

swissbit[®]

Application Note

AN4105de

Host Controlled Thermal Management (HCTM) for NVMe SSDs

© Swissbit AG 2025

  Creative-Commons-Lizenz¹

¹ Dieses Werk steht unter der Creative-Commons-Lizenz „Namensnennung 4.0 International“. Um eine Kopie dieser Lizenz zu sehen, besuchen Sie <http://creativecommons.org/licenses/by/4.0/>

Inhaltsverzeichnis

1 Zusammenfassung	2
2 Einleitung	2
3 Grundlagen	2
4 Temperaturdefinitionen im NVMe-Kontext	3
5 Funktionsweise von HCTM	3
6 Implementierungsbeispiel mit nvme-cli	3
6.1 Prüfen ob HCTM unterstützt wird	3
6.2 Setzen von Thermal-Management-Schwellen	4
6.3 Überprüfung mit Get-Feature	4
7 Vorteile von HCTM	4
8 Bewährte Verfahren	5
9 Anwendungsfälle	5
10 Fazit	5

1 Zusammenfassung

Das Host Controlled Thermal Management (HCTM) ermöglicht es, das thermische Verhalten von NVMe-SSDs an die Kühlfähigkeiten des Host-Systems anzupassen. Durch die Konfiguration von Temperaturschwellen über NVMe-Standardfunktionen können Performance, Zuverlässigkeit und Energieeffizienz in Einklang gebracht werden.

Der Host setzt die Thermal Management Thresholds (TMT1/TMT2) innerhalb der vom Controller definierten Grenzen (MNTMT/MXTMT). Damit bestimmt der Host wann thermische Maßnahmen starten, während das Gerät bestimmt, wie sie ausgeführt werden. So werden unerwartete Abschaltungen vermieden, konstante Leistung sichergestellt und plattformspezifische Optimierungen ermöglicht, vom Rechenzentrum bis zu Embedded-Systemen.

Dieses Whitepaper stellt die Grundlagen von HCTM vor, beschreibt die Implementierung mit Standardtools und zeigt relevante Anwendungsfälle.

2 Einleitung

Mit steigenden NVMe-Durchsatzraten wird das Wärmemanagement entscheidend, um Gerätetozustand und QoS zu sichern. Unzureichende Kühlung führt zu Drosselung, Alterung oder sogar Abschaltungen. Zu konservative Limits hingegen verschenken Leistungsspielraum.

Traditionell nutzen SSDs eine gerätgesteuerte thermische Drosselung. Wenn ein interner Sensor erkennt, dass das Laufwerk vordefinierte Schwellenwerte erreicht hat, reduziert das Gerät selbstständig die Leistung, um eine Überhitzung zu verhindern. Dies schützt zwar die SSD, führt aber zu unvorhersehbaren Latenz- und Bandbreitenschwankungen, die sich negativ auf das Benutzererlebnis auswirken können.

Um dieses Problem zu lösen, führt die NVMe-Spezifikation das Host Controlled Thermal Management (HCTM) ein. HCTM ermöglicht dem Hostsystem die aktive Teilnahme am SSD-Wärmemanagement und sorgt so für einen besser koordinierten und vorhersehbaren Ansatz.

HCTM erlaubt dem Host, Geräteschwellen bewusst festzulegen. Überschreiten diese Werte, führt der SSD-Controller herstellerspezifische Maßnahmen aus. Diese klare Aufgabenteilung erlaubt eine flexible Anpassung an unterschiedliche Systemumgebungen, von gut gekühlten Server-Racks bis zu lüfterlosen Embedded-Systemen.

3 Grundlagen

Die NVMe-Spezifikation unterscheidet zwei Ansätze:

- Host Controlled Thermal Management (HCTM) bei dem der Host die Schwellenwerte festlegt und diese an die Systemkühlung anpasst.

- Device Controlled Thermal Management (DCTM), bei dem die SSD ihr Thermal Management eigenständig regelt.

In der Praxis definiert der Host, wann thermische Aktionen beginnen (über TMT1/TMT2), während das Gerät definiert, wie Aktionen (Drosselung, Änderungen des Energiezustands) innerhalb des NVMe-Funktionsmodells erfolgen.

Zu den wichtigsten Parametern gehören MNTMT/MXTMT (Identify Controller Boundaries) und die Thermal Management Thresholds TMT1/TMT2, die über den Feature Identifier 0x10 mithilfe von Set Features konfiguriert werden.

4 Temperaturdefinitionen im NVMe-Kontext

NVMe-SSDs zeigen mehrere Temperaturwerte an, die jeweils einem anderen Zweck dienen:

Composite Temperature: Die Haupttemperatur des Laufwerks, in Kelvin im SMART/Health Information Log verfügbar.

Temperatursensoren: Optionale zusätzliche Sensoren (Sensor 1, Sensor 2, ...), die z. B. Controller-, NAND- oder Gehäusetemperatur darstellen.

Thermal Management Thresholds: Durch den Host konfigurierbar (innerhalb MXTMT/MNTMT) und Auslöser für thermische Gegenmaßnahmen.

