



今回はKIOXIA本社で話を伺った。左:SSD事業部 SSD応用技術部 リテールSSDマーケティング技術担当 津秦伴紀氏、右:営業本部 パーソナルストレージ企画部新規ビジネス担当 藤田祐己氏

SSD製品に記載されている「TBW」という指標。筆者を含め多くの人が、このTBWを“これだけの容量を書き込んでも問題がないことを示すもの”と漠然と捉えているかもしれないが、実際はどういった指標なのか。

今回、キオクシアでSSD製品の開発に携わっている、SSD事業部 SSD応用技術部 リテールSSDマーケティング技術担当の津秦伴紀氏と、SSD事業部 SSD応用技術部の石井賢一氏に話を聞いてきたので、紹介する。

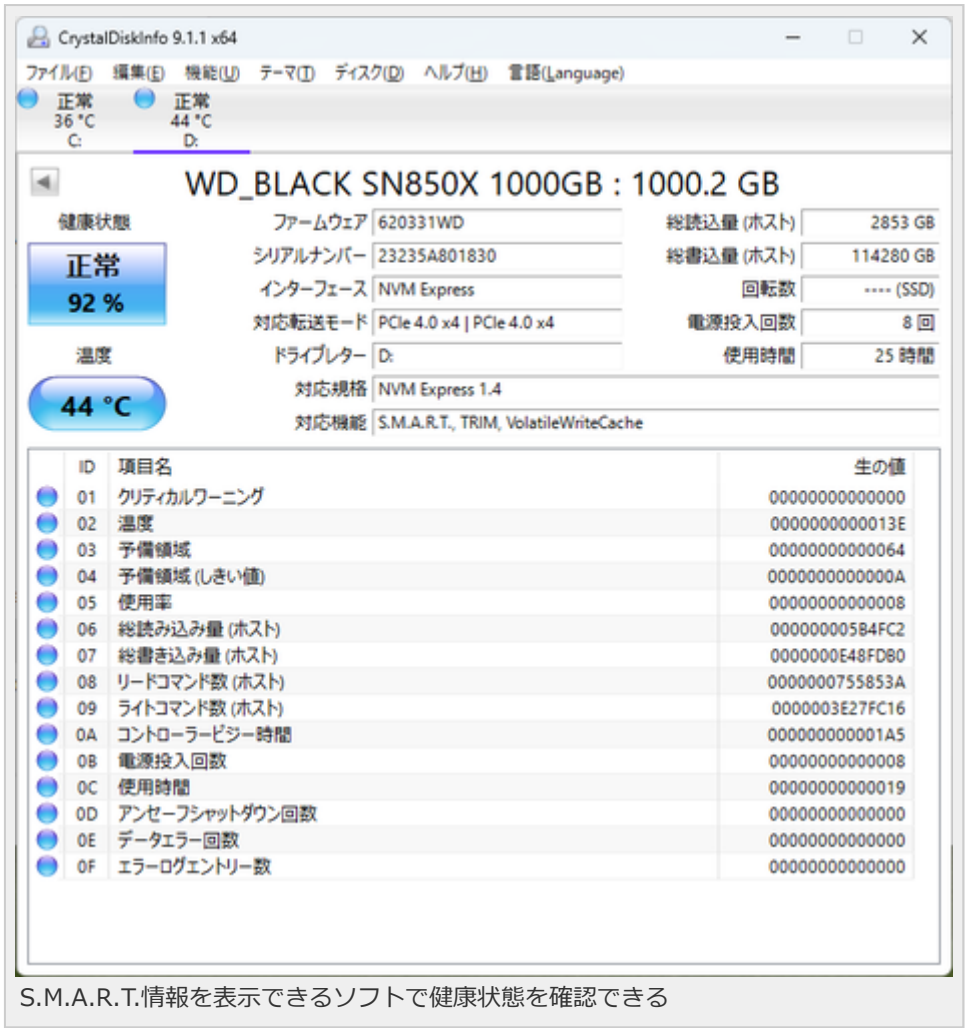
記事を要約する(AI)

