Evaluation of Standardized Testing Protocols for Fuel Cell Modules in Heavy-Duty Applications

- Measurement Results from 7 Standard Sized Fuel Cell Prototypes -

A. Balaji [1)] K.J. Mrozewski [1)] C. Bekdemir [1)] Y. Wang [2)] E. Havret [3)]

1) TNO Powertrains, Helmond, The Netherlands

E-mail: anirudh.balaji@tno.nl

2) FEV, Motor Hybrid and Fuel Cell Powertrains, Aachen, Germany

3) CEA, Energy Division, Grenoble, France

ABSTRACT: Standardization is key to accelerate fuel cell deployment into a wide array of heavy-duty markets. Therefore, a collective of heavy-duty Original Equipment Manufacturers, fuel cell manufacturers, and research institutes have proposed standards of fuel cell module size, interface, and testing protocols. The objective of this study is to evaluate the 6 tests from the standardized testing protocols by applying them to 7 PEM fuel cell modules of different fuel cell suppliers. The tested modules have a nominal power in the range of 42.5 to 125kW. As a result, key performance indicators such as efficiency and dynamic behavior of these state-of-the-art prototypes will be presented. Experiences of applying these protocols and processing the measurement data will be shared. Furthermore, recommendations will be given for experimental works that intend to apply these fuel cell module protocols in the future.

KEY WORDS: PEMFC, heavy-duty, testing, standard

1. INTRODUCTION

The development of Fuel Cell (FC) systems for heavy-duty (HD) applications has intensified in response to the growing demand for sustainable, zero-emission mobility solutions. As the technology is steadily advancing towards maturity, standardization is key to accelerate its deployment into a wide array of heavy-duty markets. In view of this, the StasHH project, funded by the European Commission, was tasked with defining a standard for Fuel Cell Module (FCM) form factors, physical and digital interfaces, and test protocols [(1) (2) (3) (4) (5)]. In particular the absence of standardized testing protocols on the FCM level remains a major hindrance to reliable and consistent evaluation of FCM performance across manufacturers and applications. Therefore the StasHH consortium, consisting of FC industry and research entities, proposed a set of testing protocols that would provide the framework for consistent validation and benchmarking of FCMs, aligning with industry requirements for scalable performance assessments [(5) (6)].

The objective of this study is to evaluate the standardized testing protocols by applying them to 7 FCMs from different suppliers. An essential aspect was to consider data post-processing and reporting protocols to ensure result comparability, regardless of tests being conducted at different locations, in different test environments or under different circumstances. To protect proprietary designs, solutions and innovations of individual FCM manufacturers, the testing protocols adopted a “black box” approach, allowing performance evaluation without compromising Intellectual Property Rights. The proposed standardized testing protocols were validated by carrying out a test campaign on 7 FCMs from manufacturers from Japan, North-America, and Europe, and tested at 3 different test locations.

In the following chapters, first, the testing protocols and test object are introduced briefly. Then, the results of the test protocol measurements are presented. Subsequently, lessons learned from applying the test protocols are discussed.

2. TESTING PROTOCOLS AND TEST OBJECT

Understanding the performance, sizing, and interfaces of FC systems is paramount to fuel cell electrified vehicle/vessel development and integration. Testing of these systems is the primary means to obtain the required information. In this chapter, a concise overview of the relevant measurement parameters and their definitions in the StasHH project are provided, thus creating a common understanding for the interpretation of the acquired FCM data.

The evaluated testing protocols consist of six tests. Detailed descriptions are freely available in a StasHH project deliverable [5] and are summarized in another EVTeC 2025 article [6]. These tests consist of: start-up and shut-down, ramp-up and ramp-down dynamics, efficiency curve, dynamic profile, static inclination, and impaired cooling. Here, to satisfy the paper page number limit, only the polarization curve test protocol is shown, see Figure 1. This test is carried out to trace the efficiency curve of the FCM as a function of the net electrical power output, measured at steady state from minimal to nominal power. It is one of the most common and relevant methods to characterize FC performance, as it highly impacts multiple techno-economic features, such as system operating costs or sizing of auxiliary systems and their components, e.g. cooling/HVAC, but also to help an Original Equipment Manufacturer (OEM) to define a hybridization strategy.

