ACS SENSORS

pubs.acs.org/acssensors

Article

Accelerating Plasmonic Hydrogen Sensors for Inert Gas Environments by Transformer-Based Deep Learning

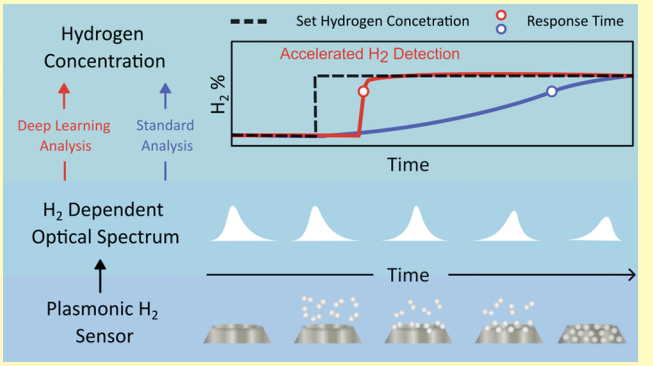
Viktor Martvall, Henrik Klein Moberg, Athanasios Theodoridis, David Tomeček, Pernilla Ekborg-Tanner, Sara Nilsson, Giovanni Volpe, Paul Erhart,* and Christoph Langhammer*

Cite This: https://doi.org/10.1021/acssensors.4c02616

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Rapidly detecting hydrogen leaks is critical for the safe large-scale implementation of hydrogen technologies. However, to date, no technically viable sensor solution exists that meets the corresponding response time targets under technically relevant conditions. Here, we demonstrate how a tailored long short-term transformer ensemble model for accelerated sensing (LEMAS) speeds up the response of an optical plasmonic hydrogen sensor by up to a factor of 40 and eliminates its intrinsic pressure dependence in an environment emulating the inert gas encapsulation of large-scale hydrogen installations by accurately predicting its response value to a hydrogen concentration change before it is physically reached by the sensor hardware. Moreover, LEMAS provides a measure for the uncertainty of the predictions that are pivotal for safety-critical sensor applications. Our results advertise the use of deep learning for the acceleration of sensor response, also beyond the realm of plasmonic hydrogen detection.



KEYWORDS: *hydrogen sensing, plasmonic sensing, nanoparticles, deep learning, neural networks*

The ability to detect, quantify, and distinguish chemical species accurately and rapidly is crucial for technologies requiring swift data capture to support well-informed decision-making, automation, and process-monitoring. Such technologies span a wide range of applications, including environmental monitoring,[1] biosensing for real-time disease diagnostics,[2] chemical process control[3] and food quality evaluation.[4] They all have in common that they critically rely on the development of sensors that are not only precise, sensitive, and selective but also respond rapidly to their target substance and are able to deliver an accurate quantitative measure of the concentration of that target.

A domain that is rapidly expanding and where sensing will play a pivotal role in facilitating safe large-scale implementation is hydrogen-based technologies, including fuel cells for heavy transport, shipping and aviation, energy storage solutions, and green steel production. They all have in common the promise of substantial reductions of greenhouse gas emissions. However, this prospect also generates new demands for active process monitoring and control and introduces safety concerns owing to the high flammability of H_2-air mixtures. All of these issues can be effectively addressed by the development of accurate H_2 sensors.

From a sensing environment perspective, two distinct settings exist, where ambient conditions characterized by an abundance of oxygen constitute the most obvious one. The second setting, which is of significant technological relevance but much less discussed in the scientific literature to date, is so-called "inert" or "oxygen-starved" environments. They are established to encapsulate/enclose large-scale H_2 installations, such as entire engine rooms on fuel-cell-powered ships, or fuel pipes on H_2-powered airplanes, to avoid the formation of flammable air-H_2 mixtures. The rapid detection of even the tiniest H_2 leaks inside these inert gas encapsulation infrastructures is critical to provide enough time for the implementation of appropriate measures to eliminate, as well as spatially localize, the leak by placing sensors at strategic locations inside the system. Specifically, in such installations, the system is continuously flushed by an inert gas, such as N_2 or Ar, to eliminate or drastically reduce the presence of molecular oxygen. Importantly, we note that the inert gas used in such systems will be of low quality from a purity perspective with respect to species, such as H_2O, CO, or SO_x, for cost reasons. This combination of lack of O_2 and the presence of sizable amounts of "poisoning" molecules that bind strongly to

Received: September 24, 2024
Revised: December 13, 2024
Accepted: December 24, 2024

many sensor surfaces poses a significant challenge because (i) established H_2 sensors of the catalytic and thermal type require O_2 to work and (ii) because the strong molecular bonds either block/poison surface sites required for H_2 dissociation and/or detection, or facilitate surface reactions that consume hydrogen species and thus prevent them from being detected.[5]

To steer the development of next-generation H_2 sensors that meet the upcoming demands of the widely implemented H_2 technologies outlined above, agencies and stakeholders have defined performance targets. The most well-known ones are defined by the U.S. Department of Energy (DOE), which identify sensor speed at ambient conditions as one of the key unresolved metrics.[6] To this end, a small number of studies exist where H_2 sensors with response times just below 1 s for a 0.1 vol % H_2 pulse have been demonstrated experimentally.[7−9] However, although they are indeed important breakthroughs, these demonstrations were made in an idealized pure H_2-vacuum environment that constitutes a severe simplification. As the main reason for this simplification, we identify the aforementioned challenge of "poisoning" molecular species in technologically relevant sensing environments due to their impact on the surface chemistry of a sensor. While it has been shown that deactivation-resistant alloys and polymer filters can mitigate sensor deactivation caused by gases such as CO, CO_2, CH_4, and NO_2,[7,10−13] and that machine learning can help alleviate sensor deactivation due to H_2O,[14] the presence of these gases still slows down the sensor's kinetics. Hence, even though these demonstrations of H_2 detection with subsecond response in idealized vacuum/pure H_2 conditions exist, it is clear that further advances in this field are necessary.[5]

Traditionally, such advances are attempted by developing new sensing materials, by nanostructuring the sensing materials and/or signal transducers, and by refinement or modification of fundamental physical sensing mechanisms.[7,9,12,15−18] Interestingly, however, only limited attention has been directed toward harnessing the potential of tailoring the treatment of output data of existing sensor platforms with the aim to improve the sensor response time, e.g., by machine learning techniques. While several studies have leveraged the potential of machine learning to enhance the *accuracy* or *sensitivity* of different kinds of gas sensors,[14,19−23] including H_2, the potential to enhance *sensor response times* has only recently started to gain attention.[24,25] Specifically, Lin et al. used a convolutional neural network (CNN) to accelerate the response time of a Pd nanocap plasmonic H_2 sensor, achieving up to a 3.7-fold reduction in response time. While proving the potential of using deep learning for this purpose, there is potential to further accelerate sensing as their CNN-based approach may not capture the full complexity of the sensor's dynamic behavior.

In response to and motivated by the high demand for faster sensors in general and H_2 sensors for inert gas environments in particular, here, we develop an approach for accelerating H_2-sensing that combines optical nanoplasmonic sensors based on hydride-forming metal nanoparticles, such as Pd and its alloys with coinage metals,[5,26,27] with transformer-based deep learning. As we show below, this combination enables sensor operation in technically highly relevant oxygen-starved environments with significantly improved sensor performance. Specifically, our approach reduces the response time for predicting the H_2 concentration by up to 40 times in inert gas environments, surpassing conventional methods which are limited by the need to reach full thermodynamic equilibration of the sensor after a change in H_2 concentration and hampered by sensor deactivation effects due to the presence of molecular contaminants in the inert gas environment. It also provides uncertainty estimates of the sensor response predictions made, which is an important feature for the safety-critical application of H_2 sensing.

To analyze the output data of plasmonic hydrogen sensors, which typically consists of a time series of scattering or extinction spectra in the visible light spectral range,[5,27] the current standard analysis (SA) widely applied in the field collapses each such measured spectrum to a single spectral descriptor, such as the spectral peak position, the full-width half-maximum or the centroid position.[28] As the key point, in this analysis, a significant amount of information contained in both the complete spectrum and its temporal evolution is not used, since it is collapsed into a single descriptor. Hence, what we argue and demonstrate here is that by utilizing *all* of this information via a tailored deep-learning model, it is possible to dramatically improve the sensor performance by analyzing temporal trends of the full spectral information to predict the thermodynamic sensor saturation level *before* this saturation is physically reached and thereby accelerate the sensor response time.

