# LEONARDO LABS

Intelligence Autonomous Systems, and Technologies for Future Rotorcraft, HPC/Big Data, Materials, Quantum Computing

**LEONARDO**

# LEONARDO LABS

Intelligence Autonomous Systems, and Technologies
for Future Rotorcraft, HPC/Big Data, Materials, Quantum Computing

## PROPRIETARY NOTICE

# contents

# editorial

This second issue of the new Polaris Innovation Journal series is dedicated to the Leonardo Labs network of technology incubators. The Labs support research and development of the most innovative technologies with a transversal approach that enables the exploration of emerging technologies. They act as a driving force for innovation throughout our Company, also aimed at anticipating future market needs.

The important investment in Labs is a key factor at Leonardo and is a pillar of the company's strategy. It is worth remembering that it is also relevant for the entire Italian Country system, because it contributes to the creation of specialised cutting-edge skills applied to advanced research, our country greatly needs, which are difficult to find today.

Every year, Leonardo invests almost two billion euros in its own Research and Development (R&D) activities - also run in collaboration with universities and external research centres - in which about ten thousand people are employed in various roles and in various capacities.

The Labs are an integral part of the R&D strategies pursued by Leonardo, and operate in conjunction with the company's other research and innovation units. However, a distinction must be made between Research and Development and Research and Technology, as the Labs are oriented towards the latter area. For this reason, too, they are characterised by a more agile and less traditional structure, which favours the freedom of action and lateral thinking that are left to the young researchers working in them, all of whom have PhDs and/or degrees in STEM disciplines.

This approach is intrinsic to Leonardo Labs, and makes them technological hubs dedicated to frontier and breakthrough technologies. They are characterised by a continuous flow of talent to ensure the turnover and renewal of professional skills and competences necessary to guarantee rapid adaptation to changing needs.

We have built a constantly evolving technological innovation ecosystem, consisting of a network of laboratories interconnected with polytechnics, universities, industrial partners and research centres. Within that ecosystem, young research fellows - also of international origin - work together with our company's experts and technicians. It is also for this reason that we have distributed the Labs over the territory, following the criterion of maximum proximity to Leonardo's main industrial sites and to their respective territories of reference.

There are twelve Labs distributed throughout the country, in which young researchers - there will be 200 by 2023 - work in transversal areas across several technological domains. This is why I defined the Labs in my opening remarks as true technology incubators, which look to the near future, investigate technologies that cannot yet be implemented in products, and can innovate with a five-year vision, without having to focus purely on business and on results to achieve in the short term. For example, quantum technologies are not yet implementable in products, but in the Labs we are already investigating how to apply them, as well as quantum computing derived from them.

This is an important advance for our Company, which broadens the field of frontier research and helps to shape the future of the Company itself on an increasingly digital basis, to make Leonardo a digital intensive company. Thanks to the Labs, Leonardo will count on a new pool of skills, capable of giving concrete applications to artificial intelligence, big data, quantum computing, robotics, sensor technology, the cloud, new materials, advanced logistics, autonomous flight systems, etc.

Some of these areas are topic of the articles in this issue of the POLARIS Journal, which are a selection of the publications produced over the past year by the young researchers at the Leonardo Labs'. Contents of their articles give the reader a measure of how far their vision is projected into the future of technologies, and demonstrate the validity of the Company's strategies in promoting the growth of this new generation of professionals.

Head of R&T and Leonardo Labs
Alessandro Massa

# POLARIS

# Exploiting Artificial Vision and Motion Monitoring to Enable Target-Guided Assisted Telemanipulation

Marco Laghi[1,2], Luigi Raiano[1,2], Fabio Amadio[1,2], Federico Rollo[1,2,3], Andrea Zunino[1,2,3], Arash Ajoudani[2]

[1]Leonardo Labs-Intelligent Autonomous System, [2]Istituto Italiano di Tecnologia-Human-Robot Interface and Physical Interaction, [3]Istituto Italiano di Tecnologia - Humanoids and Human Centered Robotics

Teleoperation combines human intelligence with robot power and endurance, making it a perfect solution for hostile-for-human environments. Nonetheless, classical teleoperation interfaces that replicate one-to-one human commands can be quite demanding in terms of physical and mental efforts. Shared-autonomy approaches foresee the combination of direct teleoperation with autonomous control. They can help the operator in overcoming such issues without sacrificing the task execution efficiency. Following this philosophy, we here propose a novel technique to assist the operators in reaching and manipulate objects with one or two arms, combining visual perception and operator's action monitoring. In particular, the operator's intention to reach a target object is detected and used to adapt robot trajectories autonomously towards a grasping pose. In addition, based on the target object size, single- or dual-arm coordinated trajectories are autonomously generated without the need for additional human actions. The experimental results confirm the potential of the proposed framework in terms of success rate and operator effort.

## INTRODUCTION

Teleoperation applications are increasingly gaining interest around the world thanks to their capacity to combine robots accuracy and power, with human versatility and problem-solving capacity in different scenarios [1]. Indeed, it has been exploited to execute delicate operations with extreme precision, as it is the case of medical surgery applications [2], as well as to avoid human presence in dangerous or inaccessible environments that characterize disaster response [3], underwater [4] and space [5] exploration. Soon, we will likely see a more pervasive use of telemanipulation also in common sectors, as industry [6] and elderly care [7]. Generally, the usability of a teleoperation solution is proportional to the level of action and perception over the remote environment provided to the user, also called telepresence [8].

From the control point of view, the most effective way to achieve perfect telepresence is to implement bilateral teleoperation, exploiting force feedback [9]. Nonetheless, it requires more complex architectures, and it is prone to instability due to communication delays, usually handled via passivity-based approaches, as in [10][11]. Simpler and more cost-effective solutions may be provided by unilateral teleoperation, enriched by shared-autonomy strategies [12] that insert some degrees of autonomy over the robot control to compensate for the lack of remote perception. These strategies also improve the user experience by unburdening the operators from taking care of the most tedious and low-level tasks [13] that can be left to the robots to solve autonomously, so diminishing their mental and physical effort. In particular, reaching-and-grasping tasks are quite a challenge for teleoperation, despite their triviality for humans.

Therefore, several solutions have been proposed to assist the user and improve usability and effectiveness of the teleoperated platforms during grasping tasks, as in [14] and [15]. However, both these works are limited to single-arm telemanipulation.

In this article, we propose a solution that covers both single-arm and bimanual teleoperated grasping of box-shaped objects. In particular, we introduce an assisted teleoperation framework that automatically detects the user's intention to approach an object and provides assistance during the reaching motion. The presented strategy exploits a user interface to capture the movements of the user's hands, and an external vision system that recognizes objects and calculates their poses and dimensions. The whole control scheme is divided in two levels (see Figure 1).
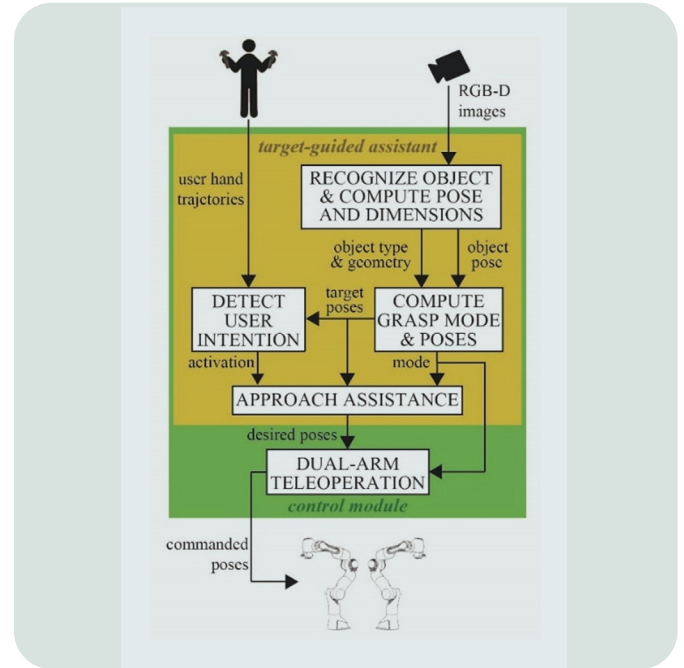
The low level is the dual-arm telemanipulation control developed in [16]. The high level is the proposed target-guided assistant that receives pose and dimensions of the detected object from the vision system, computes the most proper poses and control modality to secure the grasp, and monitors the movements of the user, guiding them towards the target.



*1 – Block Diagram representation of the implemented assisted teleoperation architecture*

In the following, we first recall the adopted dual-arm teleoperation solution introduced in [16]. Then, we present the new assistant and describe how it guides the user in the approach and grasp of the given target, in both single and dual-arm scenarios. The details of the vision system built to recognize, measure, and localize the boxes to manipulate, follow. We then describe the experimental validation and discuss the obtained results, and finally draw the conclusions.

# SHARED-CONTROL BIMANUAL TELEOPERATION

Let us denote with $\Psi_w$ the fixed world frame. The end-effector frames of the right/left robot are indicated with $\Psi_{ee_R}/\Psi_{ee_L}$, whose respective desired end-effector frames are $\Psi_{ee_R,d}/\Psi_{ee_L,d}$. An external tracking system captures the poses of user's hands, with $\Psi_{h_R}/\Psi_{h_L}$ being the frame associated to the right/left hand (see Figure 2). Finally, we indicate the position vector, the rotational matrix, and the homogenous transformation of a frame $\Psi_i$ with respect to (w.r.t.) frame $\Psi_j$ at the instant t with ${}^j p_i(t)$, ${}^j R_i(t)$, and ${}^j H_i(t)$, respectively. For ease of reading, if frame $\Psi_j$ corresponds to $\Psi_w$, the corresponding apex is neglected. In this work, the matrices $H_{h_R}(t)$ and $H_{h_L}(t)$ indicate the user's hand poses w.r.t. the world frame mapped in the manipulators workspace.

For ease of notation, no indication for left/right user/robotic arm is given in this paper, when analogous expressions are valid for both. The control architecture developed in this work exploits the one initially introduced in [16], which enables the user to choose between two control modes: independent or coordinated. In the former, the users control both the arms independently, while in the latter they can control both the robotic platforms coordinately, by using only one arm.

The operator can seamlessly switch between the two modes, depending on the task performed and the user's necessities.

## Independent Control Mode

In this control mode, the user controls the right/left robotic arm by moving the corresponding hand. The desired end-effector frame $\Psi_{ee,d}$ is then defined such that $H_{ee,d}(t) = H_h(t)$.

This is sent to the robot Cartesian Impedance controller, which calculates the desired joint torques $\tau_d(t)$ as

$$\tau_d(t) = M(q(t))\ddot{q}(t) + C(q(t),\dot{q}(t))\dot{q}(t) + g(q(t))$$

$$+ J_{ee}^T(q(t))(Ke(t) + D\dot{e}(t)), \tag{1}$$

where q is the joint positions vector, $M$ the mass matrix, $C$ the centrifugal/Coriolis term, and g the gravity compensation torque contributions. $J_{ee}$ is the end-effector Jacobian matrix, used to map the desired spring-damper cartesian system in the joint space. e(t) is the vector representation of the error between frame $\Psi_{ee,d}$ and $\Psi_{ee}$ at instant t, while $K$ and $D$ are the stiffness and damping matrices, respectively.
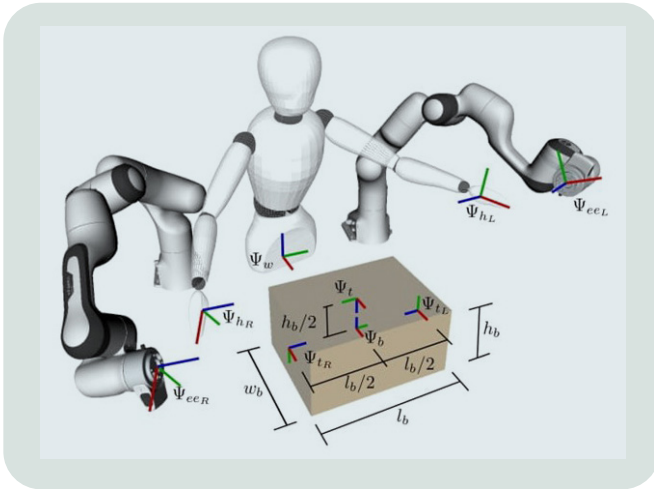
2 - Representation of box dimensions and main frames involved in the assisted teleoperation architecture. $l_b$, $w_b$, and hb are the box length, width and height. $\Psi_b$ is the box barycenter frame, $\Psi_t$ is the target frame for single-arm grasp case, while $\Psi_{t_R}$/$\Psi_{t_L}$ are the target frames for right/left arms for a dual-arm grasp case
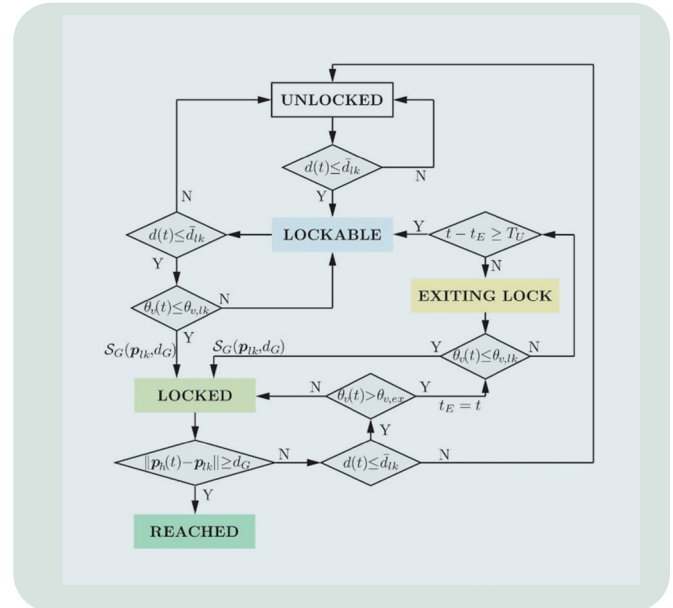


3 – Developed target assistant state machine block diagram representation

## Coordinated Control Mode

In [16], the introduced coordinated control strategy that foresees the possibility to move two robotic arms with just one hand. When activated, this strategy creates a virtual frame in between the two end-effectors, and links it to the user's hand frame. Moreover, the two end-effector desired poses are linked through a linear joint to the virtual frame, while being forced to have a symmetric pose w.r.t. the latter.

The user can control the relative distance between the two end-effectors by commanding the desired position of the virtual prismatic joints, basically by using the two arms as fingers in a pinch grasp. Both robots continue to be controlled at low level by the impedance controller (1). Please refer to [16] for a detailed description of the coordinated control mode.

## TARGET ASSISTANT

Given an object pose and its dimensions, the proposed architecture assists the user through two main phases: it first analyses the object geometry to evaluate the need for one or two robots for the grasping, and consequently calculates the target poses to secure the grasp; then it monitors the user's motion. When it detects the intention to approach the object, it guides the robot/s towards the target pose/s. This second phase is performed through a state-machine (Figure 3), which can assume one of the following states:

- UNLOCKED: the initial state in which the user has total control over the robot, and the assistant is not active;
- LOCKABLE: in this state, the robot is close enough to the target to activate the assistance, depending on the user's intention;
- LOCKED: the assistant is active and corrects the robot/s pose/s while user moves toward the target;
- EXITING LOCK: user intention to leave the target is detected. A safe time interval is awaited before switching back to LOCKABLE;
- REACHED: final state after the target has been reached. The assistant is deactivated and the user is again in full control of the robot/s;

Precise details about the state-machine switching mechanics and assistance strategy are given in the following.

## Control mode selection and target poses computation

Once the information regarding the detected object by the vision system is received, the assistant calculates the desired grasping target poses. In this work we show an algorithm for rigid parallelepipeds, which calculates the poses relying only on their pose and dimensions, but other factors can be taken into consideration if available (e.g., weight, center of mass, softness). Similar, or more advanced procedures, e.g., [17], for other classes of objects can be ad-hoc implemented and used on the need. Let $\Psi_b$ be the frame located at the barycentre of the box and $l_b$ and $w_b$ its length and width, respectively. Such dimensions determine the need for one or both robots to perform a successful grasp.

Let $\overline{L}_{gr}$ be the maximum grasping width of the used gripper, the condition for which only one arm is required is

$$\left(l_b \leq \overline{L}_{gr}\right) \vee \left(w_b \leq \overline{L}_{gr}\right) = 1, \qquad (2)$$

which means that the box has length or width short enough for the gripper to perform a grasp.
If (2) is verified, a single target frame $\Psi_t$ for the grasping end-effector pose is properly positioned in the center of the box upper face, with $\hat{z}_t$ entering the surface and $\hat{x}_t$ aligned with the box shortest side (refer to Figure 2).
If, instead, (2) is not verified, both manipulators are needed to grasp the box and two target frames $\Psi_{t_R}$ and $\Psi_{t_L}$ for the right and left end-effectors are placed each one in the center of two opposite later faces, with their $\hat{z}$ axes pointing towards $\Psi_b$, and parallel $\hat{x}$ axes (refer again to Figure 2).

## Control mode selection and target poses computation

The user's intention to approach the target is detected monitoring its position $\boldsymbol{p}_h(t)$ and linear velocity $\boldsymbol{v}_h(t)$ vectors. To avoid the assistant activates its control too often, the first condition for its activation is that the distance between the target frame origin $\boldsymbol{p}_t$ and the user frame is lower than a given threshold $\overline{\boldsymbol{d}}_k$

$$d(t) = |\boldsymbol{p}_t - \boldsymbol{p}_h(t)| \leq \bar{d}_k. \qquad (3)$$

If this first condition is verified, the target is LOCKABLE. The non-fulfilment of (3) in any instant brings back the target status to UNLOCKED.

Furthermore, we consider that the user is willing to grasp the object if its velocity tends towards the target frame $\Psi_t$, condition formalized as

$$\theta_v(t) = \left|\theta_{\boldsymbol{v}_h(t)}^{\boldsymbol{p}_t - \boldsymbol{p}_h(t)}\right| \leq \theta_{v,lk}, \qquad (4)$$

where we denote with $\theta_{\boldsymbol{v}_h(t)}^{\boldsymbol{p}_t - \boldsymbol{p}_h(t)}$ the angle between the user hand linear velocity vector $\boldsymbol{v}_h(t)$ and the vector $\boldsymbol{p}_t - \boldsymbol{p}_h(t)$, i.e., the segment going from the user position to the target position at instant t, and $0 \leq \theta_{v,lk} < \pi/2$.

Note that (4) is equivalent to check if the translational motion direction (given by the linear velocity vector) resides inside a cone with axis directed as $\boldsymbol{p}_t - \boldsymbol{p}_h(t)$ and aperture $2\theta_{v,lk}$.