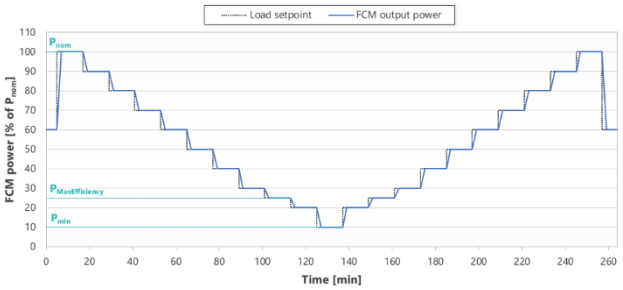

Figure 1: Test profile for efficiency curve characterization. This example assumes maximum FCM design efficiency power and minimal power at 25% and 10% of nominal power, respectively.

The test object is a FCM including a variety of sub-systems. Some of the sub-systems are optional and most of the sub-systems have an interface with the test facility. A schematic overview of a generic test object is shown in Figure 2. The large box is the boundary of the FCM, which is the same as the boundary of the test object. Outside the test object, the boxes represent the features of the test facility and the arrows that connect them to the test object indicate the exchange of fluid flows, electric current, and data. A comprehensive overview of test variables and performance parameters is presented in [5]. It comprises of about 40 parameters either measured directly using the test bench and FCM sensors or derived from measured parameters. Each parameter is described and where applicable a formula is given to compute it from other measured parameters.

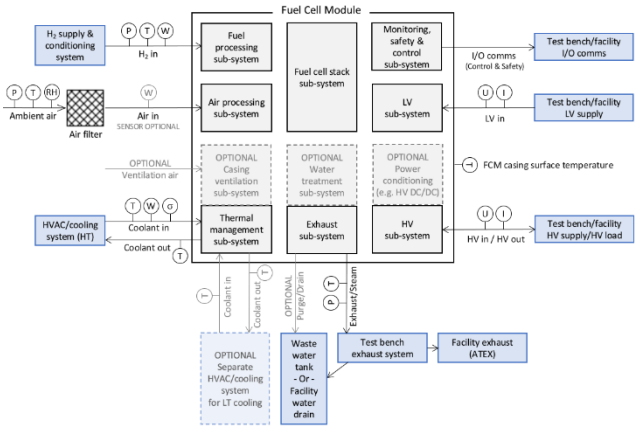

Figure 2: Generic FCM boundaries, inputs, outputs and measured signals.

3. TEST PROTOCOL RESULTS

In this chapter, the outcome of the test campaign is presented and discussed. A total of seven FCMs having a nominal power in the range of 42.5 to 125kW were tested at three test locations. The same set of test protocols are applied to all the FCMs and since the test campaigns were conducted at various locations involving different test benches, the data acquisition and data processing methods need to be standardized. This is essential to produce comparable results and to ensure that the testing protocols are repeatable across a wide array of applications.

3.1 Data acquisition and processing

The data is logged by a central data acquisition device at a prescribed frequency of 100Hz and time-stamped. The raw data is firstly filtered to reduce measurement noise and then processed.

The mean values (current, voltage, mass flow, and temperature) are calculated using an averaging window to remove the effects of measurement noise. According to the results presented in [7], a stabilization time of 30 seconds followed by an averaging window of 1 minute is sufficient to produce a polarization curve for FC stacks. However, in the case of FCMs the added system complexity may require larger windows. Therefore, a sensitivity study is performed to understand the effect of the stabilization and averaging window sizes.

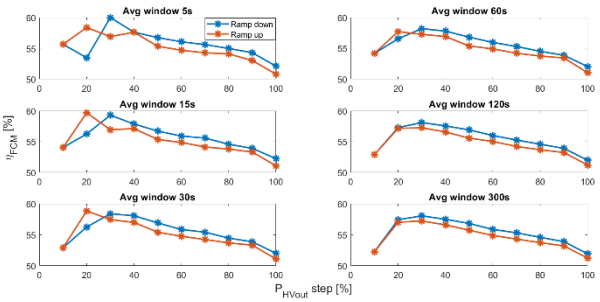

Figure 3. Effect of averaging window on efficiency calculation.

