Hailo-8 AI Acceleration Module

Thermal Design Considerations
Application Note

Rev 0.2
March 2021

Confidential and proprietary – unauthorized reproduction prohibited

Disclaimer

Copyright

© 2021 Hailo Technologies Ltd ("Hailo"). All Rights Reserved.

No part of this document may be reproduced or transmitted in any form without the express, written permission of Hailo. Nothing contained in this document should be construed as granting any license or right to use proprietary information without the written permission of Hailo.

This version of the document supersedes all previous versions.

General Notice

Hailo, to the fullest extent permitted by law, provides this document "as-is" and disclaims all warranties, either express or implied, statutory or otherwise, including but not limited to the implied warranties of merchantability, non-infringement of third parties' rights, and fitness for particular purpose.

Hailo assumes no liability for any error in this document and for damages, whether direct, indirect, incidental, consequential or otherwise, that may result from such errors, including but not limited to loss of data or profits.

The content in this document is subject to change without prior notice. Hailo reserves the right to make changes to said content without prior notification to users.

Documentation Control

Revision History

Version	Date	Description
0.1	Oct 2020	First Draft
0.2	March 2021	Early access version – approved for limited disclosure

 Confidential and Proprietary | Copyright © 2021 – Hailo Technologies Ltd.

Table of Contents

1. Introduction and Scope 5
2. Quick Start – Read This First 5
3. Thermal Requirements and Specifications 7
3.1. Heat Dissipation 7
3.2. Thermal Specification 7
3.3. Thermal PCIe Shutdown 8
3.4. Thermal Throttling 8
4. Typical System Configurations 9
4.1. Natural Convection Solution 9
4.2. Forced Airflow Solution 11
4.3. Fanless Enclosure (Direct Conductance) 12

 Confidential and Proprietary | Copyright © 2021 – Hailo Technologies Ltd.

1. Introduction and Scope

The Hailo-8 AI Acceleration Modules are a comprehensive family of PCI Express (PCIe) based acceleration modules that meet industry standards for a range of form factors and performance objectives, targeted at artificial intelligence (AI) applications.

This application note applies to the following products:

- Hailo-8 M.2 M key AI Acceleration Module (HM218B1C2FA)
- Hailo-8 M.2 M key AI Acceleration Module (HM218B1C2KA)
- Hailo-8 mPCIe AI Acceleration Module (HMP1RB1C2GA)

Due to its high performance and small form factor, most practical applications require any of the Hailo-8 modules to have a heat dissipation solution – to avoid an overheat condition and a thermal shutdown. It is critical that the hosting enclosure/chassis vendor provide a proper thermal solution.

This application note details the design considerations of the required thermal solution and provides some tested “recipes” for typical system configurations to reduce thermal design overhead.

2. Quick Start – Read This First

The Hailo AI Acceleration Module is a high-performance module and requires a heat dissipation solution.

Operating the module otherwise will result in thermal shutdown for most applications.

For operating in room temperature, the Hailo AI Acceleration Module typically requires one of these two solutions:

- An Antou HD-HB-1106 heat sink, attached to the top of the Hailo-8 device with a thermal pad (both provided in the package), and with ventilation forced around the heat sink by a fan.
  See Figure 1.

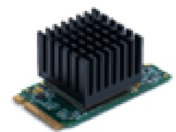

Figure 1 - Heat sink solution

- A bulk heat conductor from the Hailo-8 top to a large metal case or other heat dissipation surface, using a suitable thermal pad.
  See Figure 2.

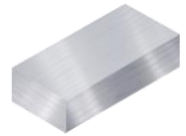

Figure 2 - Bulk aluminum heat conductor

These solutions should enable the Hailo AI Acceleration Module to operate at full capacity in room temperature conditions with sufficient margin.

For more detailed thermal design, read on.

 Confidential and Proprietary | Copyright © 2021 – Hailo Technologies Ltd.

3. Thermal Requirements and Specifications

3.1. Heat Dissipation

The heat created in the Hailo-8 AI Acceleration Module is mostly dissipated to the top surface of the Hailo-8 device on the top side of the board, where it can be further conducted to a heat sink or other heat transfer solution.

