Rack & crate considerations for Phase-2 Upgrade

Gregory Iles, 2017-12-17, **DRAFT-v1.4**

Introduction

ATCA has been identified by the ATLAS & CMS experiments as a potential replacement of the existing VME crate standard for Phase-2. It has a high speed, SerDes based, full mesh backplane (i.e. up to 4 x 10 Gb/s to/from every blade in the chassis), offering far greater bandwidth in a compact form. It has 14, rather than 21 slots, with up to 400 W per slot (front & rear cards combined). The standard is also simpler than VME, which now has multiple extensions.

The standard acknowledges that cooling 400 W may be challenging, and it recommends limiting the front board slot to 200 W. Crates are available that can cool 400W cards, but the fans consume significant power (1-2 kW) and generate noise levels exceeding 85 dBA, requiring additional health & safety protection. It is unlikely that most HEP applications will need > 200 W per board, but there are significant optical modules, which are not present on commercial cards, or at least not in such large numbers. These typically need to be operated at < 50 °C to have a reasonable longevity (i.e. less than 1% failure over 15 years) and need space for fibre routing. Furthermore, the latest generation FPGAs (e.g. Xilinx Ultrascale) dissipate significantly more heat than in the past, with more than 100W per part possible.

To gauge how the electronics in CMS has evolved over the past decade it is interesting to compare cards designed for the start of CMS, those for the Phase-1 Trigger Upgrade and future Phase-2 designs. The CMS Tracker FED (2004) has 9 large FPGAs (Virtex 2 XC2V1500 & XC2V2000), a bandwidth of 42.4 Gb/s and dissipates 80W in a 9U VME 400mm card (power density of 0.54 x 10^{-3} W mm^{-2}). The MP7 (2012), has 1 large FPGA (XC7VX690T, 45mm x 45mm footprint), a bandwidth of 1500 Gb/s, dissipates 70W on a double width AMC card (2.6 x 10^{-3} W mm^{-2}). The FED has no heatsinks while the MP7 FPGA has a heatsink with air cross-section 57mm x 18mm, dissipates approximately 35W and has an operating temperature of 60-70°C (maximum temperature 85°C). CMS deliberately selected a MicroTCA crate with full-height AMC slots so that high heatsinks and optical modules could be easily accommodated (maximum component size for VME is 13.71mm versus 22.45mm for a full height AMC card). ATCA cards can operate up to 200W-400W, a factor of 2.5 to 5 greater than 80W AMC cards. However, the ATCA card depth, and thus also air-cross-section for cooling, only increases by a factor of 1.5. To compensate for this the air-speed is increased and the area used for heat sinks must be much larger than the part foot-print. ATCA power density will increase to 2.2-4.4 x 10^{-3} W mm^{-2}.

This document reviews the mechanical dimensions of the rack setup and crates, and the impact ATCA will have on the rack volume for electronics and thus also cooling, board area, fan noise, etc. It will show that the electronics volume could be increased by up to a factor of 3 if it is deemed necessary.

Vertical Cross-section

ATCA was designed for front to back cooling, which places an intrinsic limitation on the card dimensions (i.e. for good cooling, with minimal front/rear rack space devoted to air entry/exit, the card cannot extend deep into the rack). This is not the case for the LHC experiments, which are likely to use vertical airflow. Front-to-Back cooling is considered in the appendix, but this document considers ATCA in a typical environment with vertical airflow (i.e. in a cavern, 100 m underground).

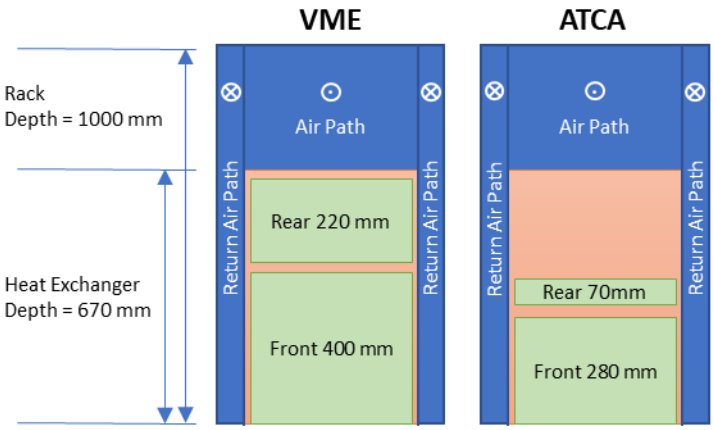


Fig. 1 A top view of the existing LHC VME rack and proposed ATCA rack. The air cross-section for cooling is far greater for VME compared to ATCA for both the front cards (43 % larger) and for combined front & rear cards (77 % larger).