If (3) and (4) are verified, the target status is changed to LOCKED, and the assistance comes into effect.
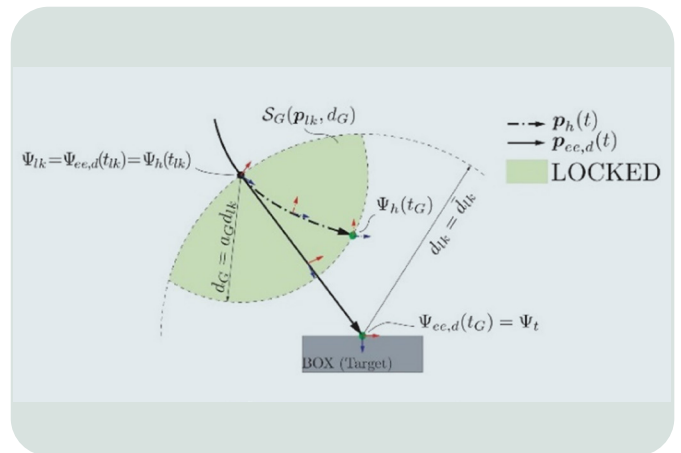


*4 - Graphical example of the guided-approach control. Light-green indicates the sphere SG (plk,dG), the volume within which, under conditions (3) and (4), the guidance is performed. When the user trajectory intersects at instant tG the edge of the assistant sphere, the guidance is completed, as $\Psi_{ee,d}(t_G) \equiv \Psi_t$*

## Guided approach towards the target

Let be t_lk the instant at which the target status is set to LOCKED. At instant $t_{lk}$, two transformations are registered: $H_{lk} = H_{ee,d}(t_{lk})$ and $^{lk}H_t = {}^{ee,d}H_t(t_{lk})$, with $\Psi_{lk}$ denoting the frame $\Psi_{ee,d}$ at instant $t_{lk}$. Then, a trajectory $H_{ee,d}(u)$ parametric on $u(t)$ $(0 \leq u(t) \leq 1)$ is defined for t $\geq t_{lk}$ as

$$H_{ee,d}(u) = H_{lk}\,{}^{lk}H_{ee,d}(u), \qquad (5)$$

with $^{lk}H_{ee,d}(u)$ defined such that

$$
\begin{aligned}
{}^{lk}H_{ee,d}(u = 0) &= I &\rightarrow\quad H_{ee,d}(u = 0) &= H_{lk}; \\
{}^{lk}H_{ee,d}(u = 1) &= {}^{lk}H_t &\rightarrow\quad H_{ee,d}(u = 1) &= H_t.
\end{aligned}
\qquad (6)
$$

At instant $t_{lk}$, also the distance $d_{lk}=d(t_{lk})$ (see (3)) is calculated, and a sphere $S_G(p_{lk}, d_G)$ centered in $p_{lk}$ with radius $d_G = a_G d_{lk}$ is created, with $0 < a_G \leq 1$ (see Figure 4).
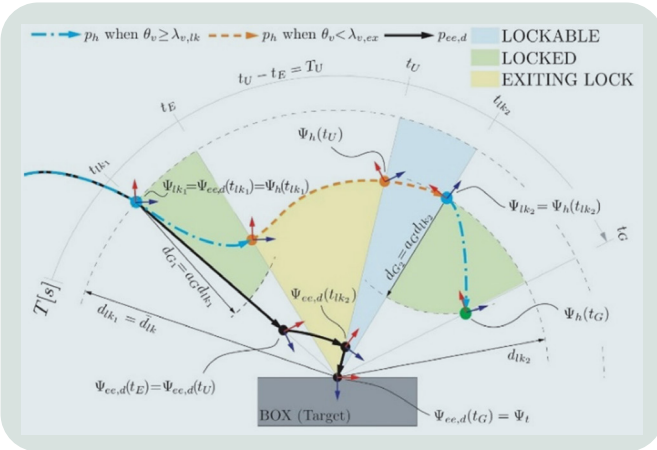


*5 - Graphical example of a possible evolution of the guided-approach mechanism 9)*

$a_G$ can be interpreted as the level of assistance provided to the user, with lower values corresponding to a higher assistance and vice-versa. As long as $p_h(t) \in S_G$ (equivalent to $|p_h(t) - p_{lk}| \leq d_G$), the parameter u of (5) is calculated as

$$u(t) = \frac{|p_h(t) - p_{lk}|}{d_G}. \qquad (7)$$

Let $t_G$ be the instant in which the user crosses the $S_G$ surface:

$$|p_h(t_G) - p_{lk}| = d_G, \qquad (8)$$

then $u(t_G) = 1$ in (7), $H_{ee,d}(u) = H_t$ in (5) (see Figure 4), and consequently the target is reached and the assistant status switches to REACHED (see Figure 3). At $t_G$, the user gains back direct control over the manipulator and can perform the grasp. It is then clear that lower $a_G$, shorter the radius $d_G$ of the sphere $S_G$, as well as the distance that the user must travel to make the robot reach the target frame.

## Target unlock and control reset

While the target is in LOCKED status, if condition

$$\theta_v(t_E) > \theta_{v,ex}, \qquad (9)$$

occurs, the target status switches to EXITING LOCK, and the command pose is freezed to $H_{ee,d}(t_E)$. $\theta_{v,ex}$ in (9) is defined such that $\theta_{v,lk} \leq \theta_{v,ex} < \pi$. If condition (9) persists for at least $T_U$ seconds, at instant $t_U = t_E + T_U$ the target status is set back to LOCKABLE, and the user regains total control over the robot. An example of this mechanism is illustrated in Figure 5.

## Dual-arm extension

When condition (2) is not fulfilled, a single-hand grasp is unfeasible due to excessive dimensions of the object detected, and both the robot arms must be employed. User's intention to grasp the detected object is inferred by monitoring the movements of both hands, by using the same method previously described. When conditions (3) and (4) are met by any of the two hands, the assistance is activated.

Without loss of generality, let us assume that the locking condition is triggered by the right robot. At locking instant $t_{lk}$, transformations $H_{lk_R} = H_{ee_R,d}(t_{lk})$ and $H_{lk_L} = H_{ee_L,d}(t_{lk})$ are registered and used to calculate the parametric trajectories for both end-effectors as in (5). The two manipulators follow their respective trajectories, both evolving on the same u parameter dependant on the user right hand movement through (7). Once the two targets are reached *(u = 1)*, the coordinated control mode is automatically selected, and the user can finalize the dual-arm grasp decreasing the distance between the two end-effectors (see [16]).

A demonstration of the assistant functioning in both single and dual-arm cases can be seen in [21].

# TARGET DETECTING VISION SYSTEM

This Section describes the vision system integrated in our assisted teleoperation architecture to recognize, which is used to localize and measure the boxes to manipulate, as the target objects in this work. We employed an RGB-D camera framing the work area from above parallel to the table. In the following, we provide the steps followed by the vision system to extract the needed information from visual data.

## RGB object segmentation

To detect the presence of an object in the working area, we adopted the YOLACT algorithm [18], a full convolutionary model for real-time instance object segmentation. The algorithm can segment over RGB images (Figure 6(a)) objects with powerful accuracy and speed.

To use such algorithm on custom data, training set of images are required, in which the objects to recognize are manually annotated. Such information is provided to the network as the ground truth objects position and shapes. Once adapted to the custom data, YOLACT is able to correctly localize the new objects over RGB images (Figure 6(b)), by providing a precise binary mask for each detected target (Figure 6(c)).

Note that even if the computer vision system is trained ad-hoc to recognize boxes for the current application, it can be easily extended to detect different and general objects by means of a labelling and training procedure.

## Object pose and dimensions computation

Computation of the boxes pose and dimensions is performed by exploiting the intrinsic characteristics of the camera that is able to acquire RGB and depth data spatially overlapped. First, we element-wise multiply the two with the refined mask and construct a point cloud of the box (Figure 6(d)), we retrieve the box vertices extracted from the mask, and map them to the corresponding 3D points in the segmented point cloud. Then, box's length $l_b$, width $w_b$ and height $h_b$ are trivially reconstructed from the obtained 3D vertices points, along with the position and orientation of the box barycentre frame $\Psi_b$ w.r.t. world frame $\Psi_w$, thereby obtaining all the information needed by the target-guided assistant to compute the target poses. Considering boxes as rigid target objects simplifies the computations. In case of considering irregularly shaped objects, more advanced algorithms tailoring precise 6D object pose estimations, *e.g.*, [19], must be adopted, but it is out of the scope of our work.
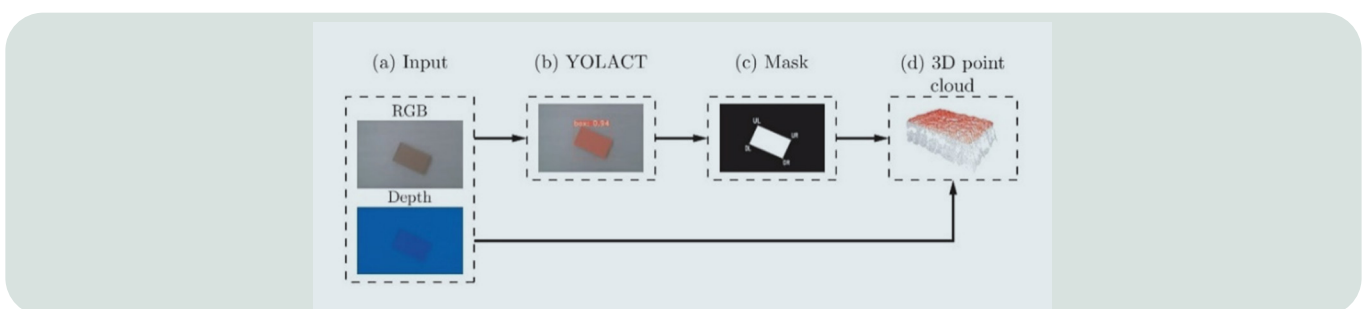


*6 - Computer vision pipeline:*
*(a) RGB-D input; (b) object detection with YOLACT;*
*(c) refinement of the object mask;*
*(d) point cloud generation for the detected region*

# EXPERIMENTAL VALIDATION

The validation of the proposed target-guided telemanipulation architecture has been carried out through a multi-subject campaign, with the main goal of comparing our novel approach against the classical telemanipulation, i.e. providing no assistance to the user during the task execution. In this section, we provide a detailed description of the experiments and their outcomes.
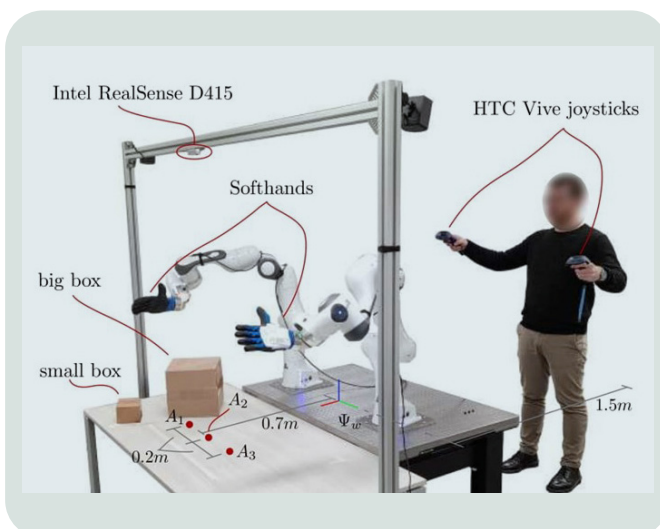
## Experimental setup

The dual-arm robotic platform used in this work is composed by two Panda arms (Franka Emika), each one provided with a SoftHand gripper (Qb Robotics) as end-effector. The two robot bases are fixed on the same plane, positioned symmetrically w.r.t. $\Psi_w$ and equally oriented (see Figure 7). The platform is controlled by an external computer that implements the whole control block (see Figure 1) in ROS and commands both robots according to (1) and the grippers closure with a refresh rate of 1kHz.

The stiffness matrix K of (1) is set diagonal with linear and rotational elements values at 500N and 15Nm, respectively, while the damping D is defined with a damping factor $\xi = 1$. These values are used both for the single and dual-arm modes. The parameters of the target-guided assistant are empirically set as $L_{gr} = 0.1m$, $d_{lk} = 0.5m$, $a_G = 0.5$, $\theta_{v,lk} = \pi/3$, $\theta_{v,ex} = 3\pi/4$, and $T_U = 1s$, which provide a reasonable trade-off between the level of assistance and a target selection precision.

The user's control interface has been implemented by using an HTC Vive Pro that allows to retrieve the joystick poses in real time at 30Hz. Specifically, a custom ROS node maps such poses to the robot workspace, resulting in the matrices $H_{h_R}$ and $H_{h_L}$, and computes the linear velocities $v_{h_R}$ and $v_{h_L}$ required to interpret the user's intention.

The artificial vision system exploits an Intel RealSense Depth Camera D415, and the YOLACT algorithm has been trained for the experimental section through a small dataset of manually labelled RGB-D images (~200), captured through the D415 camera and displaying several boxes of different dimensions located in random positions over the working table. A ROS node shares the object computed information through the whole network with a frequency of 1Hz.



7 – Setup used in the experimental sections

## Population and experimental protocol

The experimental campaign is performed by enrolling 13 volunteers (aged between 25 and 34 years) who are asked to perform manipulation tasks in two experimental sessions:

- Single-arm telemanipulation: the object to manipulate is small (a 0.09m side cube). This session is further divided into two parts: (i) single-arm without assistance (*si*) and (ii) single arm with assistance (*sa*).

- Dual-arm telemanipulation: the object to manipulate is large (a 0.24m side cube). This session is further divided into three parts: (i) dual-arm independent mode without assistance (*di*), (ii) dual-arm coordinated mode without assistance (*dc*), and (iii) dual-arm with assistance (*da*), starting in independent mode.

The different parts of the experimental protocol can be seen in/at [21].

Subjects are always asked to perform first the single-arm session and then the dual-arm, while the order of the parts within a session is randomized. Subjects are informed a priori about the active control mode, in order to be aware of the system behaviour. For each part, the volunteers are asked to perform the trial consisting in lifting the box from the picking position $A_i$ ($i = 1 \div 3$, see Figure 7) of 0.1m by teleoperating the robotic system. Each trial is repeated 3 times and is considered successful if the box does not fall. The order of the picking positions is randomized.

The performance of each participant in each part is measured by using the following parameters:

- the success rate (*SR*), measured as the number of successful trials over the total number of trials;

- the time required for completing a successful trial ($T_c$[s]);

- the scores of NASA Task Load Index (TLX) questionnaire [20], filled in by subjects at the end of each part.

## Results and discussion

The results obtained for the single and dual-arm sessions are depicted in Figure 8 and Figure 9, respectively. In regard to the single-arm session, when performing the tasks with the assistance, the volunteers significantly increased their $SR$ and reduced their $T_c$ (top-left and right plots of Figure 8, respectively). Average sa success rate is 85.5%, while in the si part only 70.1% of the grasps were successful. Regarding the average completion times, in the sa part, the subjects were able to obtain $T_c$ 32.2% shorter than the *si* case. Furthermore, they perceived smaller Mental Demand, Effort and Frustration, while no significant difference was found for the other items of the TLX scale (Figure 8, bottom plot).

Similarly to the single-arm session, in the dual-arm the di part resulted to be the worst both in terms of performance ($SR$ and $T_c$) and of TLX scale. Indeed, according to [16], in our experiments we confirmed that the dc outperforms di in terms of $SR$ (di: 82.1%, dc: 98.3%). In addition, we also found a mild average reduction of $T_c$ of 8.3%, and an improvement of the performance and effort TLX indices. Conversely, during the da, both $SR$ and $T_c$ are statistically different from di and volunteers obtained better success rates (99.1% on average) and a mean $T_c$ times reduction of a 40.9% w.r.t. di. Moreover, a statistically significant difference was found for all TLX items, except for temporal demand.

When comparing dc to da, volunteers further reduced T_c of 35.6% when receiving assistance. This leads to consider our assistance approach as the best choice among the tested ones, to implement for aiding users during telemanipulation tasks. Indeed an evident decrease of the median values of the TLX items in favour of da is shown in the bar plots of Figure 9, thus underlying also an advantage perceived by the volunteers in terms of mental and physical demand.
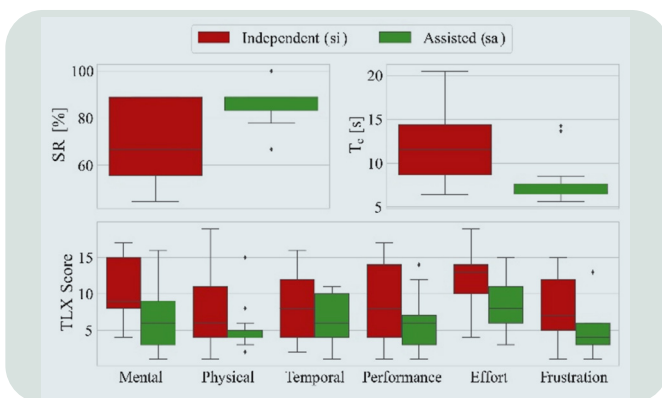


*8-Box Plots related to the results obtained during he single-arm experimental session. While for the SR (top-left) the higher the better, for Tc (top-right) and TLX Scores (bottom) the lower the better*
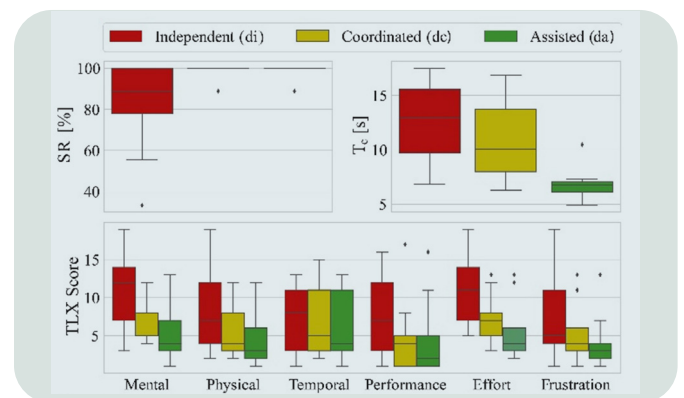


*9-Box Plots related to the results obtained during the dual-arm experimental session. While for the SR (top-left) the higher the better, for Tc (top-right) and TLX Scores (bottom) the lower the better*

## CONCLUSIONS

In this work we have introduced a novel assistant for the approaching and grasping of boxes in teleoperation scenarios, which relies on object information provided by a dedicate vision system and automatically calculates a suitable grasping pose on the detected object. Then, it monitors and interprets the user's intention by analysing his hands movements. Once the assistant detects the user's intention of reaching for the box, it guides them towards the pre-computed grasping pose, while the operator simply dictates the pace of the motion by moving his hand. Furthermore, we have extended our assisted teleoperation strategy to dual-arm grasping tasks by exploiting the shared-autonomy controller of [16] to manage coordination between the two robots. The effectiveness of the proposed solution has been assessed through a multi-subject campaign, in which each subject was asked to reach and grasp boxes of different sizes in different teleoperation settings. The obtained results have showed how the proposed assisted architecture can reduce the time to perform a grasp, while increasing the success rate. Moreover, its presence improves the overall user experience, lowering both the physical and mental effort required during teleoperation.

In the future, the assistant can be improved by including the capacity of handling numerous and various objects placed in the working area. We already made some steps in this directions, including a 'plastic enveloped' object class [22]. This object class is treated differently from the box class in the dual-arm case, for which the end effectors are oriented differently to facilitate their grasp. As shown in the previous video [22], the system automatically detects the different object classes and handle them consequently.

Furthermore, trajectories more sophisticated than the proposed linear one may be beneficial for the effectiveness of the assistant, which could also consider the approaching direction and/or orientation and be object shape-or target pose-dependent.

Marco Laghi: marco.laghi.ext@leonardo.com

## REFERENCES

[1] Adamides, G., Christou, G., Katsanos, C., Xenos, M., & Hadzilacos, T. (2014). Usability guidelines for the design of robot teleoperation: A taxonomy. IEEE Transactions on Human-Machine Systems, 45(2), 256-262.

[2] Meli, L., Pacchierotti, C., & Prattichizzo, D. (2017). Experimental evaluation of magnified haptic feedback for robot-assisted needle insertion and palpation. The International Journal of Medical Robotics and Computer Assisted Surgery, 13(4), e1809.

[3] Negrello, F., Settimi, A., Caporale, D., Lentini, G., Poggiani, M., Kanoulas, D., ... & Catalano, M. G. (2018). Humanoids at work: The walk-man robot in a postearthquake scenario. IEEE Robotics & Automation Magazine, 25(3), 8-22.

[4] Brantner, G., & Khatib, O. (2021). Controlling Ocean One: Human–robot collaboration for deep-sea manipulation. Journal of Field Robotics, 38(1), 28-51.

[5] Artigas, J., Balachandran, R., Riecke, C., Stelzer, M., Weber, B., Ryu, J. H., & Albu-Schaeffer, A. (2016, May). Kontur-2: force-feedback teleoperation from the international space station. In 2016 IEEE International Conference on Robotics and Automation (ICRA) (pp. 1166-1173). IEEE.

[6] Shukla, A., & Karki, H. (2016). Application of robotics in onshore oil and gas industry — A review Part I. Robotics and Autonomous Systems, 75, 490-507.

[7] Lv, H., Yang, G., Zhou, H., Huang, X., Yang, H., & Pang, Z. (2020). Teleoperation of collaborative robot for remote dementia care in home environments. IEEE Journal of Translational Engineering in Health and Medicine, 8, 1-10.

[8] Schloerb, D. W. (1995). A quantitative measure of telepresence. Presence: Teleoperators & Virtual Environments, 4(1), 64-80.