3行まとめ

SSDのTBWは、JEDECが定める規格に沿って書き込めるデータの総容量

Q:まずはじめに、そもそもSSD製品に記載されている「TBW」というものがどういう意味を持つ指標なのか、説明していただけませんか

A:TBWは、あくまでもその製品に対してどのぐらいの容量のデータを書き込めるか、という指標です。実際に、SSDに書き込める容量はユーザーの使い方によって変わります。しかし、ユーザーによって変わってしまうとはいえ、どの程度の容量が書き込めるか分からないというのは不安ですから、半導体素子などの規格標準化を行なっている業界団体「JEDEC」が定める規格に沿って書き込んだ場合に、これぐらいの容量が書き込めますよ、という指標となっています。



S.M.A.R.T.情報を表示できるソフトで健康状態を確認できる


キオクシアの場合、SSD製品は、製造してからの期間と、S.M.A.R.T.情報で見える健康状態の2つで保証を行なっています。ただ、それだと分かりにくいということもありますし、書き込める容量を示すにしても、ユーザーの使い方によって書き込める容量が大きく変わ

ってしまいますので、同じ条件、同じ規格で書き込んだ場合にどれだけの容量が書き込めるのか、ということ把握していただけるようにTBWを示しています。

Q:そのJEDECが定めるTBWの規格とはこういったものののでしょうか

A:JEDECが定めるTBWの規格として、クライアント向けSSD製品では「JESD218A」と「JESD219」という規格を用意しています。JESD218Aはテスト内容の詳細を、JESD219はクライアント向けSSDでの検証用のワークロードが定められています。

そして、一般ユーザー向けのクライアント向けSSDでは、この規格を利用してTBWが求められています。これら規格が定められたのが2010年頃です。




Endurance Rating (TBW Rating)


The SSD manufacturer shall establish an endurance rating for an SSD that represents the maximum number of terabytes that may be written by a host to the SSD, using the workload specified for the application class, such that the following conditions are satisfied:

- 1) The SSD maintains its capacity
- 2) The SSD maintains the required UBER for its application class
- 3) The SSD meets the required functional failure requirement (FFR) for its application class
- 4) The SSD retains data with power off for the required time for its application class

From JESD218A, Copyright JEDEC. Reproduced with permission by JEDEC




TBWの定義(出典:Flash memory Summit 2011,JEDEC SSD Endurance Workloads,Alvin Cox, Seagate)




Endurance Rating (TBW Rating)

- Establishes a rating system for comparing SSDs
- Provides ratings based on application class
- Rating for user-measurable interface activity

This rating is referred to as TBW. Requirements for UBER, FFR, and retention are defined for each application class.




TBWの定義(出典:Flash memory Summit 2011,JEDEC SSD Endurance Workloads,Alvin Cox, Seagate)



JEDEC SSD Standards

- **JESD218A: Solid State Drive (SSD) Requirements and Endurance Test Method**
 - SSD Definitions
 - SSD Capacity
 - Application Classes
 - Endurance Rating
 - Endurance Verification
- **JESD219: Solid State Drive (SSD) Endurance Workloads**
 - Client
 - Enterprise



JESD218AとJESD219(出典:Flash memory Summit 2011,JEDEC SSD Endurance Workloads,Alvin Cox, Seagate)

Q:では、実際の製品に記載されているTBWは、どのように導き出しているのでしょうか

A:SSDの中で一番消耗される部分はNANDです。NANDの場合は通電時間というよりも書き込み容量によって消耗していきますので、そちらをJEDECの規格に沿って検証します。

JEDECの規格では、SSDに対してどういう書き込みのテストを行なうのか明確に規定されていますので、その規格に沿って実際に書き込みを行ないます。その後、電源を切っても書き込んだデータがきちんと保持されるかどうかを確認することも規定されていますので、そちらもきちんと確認した上で導き出しています。