The Hailo AI Acceleration Module will also dissipate some heat via the module PCB, edge connector and socket, mounting screw and stand-off. However, this dissipation is small compared with the heat dissipation from the chip's top surface. From here on, we will neglect this heat dissipation path, which will result in a small positive margin for the thermal design. The smaller the module's form factor, the smaller this margin.

The module's power consumption is highly dependent on the resources utilized for inference, and the operating frame rate or input rate used.

The junction temperature relates to the ambient temperature according to the equation $T_J \approx T_A + P \cdot (\theta_{jc} + \theta_{ca})$, where heat dissipation via the PCB is neglected and:

- T_J is the Hailo-8 silicon junction temperature;
- T_A is the ambient air temperature;
- P is the power dissipated by the module;
- θ_{jc} is the thermal resistance from junction to device case (package top surface) and is an inherent property of the device;
- θ_{ca} is the thermal resistance from the Hailo-8 device case to ambient air.

3.2. Thermal Specification

Table 1 provides the thermal properties of the Hailo-8 device. For more information, refer to the Hailo-8 data sheet.

 Confidential and Proprietary | Copyright © 2021 – Hailo Technologies Ltd.

Symbol	Parameter	Value
T_j	Maximum operating junction temperature	$125℃$
θ_{jc}	Thermal resistance (junction to top lid)	$0.35\ ℃/W$

Table 1 – Hailo-8 Thermal Properties

3.3. Thermal PCIe Shutdown

The Hailo-8 firmware automatically shuts down the PCIe data lanes when a predefined temperature is crossed to prevent the junction temperature from rising above the absolute maximum rated (AMR).

3.4. Thermal Throttling

TBD

 Confidential and Proprietary | Copyright © 2021 – Hailo Technologies Ltd.

4. Typical System Configurations

This section presets 3 different heat dissipation solutions.

4.1. Natural Convection Solution

In this configuration, the AI Acceleration Module is mounted in a chassis that is allows ample ventilation and free air flow (an open frame computer, for example).

There should be little or no obstruction between the module and ambient air. The ambient environment should be such that ambient temperature is unaffected by the system's heat dissipation.

Figure 3 - Natural convection conditions

Under these conditions, the air flow around the module is small (typically less than 0.5 meters per second) and is mostly due to thermal convection.

The thermal solution required for this configuration is the attachment of a heat sink with a large surface area, on top of the Hailo-8 device, using a thermally conductive pad and mounting screws on opposite sides of the device.

In this configuration, θ_{ca} is the sum of:

1. The thermal resistance of the thermal pad;
2. The thermal resistance between the heat sink bulk surface and ambient air.

The thermal conductivity of low-cost, silicone based thermally conductive pads is typically around $3\frac{W}{m\cdot℃}$. A 0.5mm thick thermal pad fitted to the Hailo-8 top lid surface (12mm x 12mm) would therefore have a thermal resistance of $0.85\ ℃/_{W}$.

A small heat sink with a large surface area, such as the Antou HD-HB-1106, provides a typical thermal resistance of $18\ ℃/_{W}$ at typical natural air flow conditions.

Figure 4 demonstrates how a heat sink rated for $18\ ℃/_{W}$ under natural air flow and used with a 0.5mm thick thermal pad results in a total thermal resistance of $19.2\ ℃/_{W}$.

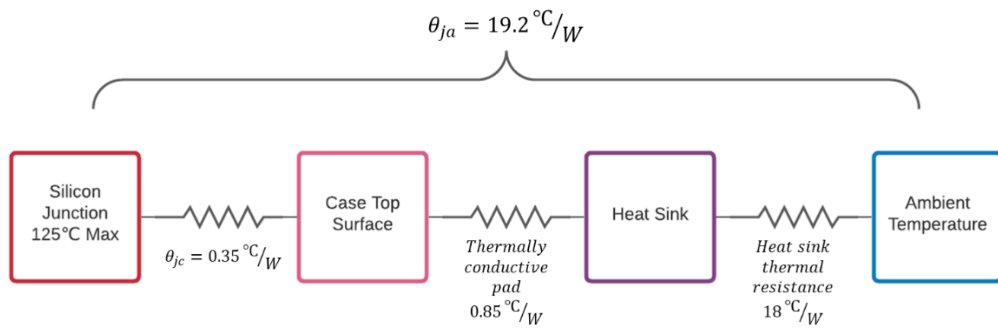

Figure 4 - Thermal Resistance for a Natural Convection Configuration

This total thermal resistance results in a thermal constraint that enables operation at 5.2W power consumption under room temperature conditions.

