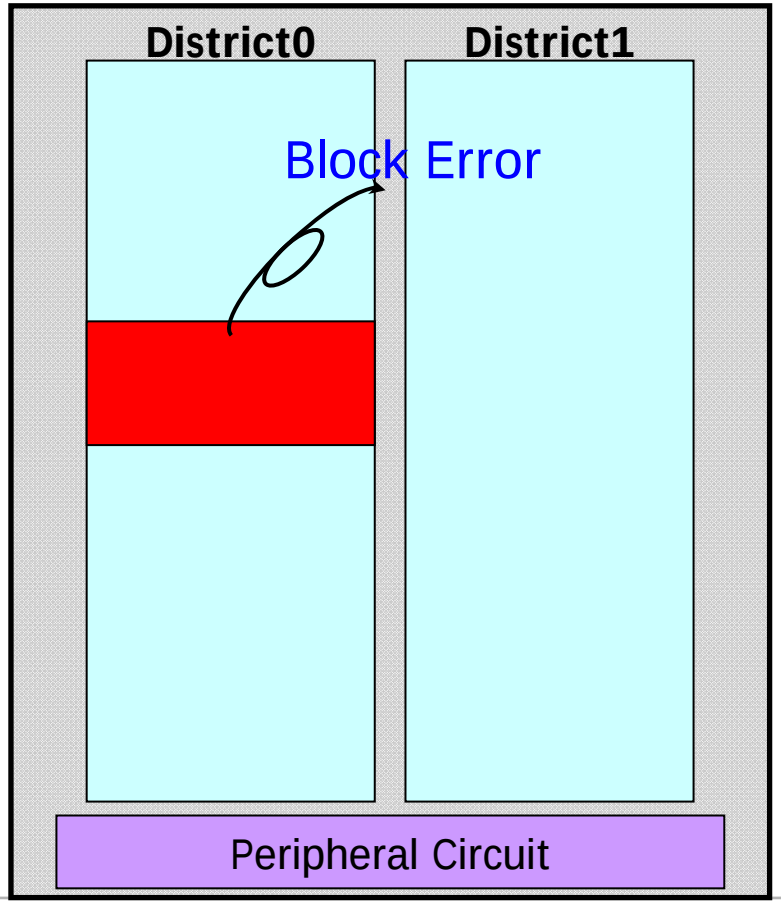
What type of error is recoverable?

Possible NAND failure modes						
No Failure	Random Bit Error	Single Page Error	Multiple Page Error	Single Block Error	District Error	Single Chip Error

- **Multiple Page Error**



- **Single Block Error**



SSDの寿命を理解する (1)

■ SSDの故障モードとは？

1 余裕容量が減る

- NANDのブロック故障が増加
  様々な対応が考えられるが、
  東芝の実装では、Read Only Modeに移行

2 データが失われる

- エラー訂正に失敗する
  HDD同様、Blue Panicの可能性も

3 動かない. アクセス出来ない.

- NANDチップ丸ごと故障
- コントローラ、DRAMの故障など

■ SSDの信頼性

9. RELIABILITY

A failure is defined as an inability of the drive to perform its specified function described in the requirements of this document when being operated under the normal conditions or conditions specified in this document. However, damages caused by operation mistake, mishandling, accidents, system errors and other damages that can be induced by the customers are not defined as failure.

.

9.1 Error Rate

9.1.1 Non- Recoverable Error Rate

1 error per 10^{14} bits read

The defective sectors allocated to the spare locations in the factory are not counted in the error rate.

9.2 Mean Time to Failure (MTTF)

1,000,000 hours

A failure means that the drive can not execute the function defined in this document under the nominal temperature, humidity and the other conditions specified in this document. Damages caused by operation mistake, mishandling, system failure and other damages occurred under the conditions which are not described in this document are not considered as the failure.

- Power on hours (note1) : 8760 hours/year
- Operating time (note2) : Active R/W Duty 20% of power on hour
- Environment : 0-60℃ (Tc)

- Do not apply electrical static discharge, vibration and shock to the drive.
- Do not press top and bottom PCBA surface of the drive.
- All others condition should be within specification show in section 6/7/8/9.

(note 1) Power on hour includes sleep and standby mode.
(note 2) Operating: writing (20GB/Day) and reading.

Applicable warranty and warranty period should be covered by the purchase agreement.

9.3 Product Life

Approximately 5 years or 20,000 power on hours whichever comes earlier under the following conditions.

- Power on hours (note 1): 8760 hours/year
- Operating time (note 2) : Active R/W Duty 20% of power on hour
- Environment : Normal (40℃ max.)

- Do not apply electrical static discharge, vibration and shock to the drive.
- Do not press top and bottom PCBA surface of the drive.
- All others condition should be within specification show in section 6/7/8/9.

(note 1) Power on hour includes sleep and standby mode.
(note 2) Operating: writing (20GB/Day) and reading.

Applicable warranty and warranty period should be covered by the purchase agreement.

SSDの寿命を理解する (2)

- SSDの故障確率は、以下の変数による関数

① Drive全体の容量
現状は、128GBが売れ筋

② ユーザ保存容量
1日に保存するGB数
ユーザにより異なる

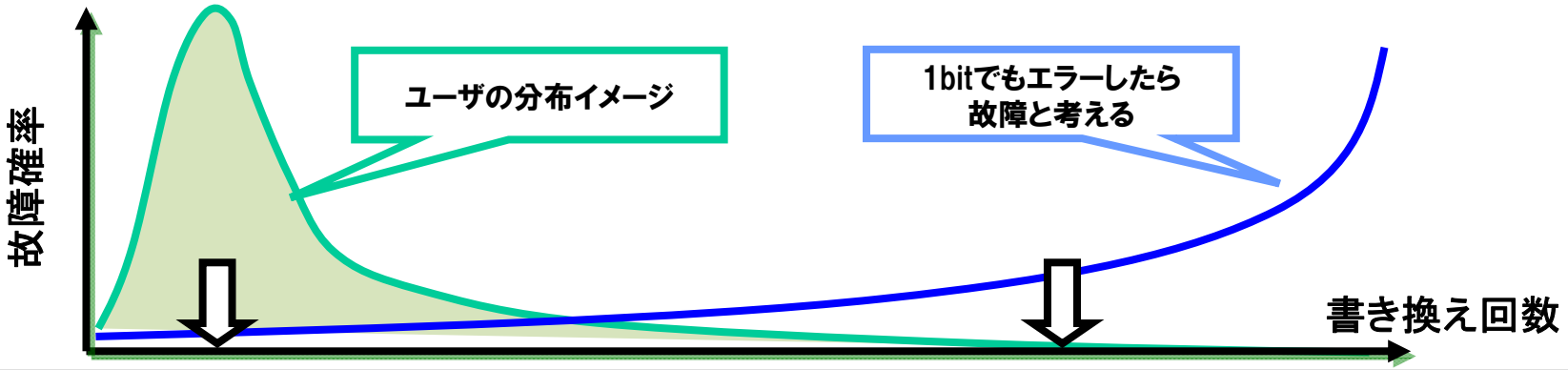
③ SSD書き込み効率
ユーザが“1”書いたときの
SSD内部での実書き込み量
（係数）

- SSDへの実書き換え回数(書き込み回数)を求める

$$\text{書き換え回数} = \frac{② \ 4\text{GB} * ③ \ 6(\text{書き込み効率}) * 1825\text{日}(5\text{年})}{① \ 128\text{GB}}$$

1日4GB、書き込み効率6を仮定

- SSDの故障確率は、ユーザによって異なってくる



SSDの寿命を理解する (3)

■ ユーザーモデル (Usage Model)

SSDの寿命を理解するうえで、”ユーザーの使い方” を理解することが重要

あなたは、1日何GB書いていますか？
- 1GB？　2GB？　5GB？　10GB以上？

あなたは、Driveの全容量を使い切っていますか？
- 100%使ってます　30%程度余裕を持ってます　半分くらい空いてます

HDDに比べ容量が少ないSSD、どうやって使いますか？
- 性能を期待して、ゲーミングマシンとして。ソフトの保存はHDD。
- 耐振動、耐衝撃性を期待してモバイルPC。