To harness this information with the aim to accelerate plasmonic H_2 sensor response in general, and in inert gas environments in particular in this work, we introduce LEMAS, short for Long Short-term Transformer Ensemble Model for Accelerated Sensing, which improves the sensor speed by learning the relationship between the time dependence of the full spectrum and the H_2 concentration, while simultaneously assessing uncertainty in the model predictions through model ensembles. The long short-term transformer (LSTR) architecture consists of long and short-term memory and has been demonstrated to be well-suited for modeling long time sequences.[29] We demonstrate that LEMAS reduces the response time of a $Pd_{70}Au_{30}$ alloy plasmonic H_2 sensor by up to 40 times when exposed to a distinct H_2 pulse down to 0.06 vol % H_2 in an inert gas environment at atmospheric pressure, in a scenario simulating a sudden large leak. Furthermore, we illustrate the ability of LEMAS to rapidly discern and quantify slow gradual changes in H_2 concentration from mere noise in a simulated scenario of detecting a small leak in an enclosed inert gas environment. This ability is critical for detecting H_2 at as low concentrations as possible as quickly as possible, allowing sufficient time to apply safety measures, such as system shutdown, before a safety-critical H_2 concentration is reached. Finally, as an ensemble model, LEMAS enables one to obtain uncertainty estimates, which is of fundamental importance for safety-critical applications, including but not limited to H_2 sensing. While we focus here on the specific case of hydrogen sensing in an inert environment, we emphasize that LEMAS is broadly applicable to nanoplasmonic sensors in any environment. Our model makes no assumptions about the specific sensing environment or the nanoparticle composition; rather, it is trained on experimental data from a particular sensor particle type obtained in a specific environment as a demonstration. The approach itself, however, is fully adaptable to other sensor types, compositions, and sensing conditions.

RESULTS AND DISCUSSION

$Pd_{70}Au_{30}$ Alloy Plasmonic H_2 Sensors. As the plasmonic H_2 sensor platform of choice, we selected the well-established $Pd_{70}Au_{30}$ alloy system, which we have investigated in detail in

https://doi.org/10.1021/acssensors.4c02616
ACS Sens. XXXX, XXX, XXX−XXX

earlier works, where parameters such as limit of detection (LOD), response times, and sensitivity have been reported.[7,30−33] This material system is especially suited for inert gas sensing environments since its sensing mechanism, the interstitial sorption of hydrogen into the lattice of the metal host, does not require O_2 to be present. The Au alloyant serves the purpose of eliminating the intrinsic hysteresis characteristic for pure Pd by lowering the critical temperature of the system.[34−37] At 30% Au the best compromise between completely eliminating hysteresis, establishing a linear optical response to H_2 and maximizing optical contrast per unit sorbed H_2 is reached. Therefore, we nanofabricated quasi-random arrays of $Pd_{70}Au_{30}$ alloy nanodisks with a mean diameter of 210 and 25 nm height onto fused silica substrates using hole-mask colloidal lithography (Figure 1a,b), following the procedures described in detail in our earlier work[38] and in Sample fabrication in Methods.

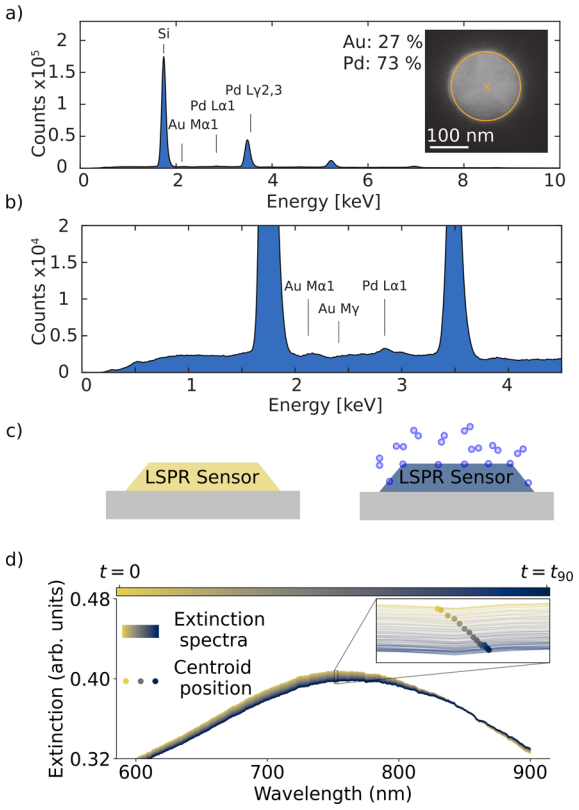


Figure 1. $Pd_{70}Au_{30}$ alloy nanoparticle plasmonic sensor characterization and operating principle. (a) Energy-dispersive X-ray (EDX) spectrum collected from a single $Pd_{70}Au_{30}$ alloy nanodisk in the quasi-random array of such disks that constitutes the active sensor surface. (b) Zoom-in of the EDX-spectrum in (a) up to 4.5 keV to focus on the characteristic Pd and Au peaks. (c) Schematic illustration of the plasmonic H_2 sensing principle, where the sorption of hydrogen into hydride-forming metal nanoparticles induces a change in their localized surface plasmon resonance frequency, which leads to a color change that is resolved in a spectroscopic measurement in the visible light spectral range. (d) Example of the spectral response of the $Pd_{70}Au_{30}$ alloy plasmonic sensor used in this work, resolved as a gradual shift in the extinction spectrum as hydrogen is absorbed the crystal lattice. Inset: Temporal evolution of the peak centroid position is one of the spectral descriptors that can be tracked to enable real time H_2 detection.

The working principle of plasmonic H_2 sensors is based on the localized surface plasmon resonance (LSPR) phenomenon, which is characteristic of metal nanoparticles irradiated by visible light. In an optical transmission, scattering or extinction spectrum, the LSPR manifests itself as a distinct peak with a maximum at a specific wavelength. The spectral position of this peak maximum, as well as related peak descriptors, such as width and intensity, exhibit a linear dependence on the H_2 partial pressure surrounding the particles and on the amount of hydrogen species absorbed into interstitial lattice sites of the Pd or Pd alloy host (Figure 1c,d).[32] Since the absorption and desorption of hydrogen into and from these interstitial lattice positions, respectively, occur spontaneously and reversibly at ambient conditions and also in oxygen-free environments, tracking of the spectral position (as well as other peak descriptors) of the LSPR peak as a function of H_2 partial pressure enables real-time H_2 detection (inset in Figure 1d). In this work, for what we refer to as the SA, we use the centroid position as a spectral descriptor, which we relate to the H_2 concentration by a calibration function (see Standard analysis in Methods and Note S3 for details).

Deep Learning Model Selection. We base our choice of an LSTR model for accelerating the plasmonic H_2 sensor response on several key characteristics of the output data generated by this type of sensor (see Figure 2 and Deep learning model in Methods for details about the architecture). The first important characteristic to take into account is that the measured extinction spectra that constitute the raw sensor response over time exhibit intrinsic noise (due to intensity fluctuations of the halogen light source and detection noise of the spectrometer used) that is comparable to the magnitude of changes induced in the spectra by small variations in H_2 concentration. Consequently, a crucial criterion for selecting the deep learning model is its ability to accurately model long temporal sequences, which the LSTR architecture is explicitly designed for.[29] This capability allows relevant temporal trends in the extinction spectrum to be differentiated from the inherent noise and used for predicting the hydrogen concentration.

The second critical aspect influencing the performance of the LSTR model, based on the characteristics of the sensor data, is the preprocessing of the measured extinction spectra. Such preprocessing is needed due to drift in the sensor response over time (mainly due to long-term variations of light source intensity), as well as small variations in the extinction spectra obtained in different measurements using the same sensor, due to slightly different placement of the sensor in the measurement chamber for each independent experiment. Here, we found that using several preprocessing methods is beneficial for the performance of the LSTR model. Therefore, we used four different preprocessing techniques (see Data preprocessing and Note S2 for details) and concatenated them into a single array.[39] As a result, the input data for the deep learning models was a time series, where each element in the sequence consisted of the concatenation of the different preprocessing techniques (see Figure 2a).

Another modeling choice that we make is to employ an *ensemble* of LSTR models. This choice is motivated by the safety-critical nature of the hydrogen sensor application and yields a more robust prediction, as well as a measure of uncertainty by aggregating the predictions of several LSTR models to compute the mean and the standard deviation (see Ensembles in Methods for details). Combining these modeling

https://doi.org/10.1021/acssensors.4c02616
ACS Sens. XXXX, XXX, XXX−XXX

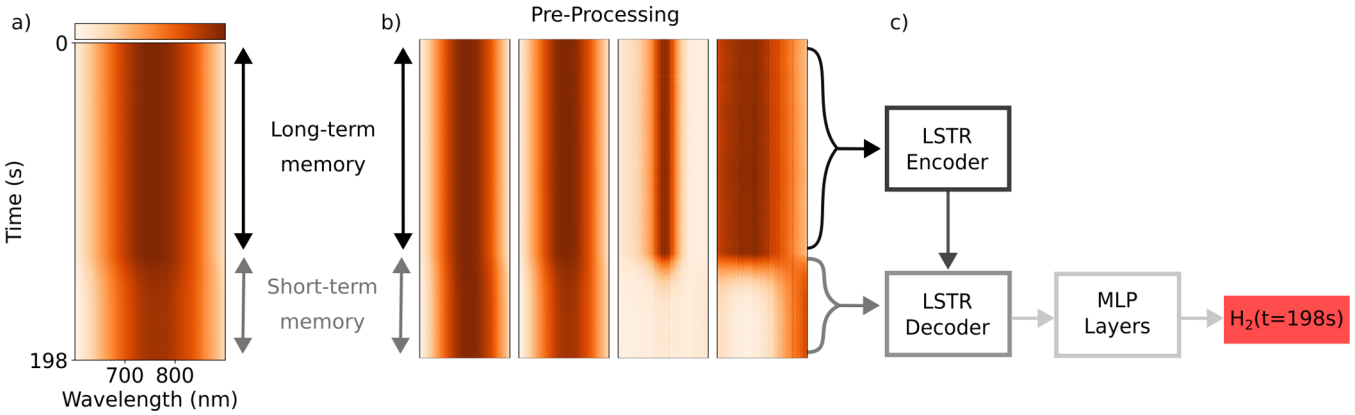


Figure 2. Long Short-term Transformer Ensemble Model for Accelerated Sensing (LEMAS). Illustration of the deep learning model used in this work based on the LSTR architecture (see Figure S16 for a more detailed illustration of the LSTR model). (a) The input data to the model consist of a time sequence of the past evolution of the spectral response of the sensor. In this figure the time sequence consists of 600 time steps corresponding to 198 s. (b) The time sequence is first split into long- and short-term memory and preprocessed using four different methods, including wavelength dependent min-max normalization, standard normal variate standardization, global min-max normalization, and level scaling, before the concatenation of the preprocessed data is being fed to the LSTR. (c) The LSTR first compresses the long-term memory to a fixed length latent representation in the LSTR encoder. Second, the LSTR decoder extracts relevant temporal features in the short term memory while also querying the compressed long-term memory. The extracted temporal features are then passed through a stack of multilayer perceptron (MLP) layers to obtain a prediction of the current H_2 concentration.

choices, we arrive at LEMAS, characterized by an ensemble of LSTR models that can both rapidly predict the H_2 concentration and provide a measure of uncertainty from a time series of preprocessed spectra.