Figure 5 shows the full envelope of thermally feasible operation with the described solution.

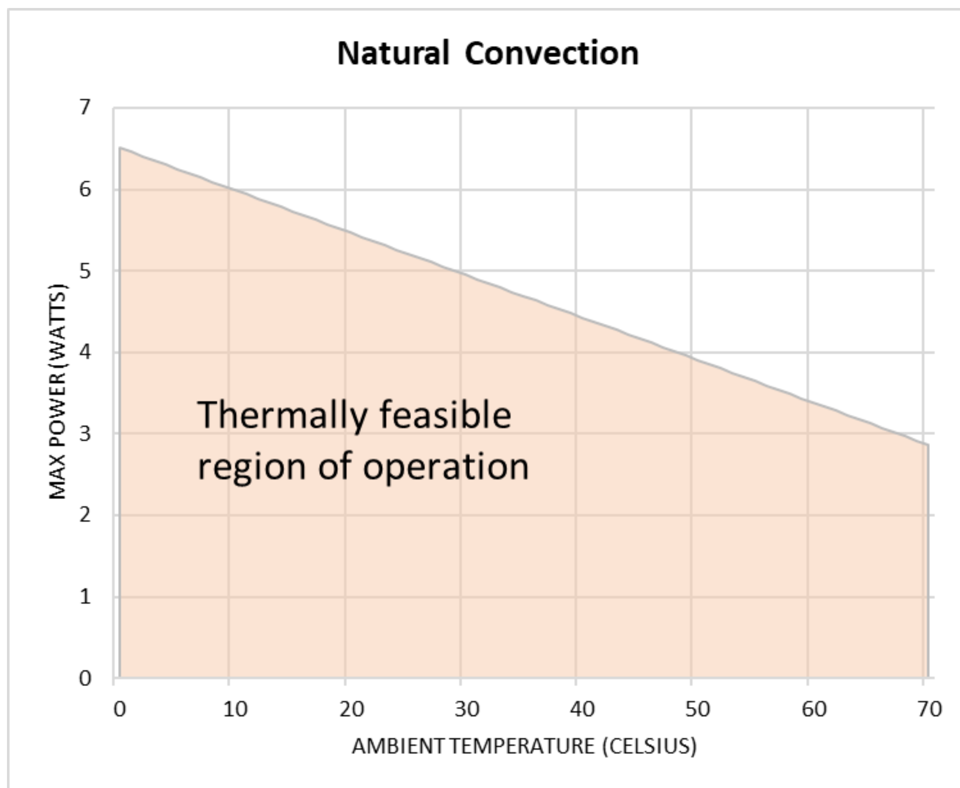

Figure 5 - Natural convection solution thermal envelope

 Confidential and Proprietary | Copyright © 2021 – Hailo Technologies Ltd.

4.2. Forced Airflow Solution

In this configuration, the AI Acceleration Module is mounted in a chassis that forces cool air flow (typically 2-4 meters per second) around the module using a fan. A standard desktop PC is a good example.

Figure 6 - Forced airflow conditions

In this configuration, θ_{ca} is the sum of:

1. The thermal resistance of a thermal pad 0.5mm thick (0.85 $^{\circ C}/_{W}$);
2. The thermal resistance between the heat sink bulk surface and ambient air, under forced air flow conditions.

The forced air flow results in a lower thermal resistance with any given heat sink.

Figure 7 demonstrates how a heat sink rated for 10 $^{\circ C}/_{W}$ under a forced air flow of 3 meters per second (Antou HD-HB-1106 or similar), used with a 0.5mm thick thermal pad, result in total thermal resistance of 11.2 $^{\circ C}/_{W}$.