キオクシアでは1カ月以上の期間をかけ実際にデータを書き込みチェックしている

Q:テストを行なう際のワークロードの内容はどういったものなのでしょうか

A:まず最初に、SSDの容量全てにデータを書き込みます。その状態で、JEDECが規定しているワークロードのパターンに従った書き込みなどを行なっていきます。

SSD中のデータは「LBA(Logical Block Addressing)」というアドレスで管理していますが、特定のアドレスに対してこれだけの容量のデータを書き込みなさい、というものが膨大なパターンとして用意されています。

データ量は大きいものから小さなものまでさまざまに指定されていますし、その途中にはガベージコレクション(物理的に消去してもいいデータのアドレスを知らせる)というコマンド「Trim」を実行することなども指定されています。

こういった、細かなワークロードのパターンを規定しているのがJESD219ですが、我々の製品では、その規格に定められたパターンのワークロードを繰り返し実施してTBWを導き出しています。



SSD Endurance Classes And Requirements

Application Class	Workload	Active Use (power on)	Retention Use (power off)	Functional Failure Rqmt (FFR)	UBER
Client	Client	40°C 8 hrs/day	30°C 1 year	≤3%	≤10 ⁻¹⁵
Enterprise	Enterprise	55°C 24hrs/day	40°C 3 months	≤3%	≤10 ⁻¹⁶

From JESD218A, Copyright JEDEC. Reproduced with permission by JEDEC



クライアントとエンタープライズでも要求が異なる(出典:Flash memory Summit 2011,JEDEC SSD Endurance Workloads,Alvin Cox, Seagate)



Client Workload – JESD219

- The client workload consists of the following:
 - Precondition phase
 - Write all user LBAs
 - Testing at 100% full still under discussion
 - Run the test trace
 - Based on standard ATA I/O commands
 - Adjusted to stay within the SSD user capacity
 - Includes all commands from a real trace capture
 - The trace only has commands, including TRIM
 - Data payload is generated separately as random data
 - Replay the test trace to verify TBW rating



JESD219のクライアントワークロード(出典:Flash memory Summit 2011,JEDEC SSD Endurance Workloads,Alvin Cox, Seagate)



JESD219 Enterprise Workload

- Adjusted SPC-1C transfer size distribution to include 10% <4K transfers
 - Trace data indicated many applications have some non-alignment characteristics
- Random data is used for the data payload

512 bytes (0.5K) 4%
1024 bytes (1K) 1%
1536 bytes (1.5K) 1%
2048 bytes (2K) 1%
2560 bytes (2.5K) 1%
3072 bytes (3K) 1%
3584 bytes (3.5K) 1%
4096 bytes (4K) 67%
8192 bytes (8K) 10%
16,384 bytes (16K) 7%
32,768 bytes (32K) 3%
65,536 bytes (64K) 3%




JESD219のエンタープライズワークロード(出典:Flash memory Summit 2011,JEDEC SSD Endurance Workloads,Alvin Cox, Seagate)

Q:非常に細かいワークロードが指定されているということですが、実際の検証で書き込むデータの内容についても規定されているのでしょうか

A:実際に書き込むデータの中身についてはJEDECは用意していませんので、SSDメーカーが独自に用意したものを利用しています。

我々の製品では、ランダムデータを用意して利用しています。SSD製品によっては、データの圧縮技術を利用しているものが存在していますので、そういった場合にはランダムデータを利用する場合と、同じ内容が続くデータでは結果が異なってしまいます。ですので、最も厳しい条件となるようにランダムデータを使っています。





Workloads & SSD Applications

Endurance is dependent on the applications being run and the workloads presented to an SSD