LEMAS Model Training and Testing. Having introduced the architecture of the LEMAS model, we discuss the training and testing data used for optimizing the sensor response in (i) a large and fast leak scenario and (ii) a slow, gradual leak scenario. These data were generated by measuring tailored time series of optical extinction spectra of the sensor localized in a custom-made measurement chamber with a small volume to enable rapid gas exchange at atmospheric pressure to expose the sensor to varying H_2 concentrations in the simulated inert gas environment as used for encapsulation of hydrogen installations (see Hydrogen sensing experiments in Methods and Figure S1 for details). Specifically, we used three different H_2 profiles for generating the training data: (i) stepwise increase/decrease of H_2 from 0.00 vol % H_2 to 0.06−1.97 vol % H_2 in inert Ar environment (Figure S4), (ii) linear increase/decrease of H_2 from 0.06 vol % H_2 to 0.09−1.97 vol % H_2 in inert Ar environment (Figure S5a) and (iii) exponential increase/decrease of H_2 from 0.06 vol % H_2 to 0.09−1.97 vol % H_2 in inert Ar environment (Figure S5b), see Note S1 for details.

For the first case of a large simulated leak characterized by a rapid stepwise increase of H_2 concentration in the sensor surroundings, we trained LEMAS using two measurements of stepwise H_2 concentration increase/decrease and subsequently tested the trained LEMAS model on a third measurement not used for training. For the second case of a simulated small slow leak, we trained LEMAS on one measurement of linear H_2 concentration increase/decrease and tested the performance of the trained model on one measurement with exponential H_2 increase/decreases (see Deep learning training for details).

In this study, all measurements were performed using the same sensor. Consequently, applying either LEMAS or SA to a different sensor of the same type may require retraining, as slight variations in sensor responses can arise due to randomness in the fabrication process. However, in a practical application, for other quality control reasons, it would be imperative that the sensor chips used exhibit identical characteristics across different batches of fabrication. This can easily be achieved by state-of-the-art nanofabrication methods, such as deep-uv photolithography or nanoimprint lithography, which all have their origins in microelectronics, where the aspect of high reproducibility is absolutely critical.

The models trained for optimizing the sensor response in a large and fast leak scenario used a total input sequence length corresponding to the past 3 min, whereas the models trained for optimizing the sensor response in a slow gradual leak scenario used an input sequence length corresponding to the past 22 min of the sensor history. These choices were made based on an analysis of the change in centroid position in the training data and an estimation of the length of the time sequence needed to differentiate the slowest occurring process in the sensor output data from the noise in the measurement (see Figures S10 and S11 for details).

Accelerating Sensor Response to a Simulated Large Leak in Inert Ar Environment. To assess the ability of LEMAS to accelerate the response of a plasmonic H_2 sensor, we first consider a scenario where a 0.06% H_2 pulse in inert Ar gas is applied to our device at 30 °C (Figure 3a). For this analysis, we define the response time t_{90} as the first point in time where the sensor response has reached 90% of its new steady state value. Applying first the SA that predicts the hydrogen concentration using the instantaneous value of the centroid position reveals that it takes on the order of 85 s to reach t_{90} − reflecting the physical time it takes for the response of the system to saturate. Deploying the LEMAS analysis on the same data shows that it is able to predict the saturated H_2 level, using the *temporal* changes in the extinction spectrum, after only 3.6 s and thus long before the response of the system has saturated, leading to a more than 20-fold reduction of the response time. This result is corroborated when comparing t_{90} values obtained by SA and LEMAS across a range of H_2 pressure pulses from 0.06 vol % H_2 to 1.97 vol % H_2 (Figures 3b and S17). While LEMAS demonstrates a significant reduction in response time compared with the SA, it is

https://doi.org/10.1021/acssensors.4c02616
ACS Sens. XXXX, XXX, XXX−XXX

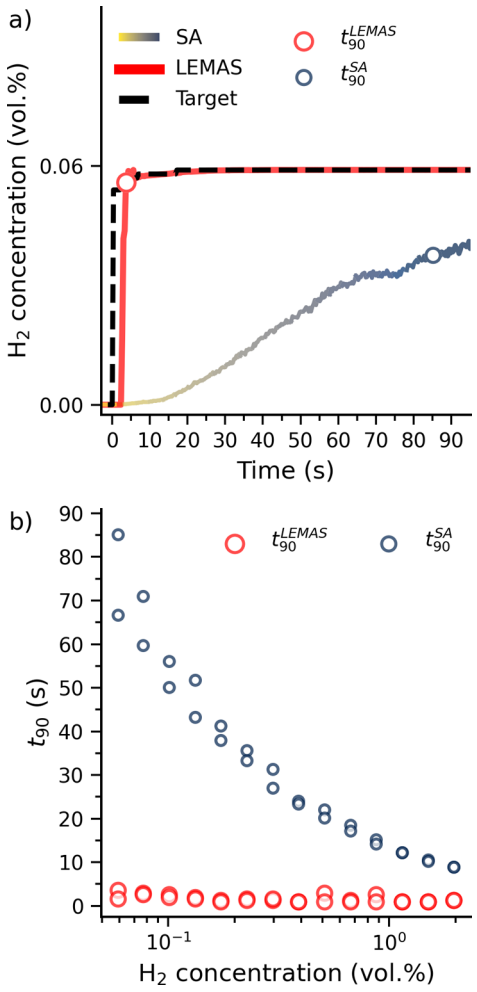


Figure 3. Accelerating sensor response to a simulated large leak in inert gas environment. (a) Comparison of the prediction of LEMAS and the centroid based SA for a pulse of 0.06 vol % H_2 in inert Ar environment. By utilizing the full time-dependent spectrum of the measured sensor response, LEMAS is able to accurately predict the final value of H_2 concentration before the sensor physically reaches its new state in equilibrium with the new H_2 concentration level. (b) Comparison of response times obtained by LEMAS and the centroid based SA as a function of H_2 concentration in inert Ar environment. Note the significant acceleration by LEMAS, in particular at the lowest H_2 concentrations, and the elimination of the concentration dependence of the response.

important to note that the SA approach is computationally more efficient. Fitting the SA model requires only a few seconds, whereas training LEMAS takes approximately 10 h. However, the primary time-consuming aspect for both approaches is not training the model itself but generating the training data; for example, a single measurement requires approximately 20 h (see Figures S4 and S5). Notably, a large number of such measurements would have to be carried out also when using the SA in a practical implementation of the sensor to generate the critically required sensor calibration curve.

Remarkably, LEMAS also achieves a response time that is practically independent of H_2 concentration whereas the t_{90} from the SA quickly increases with decreasing H_2 concentration. We attribute this behavior to the fact that LEMAS only requires a certain number of data points to identify the change in the extinction spectrum and make its prediction, the availability of which is dictated by the read-out frequency of the spectrometer rather than the H_2 pressure applied. By contrast, the SA is limited by the intrinsic kinetics of the material platform, causing a strong dependence on the H_2 pressure. As a result, LEMAS yields the largest boost in acceleration in the application critical range of lower H_2 pressures, overcoming one of the most important intrinsic limiting factors of hydride-based H_2 sensors.

Specifically, for the smallest concentrations considered in our experiment, at 0.1 vol % H_2 and below, the response times range between 1.6−3.6 s for the LEMAS analysis compared to 50−85 s for the SA. This corresponds to a 21- to 40-fold improvement compared to the SA. This improvement exceeds the one achieved by Lin et al., who used a CNN to accelerate a Pd nanocap plasmonic H_2 sensor, reducing the response time for hydrogen concentrations of 0.1 vol % H_2 and below to 6−14 s from 13−30 s.[25] At the same time, we also note that even the accelerated response obtained by LEMAS in the present inert gas conditions is slower than the state-of-the-art in vacuum/H_2 environment without acceleration.[7−9] As the main reasons, we identify the following points: (i) The traces of poisoning species, such as H_2O, CO, etc. present in the Ar inert gas used, significantly decelerate the sensor, as expected[7] (see Figures S2 and S3 for quantitative mass spectrometric analysis of the background molecular species present in the Ar inert gas used). (ii) We have used relatively large nanoparticles, and it is known that reducing size increases sensor speed due to reduced hydrogen diffusion path lengths.[7] (iii) We have not applied any polymer coatings, which are known to accelerate sensor response, as well as protect them from the poisoning molecular species present in the inert gas.[7,40] Importantly, we emphasize that the primary outcome here is not the *absolute* response time achieved by LEMAS but rather the level of *acceleration*. This distinction is important since, given the generic applicability of LEMAS, it suggests that similar acceleration can be achieved for sensors where the intrinsic physical response is faster than the sensor we use here.

