Evaporative CO$_2$ microchannel cooling for the LHCb VELO pixel upgrade

Oscar A. de Aguiar Francisco$^{a+}$, Jan Buytaert$^b$, Paula Collins$^b$, Raphael Dumps$^b$, Malcolm John$^c$, Alessandro Mapelli$^b$ and Giulia Romagnoli$^b$.

$^a$Universidade Federal do Rio de Janeiro (UFRJ), Instituto de Física, Rua Athos da Silveira Ramos 149, CP 68528 CEP 21941-972, Rio de Janeiro, Brazil

$^b$CERN, CH-1211 Genève 23, Switzerland

$^c$University of Oxford, Particle Physics, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, United Kingdom

E-mail: oscar@if.ufrj.br

ABSTRACT: The LHCb Vertex Detector (VELO) will be upgraded in 2018 to a lightweight pixel detector capable of 40 MHz readout and operation in very close proximity to the LHC beams. The thermal management of the system will be provided by evaporative CO$_2$ circulating in microchannels embedded within thin silicon plates. This solution has been selected due to the excellent thermal efficiency, the absence of thermal expansion mismatch with silicon ASICs and sensors, the radiation hardness of CO$_2$, and very low contribution to the material budget. Although microchannel cooling is gaining considerable attention for applications related to microelectronics, it is still a novel technology for particle physics experiments, in particular when combined with evaporative CO$_2$ cooling. The R&D effort for LHCb is focused on the design and layout of the channels together with a fluidic connector and its attachment which must withstand pressures up to 170 bar.

Even distribution of the coolant is ensured by means of the use of restrictions implemented before the entrance to a race track like layout of the main cooling channels. The coolant flow and pressure drop have been simulated as well as the thermal performance of the device. This proceeding describes the design and optimization of the cooling system for LHCb and the latest prototyping results.

©CERN on behalf of the LHCb VELO group

KEYWORDS: Cooling; CO$_2$ evaporative; microchannels; LHCb Upgrade.

$^+Corresponding author.

$^†On behalf of the VELO group and CERN PH-DT.
1. The LHCb VErtex LOcator

LHCb is one of the four major experiments at the Large Hadron Collider (LHC). It is a forward spectrometer designed to measure CP violation, to study rare decays of bottom and charm hadrons and to search for physics beyond the Standard Model. The VErtex LOcator (VELO) is the most upstream detector and is responsible for the vertex reconstruction. It also provides information about the track of the particles produced in the collisions. This system is composed of 88 silicon planes arranged along the beam that provide the hit information on R-φ coordinates. The closest active detecting region is just 8.2 mm from the beam. For this reason, the modules are moved away during the beam injection to prevent damage and then they are only moved in when the beams are stable. This detector can achieve a hit resolution of 4 µm at the optimal track angle. Each module produces approximately 16.5 W of power.

2. VELO Upgrade

The LHCb is going to be upgraded to operate with an instantaneous luminosity that is 5 times higher than in Run I. Furthermore, the closest distance between the beam and the active detecting region is going to be reduced to 5.1 mm. These changes are going to produce a high and non-uniform radiation (8x10^{15} n_{eq}/cm^{2} for 50 fb^{-1}) on the modules and, consequently, a data rate up to 15 Gbit/s for the central ASICs and 2.9 Tbit/s in total.

The cooling solution for the VELO upgrade should fulfill the following criteria:
• The innermost region is going to have the highest radiation damage, consequently, largest
leakage current. To prevent thermal-runaway, the entire sensor should be kept at a maximum
temperature of -20°C safely. To achieve this, it is essential to minimize the $\Delta T$ (<8°C)
between the coolant and the tip of the sensor.

• Each module can generate up to 43 W considering that each one is composed of 12 ASICs
that can generate up to 3 W (36 W/module). The hybrid can produce 5 W of additional heat
and the sensors can produce up to 2 W at the end of their lifetimes. Hence for each module
a total power of 43 W must be removed.

• To fulfill the previous requirements, the coolant should run as close as possible to the ASICs,
inside the LHCb acceptance. In addition, the amount of material on each module should be
minimized to improve the impact parameter resolution. Therefore, the choice of the cooling
technology is crucial to satisfy the thermal requirements with minimal material.

• The modules are located in vacuum, consequently no leaks are tolerated.

• It is advantageous to avoid mechanical distortions due to different coefficient of thermal
expansion (CTE) between the module components.

The preliminary R&D results show that the evaporative CO$_2$ microchannel cooling is the pre-
ferred option capable to fulfill the cooling requirements while providing optimal impact parameter
resolution due to its low mass. The Figure [1] shows the design of the upgrade modules. The advan-
tages of this cooling technique are the following:

• Evaporative cooling [5] is an isothermal process if the pressure drops are neglected, so a very
low gradient of temperature is expected across the module. It is also easy to control because
by regulating the pressure it is possible to regulate the flow of the coolant. The temperature
in this kind of system is also relatively insensitive to the variation of the heat load when the
system is not close to the dry-out$^1$ condition.