Min/Max Temperature (MXTMT/MNTMT):

Absolute Grenzwerte, die der Controller vorgibt. Das Festlegen von Features außerhalb dieser Grenzen schlägt mit einem ungültigen Feldfehler fehl.

Normalisierte Temperaturen: Manche Hersteller geben angepasste oder normalisierte Werte aus, die Rohdaten korrigieren oder Sicherheitsaufschläge berücksichtigen.

Hinweis:

- Rohwerte sind immer in Kelvin angegeben.

- Für Anwenderlesbarkeit empfiehlt sich die Umrechnung in Celsius.

- Normalisierte Werte sind herstellerspezifisch und können abweichen.

5 Funktionsweise von HCTM

1. Der Host prüft mit dem Identify Controller-Befehl, ob HCTM unterstützt wird.
2. Falls unterstützt, definiert der Host TMT1 und TMT2 über den Set Features-Befehl.
3. Sobald die SSD diese Temperaturen erreicht, aktiviert der Controller automatisch Maßnahmen wie Throttling oder den Wechsel in niedrigere Power States (siehe Abbildung 1).
4. Die konkreten Maßnahmen sind herstellerspezifisch und können je nach Modell variieren.

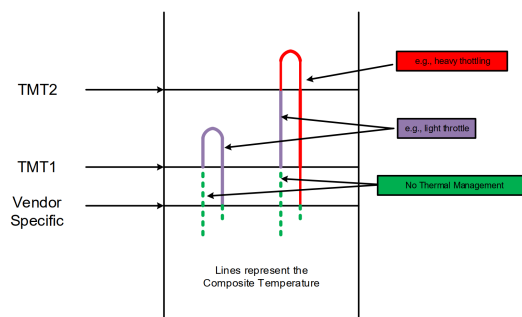


Abbildung 1: NVMe Throttle Management

6 Implementierungsbeispiel mit nvme-cli

6.1 Prüfen ob HCTM unterstützt wird

Prüfung der HCTM Unterstützung mit Identify Controller:

```
sudo nvme id-ctrl /dev/nvme0
```

Relevante Felder:

- hctma (Host Controlled Thermal Management Supported)
- mntmt (Minimum Thermal Management Temperature)
- mxmtmt (Maximum Thermal Management Temperature)

Wenn HCTMA und MXTMT/MNTMT ungleich null sind, wird HCTM vom Gerät unterstützt.

Beispielausgabe:

```
...
hctma : 0x1
mntmt : 303
mxmtmt : 373
...
```

6.2 Setzen von Thermal-Management-Schwellen

Verwenden Sie Set Features (Feature-ID 0x10), um TMT1/TMT2 zu definieren:

```
sudo nvme set-feature /dev/nvme<drive> -f 2
└─ 0x10 -v <value> [--save]
```

- value bits [31:16] = TMT1 (Kelvin)
- value bits [15:0] = TMT2 (Kelvin)

(optional persistent mit '--save')

Beispiel (90 °C / 100 °C):

```
sudo nvme set-feature /dev/nvme0n1 -f 0x10 2
└─ -v 0x016B0175
```

Detail: Umrechnung der Temperatur in den gepackten Wert:

- 90 °C = 363 K = 0x016B
- 100 °C = 373 K = 0x0175
- Kombiniertes Wert = 0x016B0175 (TMT1 in Bits 31:16, TMT2 in Bits 15:0).

Beispielausgabe:

```
set-feature:10 (Host Controlled Thermal 2
└─ Management), value:0x16b0175
```

6.3 Überprüfung mit Get-Feature

Der folgende Befehl zeigt den aktuellen gepackten Wert und die dekodierten Schwellenwerte an.

```
sudo nvme get-feature /dev/nvme0n1 -f 0x10 -H
```

Beispielausgabe:

```
get-feature:0x10 (Host Controlled Thermal 2
└─ Management), Current value:0x16b0175
Thermal Management Temperature 1 (TMT1) : 2
└─ 363 Kelvin
Thermal Management Temperature 2 (TMT2) : 2
└─ 373 Kelvin
TMT1 = 363K (90 °C)
TMT2 = 373K (100 °C)
```

Hinweis:

Die Unterstützung für Persistenz mit '--save' ist geräteabhängig. Bei Erreichen der Schwellen führt der Controller Maßnahmen wie Drosselung oder Power-State-Wechsel aus. Rückkehr-/Hysterese-Verhalten ist gerätespezifisch. Verwenden Sie Identify, um sicherzustellen, dass die Werte innerhalb der MNTMT/MXTMT-Grenze bleiben. Das Festlegen außerhalb der Grenzen führt zu einem Fehler.

Beispielausgabe für diesen Fehler:

```
NVMe Status:INVALID_FIELD: A reserved coded 2
└─ value or an unsupported value in a defined 2
└─ field(2)
```

7 Vorteile von HCTM

Warum sollte der Host die Parameter ändern wollen?