実際のところ、Windowsのユーザはどんな使い方をしているのだろうか？

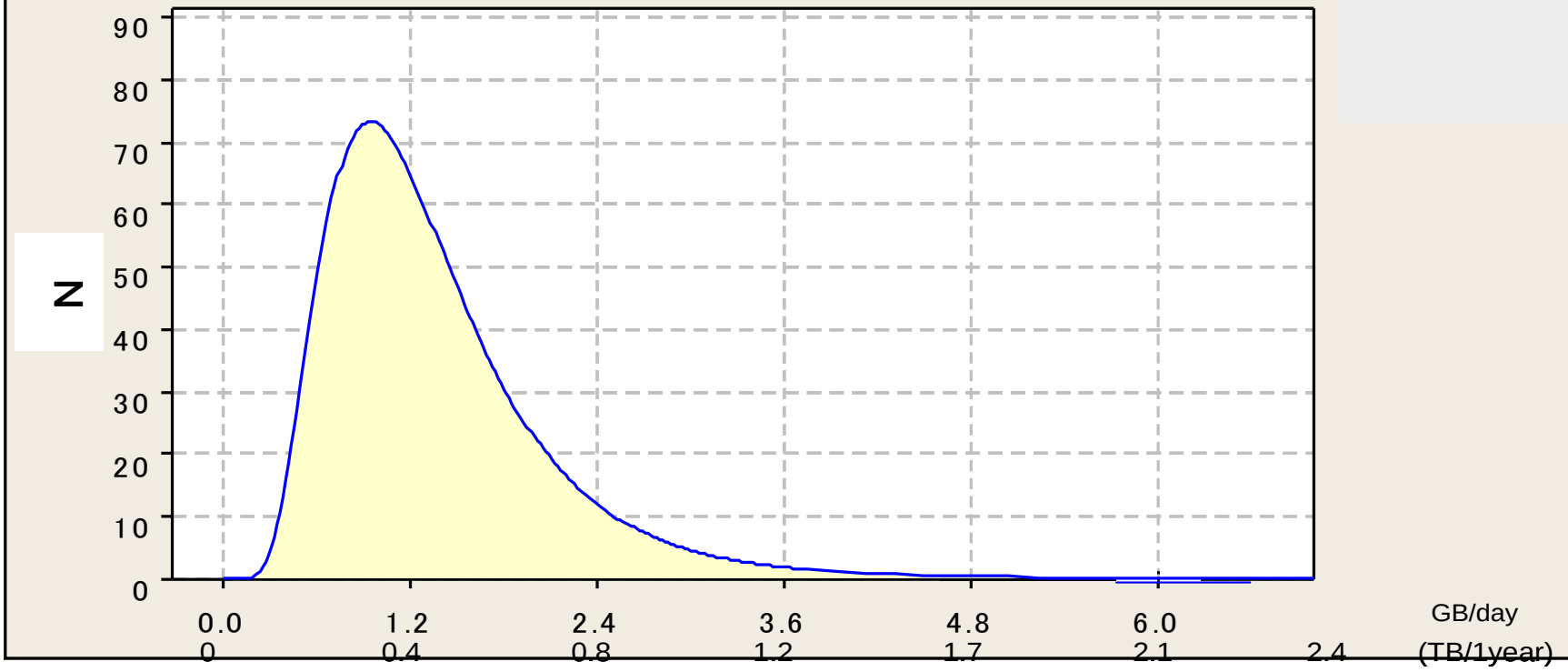
SSDの寿命を理解する (4)

■ Windowsユーザ履歴

Windows-XP環境でのモニター結果
対象　ITユーザー
期間　　約2ヶ月

1日当たりの書き込みデータ量（Hibernation除外）

平均書き込み量　XGB　　最大書き込み量　YGB



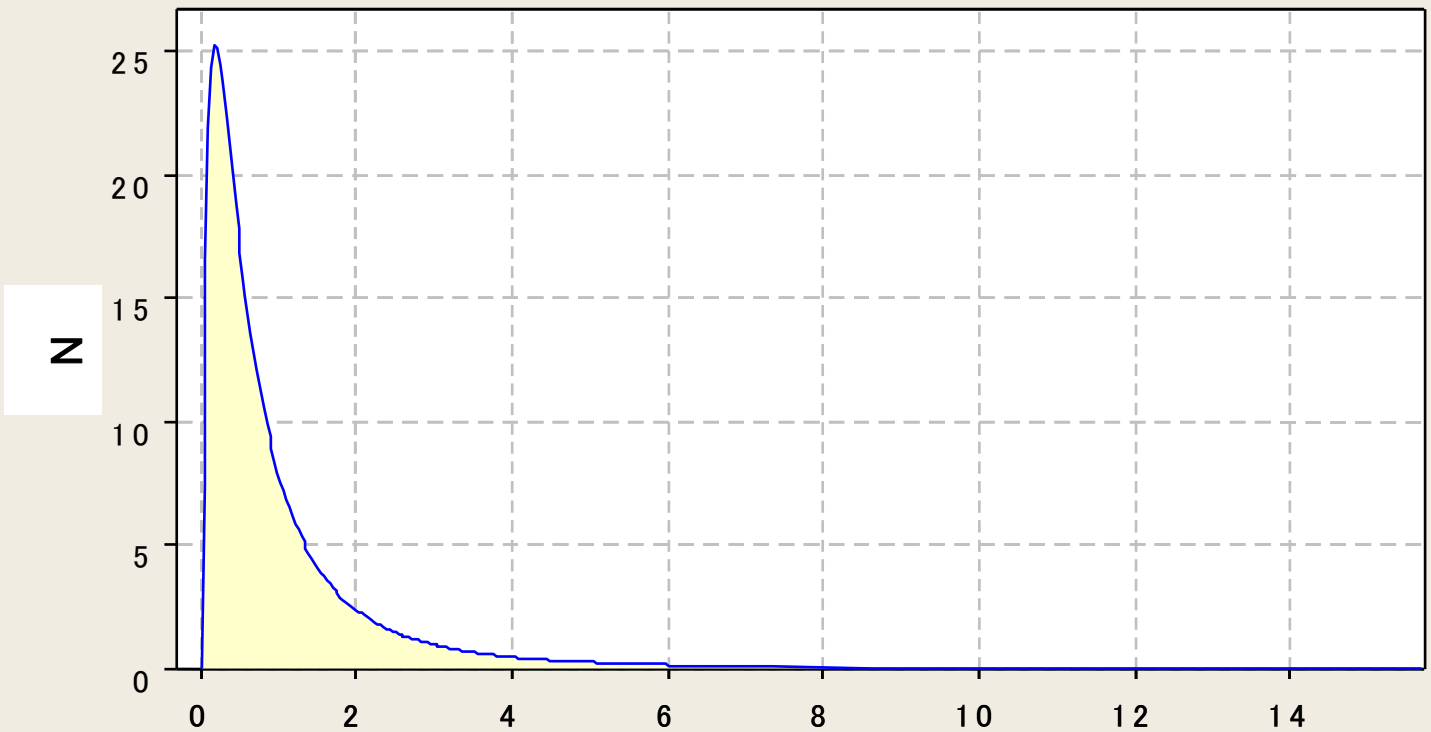
SSDの寿命を理解する (5)

■ Windowsユーザ履歴

Windows-XP環境でのモニター結果
対象　ITユーザー
期間　　約2ヶ月

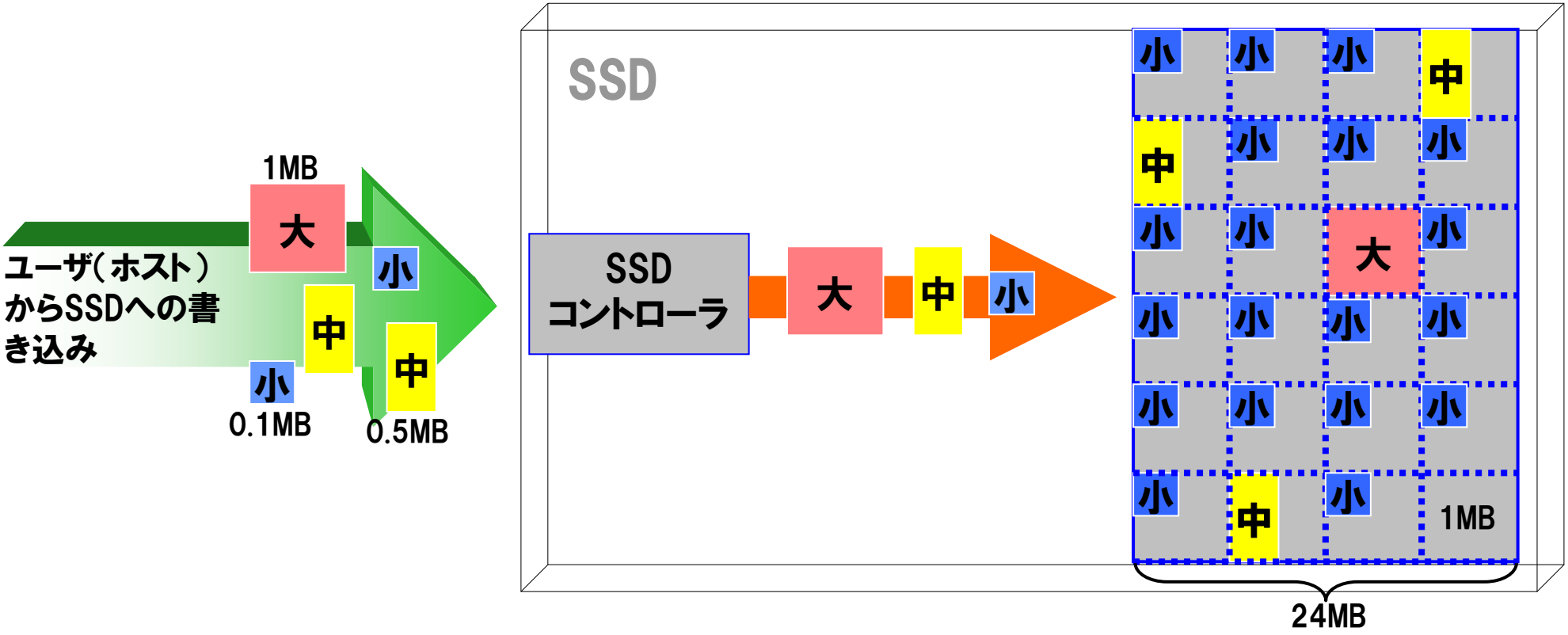
1日当たりのHibernation書き込み量

平均書き込み量　XGB　　最大書き込み量　YGB



SSDの寿命を理解する (6)