[9] Laghi, M., Ajoudani, A., Catalano, M. G., & Bicchi, A. (2020). Unifying bilateral teleoperation and tele-impedance for enhanced user experience. The International Journal of Robotics Research, 39(4), 514-539.

[10] Rebelo, J., & Schiele, A. (2014). Time domain passivity controller for 4-channel time-delay bilateral teleoperation. IEEE transactions on haptics, 8(1), 79-89.

[11] Panzirsch, M., Singh, H., & Ott, C. (2020). The 6-dof implementation of the energy-reflection based time domain passivity approach with preservation of physical coupling behavior. IEEE Robotics and Automation Letters, 5(4), 6756-6763.

[12] Selvaggio, M., Cognetti, M., Nikolaidis, S., Ivaldi, S., & Siciliano, B. (2021). Autonomy in physical human-robot interaction: A brief survey. IEEE Robotics and Automation Letters.

[13] Gholami, S., Garate, V. R., De Momi, E., & Ajoudani, A. (2020). A Shared-Autonomy Approach to Goal Detection and Navigation Control of Mobile Collaborative Robots. In 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) (pp. 1026-1032). IEEE.

[14] Namiki, A., Matsumoto, Y., Maruyama, T., & Liu, Y. (2017, May). Vision-based predictive assist control on master-slave systems. In 2017 IEEE International Conference on Robotics and Automation (ICRA) (pp. 5357-5362). IEEE.

[15] Selvaggio, M., Chen, F., Gao, B., Notomista, G., Trapani, F., & Caldwell, D. (2016, August). Vision based virtual fixture generation for teleoperated robotic manipulation. In 2016 International Conference on Advanced Robotics and Mechatronics (ICARM) (pp. 190-195). IEEE.

[16] Laghi, M., Maimeri, M., Marchand, M., Leparoux, C., Catalano, M., Ajoudani, A., & Bicchi, A. (2018, November). Shared-autonomy control for intuitive bimanual tele-manipulation. In 2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids) (pp. 1-9). IEEE.

[17] Du, G., Wang, K., Lian, S., & Zhao, K. (2021). Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: a review. Artificial Intelligence Review, 54(3), 1677-1734.

[18] Bolya, D., Zhou, C., Xiao, F., & Lee, Y. J. (2020). Yolact++: Better real-time instance segmentation. IEEE transactions on pattern analysis and machine intelligence.

[19] He, Y., Huang, H., Fan, H., Chen, Q., & Sun, J. (2021). Ffb6d: A full flow bidirectional fusion network for 6d pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3003-3013).

[20] Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage publications.

[21] https://leonardo.canto.global/b/J5PVB

[22] https://leonardo.canto.global/b/OE7GU

# Next-Generation HPC Models for Future Rotorcraft Applications

Nicoletta Sanguini[1], Tommaso Benacchio[1], Daniele Malacrida[2], Federico Cipolletta[2], Francesco Rondina[2], Antonio Sciarappa[2], Luigi Capone[2]

[1]Leonardo Labs - Future Rotorcraft Technologies, [2]Leonardo Labs - HPC/Cloud/Big Data Technologies

Rotorcraft technologies pose great challenges for numerical computing. As available computational resources approach the exascale, more accurate simulations of engineering test cases become accessible. However, shifting legacy workflows and performance optimization of existing software on new hardware are often demanding. This paper reports results of Computational Fluid Dynamics (CFD) and structural dynamics simulations on Leonardo's davinci-1 high-performance computing (HPC) facility. Time to solution and scalability are assessed for the commercial packages Ansys Fluent, STAR-CCM+, and ABAQUS, and the open-source framework PyFR. In compressible CFD runs, normalized time to solution with PyFR is found to be up to 8 times shorter than with Fluent and STAR-CCM+, with all models offering weak and strong scaling on up to 48 GPU nodes. In structural runs with ABAQUS, both the iterative solver and the direct solver provide parallel speedup. The results showcase the potential of HPC architectures for scaling cutting-edge toolkits towards aircraft certification by simulation.

## INTRODUCTION

In the past twenty years, the exponential growth in computational power supplied by HPC clusters has made an increasingly broad range of scales accessible to computational fluid dynamics and structural mechanics simulations, along with great potential efficiencies in operational workflows. In this context, aerodynamics of rotating-wing aircraft presents unique challenges to full-scale numerical representations. For external aerodynamics, the need to simulate high-Reynolds number and high-Mach number flows typically forces practitioners to use steady, averaged models, as time-varying direct numerical simulations on feature-resolving meshes imply times to solution that are currently incompatible with aircraft production schedules. However, emerging computational frameworks harnessing heterogeneous HPC systems have the potential to change that landscape and make scale-resolving simulations available [1].

This paper aims at evaluating the performance of the open-source framework PyFR and commercial packages Ansys Fluent and STAR-CCM+ in time-varying Computational Fluid Dynamics (CFD) simulations as well as of the ABAQUS framework in structural dynamics simulations on Leonardo's HPC facility davinci-1. For the CFD simulations, particular attention is given to performance and scalability on GPU nodes. In this first assessment, we consider a relatively simple geometry, i.e., a single low-pressure turbine (LPT) blade T106A[2], and focus on a preliminary comparison of software tools in terms of fidelity of results, time to solution and scalability.

The work described in this paper was carried out at Leonardo's Corporate Research Laboratory Leonardo Labs. Leonardo, and Leonardo Helicopters Division (LHD) in particular, routinely carry out CFD-based analysis and optimization for, e.g., rotorcraft design,

layout and sizing, aerodynamic characterization, and validation of wind tunnel data, which help reducing industrial costs associated with physical prototypes and wind tunnel testing. This paper is organized as follows. Section 2 provides details on the software and hardware setup for the CFD and structural simulations. Section 3 reports details on the mesh used, the test case setup, and numerical results on fluid flow simulations and structural simulations, and Section 4 contains final considerations and an outlook to future work.

## COMPUTATIONAL SETUP

This paper considers several numerical frameworks for computational fluid dynamics and structural dynamics simulations. For computational fluid dynamics simulations, commercial packages Ansys Fluent and Simcenter STAR-CCM+ were chosen as they are widely used tools in an industrial context, including at Leonardo.

They offer a complete workflow of pre-processing tools, computational models for compressible and incompressible fluid flow simulations, and post-processing capabilities with interactive GUIs as well as batch running modes. While Fluent already has support for running on GPU-based architectures, STAR-CCM+ only has GPU support for incompressible runs. Both packages employ finite-volume based numerical methods that are at most second-order accurate. In addition, the open-source framework PyFR was considered for CFD runs.

PyFR [3] solves the compressible fluid flow equations by using the element-based flux reconstruction method [4] without any use of turbulence models. It also offers the option of running at high order via the use of high-degree polynomials, and accommodates support for unstructured meshes with several element shapes, thereby lending itself well to simulations on complex geometries.

Thanks to its data locality properties - higher accuracy is achieved at higher polynomial order by using multiple degrees of freedom inside a single mesh element - the flux reconstruction method features advantageous parallelization capabilities compared to wide stencils-based higher-order extensions of finite-volume methods. In terms of implementation, PyFR's Python code base with runtime-metaprogramming and embedded domain-specific language technology was designed for hardware portability.

In particular, PyFR has native GPU support via runtime-generated OpenCL, CUDA, and HIP backends, through which the whole code base is automatically optimized for high performance on GPU-based HPC clusters. The code was successfully tested on industrial LPT blade simulations [1], showcased excellent scalability and efficiency on thousands of GPUs [5], and was previously compared with STAR-CCM+ [6].
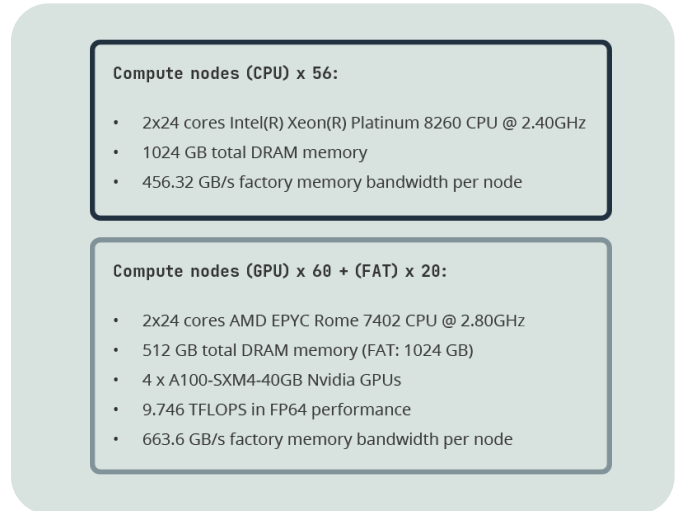
| PyFR v. 1.12.2 | |
|---|---|
| Systems | Compressible Euler, Navier-Stokes |
| Element Types | Triangles, quads, hexahedra, prisms, tetrahedra, pyramids |
| Platforms | CPU, GPU (NVIDIA and AMD) |
| Spatial Discretization | Flux reconstruction (choice of polynomial order p) |
| Turbulence model | None (DNS) |

| Ansys Fluent v2021R1 | |
|---|---|
| Systems | Compressible Euler, Navier-Stokes |
| Element Types | Tetrahedral, polyhedral, triangles, quads |
| Platforms | CPU, GPU |
| Spatial Discretization | Finite volume |
| Turbulence model | Laminar or LES |

| SimCenter STAR-CCM+ v2022.1 | |
|---|---|
| Systems | Compressible Euler, Navier-Stokes, etc. |
| Element Types | Tetrahedral, polyhedral, etc. |
| Platforms | CPU (GPU incompressible only) |
| Spatial Discretization | Finite volume |
| Turbulence model | Laminar or LES |

*Inset 1-CFD computational frameworks considered in this paper*

Inset 1, inspired by [6], provides details of the CFD software versions used in this paper. In order to provide as fair a comparison as possible, the times-to-solution and scalability figures for Fluent and STAR-CCM+ refer to runs in Laminar mode, i.e. without the use of any turbulence model (see also details on parameter choices in Section 3.1 and discussion in Section 4). For the structural mechanics simulations, the commercial software ABAQUS was chosen for its wide adoption in industrial workflows. This software offers a unified environment to generate both static and dynamic structural cases on a wide range of finite element formulations, supporting several load cases and interaction algorithms to model complex full-scale problems consisting of several hundreds of parts. In addition, its robust solvers can be fine-tuned by the user in order to solve nonlinearities, with minimal impact on computational resources.
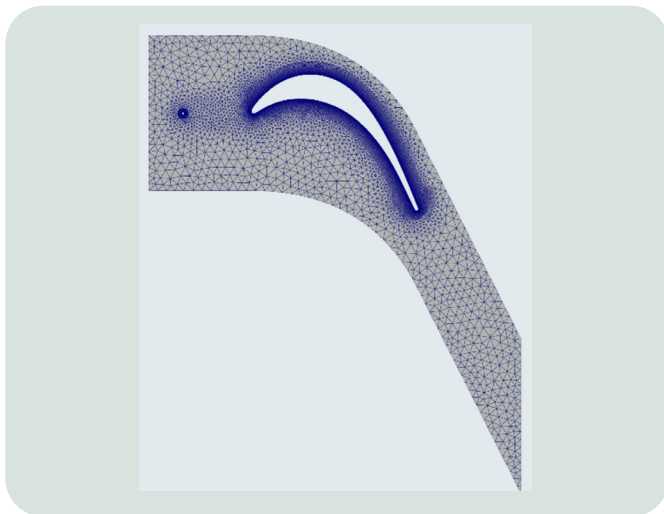
## Hardware infrastructure: davinci-1

CFD and structural mechanics simulations in this paper were performed using Leonardo's HPC facility davinci-1, which has been operative at Leonardo since 2020 to support Corporate R&D activities. Davinci-1 has 116 computing nodes with 48 CPU cores each. A partition of 80 computing nodes is accelerated with 4 NVidia A100 GPUs per node. Davinci-1 supplies a peak performance of 6.5 Petaflops, ranking #164 overall, and #4 in the AD&S sector, on the Top500 list (November 2022 standings). Due to license issues, only 10 computing nodes could be used for runs with Fluent. Figure 1 contains more details on the davinci-1 infrastructure.

**Compute nodes (CPU) x 56:**

- 2x24 cores Intel(R) Xeon(R) Platinum 8260 CPU @ 2.40GHz
- 1024 GB total DRAM memory
- 456.32 GB/s factory memory bandwidth per node

**Compute nodes (GPU) x 60 + (FAT) x 20:**

- 2x24 cores AMD EPYC Rome 7402 CPU @ 2.80GHz
- 512 GB total DRAM memory (FAT: 1024 GB)
- 4 x A100-SXM4-40GB Nvidia GPUs
- 9.746 TFLOPS in FP64 performance
- 663.6 GB/s factory memory bandwidth per node

*1-Davinci-1 computational resources*

# NUMERICAL RESULTS



*2-Unstructured mesh for the CFD experiments*

## Flow around an LPT blade

In this study, we focus on the T106A LPT blade geometry [2]. The simulation setup has a single blade exposed to uniform inflow. To consider slightly different dynamics from what already seen in literature, a cylinder was placed in front of the blade as a wake generator.

The geometry of the blade was normalized with respect to the chord c and the span h set as $h = c = 1$. The mesh was generated with the Pointwise software by extrusion starting from the two-dimensional triangle-based unstructured mesh on the side wall. To solve the boundary layers near the surface of the blade and of the cylinder, we used the T-Rex method within Pointwise and generated a mesh with 50 layers of quadrilaterals below the triangles. The minimum wall spacing was set at $6e^{-5}$ and a growth rate of 1.1 was used (Figure 2).

Along the spanwise direction, $N = 40$ elements of length $l = h/N$ were created by extrusion. The resulting three-dimensional mesh has 663,240 elements (483,400 hexahedra and 179,840 prisms). For some tests in this work, additional meshes were built by modifying the length of the span $h$, and consequently the number of elements $N$, while keeping the length of the cells l constant. For the initial conditions, the velocity was set to zero. The boundary conditions prescribe inlet and outlet velocity, and periodic upper, lower, and lateral walls.

In the first test, we performed the simulation with PyFR at several polynomial orders p. In particular, we considered $p = 1$ and $p = 2$. To set the time step, we conservatively imposed an initial Courant number equal to 0.3. As a result, for $p = 1$ the resulting initial time step was 1.5E-5, while for $p = 2$ the time step was 1E-5. We considered dimensionless dynamic viscosity $\mu = 2.77E-6$. For the compressible simulation, a purely horizontal velocity with dimensionless magnitude 0.114 was set at the inlet. As in [2], at the outflow an angle of -63.2 degrees and a velocity magnitude of 0.495 were imposed. The inlet flow was set along the streamwise direction, so that the flow arrived perturbed at the blade after passing through the wake generator.

Based on the chord $c$, and considering a velocity value of 0.5 (see below) a Reynolds number of ~250,000 was obtained. Figure 3 shows the velocity magnitude at $t$ = 100 for the two simulations. In both cases, the numerical solution provides a stable representation of the turbulent dynamics in the wake of the cylinder and the blade. The recorded maximum dimensionless velocity is for both cases 0.53 (about 170 $m/s$, maximum Mach number of around 0.5).

In addition, the higher-order solution with $p$ = 2 is more detailed than the simulation with $p$ = 1, especially in the wake of the blade. This is confirmed by a probe measurement for the vorticity magnitude at a point of the wake of the turbine blade (red crosses in the top of Figure 3). The time series (Figure 3 bottom) clearly reveals that, while the behaviour is not perfectly periodic due to the wake being underresolved at both orders, the higher order simulation yields deeper vortices. This is expected in view of the higher order of accuracy and effective resolution - the considered mesh has about 4.97$M$ degrees of freedom (DoFs) at $p$ = 1, and two and a half times as many DoFs (~12.4$M$) at $p$ = 2. Preliminary results on a higher-order run at $p$ = 4 (not shown) display deeper vortices than at lower orders.

In the second test, we compared the performance of PyFR and the two commercial packages Ansys Fluent and Star-CCM+. In this case, we considered a mesh with about 2.65$M$ elements (i.e., with a $4h$ span). For all three simulations we used the same number of computational resources, *i.e.*, 4 davinci-1 GPUs nodes.
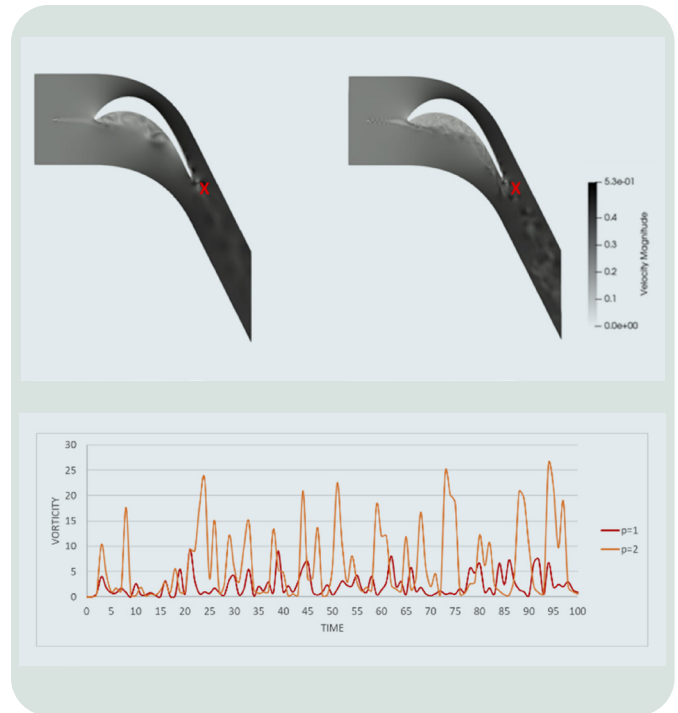


*3 - PyFR simulation on the 663.24K mesh.*
*Top: velocity magnitude at T=0.06 s using polynomial degrees p=1 (left) and p=2 (right).*
*Bottom: time series of vorticity magnitude in the wake of the blade (markers in top plots) using polynomial degrees p=1 (red)*
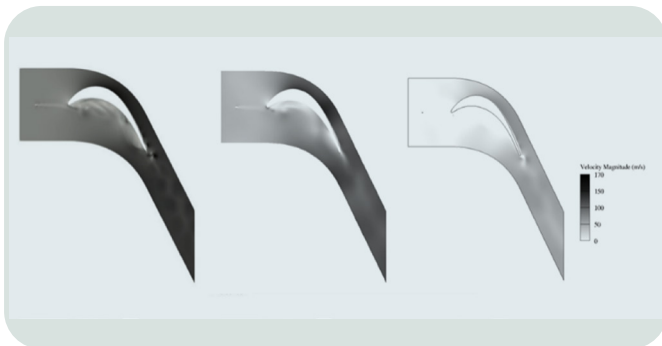


*4 - Velocity magnitude at T=0.02 s in the T106A simulation with PyFR (p1, left),*
*Fluent (center),*
*and STAR-CCM+ (right), 2.65M elements mesh*

As regards Ansys Fluent and STAR-CCM+, compressible flow was considered in the laminar configuration (no subgrid-scale turbulence model). In both cases an unsteady implicit solver was used, with a time-step of 5E-6s. The dynamic viscosity was considered constant and equal to 1.789E-5 $Pa*s$ to have the same Reynolds number of dimensionless PyFR simulations.

The results obtained by the three tools are qualitatively similar (Figure 4). StarCCM+ has a lower velocity than the other two tools, while the solution with PyFR contains more details of the dynamics than the two commercial codes. This is expected since Fluent and Star use finite volumes (1 DoF per element), while in this case PyFR runs with polynomial order $p$ = 1 (~7.5 DoF per element on the considered mesh).

In terms of time to solution, the wallclock time values of PyFR ($p$ = 1) and Fluent are very similar, while Star is about 40% faster (Figure 5 top). This result is surprising, considering that Star runs on CPUs without any accelerators for this compressible case. However, the situation is quite different when considering the wallclock time per DoF (Figure 5 bottom). Here, PyFR is more than eight times faster than Fluent and more than four times faster than Star.