- In Rechenzentren mit robuster Kühlung können höhere Schwellenwerte festgelegt werden, um die maximale Leistung aufrechtzuerhalten.
- Bei Laptops oder Mobilgeräten können Schwellenwerte gesenkt werden, um Strom zu sparen und die Akkulaufzeit zu verlängern.
- Durch die Anpassung der Schwellenwerte an die Systemkühlkapazitäten vermeidet der Host unkontrollierte thermische Abschaltungen.

- Obwohl Drosselung generell unerwünscht ist, ist sie ungeplanten Systemausfällen vorzuziehen.
 - Verbesserte Gesamtbetriebskosten durch geringeren Kühlbedarf und längere SSD-Lebensdauer.
- implementiert, über SMART-Protokolle verifiziert und unter Arbeitslast validiert. So ermöglicht es eine konsistente, vorhersehbare Optimierung von Leistung, Zuverlässigkeit und Effizienz in unterschiedlichsten Implementierungen.

8 Bewährte Verfahren

- Leiten Sie Schwellenwerte aus Luftstrom- und Temperaturmessungen ab und halten Sie den Spielraum unter MXTMT.
- Verwenden Sie NVMe Smart-Log, um composite Temperaturtrends während Stress-tests zu überwachen.
- Dokumentieren Sie thermische Richtlinien pro Formfaktor (M.2 vs. U.2/U.3) und mit/ohne Kühlkörper.
- Validieren Sie die Persistenz mit '--save' über Neustarts und Firmware-Updates hinweg.

9 Anwendungsfälle

- Rechenzentrum: Optimieren Sie die Schwellenwerte für die Rack-Kühlung, minimieren Sie unnötige Drosselung und gewährleisten Sie die Servicequalität.
- Edge/Industrie: Konservative Richtlinien für unvorhersehbare Umgebungen.
- Mobil/Embedded: Schaffen Sie ein Gleichgewicht zwischen Akustik, Komfort und Effizienz durch frühzeitige Drosselung.

10 Fazit

HCTM bietet Hosts präzise Kontrolle über den Beginn der thermischen Drosselung, während die Geräte die Ausführung definieren. HCTM wird über Identify- und Set/Get-Funktionen

Kontaktieren Sie uns

Hauptsitz	Swissbit AG Industriestraße 4 9552 Bronschhofen Schweiz	Tel. +41 71 913 03 03 sales@swissbit.com
Deutschland (Berlin)	Swissbit Germany AG Bitterfelder Straße 22 12681 Berlin Deutschland	Tel. +49 30 936 954 0 sales@swissbit.com
Deutschland (München)	Swissbit Germany AG Leuchtenbergring 3 81677 München Deutschland	Tel. +49 30 936 954 400 sales@swissbit.com
Nord- und Südamerika	Swissbit NA Inc. 238 Littleton Road, Suite 202B Westford, MA 01886 USA	Tel. +1 978-490-3252 salesna@swissbit.com
Japan	Swissbit Japan Co., Ltd. CONCIERIA Tower West 2F 6-20-7 Nishishinjuku Shinjuku City, Tokyo 160-0023 Japan	Tel. +81 3 6258 0521 sales-japan@swissbit.com
Taiwan	Swissbit Taiwan 12 F.-9, No. 268, Liancheng Rd. Zhonghe District New Taipei City 235603 Taiwan, R.O.C.	Tel. +886 912 059 197 salesasia@swissbit.com
China	Swissbit China	Tel. +886 958 922 333 salesasia@swissbit.com

Disclaimer:

The information in this document is subject to change without notice. Swissbit AG ("SWISSBIT") assumes no responsibility for any errors or omissions that may appear in this document, and disclaims responsibility for any consequences resulting from the use of the information set forth herein. SWISSBIT makes no commitments to update or to keep current information contained in this document. The products listed in this document are not suitable for use in applications such as, but not limited to, aircraft control systems, aerospace equipment, submarine cables, nuclear reactor control systems and life support systems. Moreover, SWISSBIT does not recommend or approve the use of any of its products in life support devices or systems or in any application where failure could result in injury or death. If a customer wishes to use SWISSBIT products in applications not intended by SWISSBIT, said customer must contact an authorized SWISSBIT representative to determine SWISSBIT willingness to support a given application. The information set forth in this document does not convey any license under the copyrights, patent rights, trademarks or other intellectual property rights claimed and owned by SWISSBIT.

ALL PRODUCTS SOLD BY SWISSBIT ARE COVERED BY THE PROVISIONS APPEARING IN SWISSBIT'S TERMS AND CONDITIONS OF SALE ONLY, INCLUDING THE LIMITATIONS OF LIABILITY, WARRANTY AND INFRINGEMENT PROVISIONS. SWISSBIT MAKES NO WARRANTIES OF ANY KIND, EXPRESS, STATUTORY, IMPLIED OR OTHERWISE, REGARDING INFORMATION SET FORTH HEREIN OR REGARDING THE FREEDOM OF THE DESCRIBED PRODUCTS FROM INTELLECTUAL PROPERTY INFRINGEMENT, AND EXPRESSLY DISCLAIMS ANY SUCH WARRANTIES INCLUDING WITHOUT LIMITATION ANY EXPRESS, STATUTORY OR IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

© 2025 SWISSBIT AG