A 15 minute stabilization window is considered in the test protocol and initially this is not varied. *Figure 3* shows the effect of the averaging window size on calculated efficiency for one of the tested FCMs. The window size is varied from 5 to 300s and it can be seen that the shorter windows lead to more inconsistencies in the calculations owing to a combination of measurement oscillations and response lag between fuel mass flow and current increase/decrease. By utilizing a larger window these effects can be averaged out. As seen in *Figure 3*, an averaging window of 120s is sufficient to stabilize the calculation of mean values, and similar trends were observed on all the tested FCMs.

The proposed test protocol uses a 15 minute stabilization window resulting in a total measurement effort of ~4 hours for the efficiency curve characterization test. To reduce the measurement effort, a study is performed to understand the effect of the stabilization time on calculated parameters while maintaining the 120s averaging window.

Similar to the effects seen in *Figure 3*, it was observed that smaller windows result in inconsistencies in the calculations particularly at lower loads while a window larger than 300s does not produce significant changes. Therefore, using a 300s stabilization window followed by a 120s averaging window is seen to be sufficient which can help reduce the measurement effort by up-to 50%.

3.2 Observations from applying the test protocols

The repeatability of the test protocols and applicability at different test locations is evaluated by analysing the variation in the performance of the tested FCMs on the dynamic cycles (WHSC, ISO 8178, custom cycle) which are repeated up-to 5 times for all tested FCMs and the efficiency curve test, which was repeated at different points in the test campaign for certain FCMs.

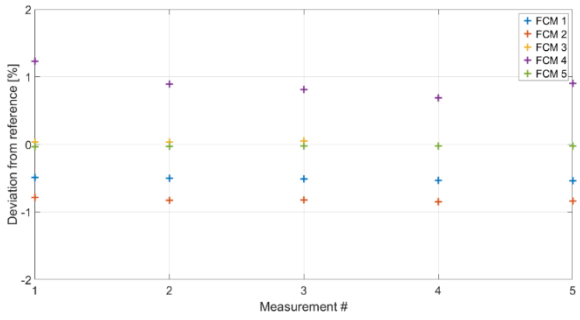

Figure 4. Deviation across WHSC measurements.

Figure 4 shows the deviation in the cumulative energy between the reference and measurement for the WHSC tests with each colour representing a different FCM. The reference energy is calculated by de-normalizing the WHSC trace for each individual FCM. While a spread in performance between the FCMs can be noted, the data shows that over the repeated tests the performance of an individual FCM varies less than 1%, irrespective of the test location.

The repeat efficiency curve characterization offers insights into whether the performance of an individual FCM is repeatable when running the same test at different times at the same location. In general, the modules tested show a good degree of repeatability with a maximum deviation of less than 2% across all loads.

The dynamic response of the FCMs is evaluated by analysing the time required for the voltage, current, and coolant temperature to stabilize after a load step. Figure 5 shows the change in current and voltage across such a measurement with the current change shown in blue and voltage change in red.

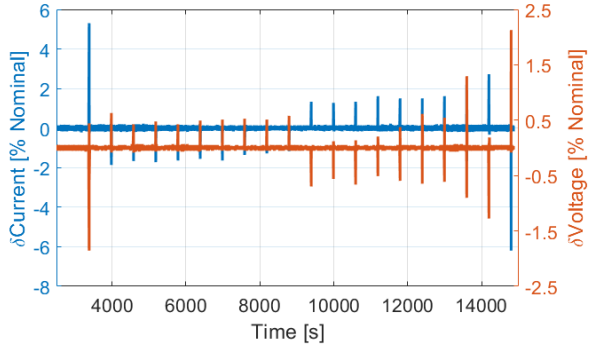

Figure 5. Δ current and voltage extracted from efficiency curve characterization test.

Figure 6 shows the current and voltage change after one of the load changes, and it can be noted that the current and voltage stabilize in approximately 1 second. While there is a difference in performance across the tested FCMs, the order of magnitude for voltage and current stabilisation time is similar.