There has been much discussion about whether the size of the underground counting rooms have sufficient size for the Phase-2 electronics, but no discussion on how efficiently that space is used. The ATCA crate only occupies 38% of the rack depth (fig. 1), not only providing limited cooling opportunities, but also limited PCB real estate for electronics. The substantial reduction in air cross-section for cooling necessitates the use of high power fans and their associated issues (i.e. noise & power, which scale very non-linearly with airspeed). For CMS, at least, the limiting constraint is probably the total rack power & cooling requirement, which is currently set at 10kW.

The simplest crate change (fig. 2: ATCA-D1) would be to remove the rear card area and extend the front card, which may allow fan-trays to remain unchanged. However, the change still leaves much of the rack unused and removes flexibility. The heat exchangers will be replaced before Phase-2 and thus there remains the possibility to extend or shorten the heat exchanger, assuming there are no mechanical restrictions apart from the rear door space requirements for plumbing. A second option (fig.2: ATCA-D2) is to extend both the front and rear cards to fully utilise the rack space.

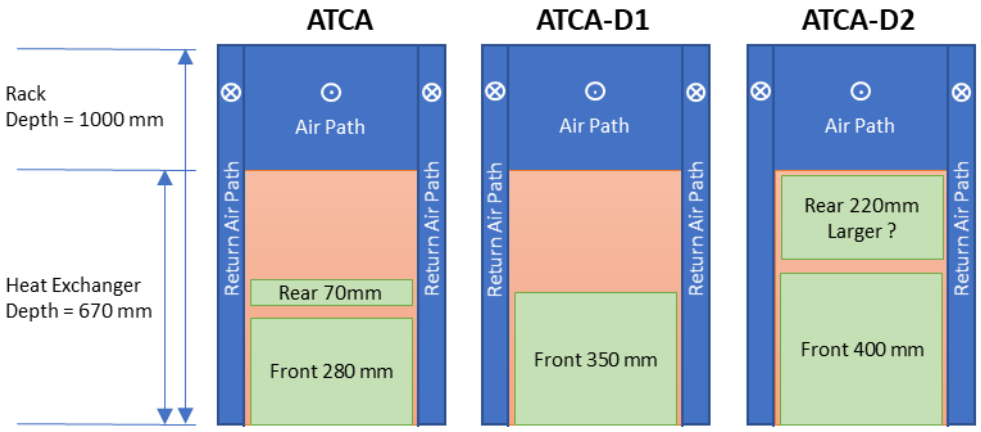


Fig. 2 Possible modifications of the ATCA crate to make better use of the rack space. Option ATCA-D1 is intended to preserve the ATCA fan-tray design. Option-D2 fully utilises the rack space, assuming the rear card area is used.

Front-Cross-Section

The present vision for CMS is to have 2 ATCA crates per rack, with the option for a 3rd crate (fig 3: rack labelled ATCA). This, combined with the limited crate depth of 280mm, results in the electronics occupying just 7% of the rack (front cards) or 9% (front & rear cards). Furthermore, the limitation to 2 crates per rack significantly reduces the total air-cross-section that could be cooled (i.e. it is possible to have 4 crates per rack, which would double the number of cooling stages).