For further comparison, we also ran a simulation with PyFR on a coarser mesh (~180$K$ elements), but at polynomial order $p$ = 4, to have the same DoFs of the $p$ = 1 case (~19.87$M$ DoFs on the 2.65$M$ element mesh). The total wallclock time in this case is equal to four times that of PyFR at $p$ = 1, but in terms of wallclock time per DoF the result is still superior and more accurate than both Fluent and Star.



*5 - Wallclock time (top left) and wallclock time / DoF (bottom right) with PyFR (p = 1 and p = 4), Ansys Fluent, and STAR-CCM+ in the T106A simulation, T=0.02 s, 2.65M elements mesh*
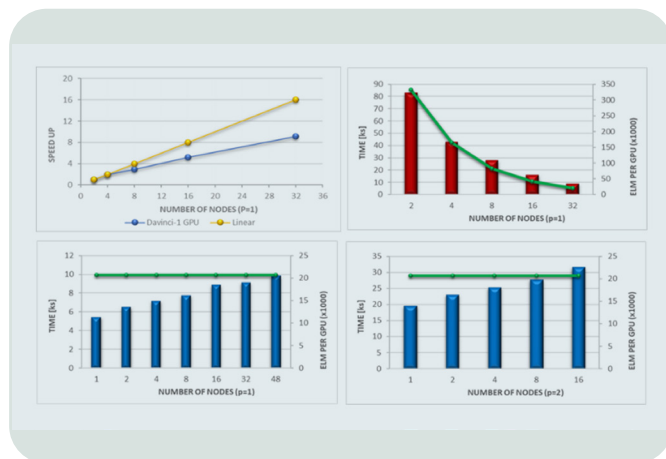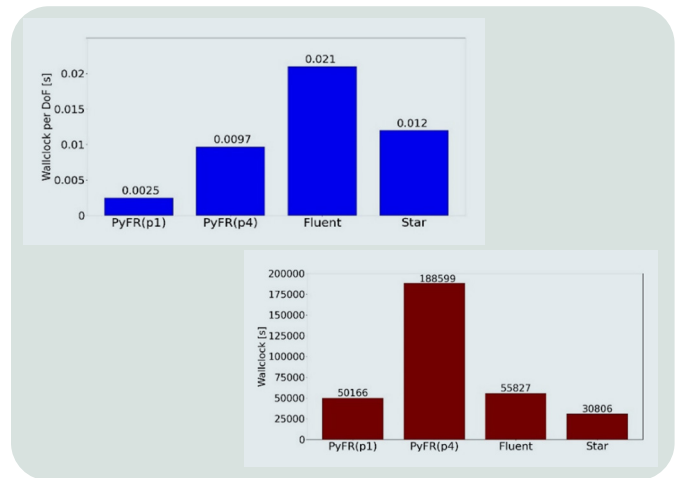


*6 - PyFR scaling.*

*Top row: strong scaling at p=1, 2.65M elements mesh. Speedup (left) and wallclock time (right) on 2,4,8,16, and 32 davinci-1 nodes, each using 4 A100 GPUs.*

*Bottom row: weak scaling at p = 1 (left, 1 to 48 nodes, 83K elements to 3.979M elements meshes), p = 2 (right, 1 to 16 nodes, 83K elements to 1.328M elements meshes). Here and in the following, the green line denotes the number of elements per GPU*

Finally, we performed scalability tests on the GPUs. For the strong scaling of PyFR we used the same mesh with 2.65$M$ elements as in the previous test and polynomial order $p$ = 1. The simulations run on 2, 4, 8, 16 and 32 GPU nodes, each node using 4 GPUs. Figure 6 shows the speedup, on the top left, and the wallclock time (red bars) and the number of elements per GPUs (green line), on the top right.

The obtained speedup is acceptable, considering that the code is run at relatively low order and without fine tuning. For the weak scaling experiment, the computational load is fixed at 20.7$K$ elm/ GPU, increasing the number of computing nodes proportionally to the problem size. Figure 6 shows at the bottom the results of the weak scaling at polynomial order $p$ = 1 (left) and $p$ = 2 (right). In particular, p=1 runs from 1 node (83$K$ elements mesh, span = $h/8$ ) up to 48 nodes (3.979$M$ elements mesh, span = $6h$), while $p$ = 2 runs from 1 node up to 16 nodes (1.328M elements mesh, span = $2h$). In both cases the weak scaling results are not optimal, because the time to solution increases with the number of nodes. This result could be due to the low computational load per GPU. Better scaling results should be obtained using higher polynomial orders on the same mesh.

For Fluent scalability, we used the best configuration tested with other benchmarks on davinci-1, *i.e.*, full node configuration (48 CPU cores/node, 4 GPUs/ node). For strong scaling we used a mesh with 21.2$M$ elements (about 10 times larger than PyFR scaling but with comparable DoFs count at $p$ = 1).

The scaling test was run on 1 node up to 9 nodes. The obtained speedup is suboptimal compared to the ideal linear trend, and in particular for 8 to 9 nodes the model antiscales (Figure 7 left). For weak scaling, we ran on 1, 2, 4 and 8 GPUs nodes, fixing the computational load at 663.24$K$ elm/GPU (Figure 7 right).
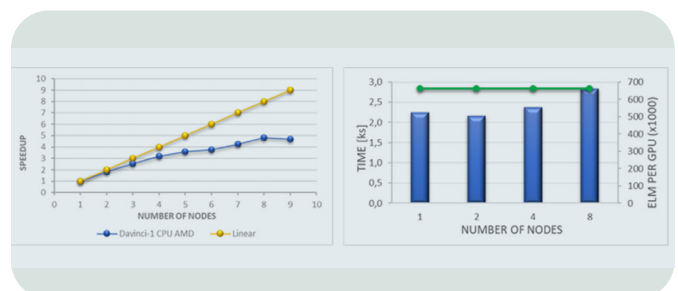


*7 - Ansys Fluent scaling results. Left: strong scaling, 2.65M elements mesh, speedup on 1 to 9 nodes, each using 4 A100 GPUs. Right: weak scaling, 1,2,4,8 nodes, 83K, 166K, 332K, 664K elements meshes*

For the first three simulations the time difference is relatively small, while a more marked difference is noted with 8 nodes. We remark that we could not perform the scalability tests on the GPUs for this compressible case with Star-CCM+, which only supports GPU acceleration for incompressible flows.

To compare the performance of all three tools, we therefore ran a case in the incompressible regime, using a 10 times smaller inflow and outflow velocity values. In this way we obtained a Mach number lower than 0.1. In this case, the load per node was fixed at $20.7K$ elm/GPU, as in the previous PyFR weak scaling, and the simulations ran on 1,2,4 and 8 GPU nodes. Figure 8 shows the comparison of the weak scaling results performed by the three tools. The values in the bar graph are relative to the normalized wallclock time with respect to the number of DoFs (PyFR is set at $p$ = 1). StarCCM+ and PyFR scale better than Fluent, and PyFR runs about 5 times faster than Star, and about 10 times faster than Fluent.
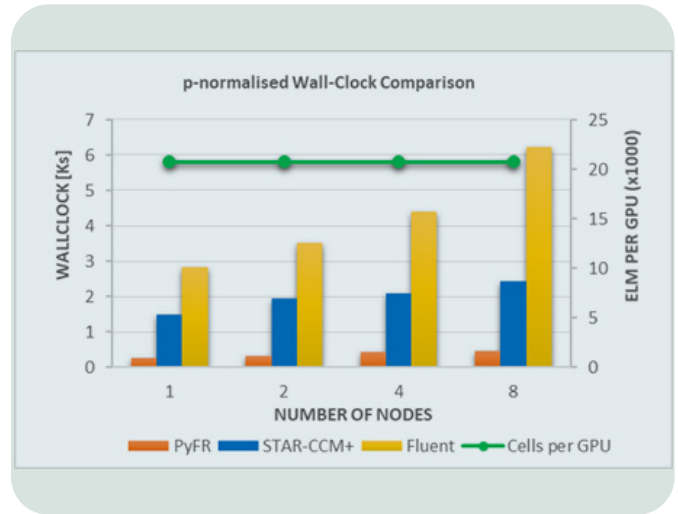


*8-Weak scaling in the incompressible flow simulations with PyFR (p = 1, orange), STAR-CCM+ (blue), and Ansys Fluent (yellow), 2.65M elements mesh, on 1, 2, 4, 8 nodes, each running with 4 A100 GPUs. The PyFR wallclock time is normalized for the DoF count*

## Linear elasticity simulations

Elasticity problems in engineering applications can contain several hundreds of physical components and include multiple features such as rigid bodies, connectors, tied links, in order to simulate connections between parts, like bolts, rivets, welding. Typical loads and boundary conditions take the form of distributed pressure and concentrated force or torque as well as imposed or prescribed displacements and are typically applied at mesh nodes. Nonlinearities in the form of geometrical formulation, constitutive models and contact formulations add complexity to the algebraic problem and require incremental solution schemes, where the convergence times becomes significantly problem dependent. Traditionally, structural finite element codes are equipped with an extensive element libraryin order to discretize geometries efficiently, avoid redundancies and minimize overhead in the analysis of specific problems.
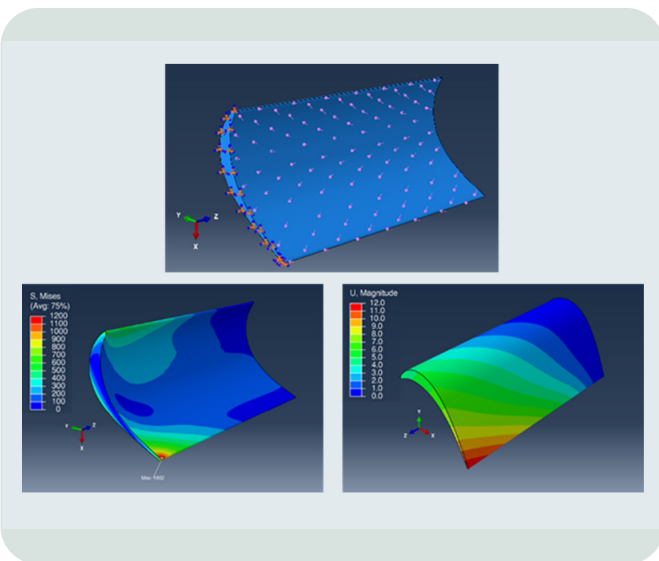


*9-Linear elasticity simulations with the T106A geometry: applied loads and boundary conditions (top), von Mises stress contour plot (bottom left), displacement magnitude (bottom right)*

Nevertheless, the advent of HPC infrastructures offers the possibility to forego several constrains and allows the analysts to solve high-fidelity models spanning multiple length scales, opening the doors to new and advanced interactions and constitutive material formulations. Thanks to this, structural models containing several millions of DoFs have become commonplace. Therefore, optimizing the usage of the available computational resources becomes a necessary step. In order to provide a fair basis for optimization and remove the influence of nonlinearities, a simple case of static linear elasticity is investigated in detail on the davinci-1 HPC infrastructure using the ABAQUS framework. The three-dimensional geometry of the T106A LPT blade is discretized with reduced integration hexahedral continuum elements (C3D8R), for a total of $1.58M$ elements and $5.0M$ DoFs. The part is loaded with distributed pressure around the profile, the root is constrained in all directions with encastre boundary conditions, while the opposite end is left free (Figure 9).

A comparison of direct and iterative solvers under several resource configurations is performed. Results up to 48 cores refer to a single CPU compute node, while data points for the 96 cores are relative to a configuration that uses two nodes. The solution clock time and memory usage are extracted from the output logs generated by ABAQUS. The absolute solution time, memory usage and speedup against the single core solution are shown in Figure 10.

Regarding the comparison between direct and iterative solver, it can be seen from Figure 10 (top left) that the iterative solver provides evident advantage in large problems like the one investigated.

Furthermore, the memory usage for the iterative solver is significantly lower at 44GB compared to 97GB for the direct solver in single core solution. This gap widens when multiple resources are requested, as the direct solver shows a significant overhead, up to +160% when 96 MPI processes are requested, whereas memory overhead is negligible when the iterative solver is used (Figure 10, top right).
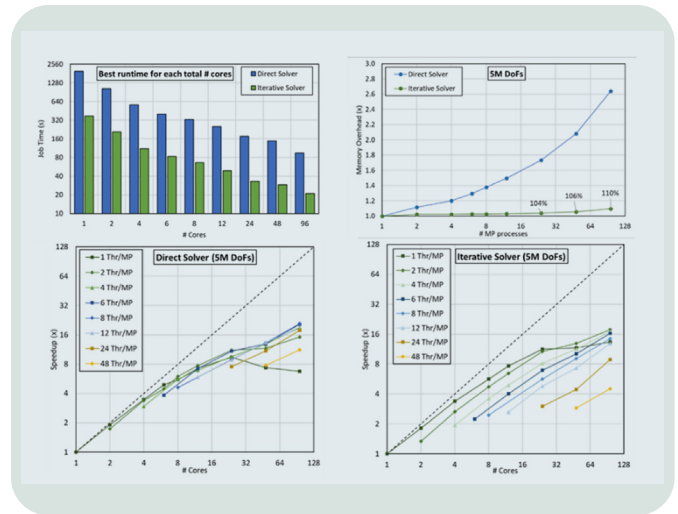


*10 - Top: strong scaling (left) and memory overhead (right) on davinci-1 CPUs in the T106A linear elasticity runs with ABAQUS. The 96 cores result uses 2 CPU nodes. Bottom: speedup at different settings of ncpus and threads_per_mpi_process for direct solver (left) and iterative solver (right)*

More specifically, along with the number of cores, declared in the *ncpus* parameter, ABAQUS allows the control of the number of MPI processes used for a solution, through the parameter t*hreads_per_mpi_process*, so that the number of MPI processes is determined as: #*ncpus* / #*threads_per_mpi_process*.
This setting is found to significantly affect the solution time, particularly for the iterative solver, which vastly benefits from a higher parallelization of the solution. The effect is less significant for the direct solver, where the speedup at different values of *threads_per_mpi_process* is similar at a given #*ncpus*. Nevertheless, the speedup obtained from the direct solver seems to hit a limit when the total number of processes is higher than 12.

## CONCLUSIONS

This paper has investigated the performance of a heterogeneous HPC architecture, Leonardo's davinci-1, on CFD and structural dynamics simulations using an idealized T106A LPT blade geometry. In simulations of compressible and incompressible flow around a wake generator and the LPT blade, the commercial packages Ansys Fluent and Simcenter STAR-CCM+ as well as the open-source high-order framework PyFR provided stable and scalable numerical solutions using hardware accelerators. Ansys Fluent provided strong and weak scaling up to 9 computing nodes with GPUs and up to 2.4 million elements per node (testing with more resources was constrained by license limitations). PyFR provided speedup up to 128 GPUs (down to about 20 thousand elements per GPU), and weak scaling up to 192 GPUs (*i.e.*, 48 GPU nodes, 60% of the whole davinci-1 GPU partition). In incompressible runs, STAR-CCM+ provided good weak scaling on up to 8 GPU nodes. In three-way time-to-solution and scalability comparisons, PyFR at lowest order *p* = 1 outperformed Fluent by a factor of more

than 8, and STAR-CCM+ by a factor of more than 4, in terms of wallclock time per degree of freedom. Notably, the coarse-grained PyFR run at *p* = 4 was also faster than the finite-volume simulations with Fluent and STAR-CCM+ in terms of wallclock per DoF. In evaluating the comparison, it is important to note that considerable effort was put in homogenizing the parameter choices in Fluent, STAR-CCM+, and PyFR to achieve a level-playing field, while differences due to, *e.g.*, time integration methods and time step size could not be quantified.

In structural mechanics' simulations of linear elasticity for the LPT blade with ABAQUS on davinci-1's CPU nodes, comparable figures were found for the speedup on up to 96 cores with the direct solver and iterative solver, but the iterative solver provided shorter time-to-solution and better usage of available memory. Regardless of the solver adopted, the accurate selection of the number of threads per MPI process through an internal solver parameter is paramount for achieving the best possible speedup when the resources available are limited.

The results reported in this paper open up a number of further avenues for investigation. First, the fidelity of simulations needs explicit validation, ideally both with experimental data and with simulations in the literature. The latter will require the use of inflow boundary conditions at an angle rather than purely horizontal. Second, PyFR results could be consolidated with runs at higher polynomial orders to fully exploit the accuracy and scalability potential of the implementation. This includes further coarse-graining of the low order meshes to obtain more accurate simulations at similar computational costs [7], a feature that is not accessible to fixed-order methods in commercial packages.

In addition, the mesh could be refined behind the trailing edge of the blade to improve the representation of the vortices in its wake. Third, the comparison between software frameworks could be extended to aircraft geometries, where scale- and boundary-layer-resolving simulations, and direct numerical simulations in particular, are bound to be computationally much more demanding. In terms of functionality, while Ansys Fluent and StarCCM+ already have capability for transonic simulations, further benchmarking on compressible aerodynamics tests will give the opportunity to test shock-capturing schemes in PyFR's high-order framework [8]. In addition, existing aerodynamics simulations chiefly concern flow over bluff-body fuselage, and implementation of moving meshes capability stands a great chance to provide higher-fidelity results in rotorcraft design, analysis, and simulation.

*Note: an earlier version of this paper appeared as a conference proceeding [9].*

Nicoletta Sanguini: nicoletta.sanguini.ext@leonardo.com

## REFERENCES

[1] A.S. Iyer, Y. Abe, B.C. Vermeire, P. Bechlars, R.D. Baier, A. Jameson, F.D. Witherden, P.E. Vincent, 2021: High-order accurate direct numerical simulation of flow over a MTU-T161 low pressure turbine blade. *Computers and Fluids*, 226, 104989.

[2] V. Michelassi, J. Wissink, and W. Rodi, 2002: Analysis of DNS and LES of flow in a low pressure turbine cascade with incoming wakes and comparison with experiments. *Flow, turbulence and combustion*, 69, 295-329.

[3] F.D. Witherden, A.M. Farrington, and P.E. Vincent, 2014: PyFR: An open source framework for solving advection-diffusion type problems on streaming architectures using the flux reconstruction approach. *Computer Physics Communication*, 185, 3028-3040.

[4] H.T. Huynh, 2007: A flux reconstruction approach to high-order schemes including discontinuous Galerkin methods. 18th AIAA *Computational Fluid Dynamics Conference*, 4079.

[5] P. Vincent, F.D. Witherden, B. Vermeire, J.S. Park, and A.S. Iyer, 2016: Towards green aviation with Python at petascale. In *SC'16: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (pp. 1-11). IEEE.

[6] B. Vermeire, F.D. Witherden, and P.E. Vincent, 2017: On the utility of GPU accelerated high-order methods for unsteady flow simulations: A comparison with industry-standard tools. *Journal of Computational Physics*, 334, 497-521.

[7] J.S. Park, F.D. Witherden, and P.E. Vincent, 2017: High-order implicit large-eddy simulations over a NACA0021 aerofoil, *AIAA Journal*, 55, 2186-2197.

[8] W. Trojak, T. Dzanic, and F.D. Witherden, 2021: Shock capturing methods in high-order flux reconstruction: Low-order graph viscosity and convex limiting approaches. *AIAA Scitech 2021 Forum*, 0496.

[9] N. Sanguini, T. Benacchio, D. Malacrida, F. Cipolletta, F. Rondina, A. Sciarappa, and L. Capone, 2022: Next-generation HPC models for future rotorcraft applications. *Proceedings of the ECCOMAS Congress, Oslo, Norway*. DOI: 10.23967/eccomas.2022.194.

# Retrofit of Hydrogen-Powered Helicopters: a Sizing Approach

Andrea Nesci[1], Massimo Brunetti[2], Giorgio Vicenzotti[2], Francesco Salucci[3]

[1]Leonardo Labs - Future Rotorcraft Technologies, [2]Leonardo - Helicopters Division,
[3]Argonne National Laboratory

Over the next few decades, hydrogen will play a critical role as an energy source able to transform aviation into a climate-neutral industry. In spite of that, research efforts about hydrogen application to rotorcraft are following with certain delay those of the fixed-wing counterpart. This is true not only because of the lower environmental impact of helicopters with respect to the commercial air transport, but also because rotorcraft performance requirements necessitate higher power-to-weight and energy-to-volume ratio solutions. In this paper, an original algorithm is presented that is capable of retrofitting helicopters across multiple categories by replacing the conventional thermal engine-driven powertrain with a hydrogen-driven powertrain that includes fuel cells and batteries as power sources. The methodology is applied to a reference helicopter that is representative of the heavy category, supposing 2030 as its entry into service date and technological assumptions projected to the year 2025, allowing for a 5-year certification and testing gap.