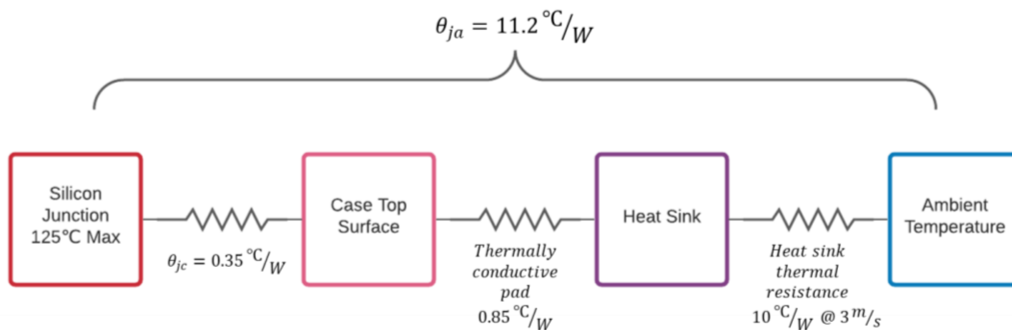

Figure 7 - Thermal Resistance for a Forced Airflow Configuration

This total thermal resistance results in a thermal constraint that enables, for example, operation at 8.2W power consumption in an ambient temperature of 33℃, or 4.9W at 70℃.

Figure 8 shows the full envelope of thermally feasible operation with the described solution.

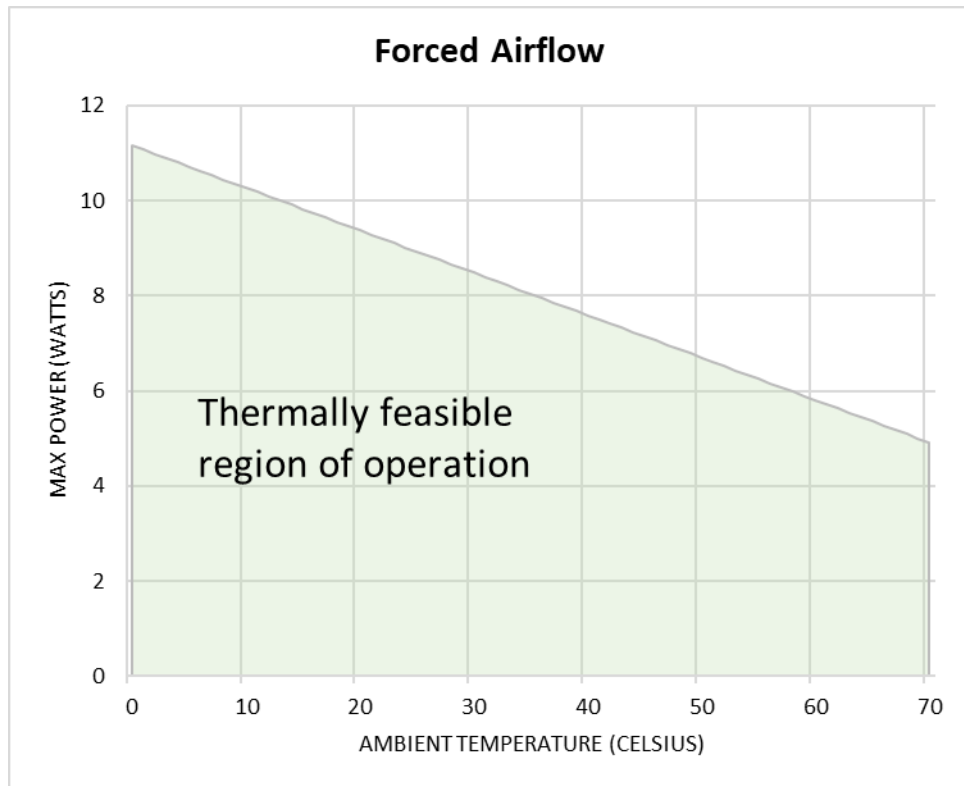

Figure 8 - Forced airflow solution thermal envelope

4.3. Fanless Enclosure (Direct Conductance)

In this configuration, the AI Acceleration Module is enclosed in a closed-box chassis that allows little or no air movement. A fanless mini-computer and some industrial/ruggedized computer models fit into this use case category.