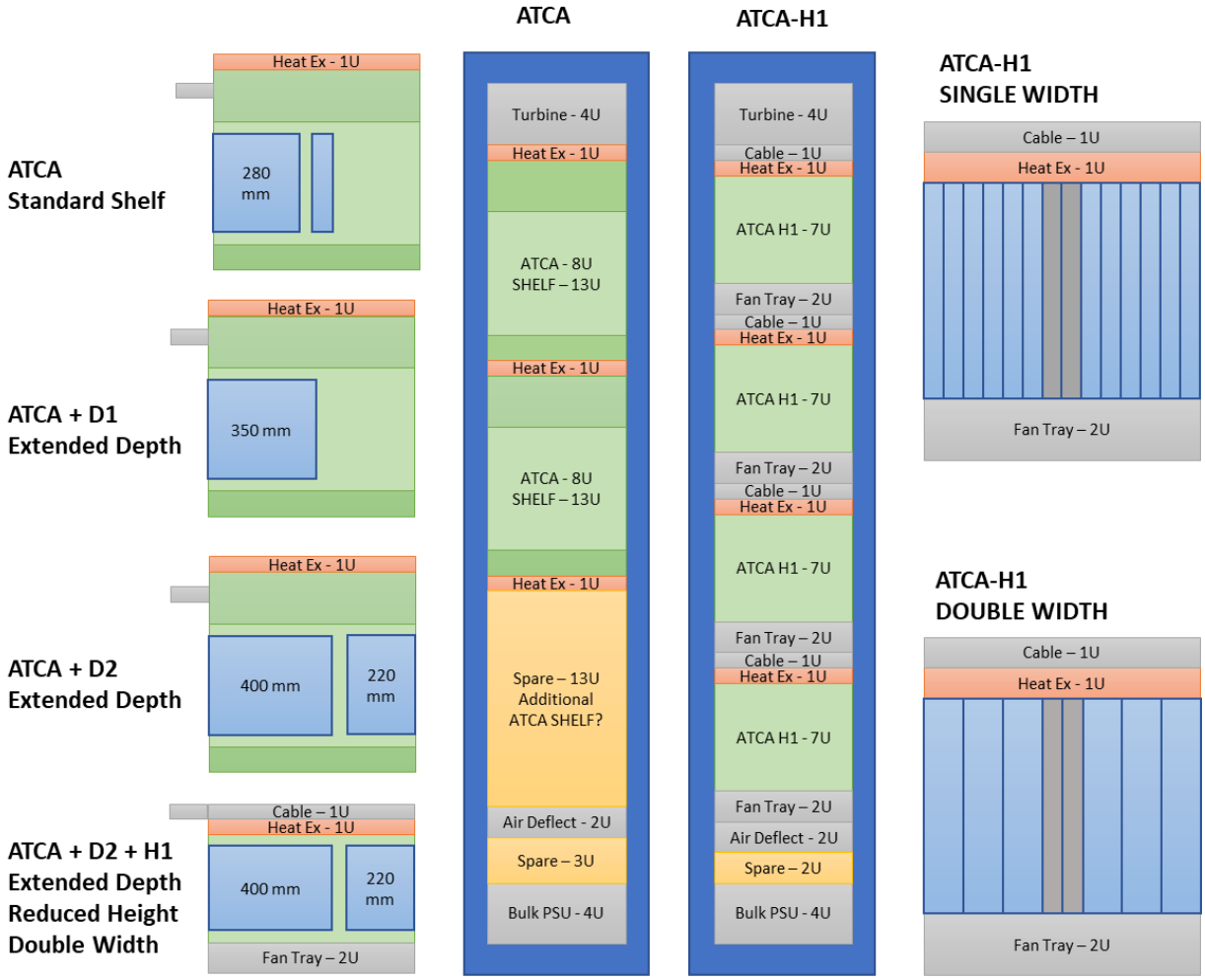


Fig. 3 The central part of the image is a front view of the proposed rack/crate standard for CMS (ATCA) and to the right of it (ATCA-H1) is an alternative that would double the number of crates per rack to improve cooling (i.e. double width cards in bottom right) or double the cards per rack (top right). On the left is a side view through the rack. All images are approximately to scale, but the images on the right (i.e. slot width options for crate ATCA-H1) are scaled by a factor of 2 less. The ATCA-H1 rack is designed for easy fan-tray extraction and drop-down cabling. The heat-exchanger is placed to remove heat as quickly as possible from the airflow (i.e. thus avoiding heat transfer to other locations) and to make card extraction easier (i.e. gap between cable tray and card cage). The bulk PSU is shown outside the vertical airflow region because they normally have a front-back airflow; however, it is more likely to be situated in the back of the rack, within the chilled air volume.

Alternatively, if rack ATCA-H1 is adopted (fig. 3), with the ATCA-D2 extended depth cards (fig. 2), the rack volume for electronics increases to 18% (front cards) or 27% (front & rear cards). A factor of 2.6 (3.0) larger than for standard ATCA front (front & rear) cards. The disadvantage of changing the card height is that standard front panels cannot be used, which slightly complicates card manufacture.