• The CO$_2$ coolant has a high latent heat. It is easy to circulate because of its low viscosity.
Since the CO$_2$ is chemical inert it doesn’t corrode the piping or the pump. It is also radiation
hard and non-toxic.

• The microchannels in silicon allow the cooling fluid to be immediately underneath the heat
source which means that the heat does not need to flow a considerable distance before reaching
the coolant. Since the mechanical support for the sensors is also the cooling substrate, the
contribution to the material budget is small. There is no mismatch of expansion coefficients
because the ASICs, sensors and the cooling substrate are made of silicon.

The operation pressure is around 20 bar but the system should be able to withstand the pressure
of the CO$_2$ at room temperature (25°C). At this temperature, the pressure of the boiling CO$_2$ is
around 65 bar. Therefore the cooling system is going to be rated to the same pressure of the current

$^1$The dry-out is the state when all the CO$_2$ that are running on the microchannels evaporates before reaching the
output (undesired).
VELO of 170 bar which corresponds to a safety margin larger than 2.5 times the pressure at room temperature.

![Figure 1.](image)

**Figure 1.** Front view (left) and cross section scheme not in scale (right) of upgraded module. There is a displacement of 5 mm between the tip of the sensor and the cooling substrate on the innermost region. The horizontal and vertical dimensions of the substrate are 80 and 104 mm, respectively.

3. Prototype

The first prototype was produced at CMI (EPFLausanne) by the CERN PH/DT. It was manufactured with an area of 4x6 cm$^2$ using 380 µm silicon wafer and 2 mm pyrex on the other side. For this reason, it is possible to see the pattern of the channels on the Figure 2. This proof-of-principle device was used to test the feasibility of this technique in 2012 [6].

The liquid CO$_2$ is injected in the system through the inlet hole (1.6 mm diameter). Then, it passes through narrow channels (70 µm × 30 µm) that represent the main fluidic resistance of the system called restrictions. The main goal of these restrictions is to trigger the boiling providing high pressure drop. It also helps to equalize the flow and prevent coupling between the channels. The CO$_2$ starts to boil due to the expansion when it reaches the main channels (70µm × 200µm)$^2$.

In the main channels, the coolant is in the two-phase state (liquid/gas). The output of the main channels are collected by a structure called manifold and oriented to the outlet of the cooling substrate.

3.1 Cooling performance

The cooling performance was evaluated using the first prototype as the cooling substrate. Two types of heaters were used to simulate the heat produced by the ASIC and the sensors. Figure 3 shows the scheme of the experimental setup and the result of this study. Since the temperature gradient to the coolant is smaller than 7°C, this technique fulfills the cooling requirement to keep the sensors at -20°C for a CO$_2$ evaporation temperature of -35°C.

A simulation using ANSYS [7] software was performed to evaluate the cooling performance of the module. It was considered that each ASIC produces 3 W of heat, the CO$_2$ is at -30°C and that there is a layer of 100 µm of Stycast 8550FT with catalyst 9 between the ASIC and the substrate.

---

$^2$The dimensions of the restrictions and the main channels are comparable to the diameter of the human hair.
Figure 2. First cooling substrate prototype manufactured with silicon and pyrex. The cooling was simulated assuming that channel walls are at the same temperature as the coolant (-30°C). In these conditions, the maximum variation of temperature across the whole module is approximately 5°C. The region overhanging shows the highest temperatures (Figure 3).

Figure 3. Scheme of the experimental setup used to evaluate the cooling performance is shown on the left. Three probes (uch1, uch2 and uch3) were used to monitor the temperature on different positions over the sensor mockups. The plot (right) shows the measured temperature difference between the sensor and the coolant with fully powered ASICs and a gradual increase in the sensor power. Considering a realistic end of lifetime expectation, half of a module should produce around 13 W (11.5 W from ASIC and 1.5 W from the inner sensor) which corresponds to a variation of temperature smaller than 7°C across the whole module.

3.2 Microchannel fabrication and resistance to pressure

The microchannels fabrication starts with the etching of the channels pattern on a silicon wafer. The second step is to close the channels by bonding a second silicon wafer on top of the first one. This procedure can be done using two types of bonding: hydrophilic and hydrophobic. In the hydrophilic bonding, the presence of water between the two silicon wafers forms an oxide layer.
that becomes the bonding plane. The hydrophobic bonding reconstructs the crystal lattice that is present in a pure bulk of silicon. This process is more difficult to perform because it requires a better surface preparation (high cleanliness and flatness) before the bonding. After the bonding, the thinning adjusts the total thickness of the sample. Plasma etching is used to create the openings for the fluidic inlets. The last step is the metallization for soldering the connectors.