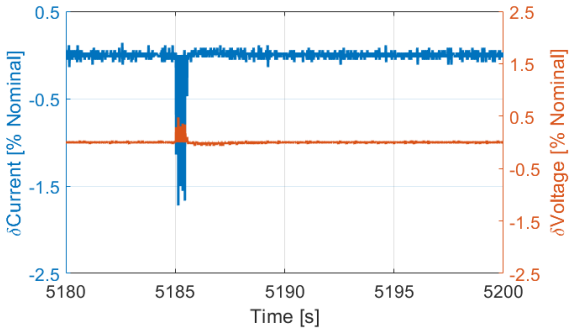

Figure 6. Zoomed in version of Δ current and voltage.

The top half of *Figure 7* illustrates the coolant circuit of the test setup, comprising of the cool-con of the test facility controls, the coolant flow line to the intermediary heat exchanger, and the FCM that controls its internal temperature. The bottom half of *Figure 7*

shows the coolant temperature from the intermediary heat exchanger to the cool-con of the test facility (T2) measured during different test campaigns (locations). Oscillations of up to 6°C can be observed. Owing to the black-box principle used in StasHH, the internal state of the FCM is not always known (i.e., T3 and T4 are not measured).

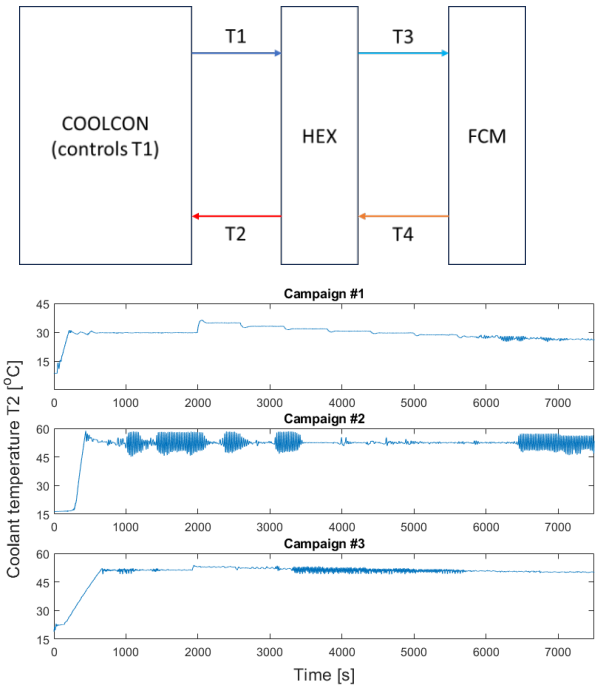

Figure 7: Coolant loop of the test bench.

A coolant liquid temperature (T3) fluctuation of 2% can result in voltage fluctuations of 0.5% according to observations reported in[7]. In the case of an FCM with coolant temperature of 60°C this represents a fluctuation of ±1.2°C. The voltage duration change observed in *Figure 6*, is limited to a few seconds and the voltage magnitude change is less than 0.5% after a load step is applied, this indicates that the internal coolant temperature of the FCM does not fluctuate much after the application of a load step.

In summary, the testing protocols are shown to be reproduceable and repeatable across FCMs, testing locations and test benches. However, certain recommendations can be drawn to help further standardize the protocols, which are elaborated in Chapter 5.

3.3 Key measurement results

In the remainder of this chapter, key test results from all the tested modules in the StasHH project are presented. As per the agreement in StasHH, data of individual FCMs are anonymized by presenting aggregated results. The bar charts represent average performance indicator value from all tested modules. Additional information about the spread of the results is supplied by black bars that indicate the minimum and maximum value within the dataset. Individual figures also indicate the number of StasHH modules that is considered in the analysis.

The start-up and shut-down test was performed on all seven modules and the results are represented with the amount of energy that is consumed by the module during the start-up and shut-down period as well as their respective durations. The bar chart in *Figure 8* shows the average and min/max values of the start-up and shut-down energy in [Wh] and duration in [s]. The tests revealed that, starting from a deactivated state, the FCMs were ready to deliver output power between 17 and 38 seconds, and the maximum needed energy supply was equal to 7 Wh. The average energy consumption during shut-down is about 17 times higher than the start-up energy, while the shut-down duration is on average 10 times higher than the start-up duration. The long shut-down duration and therefore energy consumption are not related to the capability of the FCM to ramp-down the power output. It is rather due to the FCM-internal shut-down protocol, which needs to ensure proper conditioning and deactivation of all the components within the module.