In addition to the cooling aspects there must be sufficient space on the PCB for the electronics, heatsinks and increasingly optical fibre. Given that the cards should have a planned lifetime of more than 10 years and that there may be significant airflow it is probably wise to firmly secure the fibres, rather than have them vibrating in the airflow. The PCB real estate for different standards is shown in Table 1. It is also worth noting that at present most developers are intending to use the extremely compact Samtec FireFly optical module, but new industry standards are larger, requiring more real estate. Is there the contingency in the board real estate of Phase-2 designs to switch to an alternative optical module? This may be necessary to satisfy the demands of the front-end optics, which are within the detector volume and subject to radiation damage.

	Height (mm)	Front Length (mm)	Front Area (x10^3 mm^2)	Rear Length (mm)	Rear Area (x103 mm2)
VME 400mm	366.7	400	147	220	81
ATCA	322	280	90	7	2
ATCA+D1	322	350	113	N/A	N/A
ATCA+D2	322	400	129	220	71
ATCA+D2+H1	277.6	400	111	220	61

Table 1 Card dimensions and area for 9U VME 400mm and ATCA with different crate modifications specified in Fig. 32 and Fig. 3.

Conclusions

Substantial changes to the underground electronics environment for Phase-2 can be made with at a relatively modest cost (i.e. expected to be < 100 kCHF). The changes would not only provide increased space for electronics (up to a factor of 3), but also better cooling and reduced noise. Indirectly, it could also provide more power and reduced cooling requirements because the present scheme envisages using up to 20% of total power for crate fans. Lastly, the improved cooling should improve the longevity of the electronics, such as optics, that are particularly sensitive to temperature. The increased board space will allow fibres to be properly secured and thus avoid any risk from vibration in the high-speed airflow.

Appendix

The appendix contains information not directly relevant to the discussion above, but may be of interest to the reader. It discusses front-back cooling, lists crate requirements and provides some reference plots. It shows the predicted temperature for a single large Xilinx Ultrascale+ (VU9P) and a mid-range Xilinx Ultrascale (KU115) design. It is important to note that to get reasonable cooling, the heatsinks have been extended from the part footprint (i.e. 45 mm square) to 10 mm square. It also shows the strong dependence of fan noise (and power) on air speed and the health and safety requirements.

Appendix: Front-to-Back Cooling

CMS, at least, is unlikely to opt for front-to-back cooling because it remains unclear if the room air impedance is too high. The ceilings, at least in S1, are low and the racks are close together (1 m before rear door heat exchangers are added). There are other reasons, which make vertical air cooling attractive (i.e. fire containment, re-circulated air avoids dust build up and lastly, the re-circulated air path front rack space is re-used for cable trays), but there are solutions to these issues. It is frequently stated that front-to-back airflow would not be acceptable to the LHC experiments because the rack exit air is warmer than the entry air, leaking heat into the underground caverns, which is not allowed. This, however, assumes a room temperature of 19 °C, which provides insufficient headroom above the water temperature of 15 °C for the system to operate correctly. For the rack power densities of interest to HEP (i.e. 10 kW per rack), there wouldn't be any thermal leakage into the caverns if the room temperature was 23 or 24 °C.