■ 書き込み効率とは (基本的な考え方)



$$\text{書き込み効率} = \frac{24\text{MB}}{1\text{MB} * 1 + \underset{0.5\text{MB}}{\text{中}} * 3 + \underset{0.1\text{MB}}{\text{小}} * 19} = 5.45$$

理想は倍率(効率)=1

SSDの寿命を理解する (7)

■ ユーザ履歴の結果 と TSBの想定書き込み効率から、
SSDの書き換え回数を求めてみる

1 Drive全体の容量	2 ユーザ保存容量	3 SSD書き込み効率
128GB	2.5GB	6

- Hostからは、　1年：約1,000GB　5年：約 5,000GB
- SSD内部では、　1年：約6,000GB　5年：約30,000GB

5年間毎日使用して235回程度書き換える計算

SSDの寿命を理解する (まとめ)

■ SSDの寿命とは

- NANDの寿命
- あくまで確率の問題
- 使い方が確率を左右する
- 書き換え効率大きく影響する
- エラー救済機能によっても変わる

■ SSDの故障モード

1. 余裕容量が減る
2. データが失われる
3. 動かない. アクセス出来ない.

■ SSDの信頼性

9. RELIABILITY

A failure is defined as an inability of the drive to perform its specified function described in the requirements of this document when being operated under the normal conditions or conditions specified in this document. However, damages caused by operation mistake, mishandling, accidents, system errors and other damages that can be induced by the customers are not defined as failure.

.

9.1 Error Rate

9.1.1 Non- Recoverable Error Rate

1 error per 10^{14} bits read

The defective sectors allocated to the spare locations in the factory are not counted in the error rate.

9.2 Mean Time to Failure (MTTF)

1,000,000 hours

A failure means that the drive can not execute the function defined in this document under the nominal temperature, humidity and the other conditions specified in this document. Damages caused by operation mistake, mishandling, system failure and other damages occurred under the conditions which are not described in this document are not considered as the failure.

- Power on hours (note1) : 8760 hours/year
- Operating time (note2) : Active R/W Duty 20% of power on hour
- Environment : 0-60℃ (Tc)

- Do not apply electrical static discharge, vibration and shock to the drive.
- Do not press top and bottom PCBA surface of the drive.
- All others condition should be within specification show in section 6/7/8/9.

(note 1) Power on hour includes sleep and standby mode.
(note 2) Operating: writing (20GB/Day) and reading.

Applicable warranty and warranty period should be covered by the purchase agreement.

9.3 Product Life

Approximately 5 years or 20,000 power on hours whichever comes earlier under the following conditions.

- Power on hours (note 1): 8760 hours/year
- Operating time (note 2) : Active R/W Duty 20% of power on hour
- Environment : Normal (40℃ max.)

- Do not apply electrical static discharge, vibration and shock to the drive.
- Do not press top and bottom PCBA surface of the drive.
- All others condition should be within specification show in section 6/7/8/9.

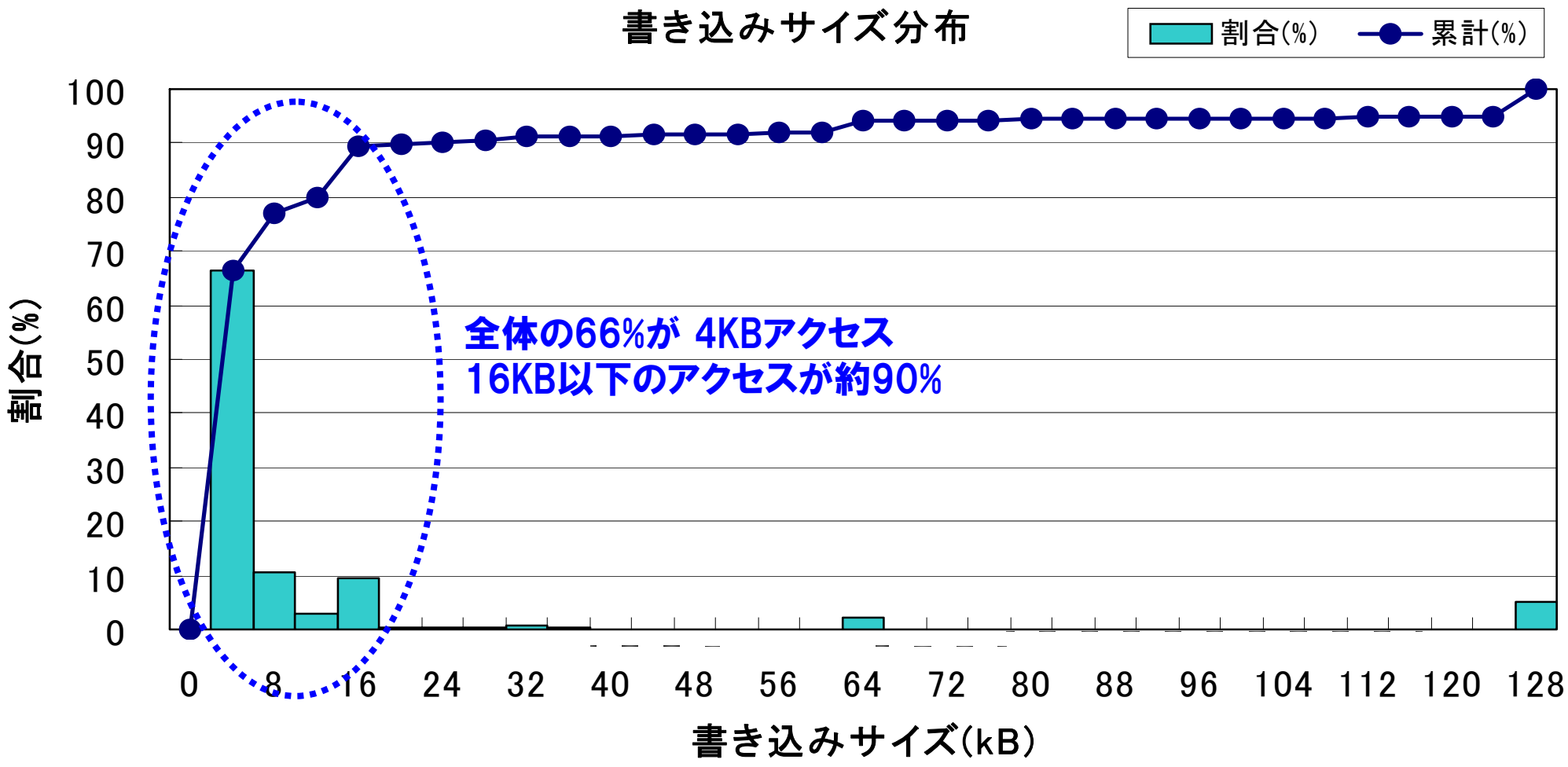
(note 1) Power on hour includes sleep and standby mode.
(note 2) Operating: writing (20GB/Day) and reading.

Applicable warranty and warranty period should be covered by the purchase agreement.