Client SSD	Enterprise SSD
<ul style="list-style-type: none">• 8/7 operation, bursty workload (considerable relaxation times)• Modest write activity• More localized access footprint, smaller than entire LBA space• Optimized cost	<ul style="list-style-type: none">• 24/7 operation, continuous workload• Heavy write activity• Mostly random accesses across the entire LBA space• Highest data reliability, availability, and integrity

Endurance ratings must reflect the expected use cases. JEDEC workloads and testing are targeted at achieving this



Workloads And SSD Applications

- Enterprise and Client SSDs have different design points, application uses and workloads
- Different endurance rating methods are required for the for the different use cases
- [JESD218A](#) and [JESD219](#) define the methods to derive a means of comparing the endurance ratings of devices within each class of device



JEDECの規格でも、アプリケーションの違いによるワークロードを想定している(出典:Flash memory Summit 2011,JEDEC SSD Endurance Workloads,Alvin Cox, Seagate)

Q:そして、その検証はどの程度の時間をかけて検証するのでしょうか

A:我々としては、製品ごとにTBWのターゲットがあらかじめ決まっていますので、そのターゲットをちょっと超えるぐらいまでの容量を書き込むまで行なっています。製品の容量にもよりますが、1カ月以上の時間をかけています。

正直に言うと、以前は容量が少なかったので、そんなに時間はかかっていなかったんですが、最近の製品は容量が大きくなっているのので、どうしても時間がかかってしまいます。

また、テスト内容がシーケンシャルライトだけだったら一気に書き込めるんですが、ランダムライトも行ないますから、書き込み速度も遅くなってしまうので、その点でも時間がかかってしまいます。

Q:製品のTBWはシミュレーションで導き出す部分が大いと思っていますが、そうではないんですね

A:もちろん、設計上このぐらいいは書けるだろうとシミュレーションで予測はします。ただ実際の試験は、TBWだけ書いても問題がないかを確認するだけではなくて、その間にいろいろな不具合が起きないか確認する長期ランニング試験も兼ねていますので、実際に長期間ずっとテストを続けます。



津秦氏

Q:ちなみに、何台ぐらいでテストを行なうのでしょうか。またテストにはこういった機材を使ってるのですか

A:製品によって変わりますが、リテール向けの製品では、数十台でテストを行なっています。

テスト機材については、OSの違いというよりもトータルでの書き込み容量や書き込みのパターンの方を重視しています。我々はテスト機材としてLinux環境の機材を利用しています。

■ データを書き込むだけでなく書き込んだデータの保持期間もチェック

Q:先ほどTBWのテストでは書き込める容量だけではなくて、データが問題なく保持されているかという部分もチェックするということでしたが、そちらはどのように検証しているのでしょうか

A:それは、書き込みのテストが終わった後に行ないます。JEDECの規定では、テストを行なった後に、SSDに通電することなく30℃の環境で1年間データが保持できるか確認するという規定がありますので、そちらを確認します。

ただ、実際に1年の時間はかけられませんので、そこは30℃で1年間が経過したことに相当する温度加速試験を行なって、その後に全てのデータを読み出して読めないデータがないか確認します。

こういった点についてもJEDECの規格で細かなところまで規定されています。



Q:ちなみに、御社の製品では、JEDECで定められているもの以上の試験も行なっているのでしょうか

A:そのあたりは詳しくお答えできませんが、我々はNANDも製造していますので、ファームウェアをどう改善していけばいいのか、コントローラをどのように使えばいいのか、というところですね。