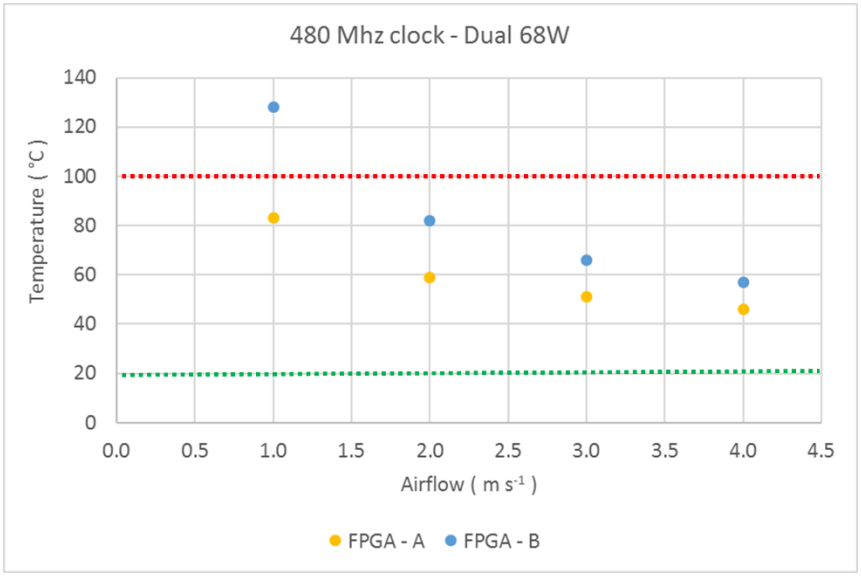
Appendix: Requirements

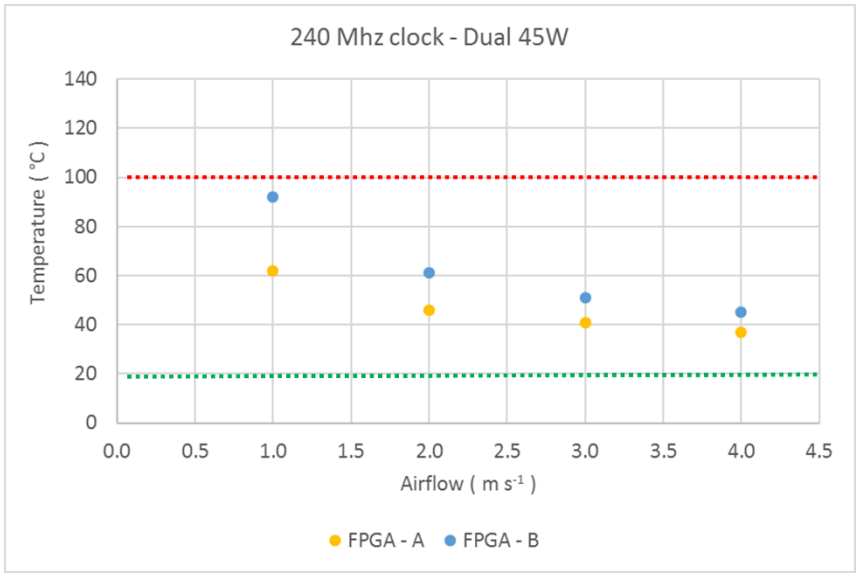
Crates must provide power, cooling and electrical interfaces (slow control, fast control/feedback, clocks, data paths). Less obvious requirements are fire containment, efficient use of space, noise, front panel space, rear access are.

Appendix: Dual High-End Kintex Ultrascale temperature versus airspeed.

Assumptions: Set to high LUT usage, high DSP usage, KU115, D1517, LUTs & FFs @ 80%, DSPs @ 80%, Clock @ 480 and 240 MHz, 64 Low Power 16G transceivers, BRAM & URAM @ 80%, No I/O or external memory. FPGA-B placed directly above FPGA-A.

The heatsink area was set to 100 mm x 100 mm, significantly beyond the 40 mm x 40 mm package size to maximise cooling, albeit at the expense of board area. If a heatsink with a footprint comparable with the package were used (i.e. as with most boards at present) the temperatures would at least double. The dashed red line is the maximum permitted temperature.