3 Technologies for SSD

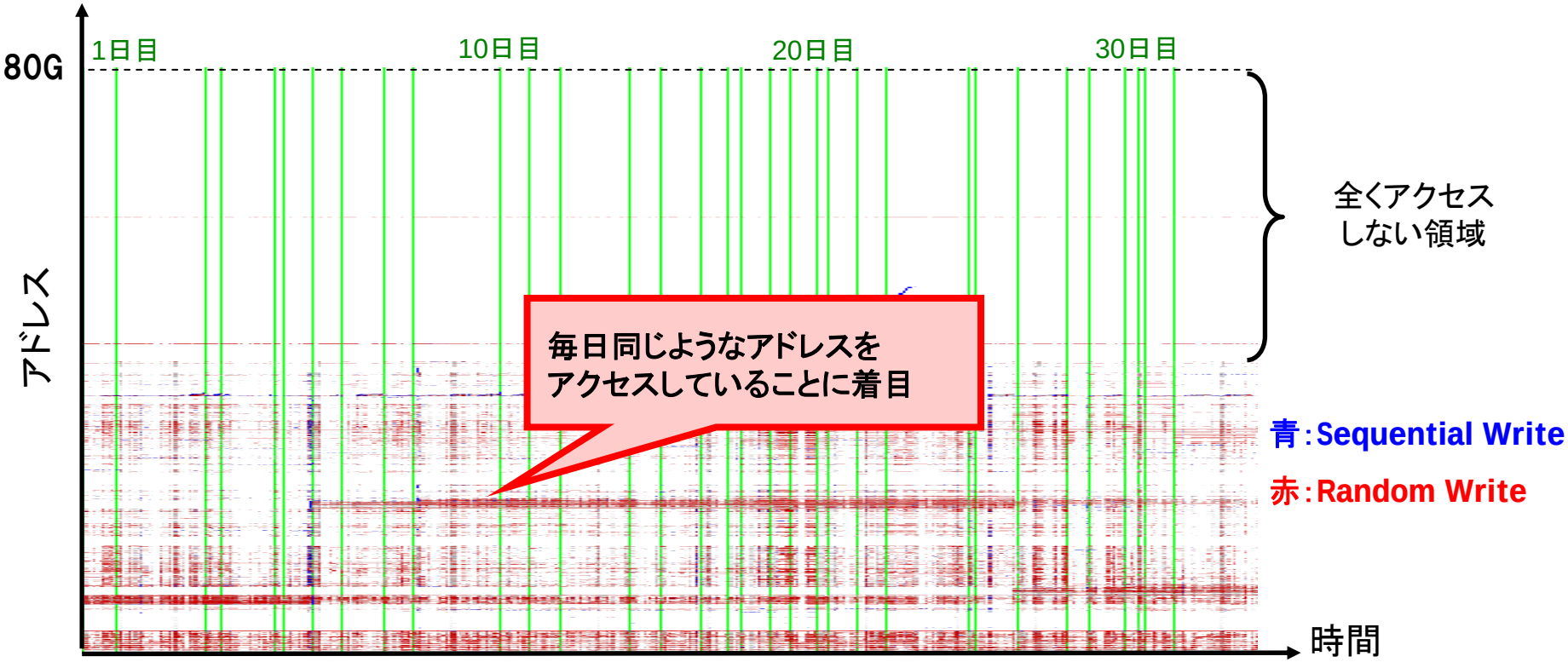
SSDを開発する (1)

■ PCからのアクセスを把握する (PCのアクセスはランダム？)



SSDを開発する (2)

■ PCからのアクセスを把握する (PCのアクセスはランダム？)



アクセスの
時間的・空間的
局所性がある

→

キャッシュが効く

→

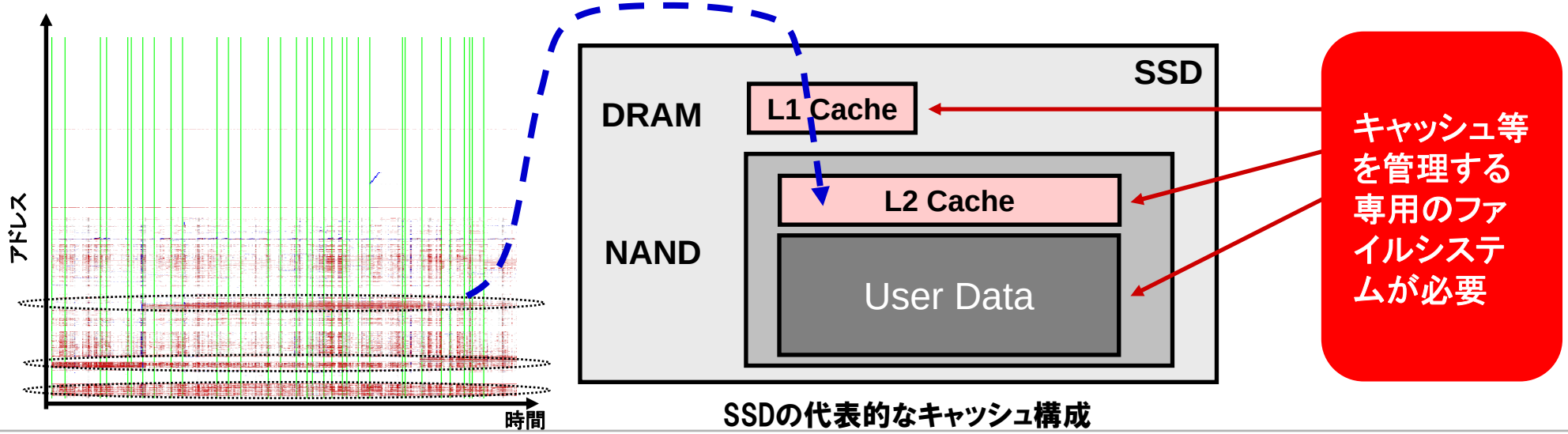
**コンピュータアーキテクチャ
技術が適用できる分野**

SSDを開発する (3)

■ SSD用キャッシュを考察する

- ランダムにキャッシュが効かないのは当然
  - ただし、フットプリントがキャッシュ容量に収まれば別
- DRAMでは収まりきらない
  - 電源切れると忘れてしまう
  - アクセス履歴結果では日を超えてローカリティがある

書き換え頻度が高いデータをうまく扱えるキャッシュを持っているSSDが速い



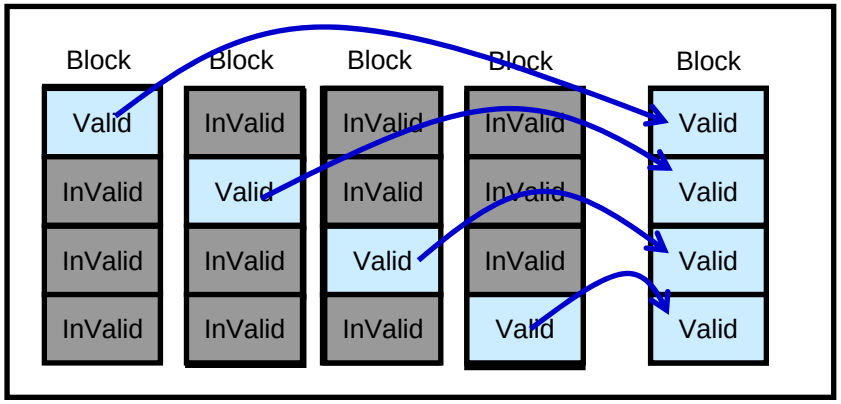
SSDの代表的なキャッシュ構成

SSDを開発する (4)

■ NAND内キャッシュの特徴

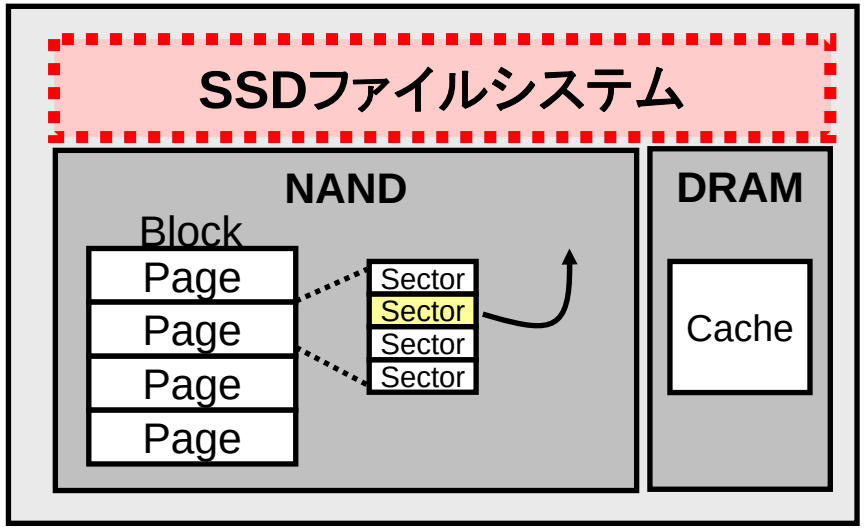
- データの書き換えに消去が必要
  - 上書きができない
  - 更新が必要なデータは消去済みの領域へ書き込み
- ガベージコレクションが必須
  - Validなデータを集めて使用可能なブロックを生成
  - ガベージコレクション頻度を下げることで書き込み効率向上
    - キャッシュ用のブロック量と制御方法が決め手