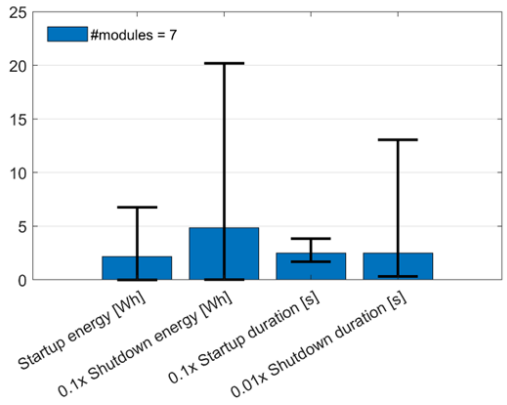

Figure 8: Start-up and shut-down energy and duration. The bars represent the average of seven FCMs including the min/max values.

The ramp-up and ramp-down test was performed on most of the modules and the results are represented with the ramp-up and ramp-down rates (from minimal to nominal power) and the duration to reach the new stationary point. The bar chart in *Figure 9* shows the average and min/max values of the ramp rates in [kW/s] and duration in [s]. Averaging is done over all the ramp-up and ramp-down events including the different temperature settings in the test profile. The coolant temperature effects (T1 in *Figure 7*) were either insignificant or changed the results without revealing a consistent pattern. The maximum recorded ramp-up

and ramp-down rates were 24.4 kW/s and 197 kW/s, respectively. Most of the tested FCMs were able to attain their nominal power output within 10 seconds, with the slowest one reaching its nominal value in 40 seconds. Ramping down from nominal to minimal power was faster and lasted between 1 to 20 seconds, depending on the module. This is in line with expectation, as, among others, ramping up involves increasing the supply of reactants and gradually adjusting flow rates to ensure they enter the FC at a proper stoichiometry, relative humidity, temperature, and pressure balance to avoid damaging the FC (e.g. reactant starvation). By contrast, ramping down primarily involves cutting off or reducing the reactant flow which is mechanically simpler and faster, that leads to limiting the electrochemical reaction rate and allowing for quicker operational stabilisation.

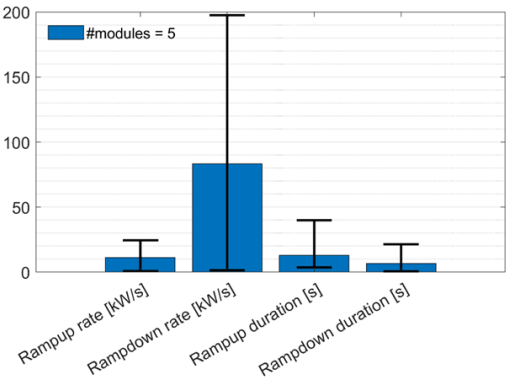

Figure 9: Ramp-up and ramp-down rate and duration. The bars represent the average of five FCMs including the min/max values.

The efficiency test was performed on all the modules and the results are represented with the FCM efficiency as a function of the normalized power output. The blue squares in *Figure 10* show the average FCM efficiency in [%] calculated across all modules at a specific normalized power output level. Averaging includes the efficiency value of each FCM at the same power output for both the upward and the downward sweep of the test profile (Figure 1). The min/max efficiencies corresponding to the downward sweep of the efficiency curve test are illustrated in red, while the upward in black. The downward efficiency is equal or higher than the upward efficiency by up to two percentage points. The hysteresis effect observed in fuel cells during upward and downward sweeps is primarily caused by differences in dynamic phenomena such as gas transport, water management, and catalyst surface conditions[8][9]. The large spread of efficiency, 5 to 8 %-point, among the different FCMs has several causes. First of all, the FCMs include a stack and BoP which may largely differ from each other because they are designed for different applications in the HD sector. Furthermore, some FCMs have a DC-DC converter integrated inside the module, so the FCM efficiency also includes losses from the converter.