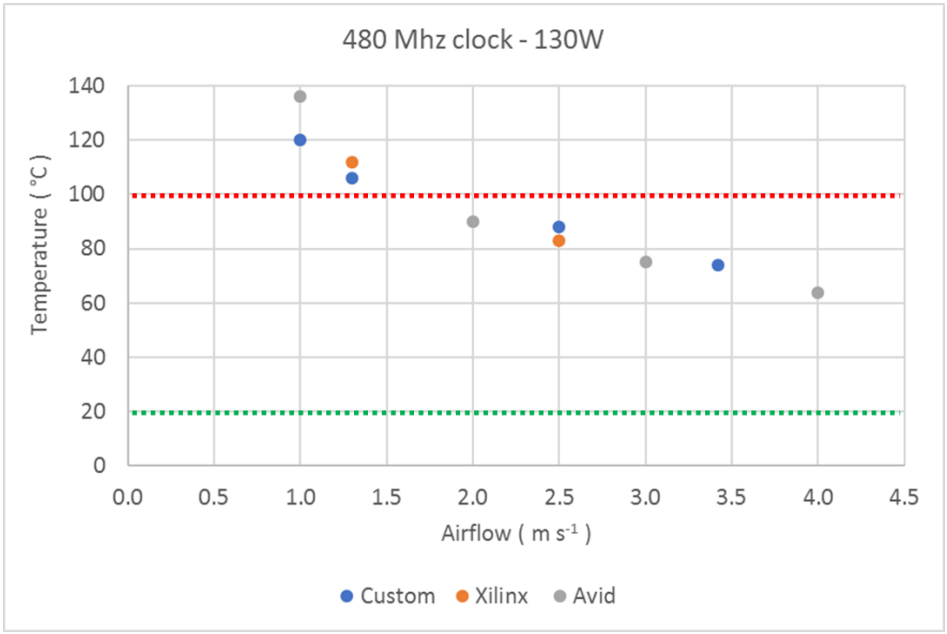


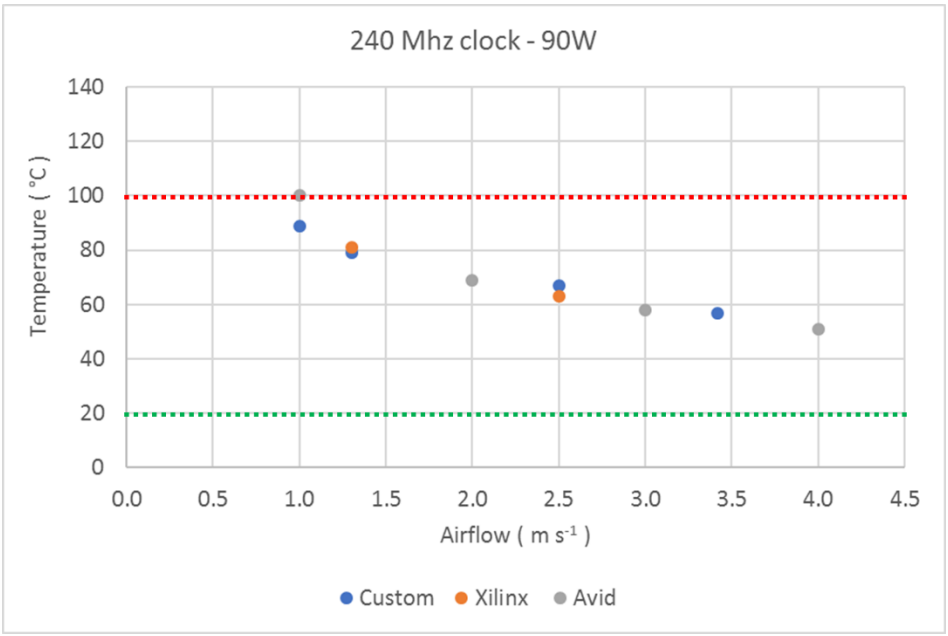


Appendix: Single Mid-range Ultrascale+ temperature versus airspeed.

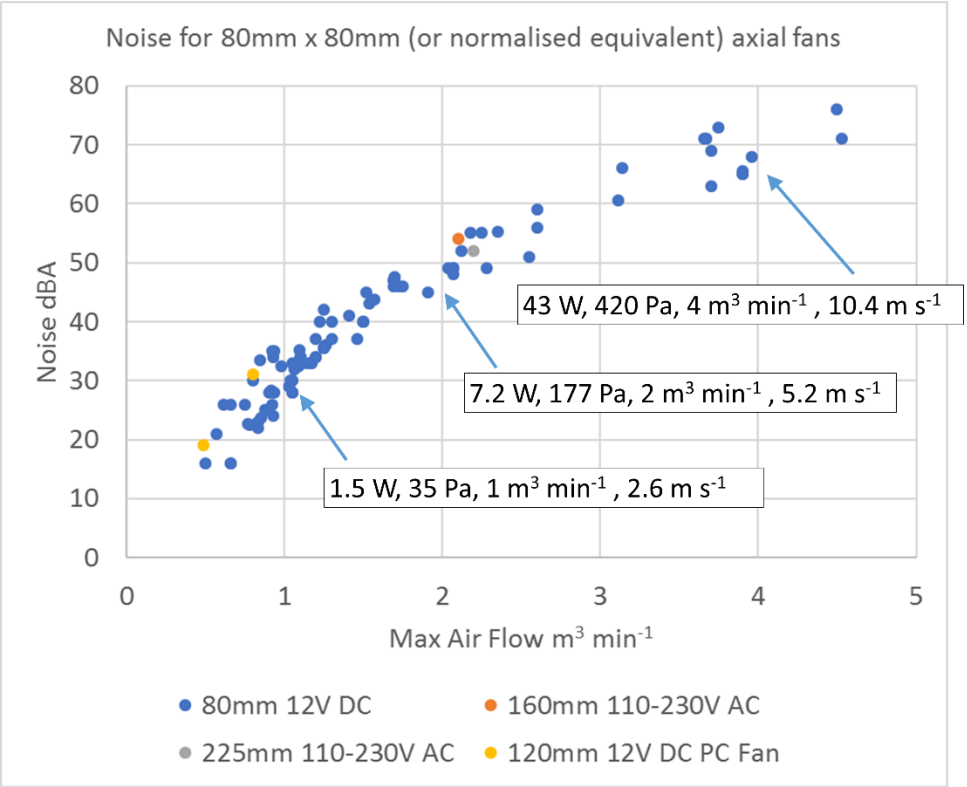
Assumptions: Set to high LUT usage, low DSP usage, VU9P, C2104, LUTs & FFs @ 80%, DSPs @ 30%, Clock @ 480 and 240 MHz, 72 Low Power 10G transceivers, 28 DFE 25G transceivers, 4 transceivers unused, BRAM & URAM @ 80%, No I/O or external memory