## INTRODUCTION

In several ongoing projects and studies, the quest for improved sustainability in aviation is promoting hydrogen as the next breakthrough technology in aircraft propulsion [1-6]. In terms of mass, hydrogen has roughly three times the energy content of gasoline or aviation fuel. Hydrogen, on the other hand, is gaseous at room temperature and, even when liquefied, it features very low density, which implies a large amount of space needed and some difficulties in efficiently storing it. Aside from the technical opportunity, there is a strong political push behind hydrogen because it can offer the very first chance of true energy independence, as it can be produced from domestic energy sources. In terms of sustainability, green hydrogen (i.e., hydrogen produced from renewable electricity) is $CO_2$-free and can be completely emission-free, depending on whether it is used as fuel for direct combustion in a Thermal Engine (TE) or oxidized in a Fuel Cell (FC) system.

FCs are similar to batteries in that the reactants are kept outside the cell. In reality, FCs are not energy storage devices, but rather electric power generation devices. The hydrogen stored in the tank represents the energy content, whereas the oxygen that activates the reaction is normally obtained from the surrounding air. This differs from batteries, which function as both an energy storage system and a power generator at the same time. Two of the most appealing FCs technologies are Proton Exchange Membrane (PEM) FC and Solid Oxide (SO) FC. PEMFCs have proven to be the most economically successful, especially for mobility applications, because of their quick start-up and low operating temperatures.

Conversely, due to their slower start-up time and greater operating temperature, SOFCs are typically selected for stationary applications. When it comes to using hydrogen in the aviation industry, current efforts
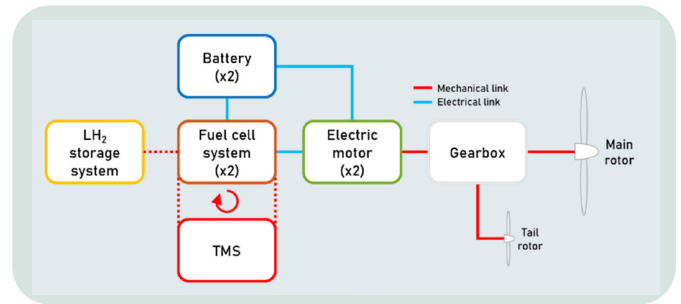
are focusing on single-aisle aircraft that use a combination of fuel cells and hydrogen direct combustion[7]. In terms of applications of hydrogen to rotorcraft, research efforts are lagging behind those of fixed-wind counterparts [8, 9]. This is the result of a number of factors. For starters, the environmental impact of helicopters is a very small portion of the overall burden of commercial air transport. Second, rotorcraft performance requirements differ significantly from those of fixed-wing aircraft, as vertical take-off/landing and hover capabilities necessitate higher power-to-weight ratios.

The ability of a helicopter to safely land after an engine failure adds the need for redundancy, which eventually turns into a weight toll. Finally, there isn't much room for large hydrogen tanks, let alone the demanding crashworthiness that helicopter fuel systems must demonstrate. However, electrification of helicopters can bring several advantages in terms of performance, safety and noise reduction [10]. In this paper, we present a methodology for the preliminary sizing of all-electric retrofit solutions of helicopters with FC-based propulsion. Then, a study case is presented showing the application of the methodology to a large helicopter where the kerosene-based powertrain is replaced with a hydrogen-based one, preserving the weights, volumes, and power ratings of the original helicopter. The retrofit solution is considered for 2030, while the parameters for technological assumptions are set for 2025, allowing for a 5-year gap for certification and testing. Finally, results are discussed in the Conclusions.
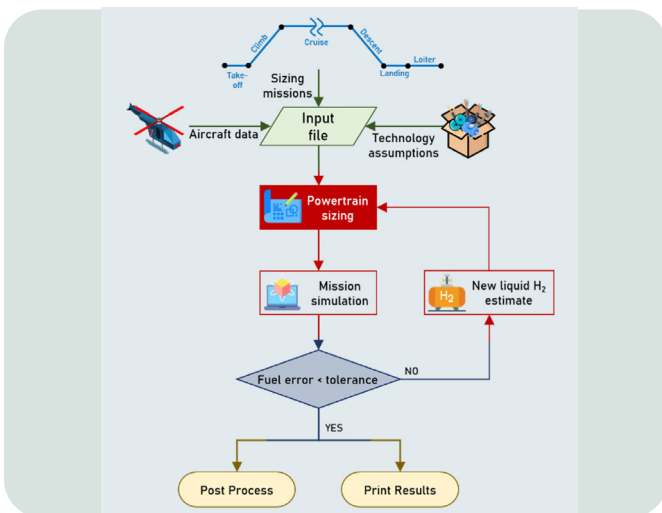
## METHODOLOGY

The methodology has been implemented in an original algorithm: Retrofit Of HydroGen powEred Rotorcraft (ROHGER). ROHGER is capable of addressing helicopters across multiple categories, from lightweight single engine to heavyweight twin engine featuring a classic main/tail rotor configuration. The solution is compliant with the airframe of the given retrofit machine, preserving its Maximum Take Off Weight (MTOW).



1–Schematic of the fuel cell-based hydrogen-powered powertrain

The main and tail rotor of the original helicopter are not changed, and the aerodynamic properties remain the same. Therefore, ROHGER's aim is to replace the conventional TE-driven powertrain with a hydrogen-driven powertrain featuring a FC system. The method includes the sizing of the single elements of the new powertrain by making use of physics-based modelling and statistical-historical regressions. A schematic of the considered series powertrain architecture is portrayed in Figure 1.

The overall ROHGER tool algorithm is outlined in the schematic of Figure 2. The input of the tool consists of three parts: helicopter data (*e.g.*, MTOW, aerodynamic coefficients, power ratings, main/tail rotor diameters, tail length), sizing mission and powertrain architecture (e.g., specific power of the FC system and battery, tank shape). Once the input data is loaded, the powertrain sizing is performed following an iterative sizing process. The first step of the process occurs within the *Powertrain Sizing* function, where a first sizing attempt is done for the drivetrain elements, as explained before. The second step of the process within the *Mission Simulation* function, simulates the sizing missions. The required power to fly is computed at each time step, and quantities referring to the hydrogen and battery system are estimated. After the mission simulation phase is over, there is a check on the estimated fuel level. If the residual fuel in the tank does not match a predetermined target level, the overall process is iterated until convergence.



2–Schematic of the ROHGER algorithm

The output of the method, consist of a weight breakdown of each of the powertrain elements at system and subsystem level, and a preliminary volume estimation for what concerns the FC system, the battery pack, the hydrogen tank and Electric Motors. Each subcomponent with its model is presented in the next subsections.

## Hydrogen storage

This element can be either a Gaseous Hydrogen (GH2) or a Liquid Hydrogen (LH2) tank, equipped with a balance of plant which allows to safely control the hydrogen mass flow rate to the FCs. For this particular case we considered a LH2 tank. The tank features a single wall construction, *i.e.* a single structural tank wall, with a series of insulation layers applied to the exterior of the tank wall itself. Foams and aerogels are typically employed for the insulation. In order to avoid fluid infiltration, a MAAMF vapour barrier is added on both sides of the foam.
A fairing is usually applied to the exterior of the tank to protect the insulation from external damages. The structural sizing of the inner tank is done by predicting the required wall thickness according to the internal bursting pressure and the limit stress of the material. For what concerns the material for LH2 applications, there is a consensus on the use of aluminium alloys in space launchers (Ariane 4 and Ariane 5) as well as in several studies from NASA regarding the design of hydrogen-powered aircraft [2]. Aluminium based alloys remain tough and ductile in the cryogenic range [5] and are known for being resistant to hydrogen embrittlement[11].

The thickness of the insulation layer, and ultimately the tank mass, can be tuned to limit the heat leakage up to a level where a target dormancy boil-off rate $r_{BO}$ is achieved. This is the quantity of hydrogen lost per unit time as a fraction of the full tank mass capacity, or rather:

$$r_{BO} = \frac{1}{T_{empty}} \qquad (1)$$

where $T_{empty}$ is the amount of time it takes for the tank to get empty due to boil off losses while the aircraft is on the ground at standstill. The resulting mass of the tank is computed by summing the weight contribution of the structural layer, the insulation layer and the fairings [12]. On top of the hydrogen tank, there are additional Balance of Plant (BoP) components that must be sized to guarantee the proper operation of the cryogenic storage system, as described in Figure 3.
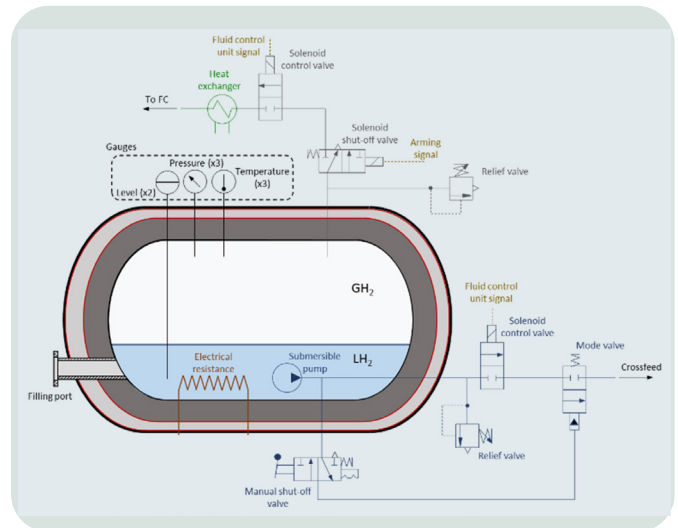


*3 – Cryogenic tank balance of plant components*

## Fuel cells

The sizing of the FC system depends on a specific power index derived from statistical regressions and predictions, while the performance of the FC system is simulated according to the polarization curve of PEM cells derived from [13]. The FC system includes a compressor for the air pressurization and other balance of plant elements such as humidifiers and power electronics.

## Battery pack

The battery's role is to provide boost power during transients and power-demanding phases, such as hover and One Engine Inoperative (OEI) conditions. On the other hand, battery can be recharged to store power in excess from the FC system. Its inherent weight inevitably has significant impact on the overall aircraft sizing. The adopted battery model is based on the Li-ion cells discharge curve. In particular, the battery cell is electrically modelled as a lumped resistance and voltage source [14]. The number of necessary cells is defined by the desired total battery power at the pack level, which is the sizing condition to use the battery as a power booster together with FCs. The mass of the overall battery pack is calculated from the overall mass of the cells considering a packing factor.

## Electric motors

The Electric Motors (EMs) are physically connected to the transmission and must provide the necessary torque to move the rotors. EMs are usually sized according to their peak torque characteristics. Indeed, motor weight depends primarily on the peak torque, proportional to peak current that is linked to the magnetic flux and to the number of windings. The sizing of EMs uses a statistical relation derived from state of the art technology [16]. During the mission simulation, the EM dynamic is neglected, and the electrical machine efficiency and the inverter efficiency are modelled as a constant [17].

## Driving system

The driving system of the original helicopter is redesigned in order to accommodate EMs instead of TE. EMs usually rotate at lower RPMs than TEs, and this simplifies the design of the transmission. The model for estimating the weight of the transmission is taken from the NASA NDARC tool [16]. The transmission can be either connected to a driving system that moves both the main and the tail rotors or only to the main rotor, with the tail rotor being directly attached to separate EMs.

## Thermal Management System

The Thermal Management System (TMS) is sized to dissipate the waste heat deriving from the FC system, EMs, inverters, as well as to warm up the LH2. Given the high powers to dissipate, both the FC and the EM cooling systems are supposed to be liquid. As regards the FC cooling system, the coolant oil flows across the FC in channels separate from those of reacting hydrogen and air. The heat absorbed by the medium fluid is subsequently transported to a Heat Exchanger (HEx) that rejects the energy to another fluid, should it be either air or hydrogen. The weight of the coolant is estimated by using a surrogate model developed by NASA [15].

## Cables

A DC power distribution system is essential to electrically connect the elements in Figure 1 [17]. This power distribution system relies on high-voltage cables that carry current loads from the FC stacks and batteries to EMs. The design of cables is linked to the maximum electrical load they have to deliver, and to the operating voltage. In particular, a DC cable conductor cross section area is primarily determined by the current load. This conductor can be classically copper or, in case of weight sensitive applications such as aeronautics, aluminium. To prevent unexpected short circuits in the cable's surroundings, the conductor must be coated with electrical insulation. The high operating voltage requires an insulation material featuring high dielectric strength. On top of that, the reduced air density in flight, the lower heat conduction, and the increased possibility of generating electrical arcs, require the application of altitude-dependent correction factors for the maximum cable current and amount of insulation [18]. Finally, an outer sheathing is necessary to protect the insulation material from physical damage. The adopted cable weight model for aeronautical applications is taken from [19] and uses aluminium as electrical conductor.

# RESULTS

As specified, ROHGER is capable of handling helicopters across multiple categories and weights. In this paper we report one application of ROHGER to the Clean Sky Green Rotorcraft Twin Engine Heavy Helicopter – 2020 Reference (TEH-R), representative of the heavy category (MTOW>8 ton). The TEH-R was chosen since, being it a large rotorcraft, it has enough room to fit large cryogenic tanks and the FC system. Data is publicly available in [20]. The FC-powered retrofit solution is designed to enter service in 2030. The powertrain specifications for the study case are summarised in [21].

The first set of results of ROHGER is represented by the mass breakdown of the retrofitted helicopter, depicted by the bar plot in Figure 4. Looking at the upper figure, from the left to the right (from cyan to flash green) bars represent the mass of payload and crew, airframe, transmission, EMs, FC system, battery packs, hydrogen, hydrogen tank, TMS and cabling. Looking at the percentages underneath the bars, we can observe that, apart from the payload and airframe mass, the largest weight toll is due to the FC system (11.3%) followed by that of the transmission (6.3%).

*4 – Mass breakdown of the retrofitted solution*
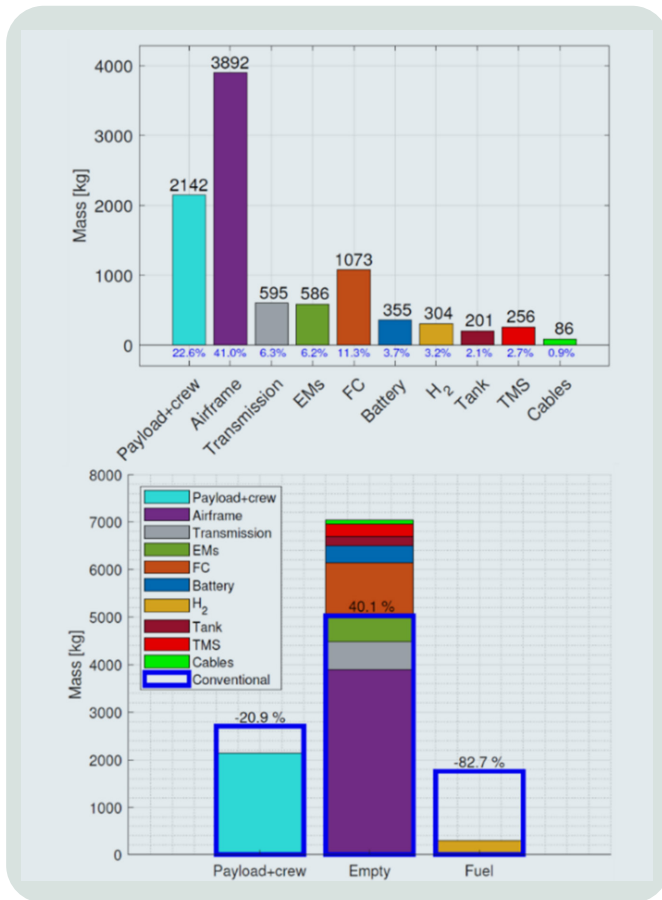
A relevant item is also the mass of the TMS that amounts to 2.7%, almost as much as the hydrogen. The lower picture in Figure 4 represents the stacked version of the upper chart comparing the values of payload, empty, and fuel weight of the retrofitted machine with respect to the original values of TEH-R. We can notice that the mass of fuel is reduced by 1,452 kg (-82.7 %) balancing out the increased empty mass of 2,015 kg (+40.1%). This reduces the residual payload by 20.9%, to 2,141 kg.

The sizing mission for this retrofit is based on 8 phases: 30 min ground hold, 1 min HOGE, 2.5 min climb, 149 min cruise, 3.2 min descent, 1 min HOGE, and 20 min loiter.
Figure 5 shows the time histories of the EMs shaft power (solid green line), battery power (solid blue line) and FC power (solid orange lines) together with the corresponding values of the battery SoC (dashed blue line), FC throttle (dashed orange line) and hydrogen level (dashed yellow line). We can see that the battery is used during the hovering phases at the beginning and at the end of the mission.

The SoC decreases to 95% during the first hover, keeps constant for the duration of the cruise and drops to 90% after the final HOGE. Looking at the battery SoC, the battery capacity seems to be oversized for the design mission showed in Figure 5. Actually, the size of battery power is dictated by the OEI condition: in the case of failure of one FC stack, the battery and the remaining FC stack can provide 92% of the take-off power for longer than 10 min, to safely fly to the closest helipad/airport and then land.

On the other hand, the tank level decreases at a different rate depending on the phase of the mission. During the ground hold phase, the aircraft is stationary on the ground with its systems turned off. However, some hydrogen must be vented to prevent from pressure in excess inside the cryogenic tank; thus, around 4.9% of the tank capacity goes lost. No more hydrogen is vented for the remaining parts of the mission, and hydrogen consumption is simply related to the FC usage, leaving a 10% fuel reserve at the end of the flight.
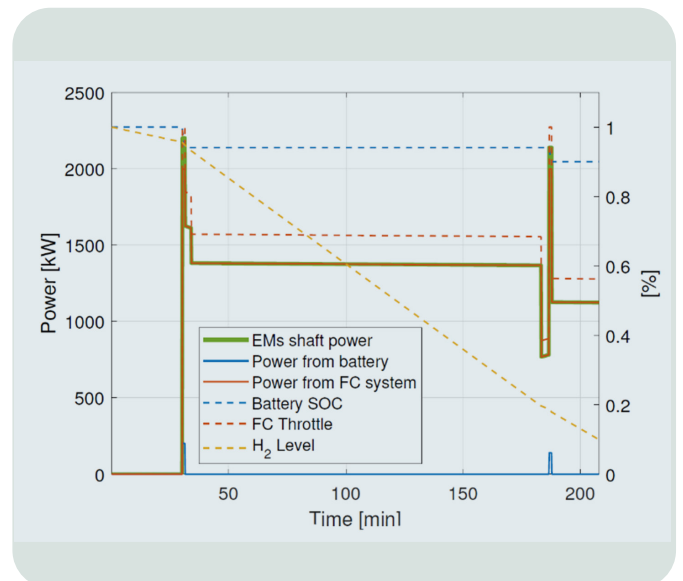


*5 – Time histories of the power of EMs, battery and FC system and SoC, FC throttle and hydrogen level*
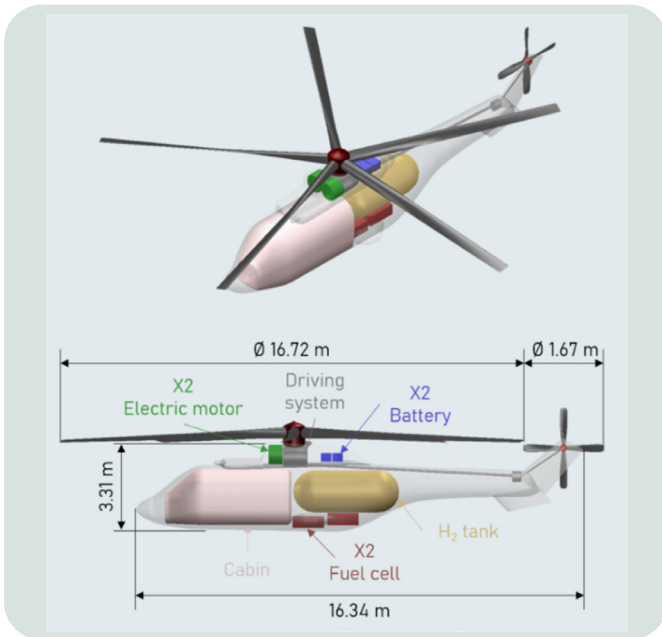
6 – Rotorcraft dimensions and powertrain components lofting

Another output is the conceptual sizing of the cryogenic tank for LH2. It achieves a gravimetric index of 61.2% including the BoP and weighs 201.2 kg. It is capable of hosting 303.9 kg (4.5 m$^3$) of LH2. The total thickness of the tank is 42.1 mm, including the structural wall (2.3 mm) and the insulation (24.0 mm). Figure 6 shows the lofting of the FC system, battery packs, EMs, and hydrogen storage for the retrofitted solution. The two orange boxes in the lower rear part of the fuselage represent the FC systems. EMs are fitted inside the canopy below the rotor, right behind the air intake. The battery packs are placed next to the EMs, towards the tail. We can observe that the volume occupied by the battery seems to be pretty small (0.089 m$^3$) when compared to that of the FCs (0.826 m$^3$). This is due to the fact that the battery operates as a power booster and not as a main energy source. For what concerns the hydrogen storage system, the large tank is located behind the cabin, right above the FC systems.