Furthermore, we highlight that the amount of response time acceleration that LEMAS can produce depends not only on the obvious intrinsic response speed of the active sensor material (in our case, the PdAu alloy nanoparticles) but also on the sampling rate of the sensor hardware, where a higher sampling rate enables a larger degree of acceleration. For our experiments discussed so far, we have used a sampling frequency of 3 Hz, which was the highest rate enabled by the used spectrometer. Consequently, in this specific implementation, LEMAS has only three data points available to identify a change in the H_2 concentration in less than 1 s. Crucially, the acceleration observed in the present case is thus not limited by LEMAS but by the underlying materials and read-out of the used light sampling device. In other words, it has the potential to be significantly improved by using a faster spectrometric device.

Having established the overall LEMAS concept and demonstrated its ability to substantially accelerate sensor speed in inert sensing environments, in particular, in the low concentration regime, it is interesting to evaluate the performance of LEMAS in more detail. To do so, we select three different H_2 concentration pulses, i.e., pulses to 0.06, 0.08, and 0.10 vol % H_2, and plot the sensor response predicted by LEMAS as a function of time, with the standard deviation of the prediction at each time point indicated in the corresponding graphs (Figures 4a and S18 for all pulses). We

https://doi.org/10.1021/acssensors.4c02616
ACS Sens. XXXX, XXX, XXX−XXX

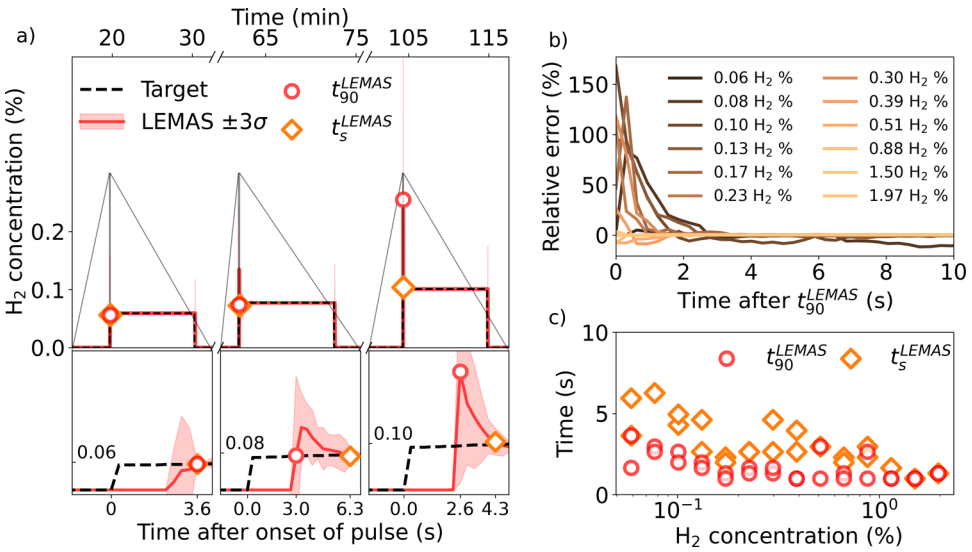


Figure 4. LEMAS prediction accuracy to H_2 pulses of different concentration. (a) The prediction and standard deviation from LEMAS for three selected H_2 concentration pulses in the test set. The lower panels display zoom-ins on the initial response to the pulse. (b) The relative error of the LEMAS predictions for the entire range of H_2 concentration pulses, starting at the response time, t_{90}^{LEMAS}, and forward. (c) LEMAS settling and response times as a function of the H_2 concentration.

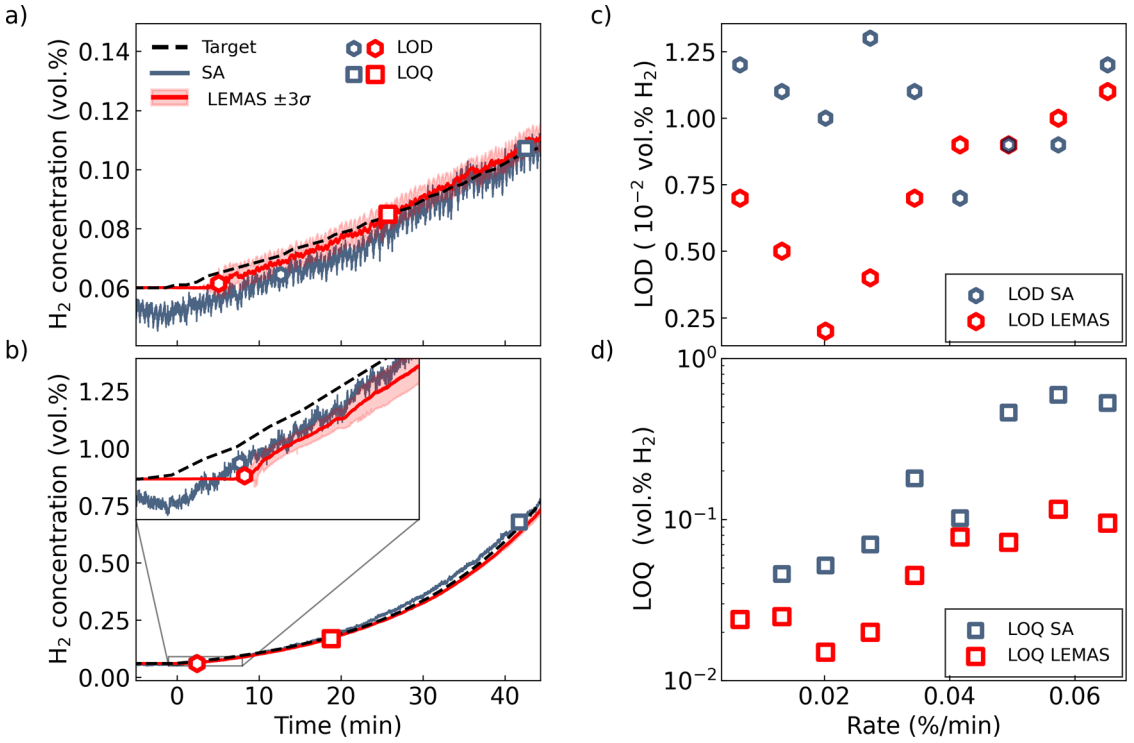


Figure 5. Small leak detection and quantification. (a,b) Time evolution of the H_2 concentration obtained using the centroid based SA and LEMAS, respectively, for exponentially increasing H_2 concentration in inert Ar environment for leak rates of (a) 1.32×10^{-3} vol % H_2 min^{-1} and (b) 5.73×10^{-2} vol % H_2 min^{-1}. The correspondingly obtained limit of detection (LOD) and limit of quantification (LOQ) are also indicated. (c,d) LOD and LOQ as a function of leak rate. Note that at the smallest leak rate, the LOQ is not reached for the centroid based SA.

also define the sensor settling time, t_s^{LEMAS}, as the first time point where the predicted H_2 concentration lies within ±10% of the target H_2 concentration and the relative standard deviation is smaller than 10%. This metric complements the response time by also considering cases where LEMAS either underestimates or overestimates the H_2 concentration after t_{90}.

First, we note that at the onset of each pulse, there is a brief interval where LEMAS predicts 0 vol % H_2, while the actual H_2 concentration has already increased. This behavior occurs since the change in the extinction spectrum induced by the presence of H_2 is not yet distinguishable from the noise level in the measurement. This initial phase is followed by an interval where a clear change in the extinction spectrum is detected but where both error (Figure 4b) and uncertainty are still rather large (red-shaded areas in Figure 4a). In the final phase, the LEMAS-predicted H_2 concentration settles at the correct value

https://doi.org/10.1021/acssensors.4c02616
ACS Sens. XXXX, XXX, XXX−XXX

once the change in the extinction spectrum is sufficiently distinct, such that all models in the ensemble predict a similar H_2 concentration, and the uncertainty becomes very small.

Finally, we note that for some pulses (illustrated by 0.08 and 0.10 vol % H_2 in Figure 4a) the predicted H_2 concentration is overestimated for a brief interval past t_{90}^{LEMAS}, before the mean prediction settles around the target value. Consequently, t_s^{LEMAS} is larger than t_{90}^{LEMAS}. Conversely, in other cases (illustrated by the pulse to 0.06 vol % H_2) the H_2 concentration is underestimated until t_{90}^{LEMAS}, at which point the uncertainty has also been reduced. As a result, t_s^{LEMAS} equals t_{90}^{LEMAS} in this (and similar) cases. To further examine the overestimations we analyze the relative error of the predicted response, starting from t_{90}^{LEMAS} (Figure 4b). Overall the relative error tends to be larger for lower H_2 concentration pulses since there is a transient overestimation in the LEMAS prediction. Consequently, t_s^{LEMAS} is generally larger for lower H_2 concentrations (Figure 4c). This is likely the consequence of the early predictions being more affected by measurement noise since lower H_2 concentrations are associated with slower absorption kinetics and smaller changes in the extinction spectrum. An important implication of these initial overestimations is that the accuracy of the sensor's initial response can be compromised if one relies on a single model, providing further evidence for the benefit of using an ensemble model, as we do with LEMAS.