Figure 9 - Fanless enclosure example

In this configuration, a thermal conductor (typically a bulk aluminum bar) must be used to conduct the heat from the Hailo-8 device to the enclosure surface, where it can dissipate to the ambient air.

In this configuration, θ_{ca} is the sum of:

3. The thermal resistance of a thermal pad 1mm thick (1.7 ${}^{\circ}\mathrm{C}/_{W}$), used to interface the Hailo-8 top lid to the aluminum bar;
4. The thermal resistance of the aluminum bar;
5. The thermal resistance between the enclosure surface and ambient air, under the rated ambient conditions.

The thermal conductivity of common aluminum alloys is typically around 170 $\frac{W}{m\cdot{}^{\circ}\mathrm{C}}$.

 Confidential and Proprietary | Copyright © 2021 – Hailo Technologies Ltd.

In a typical case of a bulk aluminum conductor with a cross section of 1cm^2 and a length of 25mm from device to enclosure surface, a thermal resistance of $1.5\ {}^{\circ}\mathrm{C}/_{W}$ is typical for the aluminum bar.

Figure 10 illustrates how an aluminum enclosure rated for a thermal resistance of $3.4\ {}^{\circ}\mathrm{C}/_{W}$ (such as the Takachi EXH14-9-10BB or similar) used with a 1mm thick thermal pad, results in a total thermal resistance of $6.95\ {}^{\circ}\mathrm{C}/_{W}$.

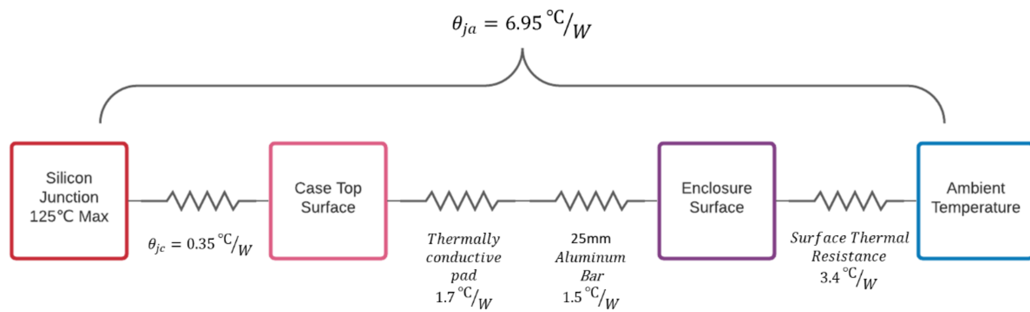

Figure 10 - Thermal Resistance for a Fan-less Enclosure

This total thermal resistance results in a thermal constraint that enables, for example, operation at 7.9W power consumption at an ambient temperature of 70℃.

Figure 11 shows the full envelope of thermally feasible operation with the described solution.