NAND内キャッシュ



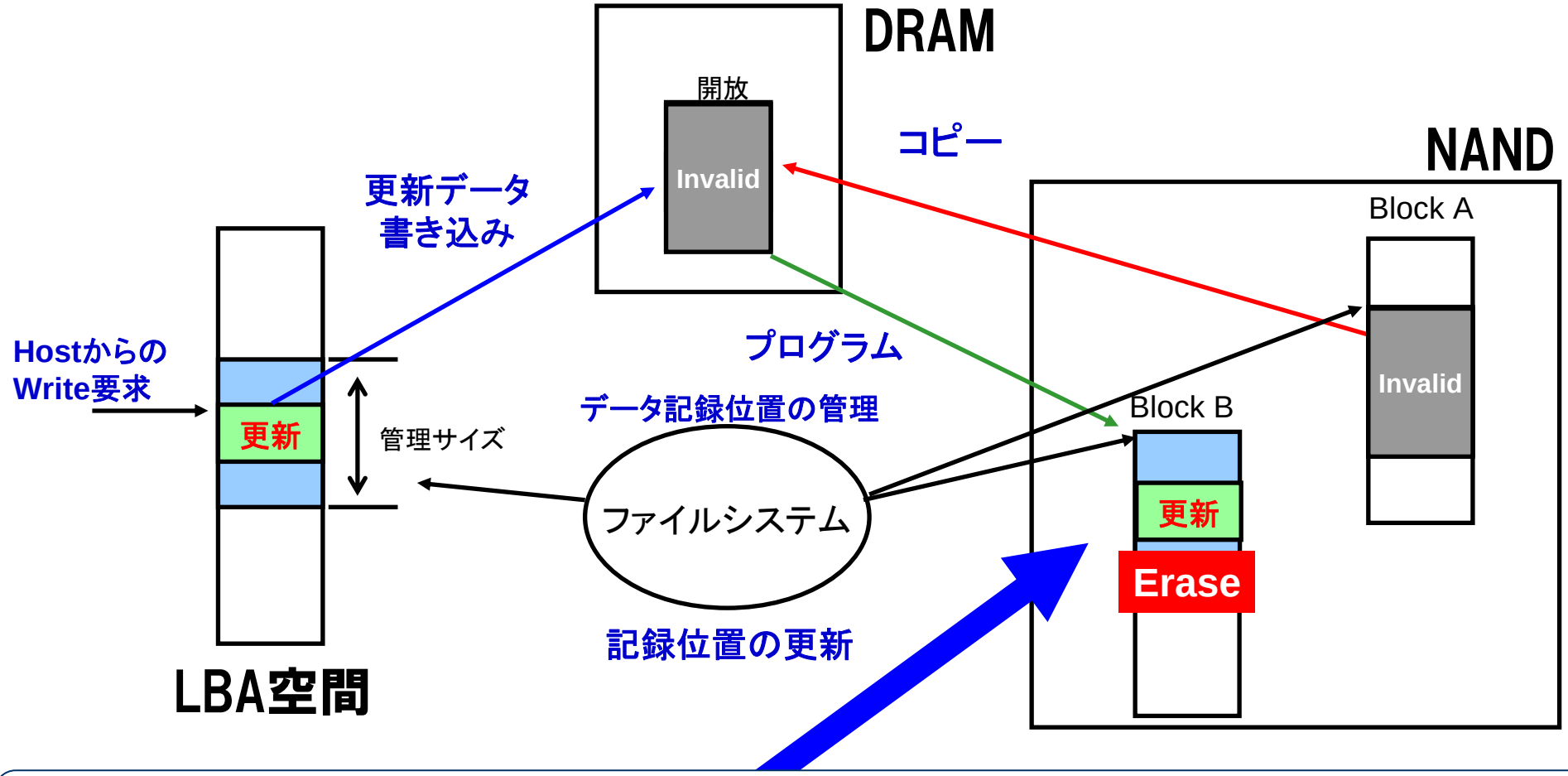
■ SSDファイルシステム

- NANDの特徴を吸収
  - ブロック単位で消去後にプログラム可能
  - ページ単位でプログラム/読み出し
- LBA空間とNAND内データの対応管理
  - セクタ単位でのランダムアクセスを可能
- NAND・DRAMキャッシュ管理
  - 書き込み効率向上



SSDを開発する (5)

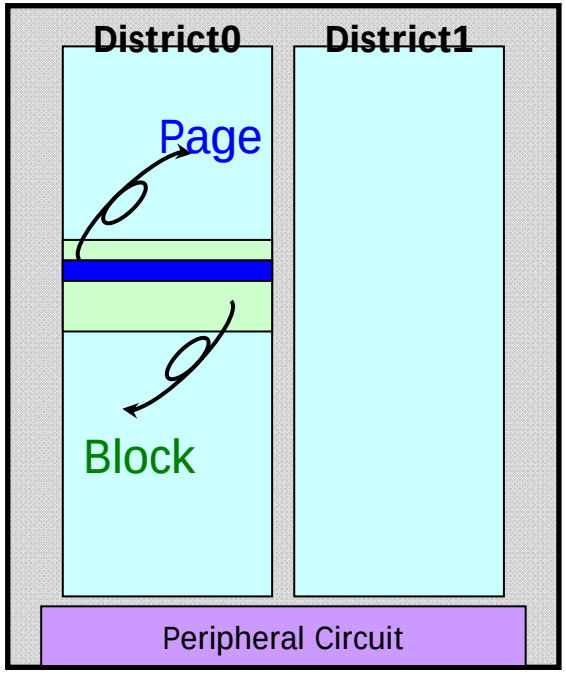
■ 代表的なデータ更新方法



更新したいデータ量と管理サイズが書き込み効率へ影響
例. 管理サイズ8KBに対して4KBの書き込みを行った場合には書き込み量は2倍

SSDを開発する (6)

■ データ管理サイズの選択



- ページ管理だと
  - テーブルがたくさん必要でDRAMに入りきらない
    - 例、8KBページで512GBを管理する場合には64Mエントリ*4B=256MB
  - 書き込み効率は良い
    - 例、8KB管理時に4KB書き込みを行った場合には書き込み効率2倍
- ブロック管理だと
  - テーブルが小さく済む
    - 例、1MBブロックで512GBを管理する場合には512Kエントリ*4B=2MB
  - 書き込み効率が悪い
    - 例、1MB管理時に4KB書き込みを行った場合には書き込み効率256倍

どうやって管理するかが各社のノウハウ

すみません。あまりお話できません

SSDを開発する (7)