Improving Sensor Response to a Simulated Small and Slow Leak in an Inert Ar Environment. In a practical application in an inert gas environment, H_2 sensors are not only required for the rapid detection of large leaks with fast and essentially instantaneous increases of H_2 concentration but also in scenarios where a small leak will lead to a slow increase in H_2 concentration in an enclosed environment over time. Technically, this translates into the challenge of being able to discern as quickly as possible a tiny sensor signal from noise. To address this scenario in the LEMAS framework, we define the LOD of a sensor as the minimal amount of H_2 required for the mean H_2 prediction to change by more than three times the standard deviation of the H_2 prediction at a baseline where the H_2 concentration is kept constant. In other words, the smallest H_2 required to discern (but not quantify) the presence of hydrogen gas with a confidence of 3σ. Furthermore, we define the limit of quantification (LOQ) as the minimal amount of H_2 for which the mean relative error of the H_2 prediction is less than 5%. The mean absolute relative error $\frac{1}{N}\sum_{t=1}^{N}\left|\frac{P_t - T_t}{T_t}\right|$ is calculated over a time window of 2 min, where N is the total number of time steps across the window, P_t is the predicted H_2 value for time step t and T_t is the true H_2 value for time step t.

We then assess the sensor response to a first scenario with a very small exponential leak rate of 1.32×10^{-3} vol % H_2 min^{-1} using both the centroid-shift-based SA and LEMAS (Figure 5a). This analysis reveals that the LEMAS-predicted H_2 concentrations contain considerably less noise compared with the SA analysis. In the SA the noise in the H_2 signal is initially comparable to the change in the H_2 concentration. As a result, the LOD is reached faster, i.e., at lower H_2 concentrations, for LEMAS due to its ability to discern changes in the H_2 concentration. We also note that the LEMAS analysis generally delivers a more accurate response, where the predicted H_2 concentration values are closer to the target values, resulting in the LOD being reached earlier for LEMAS.

We also perform a similar analysis of a second scenario with a higher exponential leak of rate 5.73×10^{-2} vol % H_2 min^{-1} (Figure 5b; see also Figure S19 for the SA and LEMAS-predicted H_2 concentrations for all exponential leak rates). Consistent with the previous analysis, LEMAS demonstrates a significantly lower LOQ, which can be attributed to its overall higher accuracy. However, despite LEMAS exhibiting less noise than the SA the LOD of the two approaches is very similar. This can be understood by noting that, in contrast to the previous case, here, the fluctuations in the prediction of the SA analysis are much smaller than the change in H_2 due to the higher exponential leak rate.

These results are further corroborated when comparing the LOD and LOQ obtained by the SA and LEMAS methods, respectively, across a range of exponential leak rates from 6.52×10^{-3} vol % H_2 min^{-1} to 6.52×10^{-2} vol % H_2 min^{-1} (Figure 5c,d). From Figure 5c we identify that for exponential leak rates at 3.44×10^{-2} vol % H_2 min^{-1} and below LEMAS has a significantly lower LOD than SA. At larger exponential leak rates, the LOD for LEMAS and SA becomes approximately equal. This occurs because, at larger rates, the change in H_2 is sufficiently large such that the initial change in the sensor signal is much larger than the intrinsic noise. Consequently, the ability of LEMAS to discern small signals from noise does not significantly contribute to decreasing the LOD. At lower leak rates, however, LEMAS indeed makes it possible to extract a discernible signal earlier, at lower leaked concentrations, thereby significantly increasing the time window from triggered sensor response to the leak has reached the flammability limit of 4 vol % H_2. Finally, in Figure 5d, we see that LEMAS has a lower LOD for all investigated leak rates, which is a consequence of the higher accuracy obtained through LEMAS (see Note S3). In summary, these results underscore, on one hand, the effectiveness of LEMAS in detecting small and slow leaks earlier, as it consistently achieves a lower LOD than the SA at small rates. On the other hand, they demonstrate that LEMAS consistently outperforms the SA in terms of leak quantification, as reflected in its lower LOQ across all rates.

DISCUSSION AND CONCLUSIONS

In this work, we leveraged plasmonic H_2 sensors with deep learning to address the crucial challenge of faster H_2 detection under technically relevant conditions. We have focused so far on the relatively unexplored yet important application area of inert gas environments intended to enclose large H_2 installations to avoid the formation of flammable air-H_2 mixtures. For such applications, hydride-forming plasmonic sensors are ideal, as they do not require molecular oxygen for their operation, unlike the more common catalytic or thermal H_2 sensors.

From the deep learning perspective, we have developed LEMAS, short for long short-term transformer ensemble model for accelerated sensing, which accelerates sensor response by learning the relationship between the time dependence of the spectral response of the plasmonic sensor and the H_2 concentration. This allows predicting the sensor's final response before it is reached physically, while also evaluating uncertainty in the prediction via model ensembles. To obtain accurate models for the ensembles and mitigate

https://doi.org/10.1021/acssensors.4c02616
ACS Sens. XXXX, XXX, XXX−XXX

artifacts from measurement noise, drift, and variations between different measurements, we found it crucial to use a sufficiently long time series and combine the results of several different preprocessing approaches.

Since LEMAS makes no assumptions about the specific environment or sensor type, the approach developed here can be readily adapted for other sensor types and operating conditions. However, for environments containing higher concentrations of "poisoning" molecular species, appropriate considerations must be made. This requires designing a sensor with deactivation-resistant alloys and/or polymer filters to mitigate the deactivation of the sensor by such molecules. In cases where such solutions are insufficient, we have previously demonstrated how a different transformer-based machine learning approach enables the accurate and stable detection of hydrogen in a highly deactivating environment, i.e., high humidity.[14] Therefore, we believe that LEMAS could also be applied successfully in such conditions with minor architectural adjustments. The key to reliable application in these scenarios lies in ensuring that the training data for LEMAS are generated under controlled conditions, with systematic variations in the concentrations of "poisonous" molecular species.

Overall, our results demonstrate the significant potential of deep learning to address current H_2 sensor limitations, such as slow response times and challenges in quantifying H_2 leaks. This is important from a practical application perspective since it provides a longer time window for the implementation of appropriate measures for handling the leak.

METHODS

Hydrogen Sensing Experiments. The measurements were conducted in a custom-built reactor chamber that is composed of a customized DN 16 CF spacer flange (Pfeifer Vacuum), equipped with a gas in- and outlet, and two fused-silica viewports (1.33 in. CF Flage, Accu-Glass). The effective chamber volume is ca. 1.5 mL. The gas flow rates were controlled by mass flow controllers (El-Flow Select series, Bronkhorst High-Tech) (Figure S1). The sample inside the chamber was illuminated by using an unpolarized halogen white light source (AvaLight-HAL, Avantes) and an optical fiber equipped with a collimating lens. The transmitted light was collected and analyzed by using a fiber-coupled fixed-grating spectrometer (SensLine AvaSpec-HS1024TEC, Avantes). The temperature was controlled with a heating coil wrapped around the chamber and a temperature controller (Eurotherm 3216) in a feedback loop manner, where the sample surface temperature inside the chamber was continuously used as the input.

All measurements were performed at 30 °C in an argon background, with a constant gas flow of 300 mL/min. The hydrogen concentration in all of the following measurements was in the range of 0.06 vol % H_2 − 1.97 vol % H_2 (detailed description of the different pulse schemes as found in Note S1). The sampling frequency of the spectrometer was set to 3 Hz.

Standard Analysis. In this work, we used the centroid position as a spectral descriptor. The centroid position is defined as $\lambda_c = \sum_\lambda \lambda I(\lambda)/\sum_\lambda I(\lambda)$, where λ is the wavelength in nm and $I(\lambda)$ is the intensity at wavelength λ. To enable comparison between the SA and LEMAS on the test measurements, we fit a calibration function, using the measured H_2 concentration in the training measurements, as

$$H_2(\Delta\lambda_c) = a\Delta\lambda_c^b \quad (1)$$

where $\Delta\lambda_c$ is the change in centroid position, taken from the smallest centroid position in each measurement. The values of parameters a and b are determined by minimizing the mean absolute percentage error between the measured H_2 and H_2 (λ_c). We fit two different calibration functions, one for the data consisting of stepwise increase/decrease of H_2 and one for the data consisting of linear/exponential increase/decrease of H_2 (see Note S3 for details).