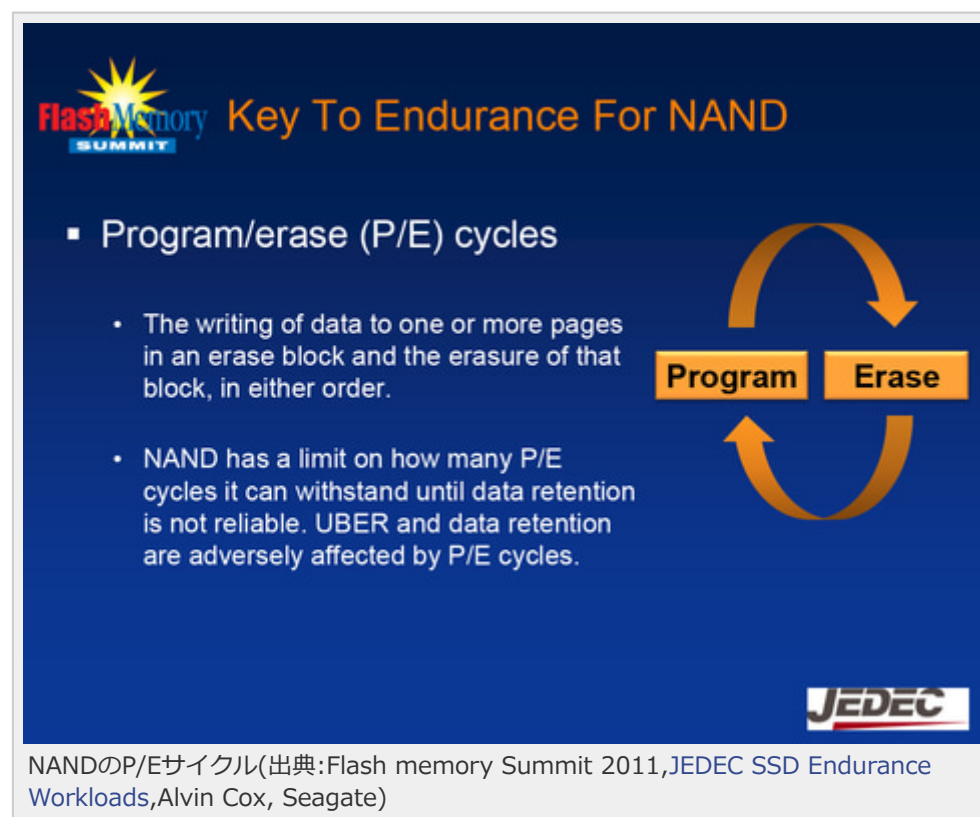
NAND自体についても、コマンドをどう使えばいいのかといったところをNANDの設計部隊と話をしながら検証を行なっています。試験の実施内容という意味ではあまり見えてこない部分ではありますが、それを支えるバックグラウンドの技術として、さまざまな知見を持っているというのが他社に負けない部分だと自負しています。

たとえば、我々は他社には出せないNANDの情報も持っていて、そちらも考慮して製品を開発しています。ですので、NANDを自社で開発しているという部分は、弊社の大きな強みとなっています。

■ TBWはSSD選択の上で重要な要素ではない

Q:TBWを決定づける要素としては、NAND、コントローラ、ファームウェアなどSSD全体に関わると思いますが、その中でも特に重要なものは何になりますか

A:TBWの値を大きくしようとするのであれば、NANDの耐久性がやはり最も重要となります。ただ、容量が増えたことで試験時間が延びてきていますし、開発段階でも“こんなにTBWっているんだっけ?”という話もあったりしますし、“TBWの数値ばかり求めても、製品としてはそんなに訴求効果ないんじゃないか”という意見もあります。



そうすると、TBWはこのぐらいで頭打ちでもいいよね、ということにして、製品開発期間を短くするために、それ以上TBWを確認しない、ということも考える必要があるかもしれません。

もちろん、競合製品と比べて劣らないところまでは確認しますが、限界まで見ることはしていない、というのが事実です。

Q:以前の製品に比べてSSDのTBWは格段に大きくなっているように思いますが、耐久性が高められている要因はどういった部分にあるのでしょうか

A:NANDが3D NANDに進化したことや、NANDへの書き込み技術が向上したということもありますが、最大の要因は、SSDの容量が大きくなった、というところですね。