The tank balance of plant equipment (air compressors, humidifier), and the hydrogen system could be fitted inside the tail cone.

## CONCLUSIONS

This paper shows the application of the original algorithm ROHGER conceptual design methodology to the retrofit of a conventional helicopter with a FC-driven powertrain and a cryogenic tank. The presented study case is a large rotorcraft concept, TEH-R, whose two TEs are replaced with two FC systems, two battery packs, and two EMs. Analogously, the original kerosene fuel system is replaced with a cryogenic tank for LH2 storage. The design process studies the feasibility of the retrofit in a short time frame (2030 entry into service). From the weight perspective, such retrofit is possible, considering a 20.9% reduction of the maximum payload to carry out the full 658km sizing mission, including climb, cruise, descent plus 20 min reserve. The payload reduction occurs despite the drop of the hydrogen weight with respect to the conventional fuel mass (-82.7%), because of the larger empty weight (+40.1%). The heaviest element in the mass breakdown is the FC system. As a consequence, trade studies on technology parameters show that the FC specific power is the parameter that most matters in reducing the payload drop. In the authors' experience, one of the biggest uncertainty in the TEH-R results lies in the accuracy of the TMS weight contribution. Despite the excellent surrogate models available in the literature, the installation of air intakes, HExs, and

coolant lines usually characterize TMS in unique and specific ways. An accurate weight estimation is possible only in a case-by-case scenario further down the design process. The weight of hydrogen must allow for an additional 4.9% to compensate for the fraction that must be vented during the 30 min ground hold phase. The dormancy boil-off rate amounts to 8.69%/h of the overall tank mass capacity in ISA+35 condition. The trade studies on the dormancy boil-off rate show that lower values are possible increasing the thickness of the insulation layer. However, values below 0.1%/h are not possible with a reasonable amount of insulation (less than 0.50 m). Below certain values, it is needed to employ active cooling strategies (cryostats) or more performing insulation techniques, such as vacuum layers. The hydrogen is hosted in a heavy and bulky cryogenic tank, that takes up 5.4 m$^3$. One way to reduce the volumetric burden linked to the cryogenic tank, would be to adapt the shape of the tank to the shape of the fuselage, i.e. using a conformal tank. Another option is to turn the tank into an integral part of the fuselage, by combining the aircraft walls with the tank insulation, which might result in an overall lighter structure. However, the possibility of inspection, maintenance and replacement activities should be kept in mind when investigating such possibilities.

An important takeaway from the results analysis is that the real problem with the introduction of a FC-powered hydrogen system lies in the lofting of the components within an existing airframe. In addition to the massive tank, space is needed for EMs, FCs, battery packs, cabling and TMS. This may require substantial modifications to parts - or to the entirety - of the existing structure. The design of the FC-powered TEH-R presented here is not optimal from a configuration point of view, but it represents a starting point and shows the design capability of the ROHGER tool. There is the need to further explore the design space that emerges from the introduction of FC all-electric powertrains on rotorcraft. For instance, more exotic fuselage shapes to fit the tanks or a switch from the conventional main/tail rotor configuration to multi rotor architectures. Finally, an easier way to introduce hydrogen within a classical helicopter layout could be achieved by resorting to hydrogen direct combustion, instead of using it within FCs. Unfortunately, hydrogen direct combustion requires twice as much the hydrogen that is required by a FC system and this further exacerbates the cryogenic tank lofting issue within the fuselage. The retrofit of a hydrogen-burning helicopter is among the future capabilities of the ROHGER tool.

Nesci Andrea: andrea.nesci.ext@leonardo.com

# REFERENCES

[1] Lapeña-Rey, N., Mosquera, J., Bataller, E., and Ortí, F., "First Fuel-Cell Manned Aircraft," Journal of Aircraft, Vol. 47, No. 6, 2010, pp. 1825–1835. https://doi.org/10.2514/1.42234

[2] Datta, A., and Johnson, W., "Powerplant Design and Performance Analysis of a Manned All-Electric Helicopter," Journal of Propulsion and Power, Vol. 30, No. 2, 2014, pp. 490–505. https://doi.org/10.2514/1.B34843

[3] Kallo, J., Flade, S., Stephan, T., and Schirmer, J., Antares DLR H2-Test bed for electric propulsion, 2015. https://doi.org/10.2514/6.2015-1305

[4] Ng, W., and Datta, A., "Hydrogen Fuel Cells and Batteries for Electric-Vertical Takeoff and Landing Aircraft," Journal of Aircraft, Vol. 56, No. 5, 2019, pp. 1765–1782. https://doi.org/10.2514/1.C035218

[5] d. Castro Ana Lidia, T., L. P., and B., M. C. H., Feasibility of Using Fuel Cell in a Small Aircraft, 2020. https://doi.org/10.2514/6.2021-3189

[6] Khan, Y., Rolando, A., Salucci, F., Riboldi, C., and Trainelli, L., "Hybrid-electric and hydrogen powertrain modelling for airplane performance analysis and sizing," IOP Conference Series: Materials Science and Engineering, Vol. 1226, IOP Publishing, 2022, p. 012071

[7] Airbus, "Airbus and CFM International to pioneer hydrogen combustion technology", 2022. URL https://www.airbus.com/en/newsroom/press-releases/2022-02-airbus-and-cfm-international-to-pioneer-hydrogen-combustion

[8] Rendón, M., Sánchez, C., Gallo, J., and Anzai, A., "Aircraft Hybrid-Electric Propulsion: Development Trends, Challenges and Opportunities," Journal of Control, Automation and Electrical Systems, 2021. https://doi.org/10.1007/s40313-021-00740-x

[9] Stickels, K., Brunetti, M., Barber, M., Brinson, P., and Roe, T., "Advances in helicopter electric tail rotor drive," Proc. of the European Rotorcraft Forum, Milan, Italy, 2017

[10] Riboldi, C. E., Trainelli, L., Mariani, L., Rolando, A., and Salucci, F., "Predicting the effect of electric and hybrid-electric aviation on acoustic pollution," Noise Mapping, Vol. 7, No. 1, 2020, pp. 35–56.

[11] Kramer, D., "Hydrogen-powered aircraft may be getting a lift," Physics Today, Vol. 73, No. 12, 2020, pp. 27–29. https://doi.org/10.1063/PT.3.4632.

[12] Winnefeld, C., Kadyk, T., Bensmann, B., Krewer, U., and Hanke-Rauschenbach, R., "Modelling and Designing Cryogenic Hydrogen Tanks for Future Aircraft Applications," Energies, Vol. 11, No. 1, 2018. https://doi.org/10.3390/en11010105.

[13] O'Hayre, R., Cha, S.-W., Colella, W., and Prinz, F. B., Fuel Cell Fundamentals, 3rd ed., Wiley, 2016.

[14] Gnadt, A. R., Speth, R. L., Sabnis, J. S., and Barrett, S. R., "Technical and environmental assessment of all-electric 180-passenger commercial aircraft," Progress in Aerospace Sciences, Vol. 105, 2019, pp. 1–30. https://doi.org/https://doi.org/10.1016/j.paerosci.2018.11.002.

[15] Chapman, J. W., Hasseeb, H., and Schnulo, S. L., Thermal Management System Design for Electrified Aircraft Propulsion Concepts, 2020. https://doi.org/10.2514/6.2020-3571,5.

[16] Johnson, W., NASA Design and Analysis of Rotorcraft, NDARC, 2015.

[17] Schefer, H., Fauth, L., Kopp, T. H., Mallwitz, R., Friebe, J., and Kurrat, M., "Discussion on Electric Power Supply Systems for All Electric Aircraft," IEEE Access, Vol. 8, 2020, pp. 84188–84216. https://doi.org/10.1109/ACCESS.2020.2991804.

[18] AS50881G, "Wiring Aerospace Vehicle," Standard, SAE International, Aug. 2019.

[19] Stückl, S., "Methods for the design and evaluation of future aircraft concepts utilizing electric propulsion systems," Ph.D. thesis, Technische Universität München, 2016.

[20] Stevens, J., Smith, C., Pachidis, V., and Thevenot, L., "Cleansky Green Rotorcraft Noise and Emissions Benefits – Maximizing the Impact of New Technologies," 2015.

[21] Nesci, A., Vicenzotti, G., Brunetti, M., Salucci, F., Retrofit of hydrogen-powered helicopters: an optimal approach, 48th European Rotorcraft Forum, 2022.

# Photonic Quantum Computing for Optimisation Problems

Massimiliano Proietti[1], Filippo Cerocchi[2], Massimiliano Dispenza[1]

[1]Leonardo Lab - Quantum Technologies, [2]Leonardo - Cyber & Security Solutions Division

Photonic quantum computers are highly robust to environmental noise. They can work at room temperature and can be easily integrated. These features, set photonic platforms as the main candidate for all those use-cases whereby an offline platform for the quantum computation is required locally, on-board, in a possibly noisy environment. However, to program a photonic quantum computer, a paradigm shift from the most common gate-based approach is required. In fact, they are programmed using the so-called measurement-based quantum computing (MBQC). In our work, we propose a novel MBQC algorithm to solve a binary optimisation problem, by introducing a procedure to design any MBQC algorithm, efficiently and recursively. To assess the performance of our algorithm we compare it to other techniques showing up to a 30-fold improvement. With our work, we contribute to make photonic quantum computing more accessible to industries.

## INTRODUCTION

Quantum computers are becoming a reality and industries are expected to exploit the promised quantum advantage in the near future. This goal can be achieved if two requirements are satisfied: the quantum algorithm should address a relevant use-case and the quantum computer should be accessible locally and easily. While many are focusing on the first requirement, not all the industries are concerned about the second one. In fact, in some cases, accessing a quantum computer remotely via cloud services might not be a problem. However, the second requirement should not be relaxed for all those industries for which data confidentiality and secrecy are unavoidable, or the quantum computer might be needed on board a vehicle to solve computational tasks locally. Among all the available quantum computing platforms, only the photonic ones meet both those requirements simultaneously. In fact, photonic quantum computers operate at room temperature, feature lower manufacturing costs as well as lower power consumption if compared to other solutions [1][2].

Moreover, thanks to advances in the integrated photonic market, they offer high compatibility with integrated solutions. Quantum advantage with a photonic platform was recently demonstrated [3][4] and industrial players are advancing the hardware and the architecture [5][6] also giving remote access to their platforms [4]. Unfortunately, the main drawback is that they are not well suited to be programmed in the most common and intuitive quantum computing model such as the gate-based framework, in which quantum information is processed by a sequence of logical quantum gates, similarly to the classical computation model we adopt today. For photonic quantum computers to be programmed a shift in paradigm is required, that is to reason in the MBQC model of quantum computing. MBQC is a framework for universal quantum computation [7] that has been developed as an alternative to the gate-based model. Rather than processing the quantum information through a deep sequence of logical gates, it requires the use of a large entangled state (technically named

The algorithm is encoded on the measurement pattern itself. Both models are universal and they map an input state into an output state of the same dimension.

The state evolution instead is fundamentally different for the two models: gate-based models use as a resource for the computation a set of 1-qubit or 2-qubits logic gates (typically provided by a superconductive circuits), and the evolution is driven by the settings for each gate. MBQC models, instead, use as a resource a large cluster state (typically provided by a photonic platform) on which single-qubit measurements are performed according to a pattern given by the user, see Figure 1 for more details.
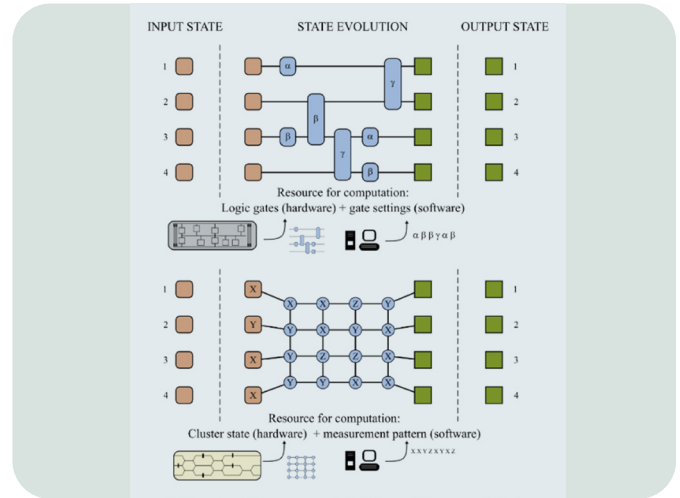


6 – Rotorcraft dimensions and powertrain components lofting

## OPTIMISATION PROBLEMS

In order to use photonic platforms for relevant use cases, developing new near-term algorithms in the MBQC model is paramount. A class of algorithms is of interest, which are the so-called hybrid quantum-classical computational algorithms [8][9]. These distribute the computational tasks among classical and quantum resources according to complexity of such tasks. For example, the evaluation of some complex parametric function is assigned to the quantum computer, while the tuning of the function's parameters is tasked to a classical computer. These algorithms are envisioned to achieve quantum advantage for practical problems in the near future. Examples of algorithms included in this category are variational quantum eigensolvers (VQEs) [10], variational quantum simulators (VQSs) [11] and quantum approximate optimization algorithms (QAOA) [12]. In particular, the QAOA was proposed to approximately solve combinatorial optimization problems as the MAX K-CUT [13], an NP-complete problem (Figure 2) that can be used as a primer to approach fundamental problems in statistical physics as well as practical problems such as data clustering,

and scheduling problems.

Figure 2 shows how problems for which an efficient classical solution is known, belong to the P class. Quantum computing algorithms target problems belong to the NP class, as not any efficient classical solution is known for them. The MAX K-CUT is NP-complete, which means that its solution can be used to solve any NP problem".

Obtaining a formulation of such algorithms in the MBQC model is not trivial. Some strategies do already exist, but they might not be optimal for QAOA algorithms. To address this point, we develop a method (see [14] for our original work) to reason natively in the MBQC framework, basing our studies on a well-known result giving the most efficient MBQC algorithm if the problem to solve is opportunely stated. The problem we address in our work is indeed the MAX K-CUT that we write in a general form, leading to the general expression of the cluster state required as a resource for computation, and the pattern of local measurements implementing the target QAOA algorithm.
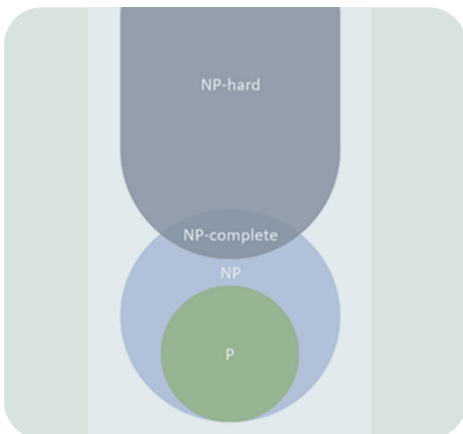


## MAX K-CUT PROBLEM FORMULATION

The MAX K-CUT problem is a combinatorial optimization problem which amounts to labelling using indices from 1 to K the vertices of a graph in such a way that there is a minimum number of adjacent vertices belonging to the same class. The graphs we consider are undirected (i.e. there is no preferred orientation to move along one edge) they do not include any loops over a single vertex and multiple edges connecting a simple pair of vertices.

2 - Complexity classes

If we assign unitary *weight* to each edge, it is equivalent to say that we want to maximize the total weight of those edges connecting vertices which are differently labelled.

By slightly stretching this second formulation of the MAX K-CUT, we can obtain a *weighted* MAX K-CUT problem. Indeed, assuming that to each edge e a given weight *w(e)* is assigned instead of a unitary weight, we can formulate the weighted MAX K-CUT in a completely analogous way. This is carried out by assigning labels from 1 to K to vertices of a given weighted graph in such a way that the total weight of edges connecting differently labelled vertices is maximized.

Let us take the formal point of view: consider a graph $\Gamma = V,E$ where $V=\{v_1,....,v_n\}$ is the vertex set and $E=\{e_1,...,e_m\}$ is the edge set and let $\underline{w}_E$ be a collection of weights for the edges of $\Gamma$ so that $\underline{w}_E(e)$ is the weight of edge e. Let $P=\{P_1, P_2,..., P_K\}$ be a partition of the vertex set *V* into *K* non-intersecting, non-empty sets. We shall call such a partition a K-cut for the graph $\Gamma$. Let $E_P$ be the subset of edges whose endpoints lay into distinct subsets $P_i,P_j$. We define the function

$$\Delta_K(P) = \sum_{e \in E_P} \underline{w}_E(e)$$

Solving the (weighted) MAX K-CUT problem over the triple (*V, E, $\underline{w}_E$*) means to find

$$argmax_{K-cuts}\Delta_K(P)$$

That is find the K-cut maximizing the function $\Delta_K$.

This abstract problem has an immediate application to clustering data in *K classes*: let's consider a dataset *D* of datapoints, represented for example as real-valued vectors. We can think of such a dataset as a weighted undirected graph in which each datapoint corresponds to a vertex, being each vertex linked to any other vertex and the weight of an edge corresponding to a pair of vertices is computed by taking the $L^2$-norm of the difference vector of the two datapoints. Finding a solution of the MAX K-CUT for the graph obtained in this way is equivalent to assign the datapoints to K clusters in such a way that the distances among the different clusters is maximal.

It is known that finding an optimal solution for the MAX K-CUT is a NP-complete problem. NP (non-deterministic polynomial time) is a complexity class characterized by the existence of a non-deterministic, polynomial time algorithm able to discriminate whether a proposed input is or is not a solution to a given decision problem. This does not mean that the algorithm is able to produce a solution in polynomial time (as opposed to what happens in complexity class P), as the algorithm is only able to check whether a given input is a solution or it is not.

NP-complete means that the problem is in NP and every problem in NP has a polynomial time reduction to it, which means that we can map instances of every problem in NP to an instance of an NP-complete problem in polynomial time. As MAX K-CUT is NP-complete, there are not efficient classical algorithm to find the optimal solution unless P = NP. Classical approximation algorithms for MAX 2-CUT usually approach the approximation task through semidefinite programming (SDP), which is a relaxation of the integer programming (IP) problem underlying the MAX 2-CUT [15]. However, it has been proved [16] that for MAX K-CUT the relative error between the optimal solution and the solution that can be found by a classical polynomial time approximation algorithm is lower bounded by $\delta$/*2K-1* where $\delta$ is the relative error for MAX 2-CUT. On the other hand approximating a solution of the MAX 2-CUT within any constant below 17/16 it is known to be NP-hard, which makes NP-hard approximating MAX K-CUT within *1 + 1/32(k-1) -* $\varepsilon$ for any $\varepsilon$ *> 0*. The classical inapproximability of this combinatorial optimization problem makes MAX K-CUT a suitable candidate for demonstrating quantum advantage through QAOA.

## MBQC ALGORITHM TO SOLVE THE MAX K-CUT

In the following, we highlight the main steps to run our algorithm. More details can be found in [14]. We recall that any MBQC model requires two ingredients:

· A cluster state |*C*> (typically a large entangled state);
· A pattern *M* of local measurements on that state.