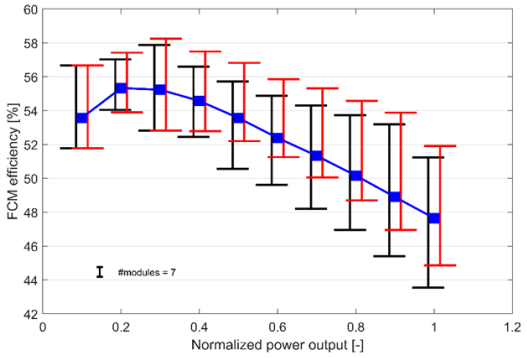

Figure 10: FCM efficiency as a function of the normalized load. The blue symbols represent the average of 7 FCMs. The min/max efficiencies corresponding to the upward and downward sweep are indicated with the black and red bars, respectively.

All the modules have been exposed to 3 dynamic cycles: WHSC, ISO 8178 (type E3) and a custom cycle. For some of the modules, it was not feasible to run all the three cycles due to strict limitation of the ramp rate by the supplier and/or limiting low load capability to properly run the defined cycle. The bar chart in *Figure 11* shows the average and min/max values of the hydrogen consumption reported in g/hkW$_{nom}$. The power cycles are normalized based on the nominal power of each FCM, therefore, the presented unit on the bar chart is grams of hydrogen consumed per hour divided by the nominal power rating of the respective FCM.

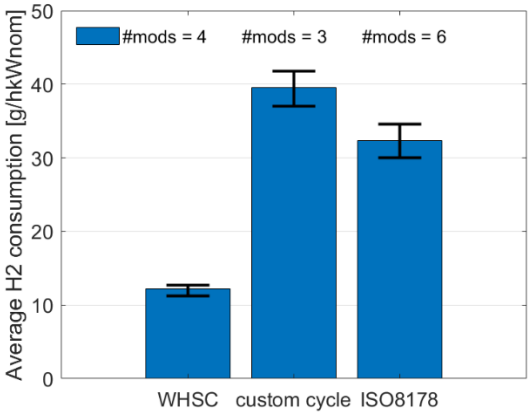

Figure 11: Hydrogen fuel consumption in g/hkW$_{nom}$ for WHSC, custom cycle, and ISO 8178 (type E3). The bars represent the average and the min/max values.

The number of modules that are included in the data is indicated above each fuel consumption bar in *Figure 11*. The overall lowest hydrogen consumption was observed for the WHSC, as it presents relatively low-load operation for most of its duration. ISO 8178 and the custom cycle have extensive high-load periods and transients, that resulted in high hydrogen consumption across modules. The relative variations among the different FCMs is observed to be between 4% and 8% for all cycles.

The inclination test was performed on five modules because the tilt table was not suitable to safely tilt all the modules. The results are processed in term of the FCM efficiency as a function of the inclination angle, both positive and negative around two axes. The results did not show any specific trend compared to the uninclined reference situation. Most often the efficiency deviates about 1%-point, however, there are also cases where the deviation is around 2.5%. Since the deviations show no specific trend over the different FCMs, they are attributed to the individual differences in internal design and operating conditions management that tend to respond differently to the measurement protocol.

Events that compromise the capacity of the cooling system to evacuate heat from the FCM are likely to happen within its lifetime. The impaired cooling test was designed to check the FCM performance under excess temperature conditions, by checking its robustness regarding such an event and verifying its ability to restart without problems. Considering that the modules in StasHH are prototypes and since this safety-oriented test is potentially degrading to the FCM, it was proposed as an optional test. Manufacturers of 4 FCMs decided to submit their modules to this test, and the findings indicate that the FCMs are capable of operating for an extended period at their specific coolant temperature limit, safely shut-down and restart without any change in performance.

4. LESSONS LEARNED FROM THE TEST PROTOCOLS

The testing campaign offered invaluable insights into the strengths and limitations of the proposed standardized protocols applied to multiple FCMs and in different testing environments. By analysing key challenges and variations encountered, this chapter outlines and discusses observations and lessons learned from applying these standardised test protocols.