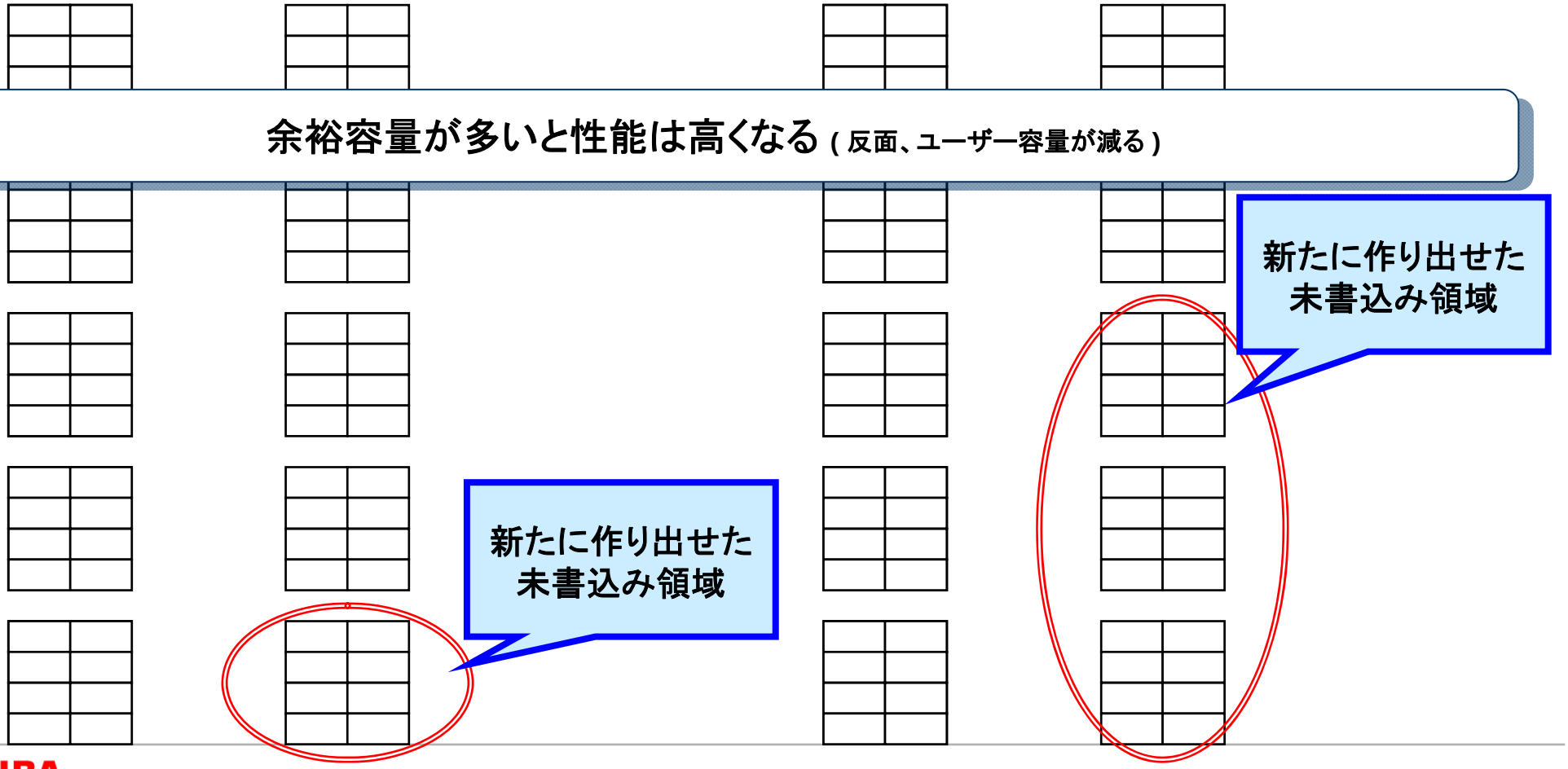
■ 余裕容量と性能

- 余裕容量が多いと、①書き込み速度向上 ②書き換え効率改善 する

Valid Data
Invalid Data
未書込み

◆ 余裕容量少ない

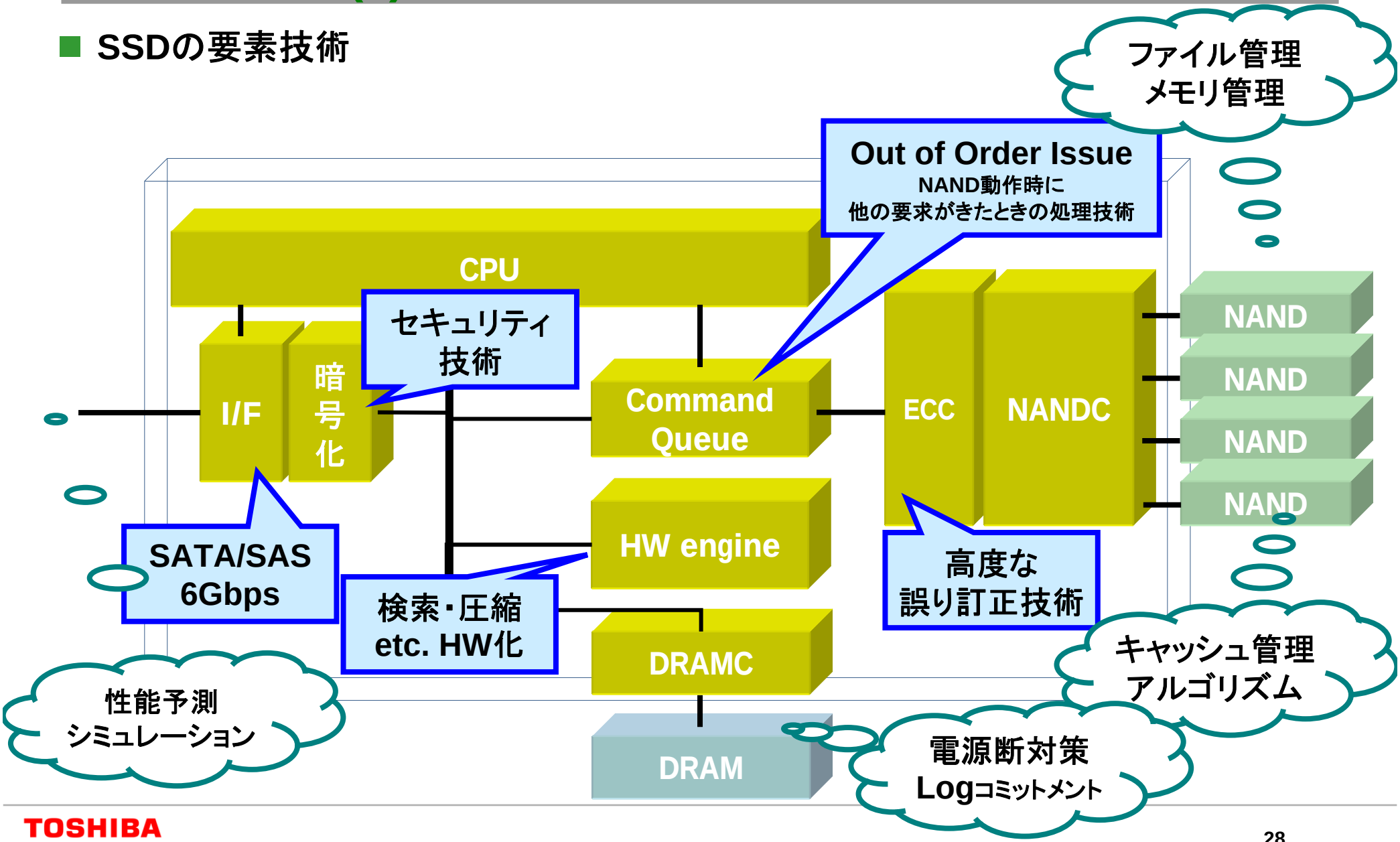
◆ 余裕容量多い



余裕容量が多いと性能は高くなる (反面、ユーザー容量が減る)

SSDを開発する (8)

■ SSDの要素技術



SSDを開発する (9)

■ SSD向けSoC開発

- **NAND世代ごとに開発**
  - エラー訂正の高度化
  - 速度向上の要求
  - I/Fの高速化
  - 先端プロセスで低価格化
- **HWと同時にFWも重要**
  - 単純な処理はHW ： 高速化のため
  - 複雑な処理はFW ： 処理の柔軟性のため
- **コンピュータアーキテクチャ技術の塊**
  - キャッシュ、誤り訂正、並列処理、シミュレーション技術 etc.
- **設計、検証技術**
  - HW/FW Co-design, SWエミュレータ上でのFW開発, Chip全体検証，ランダム検証

4 Future of SSD

SSDの今後 (1)

INVENTOR OF FLASH MEMORY

■ 今後のSSDマーケット発展の為に考えると

- **性能競争はほどほどに**
  - 速けりゃ良いってモノでもなくなりつつあるのでは？
  - 最大速度以外にやるべきことはあるはず
- **小さいのが価値だ**
  - 半導体の特徴を活かしたいろんなForm　Factorが必要とされている
  - 脱HDD互換Form Factor
- **次世代メモリの活用**
  - NANDでないデバイスでSSDを作る
  - 新しいアーキテクチャ、新しいアルゴリズム検討する

SSDの今後 (2)

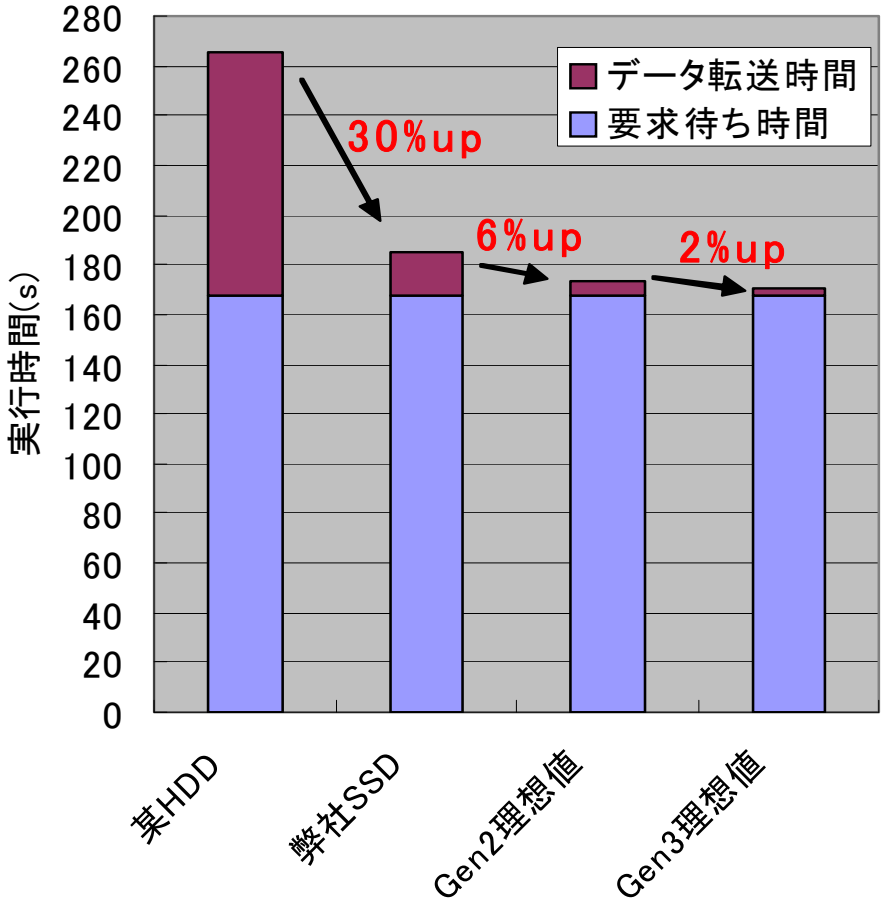
■ Ultra高速SSDの意味

PCMark05実行結果と理想値の予測

	某HDD	弊社SSD	Gen2 理想値	Gen3 理想値
score	5594	26546	90000	180000
XP Startup (MB/s)	8	76	300	600
Application Loading (MB/s)	7	64	300	600
General Usage (MB/s)	6	51	300	600
Virus Scan (MB/s)	97	144	300	600
File Write (MB/s)	67	152	300	600

【評価環境】
PC: DELL INSPIRON 530
OS: Windows XP Pro.