The heatsink area was set to 100 mm x 100 mm, significantly beyond the 47.5 mm x 47.5 mm package size to maximise cooling, albeit at the expense of board area. If a heatsink with a footprint comparable with the package were used (i.e. as with most boards at present) the temperatures would at least double. The dashed red line is the maximum permitted temperature.

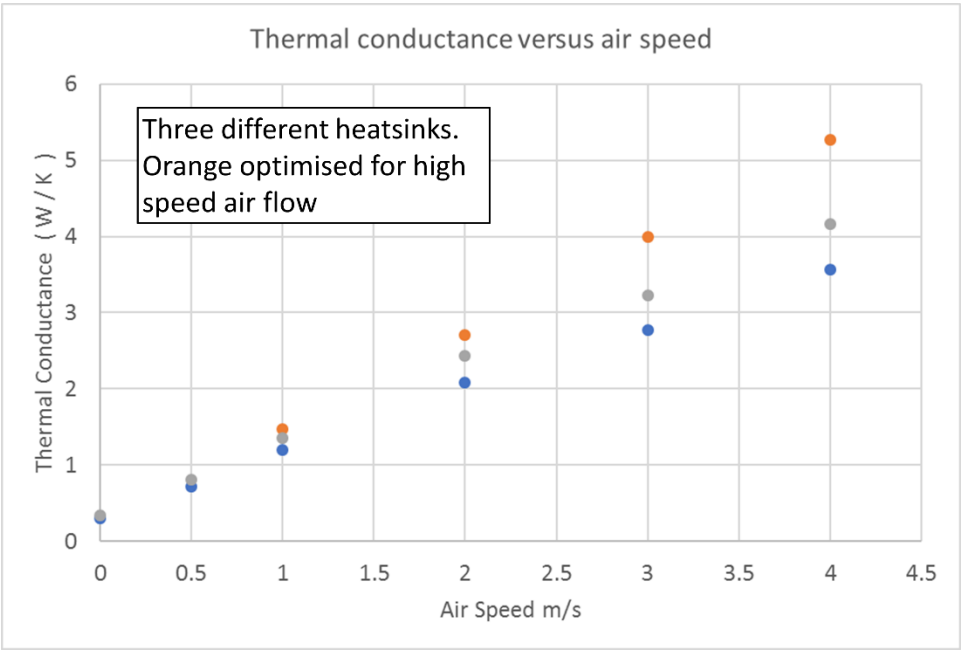




Appendix: Fan noise as a function of air speed (normalised to 80mm x 80mm fans)



Appendix: Thermal conductance versus air speed for different heatsinks



Appendix: Noise levels in USC

Assumed 64 racks, but noise not all concentrated in same location, $1/r^2$ reduction. Reflections from many hard surfaces. Decided to scale by 8 racks (9dB). Worst case noise (i.e. fans on max), assuming no noise reduction is therefore 96 dBA (i.e. 87 dBA per rack + 9dB)

https://espace.cern.ch/ph-dep-ESE-BE-ATCAEvaluationProject/Final%20reports%20public/ATLAS_ATCA_coolingEvaluationProject.pdf