昔から、実際にSSDを使う上で、書き込み回数はそれほど変わっていないんです。SSDでは書き込みを平準化する「ウェアレベリング」という機能もありますし、全体容量が大きいと書ける容量が増えますので、当然TBWも増えることになります。

性能差によるSSD選び

Q:お話しをお聞きしていると、通常我々がPCでSSDを使う範囲内であれば、製品に記載されているTBWを問題なくクリアできると思いますが、そうならないような使い方というのはどういったものがありますか

A:そうですね、たとえば4Kランダムライトをずっとやり続ける、といったことでしょうか(笑)。4Kランダムライトを全領域に対して延々とやり続けるといったものはJEDECの規格では定義されていませんから、おそらくTBWからは外れたものになると思います。

ただ、そういった使い方をするユーザーがいるかという疑問です。

Q:近年、NASなどでHDDレス、オールSSDの製品が登場していますが、そういった使い方でも特に問題はないと考えていいのでしょうか


A:NASのようなストレージですと、どちらかというと細かなデータよりも大容量のデータを扱う場面が多いと思いますので、一般のPCに比べるとSSDへの書き込みパターンはそれほど悪くないのではないかと思います。

NASは24時間動きっぱなしになるので、通電時間という点ではPCのほうが優しいことにはなります。そもそもクライアント向けSSDではそういった用途を想定していませんので、連続通電時間などの検証は行なっていません。ただアクセスパターンとしては、PCの場合はテンポラリファイルを作ったりするなどで、細かなアクセスが多くなりますので、PCのほうが厳しいと思います。

NANDは、16KB単位でページを持っていて書き込みを行っていますが、細かい書き直し、たとえば64Byteの書き戻しという作業はNANDに対して非常に負荷が高いのです。そういったリードモディファイライト命令のようなワークロードが多い方がNANDとしては苦手です。

ホストからはランダムデータの容量の小さいデータしか書いていないんですが、SSDの中は16KB単位で管理していますので、どうしても作業が増えてしまいます。


データを書き換える場合、それが16KBのページより小さいサイズの場合には、その部分だけを新しく書き換えることができないので、ページ全体を読み出して特定部分を更新し、ページ全体で書き戻すことになります。その作業では、ホストから見えないコピー作業が発生していますので、内部の書き込み容量が増えることになるのです。



WAF – A Key Endurance Factor

The higher the WAF, the faster the Flash wears out

- **Write Amplification Factor (WAF)**
 - The amount of data written to the NVM divided by the amount of data written by the host to the SSD
- WAF is directly associated with P/E cycles
 - Higher WAF = More P/E cycles happen
- An SSD usually writes more data to the memory than it is asked by the host to write
- The nature of the workload has a significant impact on WAF



Write Amplification Factor(書き込み効率)がNANDの寿命に直結する(出典:Flash memory Summit 2011,JEDEC SSD Endurance Workloads,Alvin Cox, Seagate)

Q:では、ユーザーがSSDを選ぶ際に、こういった部分をチェックすべきとお考えでしょうか

A:そのあたりは、ユーザーさんがSSDを何に使いたいのか、というところが一番重要だと思っています。たとえば、ゲームをしたいです、動画編集がしたいですといった場合に、どれだけの性能があれば十分なのかを考えて、製品を選んでいただければいいと思います。

たとえば性能ですが、DRAMキャッシュを搭載している製品では性能が高くなりますので、ハイエンド向けの製品がいいという場合にはそちらを選択すればいいと思います。DRAMの有無は速度が変わるだけで品質には差がでませんので、DRAMの有無で品質について気にする必要はありません。

弊社の製品ですと、エントリー向けの製品はオレンジ、ハイエンド向けの製品はマゼンタと色分けしていますので、選びやすいかと思います。とにかく、ユーザーさんのユースケースによってそのあたりは大きく変わってくると思います。