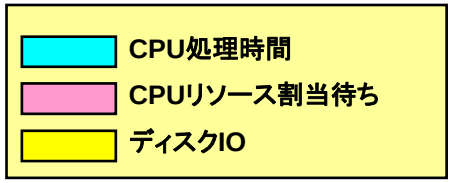
ベンチマーク実行時間の比較



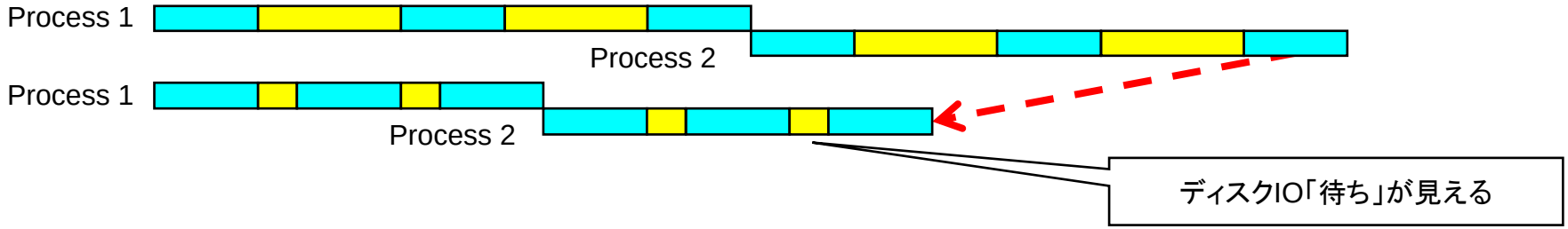
SSDの今後 (3)

■ Ultra高速SSDの意味

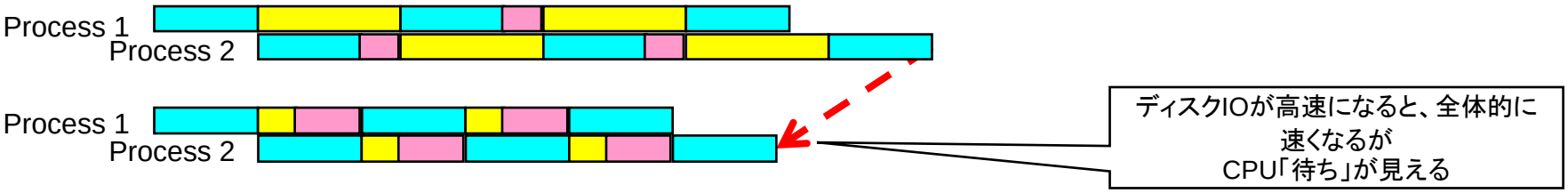
ウイルススキャンやバックグラウンドサービスが動作中に
アプリケーションを起動する場合など。



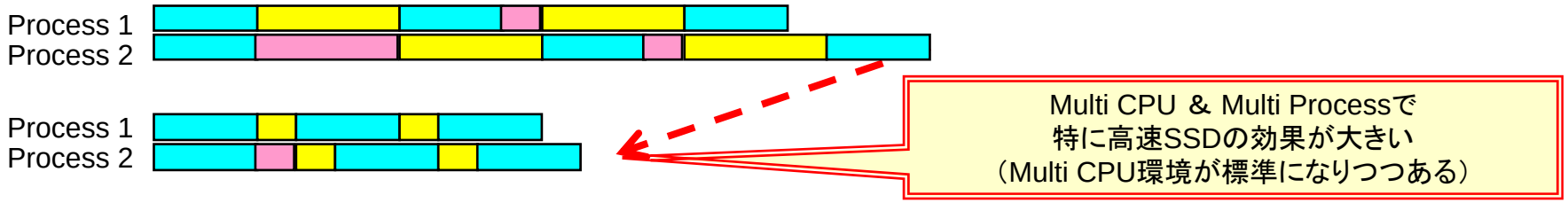
Single Process



Single CPU :: Multi Process



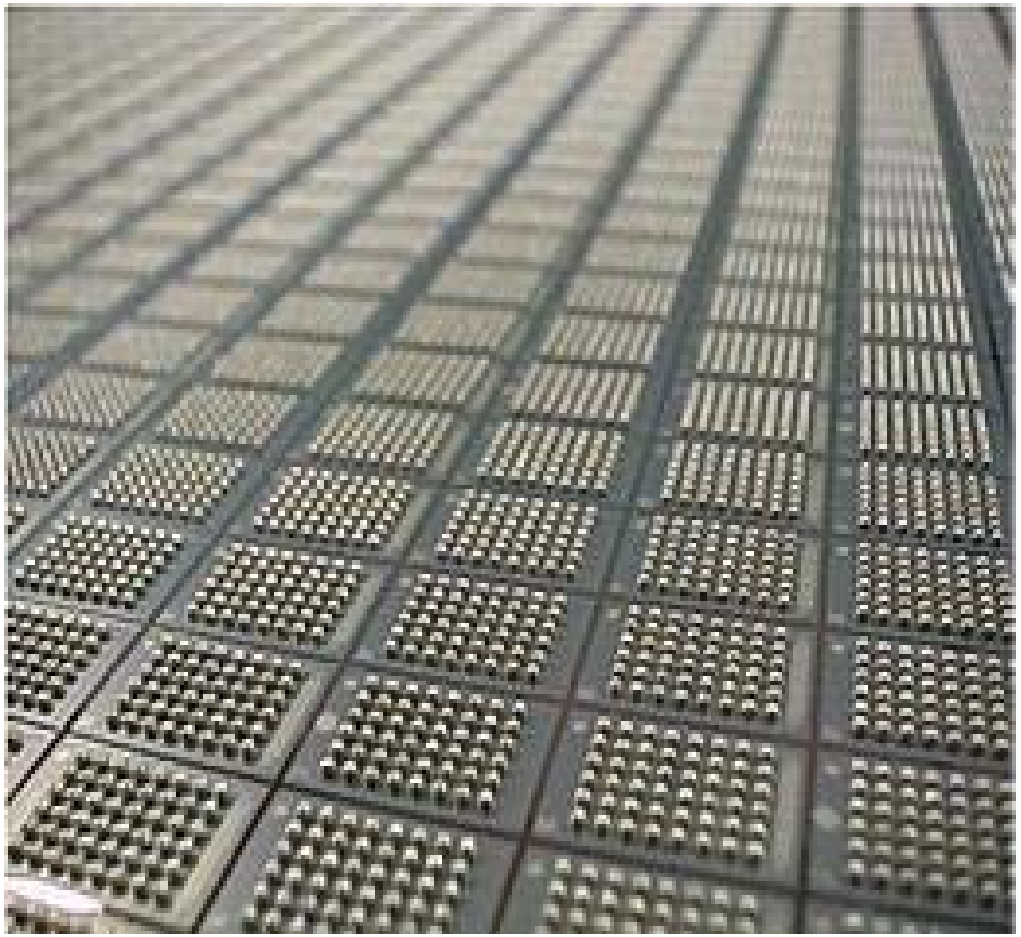
Multi CPU :: Multi Process

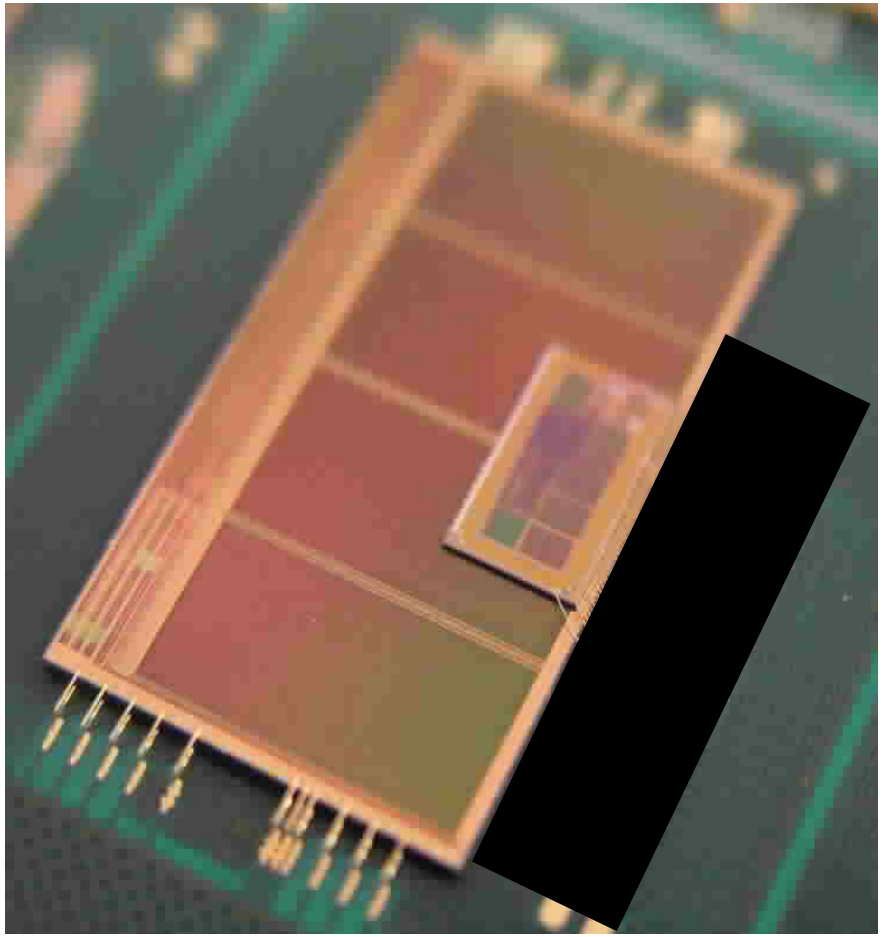


SSDの今後 (4)