The general feedback from the application of the standardised FCM test protocols in three locations, is that the proposed tests are sound. While being prescriptive, the protocols still offer the required flexibility, for instance the efficiency curve test includes the measurement of efficiency at minimum output power and measurement at peak system-level efficiency, which often vary between modules. Nevertheless, an observation was that test conditions should be stated more explicitly in order to strictly impose and validate if minimum performance acceptance criteria for the target application are met. As an example, for mobility applications, the minimum tilt angle for the “Performance under inclination” should be set according to HD on-road automotive requirements of 8° tilt, while for maritime an angle of 22.5° tilt is required [10] [11].

Furthermore, it was observed that the in-built settings of some tested FCMs were too constraining, preventing them from realizing the test as described within the protocols. In one case, the dynamic response of the FCM was slower than the ramp times imposed during the semi-transient dynamic load profile tests. By consequence, the duration of the full profile is longer, extending the time window in which the FCM is operational. In practice, a slower ramp time leads to an additional penalty for the FCM by increasing its average hydrogen consumption calculated for that dynamic profile.

To minimize variability and ensure comparability of results generated across various testing locations, setups, and for different StasHH modules, we aimed to identify standardisation opportunities and possible best practices within the data post-processing and result reporting pipeline.

Regarding the overall economy of the testing protocols, 50% reduction of the overall measurement effort can be achieved by optimising the measurement time needed to achieve a reliable steady state measurement. It was observed that, after each load step, using a 300 second stabilization window, followed by a 120 second averaging window for the calculation of mean parameters (power, efficiency, flows) is sufficient to eliminate the effects of measurement noise. Utilizing a moving mean filter with a frame length of 50 on the raw measurement produces signals that are suitable for computation of parameters and comparison across FCMs.

The application of the first concept of standardised testing protocols helped in identifying a number of potential improvements. For the determination of the start-up duration and energy consumption, measured during the “Start-up and shut-down” test, it is important to take into account the fact that upon sending the “Ignition ON” request, each FCM can follow different internally programmed strategies for start-up and shut-down. Despite the standardisation of the digital interface (in particular the State Machine, as defined in [1], achieving the “In-operation” and “Off” states could not be controlled via a singular demand from

the test rig. These differences were not foreseen in the testing protocols or the digital interface standard, leading to additional test configuration effort.

To reduce the total testing time and hydrogen consumption, it is recommended to limit the number of individual cycle reruns from five to two during the “Dynamic load profile performance” test. Here, the first measurement is used for the calculation of key performance metrics, and the second used for measurement validation. The StasHH test campaign revealed that all FCMs exhibited stable behaviour as the differences in hydrogen consumption measured across the five cycle repetitions were marginal (< 2%). This stability suggests that, under normal conditions, two repetitions suffice for assessing FCM performance in dynamic operation.

However, additional cycle repetitions should be considered if the FCM behaviour appears unstable, indicated by deviations exceeding a predefined threshold, e.g., more than 5% variance in hydrogen consumption between cycles, module temperature instability exceeding ±10°C from the setpoint. In scenarios, where instability is observed, determining which cycles to include in the average is critical. If early cycles, e.g. the first or second, show significant deviation due to transient effects, such as no preconditioning procedure, system warm-up or unexpected anomalies, they may need to be excluded from the final calculation. In such cases, the most stable consecutive cycles, e.g. cycles 3-5, should be used to derive an average hydrogen consumption value. Establishing a standard method for defining and excluding outlier cycles will enhance consistency and comparability across tests.

5. CONCLUSIONS

The StasHH project proposed a set of standardised testing protocols, designed to ensure consistent performance benchmarking at the fuel cell module level. These protocols are successfully applied and validated on seven modules from six manufacturers, tested at three locations and using different testing equipment.