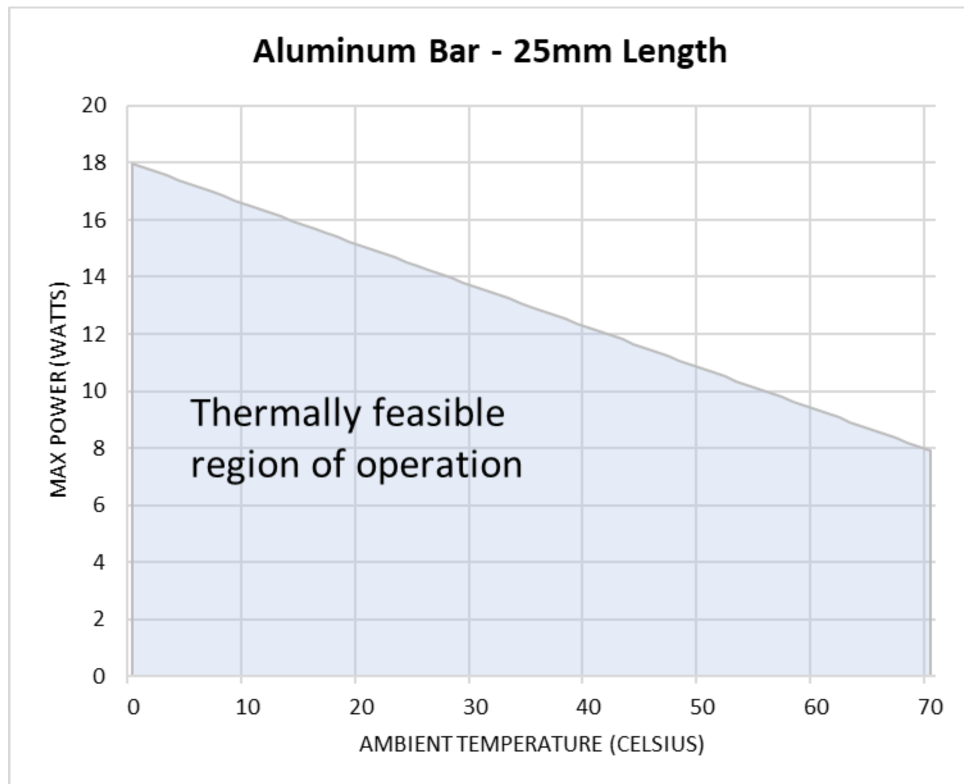

Figure 11 - Bulk thermal conductor solution thermal envelope

 Confidential and Proprietary | Copyright © 2021 – Hailo Technologies Ltd.

The thermal resistance varies with thermal conductor dimensions. For example, Figure 11 demonstrates the different thermal operation envelopes for using different lengths of bulk aluminum thermal conductors with a 1cm^2 cross section:

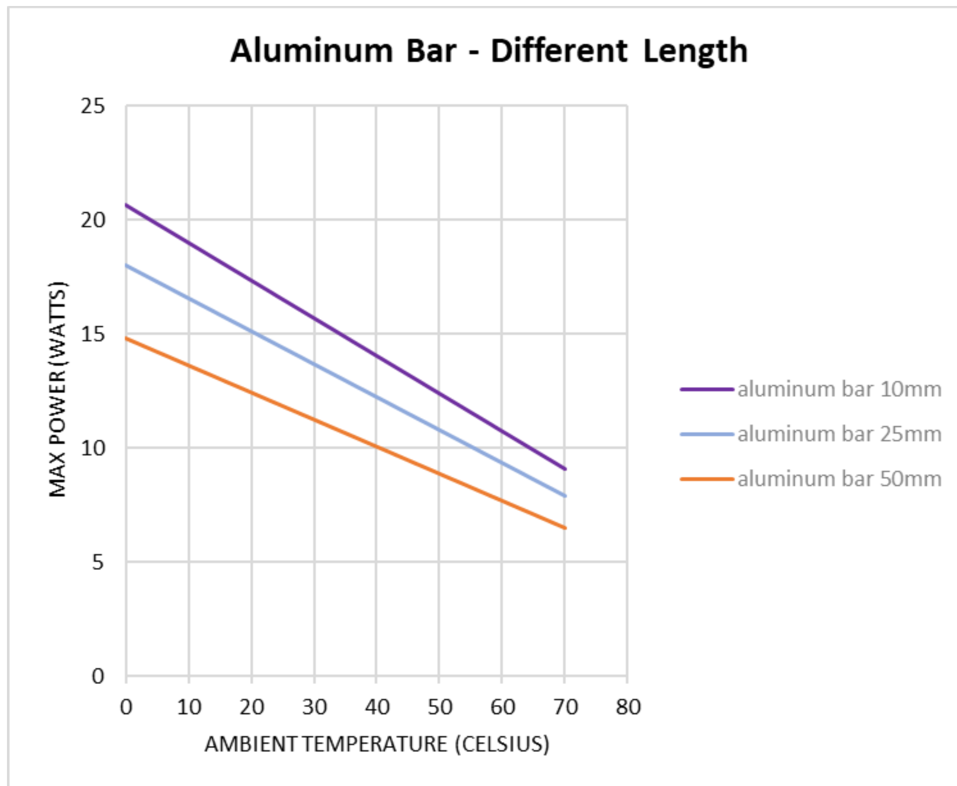

Figure 12 - Thermal envelope with different aluminum bar lengths