Data Preprocessing. The spectra recorded in different measurements varied in total intensity due to changes in the light source but also due to slightly different positioning of the sensor chip inside the chamber which causes different particles to be probed at a given measurement (see Figure S14). To address this, we preprocessed the data before it was fed into the deep learning model using four different methods, and the concatenation of these methods was fed as input to the deep learning model. This to ensure that spectra corresponding to the same hydrogen concentration were similar across all measurements. Each measurement was preprocessed individually by using the initial sequence of 5 pulses of 1.97 vol % H_2, for each measurement, to estimate the minimum/maximum/and mean intensity. The preprocessing methods were (i) wavelength-dependent min-max normalization: for each wavelength subtracting the estimated minimum intensity at the corresponding wavelength and dividing by the difference between the estimated maximum intensity of all wavelengths and the estimated minimum intensity measured at the specific wavelength, (ii) standard normal variate standardization: scaling each spectrum using its mean and standard deviation, (iii) global min-max normalization: subtracting the estimated minimum intensity and dividing by the difference between the estimated maximum and minimum intensity in, and (iv) level scaling: subtracting and dividing each wavelength in each spectrum in the measurement by the estimated mean intensity of each wavelength (see Note S2 for details).

Deep Learning Model. The deep learning architecture that was used in this work was a long short-term transformer (LSTR)[29] which operates as illustrated in Figure S16. First, each temporal feature consisting of the concatenation of the preprocessed spectrum is linearly mapped to a vector of size $d_{model} = 256$. Subsequently, positional encoding is added and the data is split into short-term memory and long-term memory. Here, we down-sample the long-term memory using a stride of 4. To process the time series, the long-term memory first undergoes a two-stage memory compression through the LSTR encoder, using a set of learnable token embeddings of dimensions $d_{model} \times n_1$ and $d_{model} \times n_0$. Here, we used $n_0 = 8$ and $n_1 = 4$ and the encoder consisted of 4 transformer decoder units.[41] Second, the LSTR decoder extracts relevant temporal features in the short-term memory while also querying the encoded long-term memory to retrieve useful information from the history of the sensor. Here, the decoder consisted of 8 transformer decoder units. The extracted temporal features are then passed through $n_{MLP} = 8$ MLP layers of dimension $d_{MLP} = 512$ to obtain H_2 concentration predictions. Here all the transformer decoder units performed multi-head attention with $h = 8$ heads and $d_k = d_q = d_v = d_{model}/h = 32$, and the dimension of the MLP inside the transformer decoder units was $d_{ff} = 512$. Furthermore, in the LSTR decoder, masked multihead attention was performed such that during training the H_2 concentration corresponding to each time step in the short-term memory could be used for supervision during training. We used dropout with a value of 0.1 in all layers, except the last n_{MLP} MLP layers. These hyperparameters were chosen based on those in the original paper,[29] with adjustments including additional MLP layers and an increased number of transformer layers in the decoder, which we found beneficial for reducing the noise in the LSTR's predictions and a smaller value of d_{model} and h to reduce computational complexity.

Ensembles. The constructed ensembles comprised ten models, each varying in the lengths of short-term and long-term memory. This variation was designed to induce diversity in the predictive capabilities of the models in the ensemble. Specifically, two models were designated for each combination of long- and short-term memory lengths, while the total length of the input sequence was the same for all models. For the ensemble tailored for leak detection, the length of the time series was 4000 time steps, while the ensemble model aimed at minimizing the response time used an input sequence of 600-time steps. As we mentioned in LEMAS model training and testing, these choices were based on estimating the time sequence length needed to

https://doi.org/10.1021/acssensors.4c02616
ACS Sens. XXXX, XXX, XXX−XXX

distinguish the slowest process in the sensor data from measurement noise (see Figures S10 and S11 for details). For both ensembles, the selected lengths for short-term memory were 20, 40, 60, 80, and 100 time steps, respectively.

The values of the short-term memory length were selected based on the idea that the short-term memory should be long enough such that the LSTR decoder can analyze the most recent temporal trends but short enough such that we still utilize the LSTR encoder's ability to summarize important long-term history. The choice of the range for these values is based on our understanding of the kinetics of the underlying physical response of the sensor device studied here. The variation in the length of the memories between the models in the ensembles allows each model to capture slightly different patterns and aspects of the data. The length of the short-term memory controls how much of the recent portion the model can access directly, while the LSTR decoder summarizes the rest of the time series, i.e., the long-term memory. This setup should intuitively allow models with different short-term memory lengths to respond slightly differently to recent changes in the data. By aggregating the predictions across all models, using the mean and standard deviation to compute a prediction and measure of uncertainty, we limit the potential that individual model limitations, such as oversensitivity to noise due to the specific choice of short-term memory, cause inaccurate prediction. Instead, since we average over predictions from models with varied memory configurations, the ensemble is less likely to be biased by the limitations of any single model. Furthermore, the uncertainty estimate obtained from the ensemble offers insight into the prediction's reliability.

It is important to note that apart from the variation in memory lengths, all models shared identical hyperparameters (see Table S1). To make the prediction of the ensemble more robust to potential outliers, we included only predictions that fall between the first and third percentile to compute the ensemble prediction and uncertainty.

Deep Learning Training. The models were implemented using TensorFlow[42] and were trained for 100 epochs on Nvidia A100 graphical processing units using the AdamW[43] optimizer with weight decay 5×10^{-5}, a batch size of 128, and mean-absolute-error loss. The learning rate was increased linearly from zero to 5×10^{-5} during the first 15 epochs and then decayed to zero following a cosine curve. To analyze the impact of different preprocessing methods we used the first half of the data from measurement Figure S4a,c as training data and the other half as validation data (see Note S2).

For the first case of a large simulated leak characterized by a rapid stepwise increase of H_2 concentration in the sensor surroundings, we trained LEMAS using the data from measurement Figure S4a,c and used the data from measurement Figure S4b to test the model. For the second case of a simulated small slow leak, we trained LEMAS on the data from measurement Figure S5a and used the data from measurement Figure S5b to test the model. Furthermore, during the training phase, each model in the ensemble was exposed to a distinct subset of the training data, comprising a random 90% of the total data set.

Sample Fabrication. Quasi-random PdAu alloy (nominal 70:30 at. %) nanodisk arrays with 210 nm average disk diameter and 25 nm height were fabricated using hole-mask colloidal lithography (HCL).[44] The metals were deposited layer-by-layer via electron beam evaporation onto 1 cm × 1 cm fused silica substrates (Siegert Wafer GmbH). Subsequent annealing was performed at 500 °C for 18 h under a flow of 4 vol % H_2 in Ar to induce alloy formation. A more detailed description of the nanofabrication procedure can be found in our earlier work.[45]

■ ASSOCIATED CONTENT

Data Availability Statement

Data is available from the authors by reasonable request.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssensors.4c02616.

Experimental details: hydrogenation protocols, quadrupole mass spectrometer measurements, and experimental setup; method details: preprocessing methods, standard analysis information, details on the LSTR architecture and hyperparameters; and results: LEMAS predictions for all stepwise and exponential increases (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Paul Erhart − *Department of Physics, Chalmers University of Technology, SE-41296 Göteborg, Sweden;* orcid.org/0000-0002-2516-6061; Email: erhart@chalmers.se

Christoph Langhammer − *Department of Physics, Chalmers University of Technology, SE-41296 Göteborg, Sweden;* orcid.org/0000-0003-2180-1379; Email: clangham@chalmers.se

Authors

Viktor Martvall − *Department of Physics, Chalmers University of Technology, SE-41296 Göteborg, Sweden*

Henrik Klein Moberg − *Department of Physics, Chalmers University of Technology, SE-41296 Göteborg, Sweden*

Athanasios Theodoridis − *Department of Physics, Chalmers University of Technology, SE-41296 Göteborg, Sweden;* orcid.org/0000-0002-4170-4325

David Tomeček − *Department of Physics, Chalmers University of Technology, SE-41296 Göteborg, Sweden*

Pernilla Ekborg-Tanner − *Department of Physics, Chalmers University of Technology, SE-41296 Göteborg, Sweden;* orcid.org/0000-0002-9427-4816

Sara Nilsson − *Department of Physics, Chalmers University of Technology, SE-41296 Göteborg, Sweden*

Giovanni Volpe − *Department of Physics, University of Gothenburg, SE-412 96 Göteborg, Sweden;* orcid.org/0000-0001-5057-1846

Complete contact information is available at: https://pubs.acs.org/10.1021/acssensors.4c02616

Notes

The authors declare the following competing financial interest(s): C. L. is co-founder and scientific advisor at Insplorion AB who markets plasmonic hydrogen sensors. D.T. was postdoc in C.L. group when contributing to this work and is now employed at Insplorion AB.

■ ACKNOWLEDGMENTS

This work was funded by the Vinnova project 2021-02760, the Swedish Research Council (grant numbers 2018-06482, 2020-04935, 2021-05072), the Swedish Energy Agency (grant No. 45410-1), the Area of Advanced Nano at Chalmers, and the Competence Centre TechForH2. The Competence Centre TechForH2 is hosted by Chalmers University of Technology and is financially supported by the Swedish Energy Agency (P2021-90268) and the member companies Volvo, Scania, Siemens Energy, GKN Aerospace, PowerCell, Oxeon, RISE, Stena Rederier AB, Johnsson Matthey, and Insplorion. The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at C3SE partially funded by the Swedish Research Council through grant agreement no. 2022-06725. This work was performed in part at Myfab Chalmers and the Chalmers Materials Analysis Laboratory (CMAL).

https://doi.org/10.1021/acssensors.4c02616
ACS Sens. XXXX, XXX, XXX−XXX

REFERENCES

(1) Zhang, M.; Yuan, Z.; Song, J.; Zheng, C. Improvement and mechanism for the fast response of a Pt/TiO2 gas sensor. *Sens. Actuators, B* **2010**, *148*, 87−92.