ハイエンド向けとエントリー向けでパッケージの色を変えているKIOXIA製品

日常的なオフィスワークが中心ということでしたら、エントリーからメインストリームの製品を使っていただく、また動画編集とかゲームなど、大量に速くデータを書き込んだり速くデータを読み出したいというのであればプロ向けがお勧めです。

製品のグレードによって品質はほとんど差がありません。違いがあるのは性能と消費電力です。エントリー向けの製品は消費電力が少ないですから、ノートPCに搭載する場合にはエントリー向けの製品の方がバッテリー駆動時間が伸びると思います。

■ 放置していたSSDが認識されなくなった理由は



Q:個人的に、過去に何度かSSDが死んだことがあります。その中で、使わずに長期間置いていたSSDをPCに搭載しても全く認識されなかった、といったことがありました。SSDを保管する場合には何に気を付ければいいのか

A:SSDをどれだけ保管できるかは、温度によります。NANDは温度が高くなるとデータの保持能力が下がる特性があります。ですので、長期間SSDを保管するのであれば、温度の低い状態で保管した方がいいです。

長期間置いておいたSSDが認識されなかったということですが、SSDは電源から外して保管する状態が一番持たないのです。

SSDのファームウェアはNANDの中に保存されています。基本的にファームウェアは多重化されていて一方に問題が発生しても大丈夫なようになっていますし、そもそもファームウェアを保存している領域は全く書き換えが発生しない部分なので、通常はかなりの耐久性があります。

しかし、もし長期間保存している間に保存環境が悪くてNANDの保持能力が下がり、運悪くファームウェアを保管している部分のデータが保持できなくなった場合には、SSDとして機能しなくなってしまうのです。おそらくそういったことが発生したのだと思います。

PC Watchが過去に実施したSSD耐久性テストは意味があったのか？

Q:ところで、PC WatchであるSSD製品の耐久性テストを行なった記事を書いたのですが、そちらをご覧になってどのように思われましたか

関連記事



米中SSDの耐久対決、3カ月書き続けてついに決着。勝ったのは……



安い中国製SSDってどう?米大手とガチンコ対決!

A:あの試験内容では、TBWを語ることはできないと思っています。信頼性については、あの試験内容を見る限りではどちらもNANDの限界には達していないので、あの時点でNANDの信頼性は分からないと思います。

そもそもNANDはデータを書けなくなるよりも、データを保持できなくなるものなのです。データを書けなくなるというのは相当なことです。偶発的な不良で書けなくなることはあるんですが、疲弊して書けなくなることはまずほとんどないです。

多くの場合はデータが保持できなくなって読み出したときにデータが化けるので、たとえば30℃で1年持つものが30℃で1カ月ぐらいしか持たなくなるとか、そういったところを確認しないといけないと思います。


またやるとしたら、やはりJEDECが規定しているワークロードの試験を行なうのがいいと思います。本当なら温度加速試験で保持期間もチェックすべきですけど、そちらは個人レベルではなかなか難しいでしょう。

あと、中国製SSDのほうは書き込み速度が低下していましたが、そちらはSLCキャッシュが使えなくなってTLC NAND直書きになったためでしょう。SLCキャッシュは、静的にSLCキャッシュ領域を確保しているものと、動的にSLCキャッシュ領域とTLCのデータ領域を切り替えている場合があって、SLCだと耐久性は10倍以上になりますが、後者だとTLCとして使う場合の品質も確保する必要があるので、SLCキャッシュとして使える期間が短くなります。これが速度低下の要因と考えられます。

Q:実際にあのテストは、データの保持については全く検証できていけませんので、あくまでもこれだけ書けましたよ、ということを示しているだけ、ということですね。

今後同様のテストを行なう場合には参考にさせていただきます。本日はどうもありがとうございました。

編集部のおすすめ記事



米中SSDの耐久対決、3カ月書き続けてついに決着。勝ったのは……