■ 小型化への挑戦　　フラッシュカード,MCPでは既に始まっている

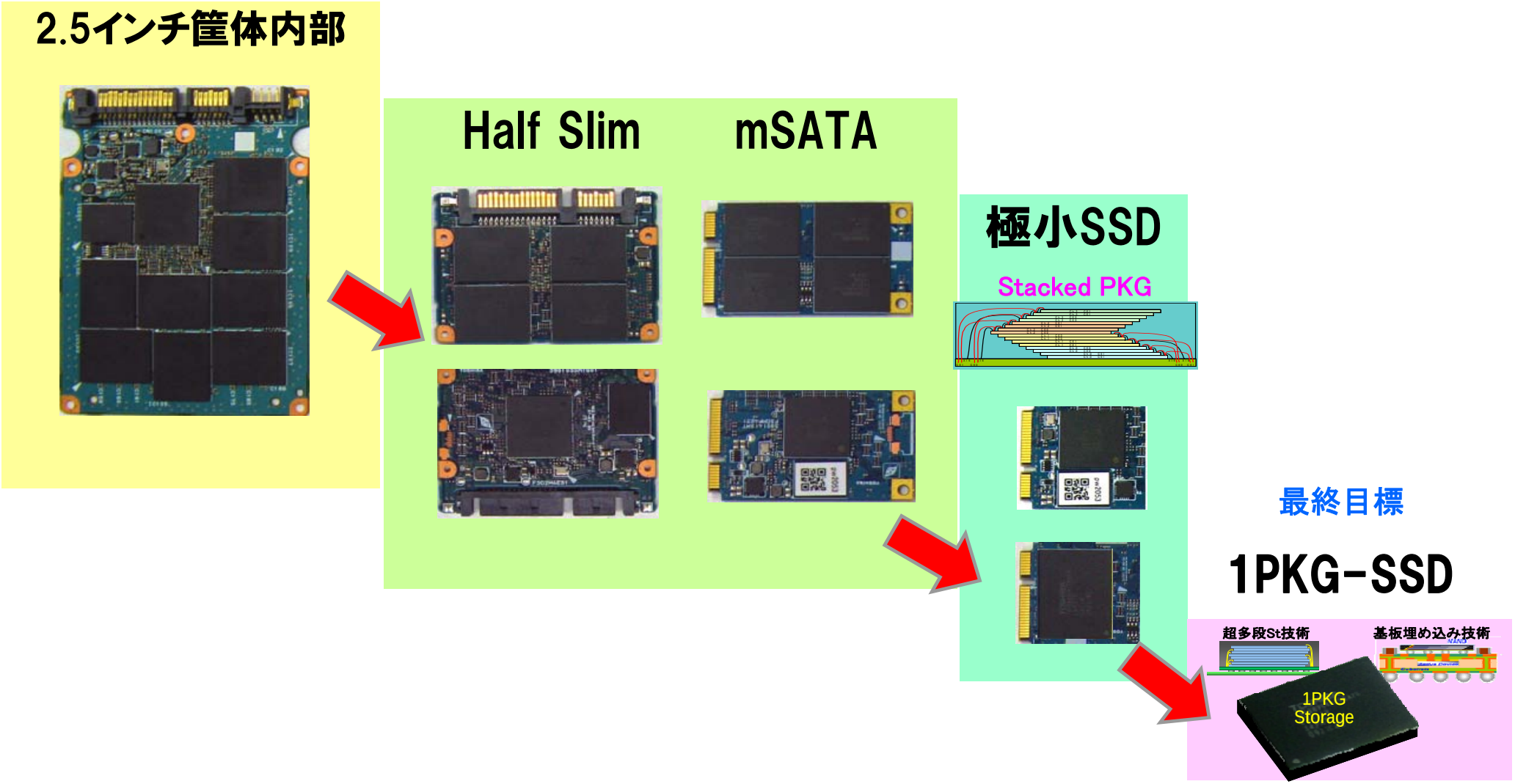
WAFER LEVEL CSP ， Siチップ上への再配線技術を応用





SSDの今後 (5)

■ 小型化への挑戦



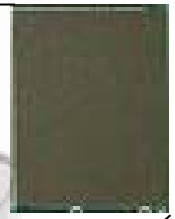
SSDの今後 (6)

■ 小型化への挑戦 (1パッケージ ソリューション)

現状のSSD

実装技術活用

1PKG SSD



14x18mm
252mm²

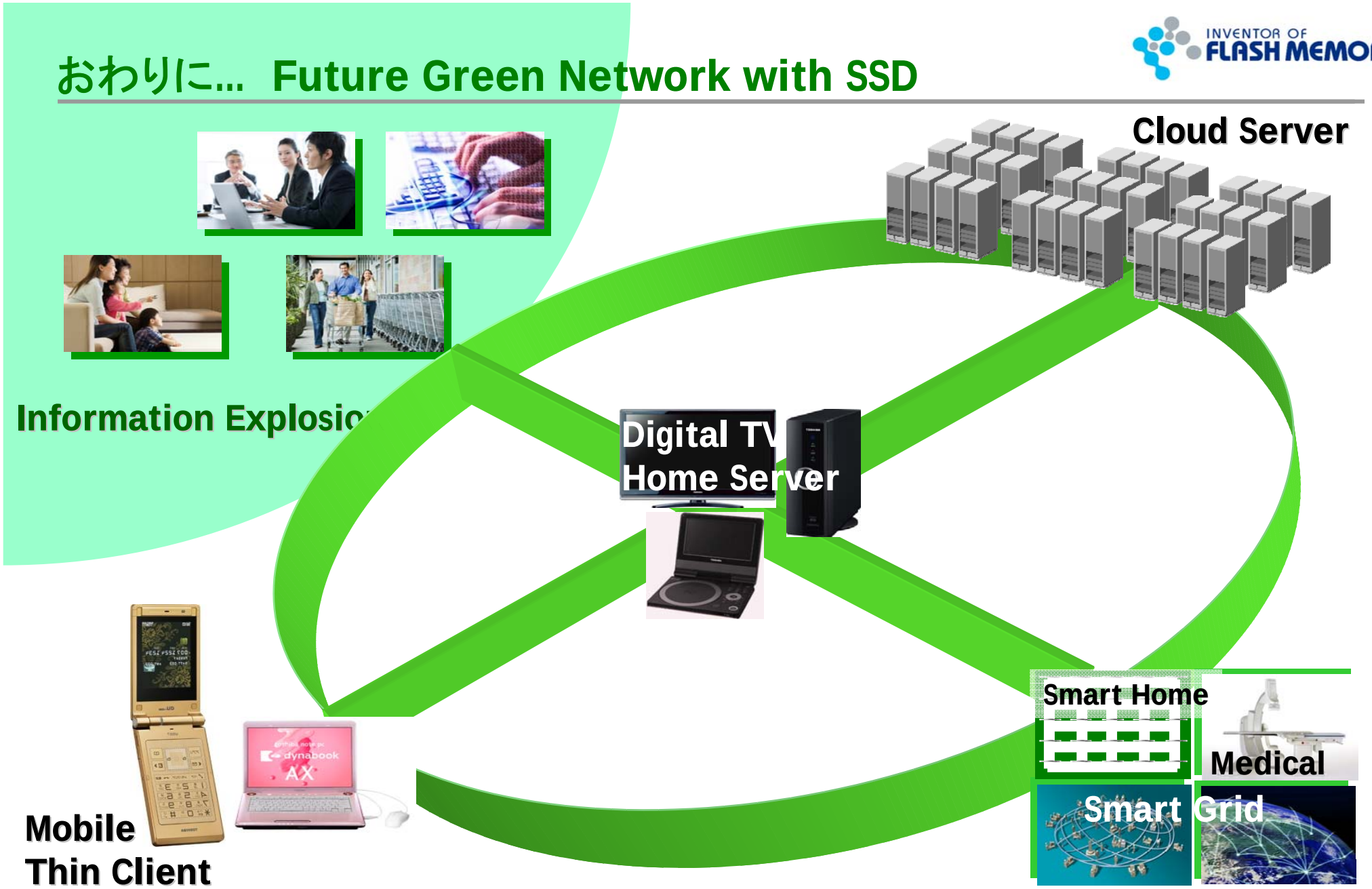
実装面積

$$\frac{1}{15}$$

極薄技術活用

実装技術のみならず、極薄技術、
歩留まり、Test技術、熱問題 etc
様々な技術の開発、課題の克服
が今後のSSD発展に不可欠である

おわりに… Future Green Network with SSD



ご清聴有難うございました

Last Page