(2) Ribet, F.; Stemme, G.; Roxhed, N. Real-time intradermal continuous glucose monitoring using a minimally invasive microneedle-based system. *Biomed. Microdevices* **2018**, *20*, 101.

(3) Rolinger, L.; Rüdt, M.; Hubbuch, J. A Critical Review of Recent Trends, and a Future Perspective of Optical Spectroscopy as PAT in Biopharmaceutical Downstream Processing. *Anal. Bioanal. Chem.* **2020**, *412*, 2047−2064.

(4) Poghossian, A.; Geissler, H.; Schöning, M. J. Rapid methods and sensors for milk quality monitoring and spoilage detection. *Biosens. Bioelectron.* **2019**, *140*, 111272.

(5) Darmadi, I.; Nugroho, F. A. A.; Langhammer, C. High-Performance Nanostructured Palladium-Based Hydrogen Sensors−Current Limitations and Strategies for Their Mitigation. *ACS Sensors* **2020**, *5*, 3306−3327.

(6) U.S. Department of Energy *2015 Safety, Codes and Standards Section*, 2015. https://www.energy.gov/sites/default/files/2015/06/f23/fcto_myrdd_safety_codes.pdf.

(7) Nugroho, F. A.; Darmadi, I.; Cusinato, L.; Susarrey-Arce, A.; Schreuders, H.; Bannenberg, L. J.; da Silva Fanta, A. B.; Kadkhodazadeh, S.; Wagner, J. B.; Antosiewicz, T. J.; et al. Metal−polymer hybrid nanomaterials for plasmonic ultrafast hydrogen detection. *Nat. Mater.* **2019**, *18*, 489−495.

(8) Luong, H. M.; Pham, M. T.; Guin, T.; Madhogaria, R. P.; Phan, M.-H.; Larsen, G. K.; Nguyen, T. D. Sub-second and ppm-level optical sensing of hydrogen using templated control of nano-hydride geometry and composition. *Nat. Commun.* **2021**, *12*, 2414.

(9) Bannenberg, L.; Schreuders, H.; Dam, B. Tantalum-Palladium: Hysteresis-Free Optical Hydrogen Sensor Over 7 Orders of Magnitude in Pressure with Sub-Second Response. *Adv. Funct. Mater.* **2021**, *31*, No. 2010483.

(10) Darmadi, I.; Stolaś, A.; Östergren, I.; Berke, B.; Nugroho, F. A. A.; Minelli, M.; Lerch, S.; Tanyeli, I.; Lund, A.; Andersson, O.; Zhdanov, V. P.; Liebi, M.; Moth-Poulsen, K.; Müller, C.; Langhammer, C. Bulk-Processed Pd Nanocube−Poly(methyl methacrylate) Nanocomposites as Plasmonic Plastics for Hydrogen Sensing. *ACS Applied Nano Materials* **2020**, *3*, 8438−8445.

(11) Östergren, I.; Darmadi, I.; Lerch, S.; da Silva, R. R.; Craighero, M.; Paleti, S. H. K.; Moth-Poulsen, K.; Langhammer, C.; Müller, C. A surface passivated fluorinated polymer nanocomposite for carbon monoxide resistant plasmonic hydrogen sensing. *J. Mater. Chem. A* **2024**, *12*, 7906−7915.

(12) Darmadi, I.; Nugroho, F. A. A.; Kadkhodazadeh, S.; Wagner, J. B.; Langhammer, C. Rationally Designed PdAuCu Ternary Alloy Nanoparticles for Intrinsically Deactivation-Resistant Ultrafast Plasmonic Hydrogen Sensing. *ACS Sensors* **2019**, *4*, 1424−1432.

(13) Darmadi, I.; Khairunnisa, S. Z.; Tomecek, D.; Langhammer, C. Optimization of the composition of PdAuCu ternary alloy nanoparticles for plasmonic hydrogen sensing. *ACS Applied Nano Materials* **2021**, *4*, 8716−8722.

(14) Tomeček, D.; Moberg, H. K.; Nilsson, S.; Theodoridis, A.; Darmadi, I.; Midtvedt, D.; Volpe, G.; Andersson, O.; Langhammer, C. Neural network enabled nanoplasmonic hydrogen sensors with 100 ppm limit of detection in humid air. *Nat. Commun.* **2024**, *15*, 1208.

(15) Alenezy, E. K.; Sabri, Y. M.; Kandjani, A. E.; Korcoban, D.; Abdul Haroon Rashid, S. S. A.; Ippolito, S. J.; Bhargava, S. K. Low-Temperature Hydrogen Sensor: Enhanced Performance Enabled through Photoactive Pd-Decorated TiO2 Colloidal Crystals. *ACS Sensors* **2020**, *5*, 3902−3914.

(16) Kabcum, S.; Channei, D.; Tuantranont, A.; Wisitsoraat, A.; Liewhiran, C.; Phanichphant, S. Ultra-responsive hydrogen gas sensors based on PdO nanoparticle-decorated WO3 nanorods synthesized by precipitation and impregnation methods. *Sens. Actuators, B* **2016**, *226*, 76−89.

(17) Hassan, K.; Iftekhar Uddin, A.; Chung, G.-S. Fast-response hydrogen sensors based on discrete Pt/Pd bimetallic ultra-thin films. *Sens. Actuators, B* **2016**, *234*, 435−445.

(18) Losurdo, M.; Gutiérrez, Y.; Suvorova, A.; Giangregorio, M. M.; Rubanov, S.; Brown, A. S.; Moreno, F. Gallium plasmonic nanoantennas unveiling multiple kinetics of hydrogen sensing, storage, and spillover. *Adv. Mater.* **2021**, *33*, No. 2100500.

(19) Huang, S.; Croy, A.; Ibarlucea, B.; Cuniberti, G. *Machine Learning for Advanced Functional Materials*; Joshi, N.; Kushvaha, V.; Madhushri, P., Eds.; Singapore: Springer Nature Singapore, 2023; pp 21−41.

(20) Hayasaka, T.; Lin, A.; Copa, V. C.; Lopez, L. P., Jr; Loberternos, R. A.; Ballesteros, L. I. M.; Kubota, Y.; Liu, Y.; Salvador, A. A.; Lin, L. An electronic nose using a single graphene FET and machine learning for water, methanol, and ethanol. *Microsyst. Nanoeng.* **2020**, *6*, 50.

(21) Wijaya, D. R.; Afianti, F.; Arifianto, A.; Rahmawati, D.; Kodogiannis, V. S. Ensemble machine learning approach for electronic nose signal processing. *Sensing and Bio-Sensing Research* **2022**, *36*, No. 100495.

(22) Thai, N. X.; Tonezzer, M.; Masera, L.; Nguyen, H.; Duy, N. V.; Hoa, N. D. Multi gas sensors using one nanomaterial, temperature gradient, and machine learning algorithms for discrimination of gases and their concentration. *Anal. Chim. Acta* **2020**, *1124*, 85−93.

(23) Kanaparthi, S.; Singh, S. G. Discrimination of gases with a single chemiresistive multi-gas sensor using temperature sweeping and machine learning. *Sens. Actuators, B* **2021**, *348*, No. 130725.

(24) Kang, M.; Cho, I.; Park, J.; Jeong, J.; Lee, K.; Lee, B.; Del Orbe Henriquez, D.; Yoon, K.; Park, I. High Accuracy Real-Time Multi-Gas Identification by a Batch-Uniform Gas Sensor Array and Deep Learning Algorithm. *ACS Sensors* **2022**, *7*, 430−440.

(25) Lin, X.; Cheng, M.; Chen, X.; Zhang, J.; Zhao, Y.; Ai, B. Unlocking Predictive Capability and Enhancing Sensing Performances of Plasmonic Hydrogen Sensors via Phase Space Reconstruction and Convolutional Neural Networks. *ACS sensors* **2024**, *9*, 3877−3888.

(26) Ai, B.; Sun, Y.; Zhao, Y. Plasmonic Hydrogen Sensors. *Small* **2022**, *18*, No. 2107882.

(27) Wadell, C.; Syrenova, S.; Langhammer, C. Plasmonic Hydrogen Sensing with Nanostructured Metal Hydrides. *ACS Nano* **2014**, *8*, 11925−11940.

(28) Dahlin, A. B.; Tegenfeldt, J. O.; Höök, F. Improving the Instrumental Resolution of Sensors Based on Localized Surface Plasmon Resonance. *Anal. Chem.* **2006**, *78*, 4416−4423.

(29) Xu, M.; Xiong, Y.; Chen, H.; Li, X.; Xia, W.; Tu, Z.; Soatto, S. Long Short-Term Transformer for online action detection. *arXiv* **2021**.

(30) Wadell, C.; Nugroho, F. A. A.; Lidström, E.; Iandolo, B.; Wagner, J. B.; Langhammer, C. Hysteresis-Free Nanoplasmonic Pd−Au Alloy Hydrogen Sensors. *Nano Lett.* **2015**, *15*, 3563−3570.