Given a target problem to solve, the challenge is to find which cluster state and which measurement pattern encode the unitary evolution solving that problem. To avoid confusion, in the following when referring to the terms graph, vertices and edges, we always mean the graph for which we want to find the MAX K-CUT. When instead we refer to nodes and links, we refer to the cluster state of the MBQC. In our specific scenario the target unitary evolution solving the MAX K-CUT problem is described by the following equation:
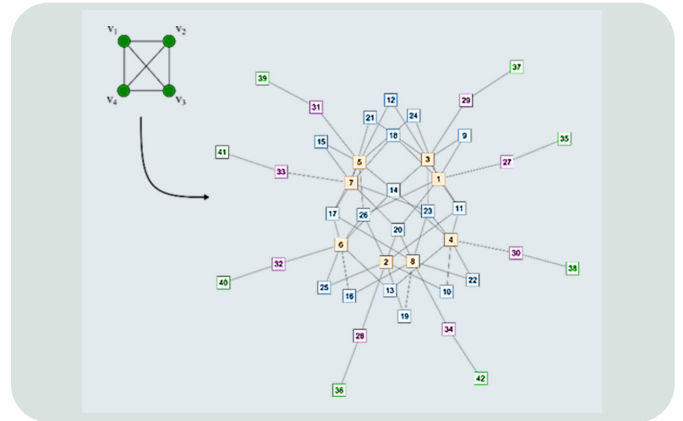


3 - Max 4-CUT of a complete graph

$$|\vec{\gamma},\vec{\beta}\rangle = \prod_{p=1}^{\infty} e^{-i\gamma_p H_t} e^{-i\beta_p H_m} |+\rangle^{\otimes n} \qquad (1)$$

Where $H_t$ and $H_m$ are to Hamiltonian operators encoding the state evolution. In particular the average value of $H_t$ is the cost function computed by the quantum computer and is minimized by a classical computer. The parameters $\gamma_p$ and $\beta_p$ are the classical parameters tuned by the classical computer to minimize the cost function. The n-qubit quantum state $|+\rangle^{\otimes n}$ is the state the quantum computer is initialised with. While the number of parameters in the equation is infinite, in practice we can only have a limited number of parameters i.e. the value of p is fixed.

Moreover, in practice, finding $\gamma_p$ and $\beta_p$ is not straightforward and is the main challenge for the classical optimization algorithm. The possibility of always finding a solution is currently an open debate and it could be a strong limitation to achieve quantum advantage. In the following we set $p = 1$ and we explain how to construct the cluster state for the MBQC algorithm: first we prepare $|V|m$ nodes where $|V|$ is the number of vertices of the graph we want to find the solution for, and m depends on $K$. Second, we add the ancillary qubits required for the local measurements. These are $K$-1$|E|$ nodes, where $|E|$ is the number of edges in the MAX K-CUT graph. We need now to connect the nodes. Notably, there exists a simple rule [17] to find such links if the unitary evolution is diagonal as in Eq (1). In particular, any pair of qubits coupled by a Z operator, gives rise to a set of CZ-gates controlling a common ancilla. Finally, the $|V|m$ nodes are also connected to two more ancillae in a linear chain. An example of such constructed cluster state is shown in Figure 3. The cluster state required for the MBQC-QAOA with $p = 1$ is shown in the figure.
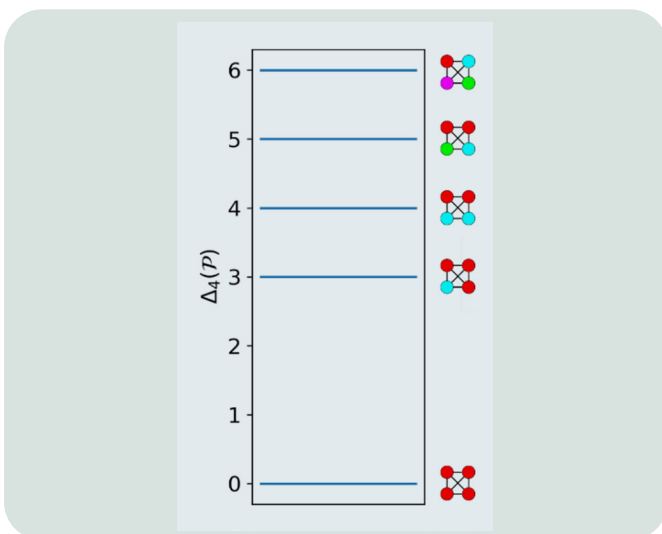


4 - MAX 4-CUT solutions. The energy levels of the target Hamiltonian (given by the average value of the target Hamiltonian) are shown in blue. The level with maximum energy represents the solution where each vertex is assigned to a different class, leading to maximum number of cuts

The orange qubits (labelled from 0 to 8) are the input and output of the algorithm. The light blue nodes (labelled from 9 to 26) are instead the ancillae which have to be measured on the ZY -base at angle depending on $\gamma$. The purple qubits (labeled from 27 to 34) extend the computation, mapping the output orange qubits into the green ones (from 35 to 42). The 8-qubit output state encodes the solution of the MAX 4-CUT. more details on the protocol can be found in [14]. After running the algorithm, the resulting $|V|m$ qubit state encodes the solution of the problem.

To decode such a solution, it should be measured; the average value of the target Hamiltonian is computed from the data obtained.
This process consisting of state preparation, evolution and measurement, is iterated many times until the average value of the Hamiltonian converges. An example of the possible values we can obtain for the average value, and their corresponding MAX 4-CUT solution is shown in Figure 4.
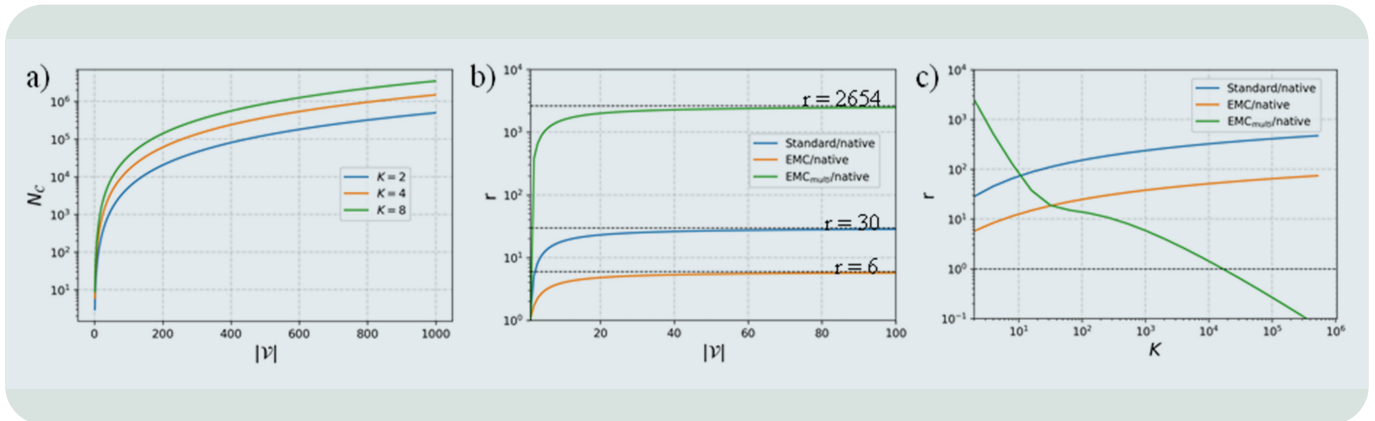
*5-Results.*
*Resource cost comparison for a complete graph with |V| vertices.*
*a) We plot the size of the cluster state NCnative for different values of K as a function of |V|. In b) and c), instead, we plot the ratio r of the cluster state size required in the standard method (orange) and in the EMC method (blue) also applied to the optimised circuit with multi-control gates (green), over the size of the cluster state required for our native approach.*
*In b) we fix K = 2 and vary |V|, while in b) we fix |V| = 100 and vary K. Up to K ≈ 10⁴ implementing our native approach is more efficient than any strategy. Thereafter, translating the circuit with multi-control gates via the EMC method is a better option, at least in terms of cluster state dimension*

## RESULTS

In the following, we compare our method with other two different MBQC algorithms to implement the evolution $U_t = e^{-i\beta H_m}e^{-i\gamma H_t}$, without including the penalty Hamiltonian and assuming $K = 2^m$, meaning that we only consider $K=\{2,4,8,...,2^m\}$ for integer values of m.
These two algorithms are obtained by translating into a MBQC pattern a gate-based circuit. This, for $K = 2^m$, implements $U_t$ when $H_t$ is composed by a sequence of Z operators. The first MBQC algorithm is derived by applying the rules in [7], we shall refer to this approach as the standard one.

The second is again obtained by translating $U_t$, but by applying rules in [18], we shall refer to this as EMC due to its structure in terms of entanglement (E), measurement (M) and correction (C) commands. The main difference is that in the standard approach, the CNOT gate is implemented with a 15-qubit cluster where input and output qubits do not match, while in the EMC case, the CNOT gate is implemented with a 4-qubit cluster state where one input overlaps with one output. We compare these two with our algorithm, which we shall refer to as the *native*. In our case, there is no translation of logic gates, but we rather implement directly on the cluster state the pattern of local measurements to evolve the state according to the target Hamiltonionian. We recall that in our native implementation the cluster state is composed by $N_C^{native}$ qubits given by

$$N_C^{native} = 3|V|\log_2 K + |E|(K - 1)$$

In Figure 5-a), we show $N_C^{native}$ for different values of K and varying |V|. In Figures 5 b) and 5 c) we show the ratio $r = NCGate{\rightarrow}MBQC/NCnative$ for both the standard (blue curve) and EMC (orange curve) methods.

We do this by considering a complete graph, for which it holds $|E| = |V|(|V| − 1)/2$, in Figure 5-b) by fixing $K = 2$ while varying V and in Figure 5-c) by fixing |V| = 100 while varying K.

Notably, as shown in both Figures 5-b) and -c), our native method shows large improvement in terms of resource cost. It requires up to a factor 30 less qubits for the generation of the cluster state when compared to the standard approach, and a factor 5 compared to the EMC method. For increasing K this advantage can be even greater as shown in c).
This is obtained for complete graphs, we note that in applications of the MAX K-CUT for data clustering a complete graph is in fact required, making our result relevant for practical applications of the QAOA algorithm. For completeness we also apply the EMC rules to translate a K-optimised circuit [19], and compare with our native MBQC algorithm.This is shown with a green curve in Figure 5 b) and c). It is interesting to see that, for *K < 16384* the native method is still more efficient, which makes it more convenient for translation of the circuit via the EMC rules.

# LEONARDO USE-CASES

From a general review of the full process flow quoted in the above paragraph, it is clear that first of all any use case that is in any way connected to clustering of data in an unsupervised scheme, may represent a useful application for the MBQC algorithms developed for MAX K-cut problem in this paper, as MAX K-cut connection to data clustering is very well established. Data clustering impacts several application fields through the Leonardo market domains throughout its different Divisions and Lines of Business. Just to name a few of those, such need is found when extracting information from sensors (radars, vision systems, etc.) as well as in Cyber applications when data coming from Internet users and their communications stream have to be analysed for threat identification or anomaly detection. Similar use cases can be found in the development of Large Systems for Situational Awareness in Maritime, Space (SSA – Space Situation Awareness), Anti-Ballistic or Hypersonic Missiles or related domains (C2ISTAR, CMS, etc).

Moving on, we can also consider that the above R&D activity includes the exploitation of QAOA building block algorithms that we have also represented in MBQC, and that it may also find applications in several specific domains or problems categories:

- Traffic routes optimisation (e.g. for urban vehicles as well as for swarms of drones) [20];
- The tracking problem in Radar domain in which the well known Multi Hypothesis Tracking – MHT algorithm, when formulated as an optimisation problem (maximisation of probability of specific trajectories) may represent a killer application, as it effectively scales exponentially with data, as it is the common situation to find quantum advantage in quantum computing solutions [21];
- Other QUBO formulated problems, which may happen in aerostructures design, material synthesis and optimisation of their performances, etc.

This large set of Use Cases is, in fact, just what we preliminarily identified from broad internal analysis, which is going to become the basis for the requirements of our future activity for closer specification of the current results in solving an actual company critical problem. We remark that concurrently with the technical results presented in this work, effort for implementing such results in real hardware is required.

On this direction, we are currently building a network of collaborations with strategic partners developing photonic quantum computing hardware as QUANDELA (a French start-up already offering cloud quantum computing solutions based on photonics) and Xanadu (a Canadian SME) as well as QUIX (from NL). The implementation branch of our future efforts will surely on a selected subset of such identified partnerships.

# CONCLUSIONS

In our work, prompted by the fast development of photonic platforms requiring algorithmic reasoning in the MBQC model, we showed a new measurement-based protocol to implement the QAOA algorithm applied on the generalised MAX K-CUT. The main contribution is the expression of the QAOA algorithm natively in the MBQC model, showing up to a 30 fold improvement with respect to standard MBQC methods and a 6 fold improvement with respect to optimized MBQC methods. In practical use cases, in which for example the MAX K-CUT is exploited for data clustering, K is typically on the order of 10 and our native method outperforms other strategies.

This is highly relevant in the Noisy intermediate-scale quantum (NISQ) era, where only a limited amount of resources is available, and methods to optimize their use are the key to achieve useful results.
For this reason, we expect our approach should stimulate theoretical and experimental investigations of native MBQC approaches.

On this line, we recall that a major roadblock for useful quantum computation is the computational depth. This is related to the circuit depth in the gate-based model, and to the measurement rounds in the MBQC framework. Regarding the comparison with other MBQC algorithms, we expect a significant improvement, as in our case the number of measurements is reduced to the minimum. A more detailed and quantitative comparison is left to future work.

More generally, it is certainly interesting to also compare MBQC and gate-based models. Intuitively, we can see the former as algorithms requiring high computational cost in terms of space (in the form of the cluster state dimension) while manifesting very few computational rounds (in the form of measurements depending on previous outcomes). The latter, instead, seems to behave in the opposite direction. In practice, noise imposes constraints on time and space, and it is therefore interesting to compare the two approaches on practical use cases and scenarios.

Massimiliano Proietti: massimiliano.proietti.ext@leonardo.com

## REFERENCES

[1] P. Walther, K. J. Resch, T. Rudolph, E. Schenck, H. Weinfurter, V. Vedral, M. Aspelmeyer, and A. Zeilinger, Experimental one-way quantum computing, Nature 434, 169 (2005).

[2] K. Chen, C.-M. Li, Q. Zhang, Y.-A. Chen, A. Goebel, S. Chen, A. Mair, and J-W. Pan, Experimental realization of one-way quantum computing with two-photon four-qubit cluster states, Physical review letters 99, 120503 (2007).

[3] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, et al., Quantum computational advantage using photons, Science 370, 1460 (2020).

[4] L. S. Madsen, F. Laudenbach, M. F. Askarani, F. Rortais, T. Vincent, J. F. Bulmer, F. M. Miatto, L. Neuhaus, L. G. Helt, M. J. Collins, et al., Quantum computational advantage with a programmable photonic processor, Nature 606, 75 (2022).

[5] J. Arrazola, V. Bergholm, K. Bra´dler, T. Bromley, M. Collins, I. Dhand, A. Fumagalli, T. Gerrits, A. Goussev, L. Helt, et al., Quantum circuits with many photons on a programmable nanophotonic chip, Nature 591, 54 (2021).

[6] S. Bartolucci, P. M. Birchall, M. Gimeno-Segovia, E. Johnston, K. Kieling, M. Pant, T. Rudolph, J. Smith, C. Sparrow, and M. D. Vidrighin, Creation of entangled photonic states using linear optics, arXiv preprint arXiv:2106.13825 (2021).

[7] R. Raussendorf and H. J. Briegel, A one-way quantum computer, Physical Review Letters 86, 5188 (2001).

[8] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, et al., Variational quantum algorithms, Nature Reviews Physics 3, 625 (2021).

[9] N. Moll, P. Barkoutsos, L. S. Bishop, J. M. Chow, A. Cross, D. J. Egger, S. Filipp, A. Fuhrer, J. M. Gambetta, M. Ganzhorn, et al., Quantum optimization using variational algorithms on near-term quantum devices, Quantum Science and Technology 3, 030503 (2018).

[10] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'brien, A variational eigenvalue solver on a photonic quantum processor, Nature communications 5, 1 (2014).

[11] M. Benedetti, M. Fiorentini, and M. Lubasch, Hardwareefficient variational quantum algorithms for time evolution, Physical Review Research 3, 033083 (2021).

[12] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, arXiv preprint arXiv:1411.4028 (2014).

[13] D. R. Gaur, R. Krishnamurti, and R. Kohli, The capacitated max k-cut problem, Mathematical Programming 115, 65 (2008).

[14] Proietti, M., Cerocchi, F., and Dispenza, M. Native measurement-based quantum approximate optimization algorithm applied to the Max K-Cut problem. Physical Review A, 106(2), 022437.

[15] G. M. X. and W. D. P., Approximation algorithms for maxcut and max2sat, in *ACM Proc. Ann. Sym. on Theory of Computing* (ACM press, 1994) pp. 422–431.

[16] V. Kann, S. Khanna, J. Lagergren, and A. Panconesi, On the hardness of approximating max k-cut and its dual, in Chic. J. Theor. Comput. Sci. (1996).

[17] D. Browne and H. Briegel, One-way quantum computation, Quantum Information: From Foundations to Quantum Technology Applications, 449 (2016).

[18] V. Danos, E. Kashefi, and P. Panangaden, The measurement calculus, Journal of the ACM (JACM) 54, 8 (2007).

[19] Y. Liu, G. L. Long, and Y. Sun, Analytic constructions of general n-qubit controlled gates, arXiv preprint arXiv:0708.3274 (2007).

[20] Inoue, D., Okada, A., Matsumori, T., Aihara, K., &amp; Yoshida, H. (2021). Traffic signal optimization on a square lattice with quantum annealing. Scientific reports, 11(1), 1-12.

[21] Kim, C., Li, F., Ciptadi, A., &amp; Rehg, J. M. (2015). Multiple hypothesis tracking revisited. In *Proceedings of the IEEE international conference on computer vision* (pp. 4696-4704).

# 3D Printing Technology: how Printing Strategies Influence Mechanical Performances of Carbon PEEK

Sara Varetti[1], Nicola Gallo[2], Ignazio Scavo[2], Stefano Corvaglia[2]

[1]Leonardo Labs - Material Technologies, [2]Leonardo – Aerostructures Division

3D printing technologies, and in particular Fused Deposition Modelling (FDM), is getting more widespread use in the aerospace industry. Nowadays, there are many materials that can be processed by the FDM technology and, with the most advanced machines, it is possible to produce components using high-performance polymers such as PEEK, Carbon PEEK, PEKK and PEI. Many process parameters can influence the mechanical performance of manufactured parts, among them there are: printing strategy, material deposition rate, nozzle temperature, chamber temperature, layer thickness and many others. In this paper the influence of different printing strategies (0deg, 90deg, ±45deg and 0/90deg) on mechanical performances of Carbon PEEK specimens is investigated. A Roboze Argo500 machine was used to produce the samples, which were subjected to tensile tests. The results indicate that 0deg is the strategy that allows higher performances, followed by 0/90deg, ±45deg and 90deg

## INTRODUCTION

The group of technologies known as Additive Manufacturing (AM) or 3D printing allow the fabrication of layer-by-layer components in different materials (polymeric, metallic, ceramic and composite) [1]-[6]. The printing procedure involves the use of a CAD file of the component to produce. The file is processed by special software to prepare it for printing: the part is oriented on the printing table, sliced into horizontal layers of the required thickness, and supports are inserted. Once all the other process parameters are set in the software, a file with all the information is obtained and fed into the machine for printing. 3D printing offers several advantages, including the ability to produce components with complex shapes, reduced waste, and the possibility to produce parts in situ. In addition, part designs can be improved specifically for Additive Manufacturing through topological optimization: a FEM analysis is conducted on the part to print and the amount of material is reduced as it is deposited only on those areas where it is actually needed. By printing the new optimized CAD, the weight of parts can be greatly reduced.