The test campaign demonstrated the general performance characteristics of the StasHH FCMs and the data processing provided insights into test repeatability and further protocol improvement. From a deactivated state, all the modules were ready to deliver power in well under a minute, some within 20 seconds and requiring maximally 7 Wh of energy supplied during start-up. Nominal power outputs were reached between 10 and 40 seconds, with a maximum measured ramp-up rate of 24.4 kW/s. Ramping down from nominal to minimal power output took between 1 and 20 seconds, with a maximum ramp-down rate of 197 kW/s. Despite the fact that individual modules were designed to different applications in the HD sector, the dynamic cycle tests revealed only a 4 to 8% difference in average hydrogen consumption between them. Using a 300s stabilization window followed by a 120s averaging window is sufficient to eliminate the effects of measurement noise. This approach can result in a 50% reduction of the overall measurement effort compared to the testing protocols. The efficiency curve test was repeated for individual FCMs at different times during the test campaign and the results were found to be repeatable, with a maximum deviation of less than 2% across all loads, indicating that the testing protocol gives reproducible outcomes.

The adoption of the testing protocols will promote benchmarking as an integral part of standardisation efforts to ensure comparability and scalability across the industry, facilitating comparative assessments and information exchange, and helping address the current gap in module-level standards for HD applications. Insights gained from this work highlights the importance of continued collaboration between stakeholders and may serve as a guideline for future FCM standardisation initiatives. Future efforts should focus on remaining standardization gaps. Firstly, harmonization of electrical power output should be prioritized, including limiting the range by introducing a universal connector system and standardised voltage ranges. Secondly, defining relevant and unified dynamic load cycles, as the application base in the HD sector for the standard FCMs is broad and is expected to become broader in the future. Thirdly, further effort should be put in the development of unified methods for data processing standards that can enhance the reliability of inter-laboratory comparisons and supports objective benchmarking across diverse FCM tests. Lastly, expanding engagement with stakeholders across the supply chain and further standard promotion among OEMs should be done in order to avoid requests for tailor-made solutions. By doing so, the industry can foster a more unified and efficient pathway toward the large-scale adoption of FC technologies in HD applications.

ACKNOWLEDGMENT

The StasHH project has received funding from Fuel Cells and Hydrogen 2 Joint Undertaking (now Clean Hydrogen Partnership) under Grant Agreement No. 101005934. This Joint Undertaking receives support from the European Union’s Horizon 2020 Research and Innovation Program, Hydrogen Europe and Hydrogen Europe Research.

REFERENCES

(1) F. Zenith, R. Bouwman, H. Lundkvist, The StasHH Fuel-Cell Module Standard, IEEE VPPC 2022, DOI: 10.1109/VPPC55846.2022.10003465

(2) Ruud Bouwman, “Standard Size definition”, StasHH deliverable D3.2, 2022, https://stashh.eu/stashh-standard.

(3) Ruud Bouwman, “Standard Interface definition”, StasHH deliverable D3.3, 2022, https://stashh.eu/stashh-standard.

(4) Henrik Lundkvist, “Standard API definition”, StasHH deliverable D3.4, 2022, https://stashh.eu/stashh-standard.

(5) K. Mrozewski, C. Bekdemir, “Test Protocols”, StasHH deliverable D5.6, 2024, https://stashh.eu/stashh-standard.

(6) K.J. Mrozewski, A. Balaji, C. Bekdemir, R. Bouwman, H. Lundkvist, “The StasHH Size, Interface, and Testing Protocol Standards for Fuel Cell Modules in Heavy-Duty Applications”, Submitted to EVTeC 2025

(7) R. Bove, T. Malkow, A. Saturino and G. Tsotridis, “PEM fuel cell stack testing in the framework of an EU-harmonized fuel cell testing protocol: Results for an 11 kW stack”, Journal of Power Sources, pp. 4522-460, 2008.

(8) Y. Hou, B. Wang, Z. Yang, “A method for evaluating the efficiency of PEM fuel cell engine”, Applied Energy Volume 88, Issue 4, pp. 1181-1186, 2011.

(9) A. Iranzo, S.J. Navas, J. Pino, N.A. Althubiti, M.R. Berber, “Influence of the dwell time in the polarization hysteresis of polymer electrolyte membrane fuel cells”, Electrochimica Acta Volume 426, 2022.

(10) Lloyd’s Register, “Type Approval System Test Specification Number 1, 2021”

(11) International Electrotechnical Commission, “Electrical installations in ships – Part 504: Automation, control and instrumentation”, IEC 60092-504;2016, 2016.