On the high pressure tests, the hydrophilic samples exploded with a maximum pressure around 400 bar. However, the hydrophobic samples resisted up to 700 bar. Both techniques satisfy the maximum required pressure of 170 bar.

The rupture pressure was also studied as function of the silicon cover thickness to optimize the thickness of the cooling plate. Three channels widths (100 µm, 200 µm and 500 µm) were studied as shown on the Figure 5. Those silicon samples were bonded to 2 mm glass. The cover thickness of 140 µm was chosen because the cooling plate is also going to be the mechanical support for the sensors, ASICs and the hybrid. For the nominal LHCb cover thickness and channel width, this structures can hold more than 350 bar well above the required pressure (170 bar).

### 3.3 Cyclic tests

The samples were subjected to cyclic tests to evaluate the long term effects such as accumulated stress and fatigue. On these tests, the experimental setup can provide a pressure between 1 and 200 bar and a variation of temperature between -40°C and +40°C. The high pressure is provided by a bottle of compressed dry air and the temperature is controlled by a couple of Peltier cooling elements attached to a cooling plate. After thousands of cycles, no long term effect was observed.
4. Towards the final microchannels layout

4.1 Optimizations

The main goal of the optimizations is to reduce the overall fluidic resistance while maintaining the resistance to high pressure. Hence the restrictions were made squared (60 µm × 60µm) to avoid clogging while the main channels have increased depth (120 µm × 200 µm). These changes reduced the fluidic resistance on the channels by a factor 4. Each channel has its own inlet and outlet and the manifold is moved to the metallic connector as it is shown in the Figure 6. Considering the CO$_2$ is at -20°C and assuming 30% output vapour quality$^3$, a flow of 0.52 g/s is necessary to dissipate 43 W/module. The total pressure drop is estimated to be ∼3 bar.

4.2 Fluidic connector attachment

The soldering of the connector on the microchannels plate is crucial for the long term reliability. To achieve a reliable attachment, the connector and the plate are metallized with layers of titanium, nickel and gold. A 0.2 µm titanium layer is used on the cooling plate because it has a good attachment to silicon and to prevent the diffusion of metals. The nickel layer of a few microns (1-4 µm) is mixed with the solder to create the intermetallic (soldered layer) during the reflow and a gold layer of 1 µm prevents the oxidation of the samples. The gold layer also dissolves in the solder during the reflow and it will be present on the solder layer. A solder foil of 55 µm made of PbSn is used to attach the connector and the cooling substrate. Figure 7 shows the cross section of the metallization layers and the solder layer.

To avoid the presence of impurities on the surface, the samples are plasma cleaned in advance and they are stored in a vacuum chamber. The reflow is performed in vacuum (∼10$^{-3}$ mbar) with heaters stacked with temperature probes to monitor the sample. The heaters provides energy to

$^3$Vapour quality is mass fraction of vapour with respect to the total mass of the mixture where 1 corresponds to pure saturated gas.
ramp-up the temperature with a rate of approximately +0.7°C/s. The cooling is done by conduction through a teflon layer of controlled thickness that is between the oven and a cooling plate. The cooling rate is around −0.7°C/s. The sample is kept above the melting point of the solder for 45-90 seconds with a maximum temperature of 195°C. During the reflow procedure, small voids are created on the soldering plane. The Figure 8 shows three soldering test samples.

The soldering procedure was already validated using circular simplified samples. The solder on these samples are washer shaped with inner radius of 1 mm, outer radius of 5 mm and 55 µm high. These samples resisted to high pressures up to 400 bar and to a maximum pull force of 580 N (8 N/mm²). The adherence force required on the realistic design is 106 N considering that the CO₂ is at 65 bar. Hence, the realistic design of the solder with an area of 164 mm² can hold more than 1200 N, 10 times higher than the required adherence force. On the other hand, the voids can create a potential leak path or increase the area of the slits reducing the tolerance to high pressure. For this reason, these voids should be reduced in size and number.

Figure 7. Cross section cut scheme of the soldering samples.

Figure 8. X-ray pictures from the solder samples. The small dark blobs are the voids created during the reflow. The input and output slits are 1 mm wide and 10 mm long.
5. Conclusion

The microchannel CO$_2$ evaporative cooling fulfills the requirements for the LHCb VELO upgrade. The channels etched in 400 $\mu$m silicon can safely hold high pressure of more than 170 bar. It can provide the necessary cooling performance to keep the ASIC and sensors at $-20^\circ$C. No fatigue or accumulated stress effects are observed by performing cyclic stress tests. A new optimized layout of the microchannels has been developed to reduce the fluidic resistance and consequently improve its cooling performance. A fluidic connector and a soldering technique under vacuum have been developed. The voids created during the reflow must be reduced in size and number, and the reliability tests are still on going. These reliability tests are focused on the evaluation of the long term effects under the influence of mechanical stress and by performing thermal cycles.

References