For printing thermoplastic polymers and thermoplastic matrix composites, the most popular technology is the Fusion Deposition Modeling (FDM) or Fused Filament Fabrication (FFF). In the FDM, material is fed in the form of filament, which is extruded through a heated nozzle. The temperature reached by the nozzle depends on the material. It is sufficient to soften it and allows it to be distributed over the layers. The material is deposited layer by layer by the movement of a deposition head on which the nozzle is mounted [7]-[9]. This movement is guided by the coordinates entered in the file prepared by the software (.gcode). Some of the parameters that can be varied in the software are: printing strategy, material deposition rate, nozzle temperature, chamber temperature and layer thickness. The term printing strategy means the direction in which the filament is deposited in each layer.
Some examples used in this paper are shown in Figure 1. The 0deg strategy consists of depositing the filament always along the direction concordant to the application of the load.
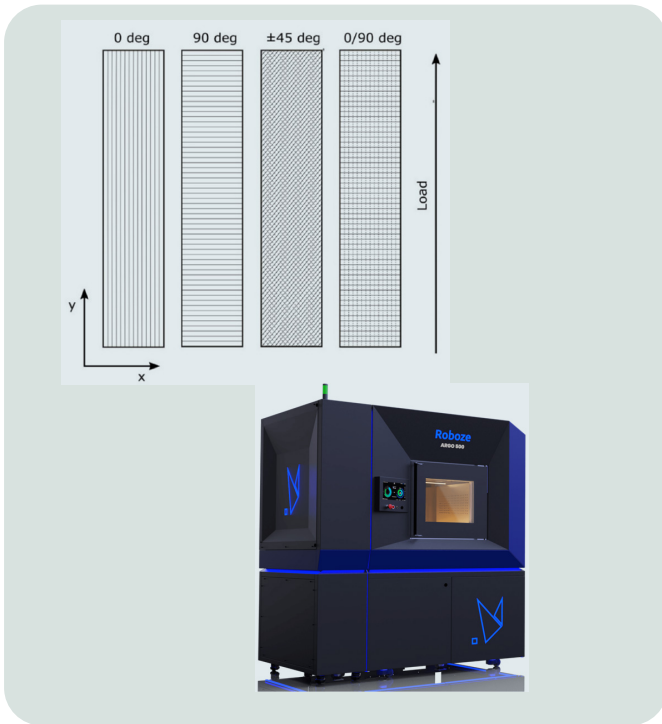
*1 (top) - Printing strategies*
*2 (bottom)- Roboze Argo 500 machine*

In the 90deg strategy the filament is always deposited in the direction transverse to the application of the load. In the ±45deg strategy the filaments are all deposited at an angle of 45° to the application of the load, but in an alternating manner (one layer at +45deg and one layer at -45deg).

Finally, the 0/90 strateg y layers with 0deg strategy and with 90deg strategy alternating. The common materials used in low-cost FDM machines are PLA (polylactic acid) and ABS (acrylonitrile-butadiene-styrene) [10]-[12], but in the machines for industrial use it is also possible to print high performances polymers as PEEK (polyetheretherketone), C-PEEK (polyetheretheretherketone with short carbon fibres), PEI (polyetherimide) and Nylon. Due to the higher mechanical performances also at high working temperatures, they are interesting for the aerospace sector [13]-[20].

In this work, mechanical characterization was carried out on C-PEEK specimens produced with a Roboze Argo500 machine (Figure 2) with different printing strategies (0deg, 90deg, ±45deg, 0/90deg) in order to study this process parameters and compare the effect of these strategies on the behaviour of the specimens.

## MATERIALS AND METHODS

All the specimens produced have been printed with a Roboze Argo 500 machine using Carbon PEEK filament as raw material. This is made of a PEEK matrix with 10% of short Carbon fibres. They are produced in four different printing strategies (0deg, 90deg, ±45deg, 0/90deg), setting a printing speed of 1800mm/min, 450°C of nozzle temperature, 180°C of chamber temperature, 100% infill, layer thickness of 0.225mm and a nozzle size of 0.4mm.

The samples are 4.05mm x 25.4mm x 254mm, according with ASTM 3039 [21] (Figure 3), and tabs in Aluminium alloy (EN AW 6082) have been attached to them through a two-component epoxy glue (Araldite 2031-1 Huntsman), as shown in Figure 4a.
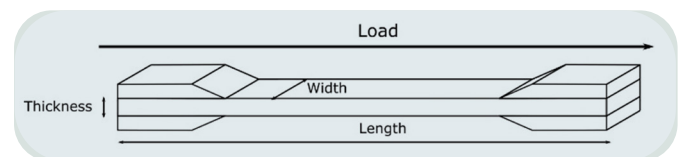


*3- Standard specimens for tensile tests*
*according to ASTM D3039 [21]*
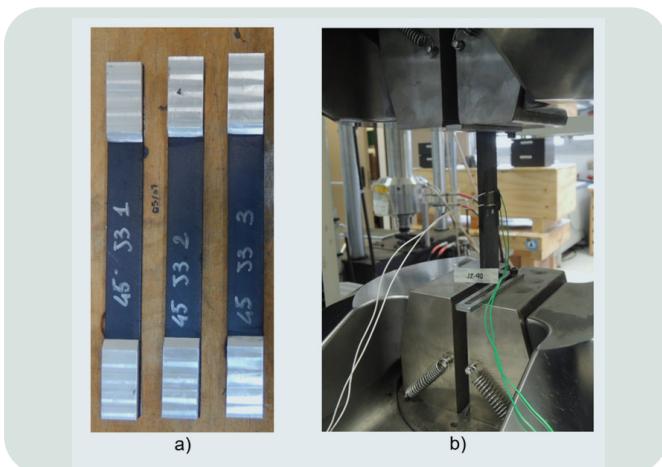
The samples have been subjected to tensile tests with an Instron machine (model 5582) at constant axial speed of 2 mm/min, according with the standard; a two-way Strain Gage (SG) has been used to calculate both the modulus and the Poisson ratio.

The setup is shown in Figure 4b. The ultimate strength has been calculated as maximum load on nominal cross section. The specimens have been also analysed by a digital Optical Microscope Hirox RH – 2000, after cutting and polishing, and fracture surfaces have been observed by SEM Zeiss. The test matrix is schematised in Table 1.



*4-a) Tensile specimens, b) Instron machine setup with Strain Gage*

| Printing strategy [deg] | Measured thickness [mm] (std.dev. [%]) | Measured width [mm] (std.dev. [%]) | N. of specimens |
|---|---|---|---|
| 0 | 4.12 (2.87) | 26.00 (5.00) | 6 |
| 90 | 4.08 (1.20) | 25.21 (0.47) | 6 |
| ±45 | 4.01 (0.47) | 25.24 (0.24) | 6 |
| 0/90 | 3.99 (0.91) | 25.42 (0.36) | 6 |

*Table 1-Accuracy and Time for an estimate on complex systems*

| Property | 0deg | 90deg | ±45deg | 0/90deg |
|---|---|---|---|---|
| Ultimate Strength [MPA] | 155 | 47 | 85 | 112 |
| StD. UTS [%] | 4.3 | 10.4 | 1.3 | 1.1 |
| E modulus [MPa] | 10589 | 4619 | 5196 | 7636 |
| StD. E [%] | 5.7 | 4.2 | 7.7 | 1.7 |
| Poisson ratio | 0.383 | 0.129 | 0.509 | 0.191 |
| StD. ν [%] | 3.4 | 7.1 | 2.1 | 3.6 |

*Table 2 - Tensile test results*

# EXPERIMENTAL RESULTS

The results obtained from the tensile tests on standard specimens for all printing strategies are analysed by fracture mechanisms, Ultimate Strength, Elastic Modulus and Poisson ratio, that are listed in Table 2.

It is possible to see that the best performance relates to specimens printed at 0deg, followed by specimens printed at 0/90deg, at ±45deg, and finally by those printed at 90deg. The use of bidirectional SG has allowed the Poisson ratio to be calculated as the strategies varied. By comparing the Poisson ratio ($\nu$) values, the most evident thinning occurs in the ±45deg specimens, due to the alignment of the filaments during the test. This result is followed by the 0deg specimens, whose value is given by the thinning of the single connected filaments between them by the intermediate weld. The 90deg and 0/90deg specimens exhibit low thinning during the test and their values of Poisson ratio are very similar.
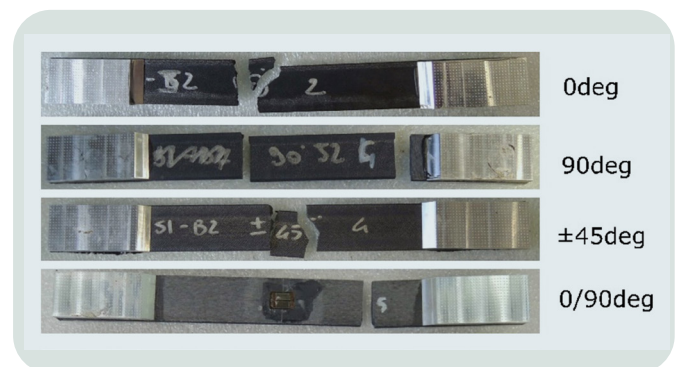


*5-Photos of specimens after failure for all printing strategies*

## 0deg strategy

The best performances are for 0deg samples: the deposition direction of the C-PEEK filament coincides with the direction of load application. In this way, all the deposited filaments help reinforcing the specimen in the direction of load application. Moreover, the good welding between filaments within the same layer and between one layer and the following one allows the high values of Ultimate Strength and modulus. As shown in Figure 5, the 0deg specimens has failed mainly along 45° with respect to the load direction, after a plastic deformation. In the same samples, the break propagates longitudinally to the length of the sample itself and sometimes the failure occurs explosively, with separation of small pieces.

## 90deg strategy

It is possible to see that the best performance relates The 90deg samples show the worst mechanical performances compared to the other strategies. The mechanical strength is given by the lateral welding of the deposited filaments that are perpendicular to the load. By extruding one filament after another, the welding between them occurs because of the contact and pressure given by the extruder on the material. The contact occurs between a filament that was already deposited and already begun to cool and a newly extruded filament. The build chamber is heated to 180°C, while the extrusion temperature of the filament is 450°C, so the filament begins to cool as soon as it is extruded. The colder the filament, the weaker the weld, and for this reason the 90deg specimens are much weaker than the filament along the deposition direction (0deg).

When the specimen is subjected to traction, the welds between filaments are the first to fail, causing a clear break in the specimen with the fracture surface coinciding with the cross-section.

Between one filament deposited and the next one, or between filaments of two adjacent layers, the welding has also a higher concentration of defects that reduce the real contact area between filaments, as shown in Figure 5. The presence of these pores can be reduced by optimization of the process parameters and by a better management of the printing temperature.

## ±45deg strategy

The ±45deg printing strategy is generally one of the most widely used for infills because it is the most balanced in stratification and behaviour. The mechanical performances are lower than those of 0deg but are higher than 90deg ones. As shown in Figure 4, the specimens have failure in several points after a plastic deformation.



## Stress-strain curves

The graph in Figure 6 shows the stress-strain curves for all the printing strategies analysed. It can be seen that the specimens at 90deg and 0/90deg show a barely noticeable plastic stretch and a net failure of the specimen at maximum load. For the 90deg specimens, a failure occurs at lower strains than for the 0/90deg specimens, due to the presence of the 0deg layers of the specimens.

From the comparison, it can be seen that printing strategies affect not only the final tensile strength, but also the way the specimen is broken. The choice of infill for the production of a part has to be evaluated by taking into account all these characteristics.

## 0/90deg strategy

As in the previous case, also the 0/90deg printing strategy is an intermediate situation between the 0deg and the 90deg ones. The specimens feature higher ultimate strength and modulus value than the samples of the ±45deg. This is due to the presence of the 0deg oriented layers that reinforce the specimen in the direction of load application. However, the results are worse than the results of 0deg specimens, due to the presence of 90deg layers, from which the break starts. The breaking mechanism is similar to the one observed in the 90deg specimens: there is a net failure along the cross section and little deformation before breaking.

*6 - Cross-section 90deg specimen*



*7 - Stress-strain curves*

## Fibres orientation

The presence of short Carbon fibres and their contribution within the PEEK matrix were evaluated by OM and SEM observations of the cross-sections. Figure 7 shows an OM image that allows to observe the orientation of the Carbon fibres within the specimen.

All the fibres are strongly oriented in the direction of extrusion of the filament. The filament was subjected to two different extrusion processes: the first one is due to the production of the filament, the second one is due to the 3D printing process and the extrusion of the filament through the nozzle.

The orientation contributes a lot in the improvement of mechanical performances of the matrix for all the printing strategies which have a filament deposition component concordant with the direction of load application, such as 0deg, ±45deg and 0/90deg.

The fibres do not provide benefits to the 90deg specimens, because the mechanisms for the 90deg strategy is only correlated to the adhesion between the filaments. Here only the matrix acts to have the welding and the fibres do not give any contribution.

In general, the presence of fibres contributes to increase mechanical properties and also increases brittleness of the material, greatly varying the fracture mechanisms compared to the matrix alone.



8 - OM images of fibres in 90deg specimens in the longitudinal section with respect to the direction of filament deposition
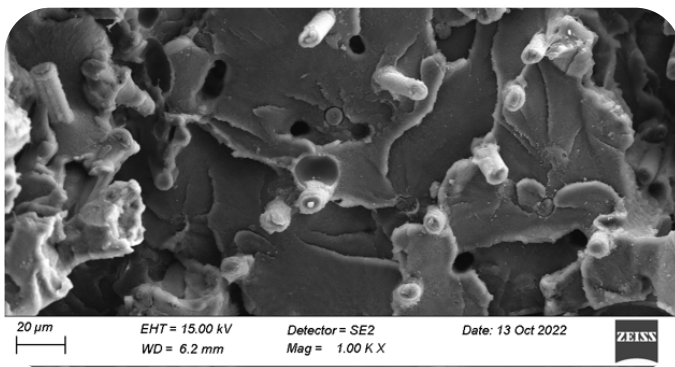


9 - SEM image of a failure surface of a 0deg printed specimen

Figure 8 shows a SEM image of the fracture surface of a specimen at 0deg. As seen from OM, it is possible to note that there is a strong orientation of the fibres in the direction of extrusion. The adhesion between matrix and fibres is good, but in many places there are holes of pull-out. The holes correspond to the points where the fibres pull out during failure and remain attached to the other half of the specimen. The optimal adhesion between the Carbon fibres and the PEEK matrix is difficult due to the high shrinkage of the matrix at cooling, which causes a detachment. An increase in the extrusion temperature or a surface treatment of the fibres could improve adhesion and therefore also the mechanical properties [13].

## CONCLUSIONS

This work presents a preliminary study of the influence of printing strategy on the production of Carbon PEEK printed with the Roboze Argo500 machine. It compares four different flat printing strategies of the specimens. The specimens with the best mechanical performance are those printed at 0deg, followed by those with 0/90deg, at ±45deg and finally by 90deg. Not only the mechanical performance but also the fracture modes of the specimens are strongly influenced by the printing strategy. Specimens at 0deg and ±45deg show clear plastic stretch before failure and high elongation at failure, while specimens at 90deg and 0/90deg have net failure without plastic deformation and low elongation at failure. So, the choice of the infill strategy for the production of a part with FDM has to be evaluated by taking into account all these characteristics.

The short Carbon fibres within the matrix are strongly oriented in the direction of extrusion of the filament and contribute a lot in the improvement of mechanical performances along the direction of load application. Adhesion between fibres and matrix is good and maybe could be further improved by increasing the extrusion temperature. The presence of pores in the cross section should be improved by optimizing the process parameters.

Sara Varetti: sara.varetti.ext@leonardo.com
Nicola Gallo: nicola.gallo@leonardo.com
Ignazio Scavo: ignazio.scavo@leonardo.com
Stefano Corvaglia: stefano.corvaglia@leonardo.com

# REFERENCES

[1] Thiruchitrambalam M, Kumar BD, Shanmugam D, et al. A review on PEEK composites – Manufacturing methods, properties and applications, Mater. Today: Proc. 2020, 33: 1085–1092.

[2] Jafferson J M, Chatterjee D. A review on polymeric materials in additive manufacturing, Mater. Today: Proc. 2021, 46: 1349–1365.

[3] Dul S, Fambri L, Pegoretti A. Fused deposition modelling with ABS–graphene nanocomposites, Compos. Part A 2016, 85: 181–191.

[4] Lay M, Thajudin NLN, Hamid ZAA, et al. Comparison of physical and mechanical properties of PLA, ABS and nylon 6 fabricated using fused deposition modeling and injection molding, Compos. Part B 2019, 176: 107341.

[5] Sheoran AJ, Kumar H. Fused Deposition modeling process parameters optimization and effect on mechanical properties and part quality: Review and reflection on present research, Mater. Today: Proc. 2020, 21: 1659–1672.

[6] Liu Z, Lei Q, Xing S. Mechanical characteristics of wood, ceramic, metal and carbon fiber-based PLA composites fabricated by FDM, J. Mater. Res. Technol. 2019, 8(5): 3741–3751.

[7] Giri J, Chiwande A, Gupta Y, et al. Effect of process parameters on mechanical properties of 3d printed samples using FDM process, Mater. Today: Proc. DOI: https://doi.org/10.1016/j.matpr.2021.04.283

[8] Berretta S, Davies R, Shyng YT, et al. Fused Deposition Modelling of high temperature polymers: Exploring CNT PEEK composites, Polym. Test. 2017, 63: 251-262.

[9] Patil P, Singh D, Raykar SJ, et al. Multi-objective optimization of process parameters of Fused Deposition Modeling (FDM) for printing Polylactic Acid (PLA) polymer components, Mater. Today: Proc. 2021, 45: 4880–4885.

[10] Chen K, Yu L, Cui Y, et al. Optimization of printing parameters of 3D-printed continuous glass fiber reinforced polylactic acid composites, Thin-Walled Struct. 2021, 164: 107717.

[11] Roy R, Mukhopadhyay A. Tribological studies of 3D printed ABS and PLA plastic parts, Mater. Today: Proc. 2021, 41: 856–862.

[12] Gunasekaran KN, Aravinth V, Kumaran CBM, et al. Investigation of mechanical properties of PLA printed materials under varying infill density, Mater. Today: Proc. 2021, 45: 1849–1856.

[13] Ding S, Zou B, Wang P, et al. Effects of nozzle temperature and building orientation on mechanical properties and microstructure of PEEK and PEI printed by 3D-FDM, Polym. Test. 2019, 45: 105948.

[14] Chang B, Li X, Parandoush P, et al. Additive manufacturing of continuous carbon fiber reinforced poly-ether-ether-ketone with ultrahigh mechanical properties, Polym. Test. 2020, 88: 106563.

[15] Magri AE, Vanaei S, Vaudreuil S. An overview on the influence of process parameters through the characteristic of 3D-printed PEEK and PEI parts, High Perform. Polym. 2021, 33(8): 1–19, DOI: https://doi.org/10.1177/09540083211009961

[16] Jayaraghul TK, Karthik K, Yaswanth A, et al. Nozzle flow characteristics of P.E.E.K (Poly-ether ether ketone) material used in 3D-printing, Mater. Today: Proc. 2021, 44: 2963–2967.

[17] Stepashkin AA, Chukov DI, Senatov FS, et al. 3D-printed PEEK-carbon fiber (CF) composites: Structure and thermal properties, Compos. Sci. Technol. 2018, 164: 319–326.

[18] Wang P, Zou B, Ding S, et al. Effects of FDM-3D printing parameters on mechanical properties and microstructure of CF/PEEK and GF/PEEK, Chinese J. Aeronaut. 2020, 34(9): 236-246. DOI: https://doi.org/10.1016/j.cja.2020.05.040

[19] Meng L, Xiaoyong T, Junfan S, et al. Impregnation and interlayer bonding behaviours of 3D-printed continuous carbon-fiber-reinforced poly-ether-ether-ketone composites, Compos. Part A 2019, 121: 130-138.

[20] Yang D, Cao Y, Zhang Z, et al. Effects of crystallinity control on mechanical properties of 3D-printed short-carbon-fiber-reinforced polyether ether ketone composites, Polym. Test. 2021, 97: 107149.

[21] ASTM D3039/D3039M-08. Standard Test Method for Tensile Properties of Polymer Matrix Composite Materials.

# POLARIS

Scan this QR code to access the web version

https://www.leonardo.com/polaris_2023_47/