(31) Bannenberg, L. J.; Nugroho, F. A. A.; Schreuders, H.; Norder, B.; Trinh, T. T.; Steinke, N.-J.; van Well, A. A.; Langhammer, C.; Dam, B. Direct Comparison of PdAu Alloy Thin Films and Nanoparticles upon Hydrogen Exposure. *ACS Appl. Mater. Interfaces* **2019**, *11*, 15489−15497.

(32) Nugroho, F. A. A.; Darmadi, I.; Zhdanov, V. P.; Langhammer, C. Universal Scaling and Design Rules of Hydrogen-Induced Optical Properties in Pd and Pd-Alloy Nanoparticles. *ACS Nano* **2018**, *12*, 9903−9912.

(33) Ekborg-Tanner, P.; Rahm, J. M.; Rosendal, V.; Bancerek, M.; Rossi, T. P.; Antosiewicz, T. J.; Erhart, P. Computational Design of Alloy Nanostructures for Optical Sensing of Hydrogen. *ACS Applied Nano Materials* **2022**, *5*, 10225.

(34) Lee, S.-M.; Noh, H.; Flanagan, T. B.; Luo, S. Hydrogen-induced lattice rearrangement of a $Pd_{0.81}Au_{0.19}$ alloy. *J. Phys.: Condens. Matter* **2007**, *19*, 326222.

(35) Luo, S.; Wang, D.; Flanagan, T. B. Thermodynamics of Hydrogen in fcc Pd-Au Alloys. *J. Phys. Chem. B* **2010**, *114*, 6117−6125.

https://doi.org/10.1021/acssensors.4c02616
ACS Sens. XXXX, XXX, XXX−XXX

(36) Mamatkulov, M.; Zhdanov, V. P. Suppression of hysteresis in absorption of hydrogen by a Pd−Au alloy. *Phys. Rev. E* **2020**, *101*, No. 042130.
(37) Rahm, J. M.; Löfgren, J.; Fransson, E.; Erhart, P. A tale of two phase diagrams: Interplay of ordering and hydrogen uptake in Pd-Au-H. *Acta Mater.* **2021**, *211*, No. 116893.
(38) Nugroho, F. A. A.; Iandolo, B.; Wagner, J. B.; Langhammer, C. Bottom-Up Nanofabrication of Supported Noble Metal Alloy Nanoparticle Arrays for Plasmonics. *ACS Nano* **2016**, *10*, 2871−2879.
(39) Mishra, P.; Passos, D.; Marini, F.; Xu, J.; Amigo, J. M.; Gowen, A. A.; Jansen, J. J.; Biancolillo, A.; Roger, J. M.; Rutledge, D. N.; Nordon, A. Deep learning for near-infrared spectral data modelling: Hypes and benefits. *TrAC Trends in Analytical Chemistry* **2022**, *157*, No. 116804.
(40) Delmelle, R.; Ngene, P.; Dam, B.; Bleiner, D.; Borgschulte, A. Promotion of hydrogen desorption from palladium surfaces by fluoropolymer coating. *ChemCatChem.* **2016**, *8*, 1646−1650.
(41) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *arXiv* **2017**.
(42) Abadi, M. et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015. https://www.tensorflow.org/.
(43) Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. *arXiv* **2019**.
(44) Fredriksson, H.; Alaverdyan, Y.; Dmitriev, A.; Langhammer, C.; Sutherland, D. S.; Zäch, M.; Kasemo, B. Hole-mask colloidal lithography. *Adv. Mater.* **2007**, *19*, 4297−4302.
(45) Nugroho, F. A. A.; Iandolo, B.; Wagner, J. B.; Langhammer, C. Bottom-Up Nanofabrication of Supported Noble Metal Alloy Nanoparticle Arrays for Plasmonics. *ACS Nano* **2016**, *10*, 2871−2879.

https://doi.org/10.1021/acssensors.4c02616
ACS Sens. XXXX, XXX, XXX−XXX

Supporting Information

Accelerating Plasmonic Hydrogen Sensors for Inert Gas Environments by Transformer-Based Deep Learning

Viktor Martvall[1], Henrik Klein Moberg[1], Athanasios Theodoridis[1], David Tomeček[1], Pernilla Ekborg-Tanner[1], Sara Nilsson[1], Giovanni Volpe[2], Paul Erhart[1,*], and Christoph Langhammer[1,†]

[1] *Department of Physics, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden*
[2] *Department of Physics, University of Gothenburg, SE-412 96 Gothenburg, Sweden*
[*] *erhart@chalmers.se*
[†] *clangham@chalmers.se*

Contents

Supplementary Notes

Supplementary Note S1: Hydrogenation protocols

The hydrogenation protocols were as follows: 5 pulses of 1.97 vol.% H_2 concentration were introduced at the beginning of each measurement, in order to activate the nanodisks. For the step wise hydrogen profile (Fig. S4), 3 different measurements were performed which consisted of a logarithmic distribution of H_2 pulses (14 steps increase/decrease, 0.06 vol.% H_2 minimum and 1.97 vol.% H_2 maximum concentration), where the pulse durations were set to 5 min, 10 min and 15 min. The time between pulses was 30 min for all measurements.

For the linear hydrogen profile (Fig. S5a), the H_2 concentration was slowly increased linearly to a maximal concentration. For each of the ramp up/down cycles, the maximum concentrations were: 0.09, 0.13, 0.19, 0.28, 0.41, 0.61, 0.90, 1.33, 1.97 vol.% H_2.The ramp-up/down duration was 1 hour. The maximal concentrations were increasing after each ramping up/down and therefore the hydrogen increase rate was changing as well. For the exponential hydrogen profile (Fig. S5b) the H_2 concentration was increased exponentially to a maximal concentration instead of linearly. The duration of the ramp-up/down was 1 hour and the same maximum concentrations were used.

Supplementary Note S2: Data pre-processing

Due to variations in the spectra from different measurements (see Fig. S12 and Fig. S14), it was necessary to pre-process the data before training a model (see Data pre-processing of the main paper). Initially, we only considered one pre-processing method (wavelength dependent min-max normalization), however, as seen by the blue curve in Fig. S13 this caused errors in the model predictions, which are likely due to the pre-processed spectra for similar H_2 concentration being relatively similar. Sequentially, we therefore added standard normal variate standardization (orange curve in Fig. S13), global min-max normalization (green curve in Fig. S13) and level scaling (red curve in Fig. S13). By using all four pre-processing methods the model accuracy increases substantially.

The reason several pre-processing methods improve the accuracy of the deep learning model, we believe, is that it is difficult to correct for the differences in spectra from different measurements using only one pre-processing method. This because spectra from different measurement differ not only in total amplitude but also slightly in shape (see Fig. S14). This causes spectra from different measurements that correspond to different hydrogen concentrations to overlap at certain wavelength intervals, making it more challenging for the deep learning model learn to distinguish between spectra corresponding to different hydrogen concentrations. However, by introducing various pre-processing methods, which enhance different features of the spectra, the deep learning model is able to learn more robust features which are not as sensitive to these slight variations.

In Fig. S12 and Fig. S15 we illustrate how the different pre-processing methods used alter the spectra. Wavelength dependent min-max normalization scales the data to the interval $[0, 1]$, sharpens the peak of the spectrum and enhances differences between spectra corresponding to different hydrogen concentrations. Standard normal variate standardization centers the spectrum to have zero mean and also accounts for potentially different noise levels by dividing each spectrum by its standard deviation. This pre-processing method enhances the red-shift of the spectrum; however, it also reduces the relative change of the peak position. Global min-max normalization, similar to wavelength dependent min-max normalization scales the data to the interval $[0, 1]$. This method maintains the shape of the spectrum, as it subtracts and divides all wavelengths of the spectrum by the same scalar. Level scaling focuses on the relative difference (per wavelength) of the spectra; however, information about the spectral shape is lost.

Supplementary Note S3: Standard Analysis

To obtain the calibration curve for the step-wise increases we used the measurement illustrated in Fig. S4a. For each pulse, we extracted the steady-state centroid shift as well as the H_2 concentration (Fig. S6). The motivation behind only using the steady-state centroid shift is that a unique mapping between the centroid shift and the H_2 concentration does not exist during the transient state. Then we minimized the mean absolute relative error and obtained $a = 0.334$ and $b = 2.91$ (see Eq. (1)).

In Fig. S6 we illustrate the use of the obtained calibration function for predicting the H_2 concentration on the measurement used for testing LSTR Ensemble Model for Accelerated Sensing (LEMAS) (Fig. S4a). The mean absolute relative error of the standard analysis (SA) for this measurement is 14.73×10^{-3} vol.% H_2, which is higher than the mean absolute relative error of LEMAS being 1.048×10^{-3} vol.% H_2, showing that LEMAS has a higher accuracy.

To obtain the calibration curve for the linear/exponential increases we used the measurement consisting of linear increases/decreases (Fig. S5a). Here, we used the entire measurement to obtain a calibration curve, as the H_2 concentration changes slower than the response time of the system (Fig. S8). Then we minimized the mean absolute relative error and obtained $a = 0.40$ and $b = 2.51$. In Fig. S9 we illustrate the use of the obtained calibration function for predicting the H_2 concentration on the measurement used for testing LEMAS (Fig. S5b). The mean absolute relative error of the SA for this measurement is 13.43×10^{-3} vol.% H_2, which is higher than the mean absolute relative error of LEMAS being 7.4×10^{-3} vol.% H_2, confirming again that LEMAS has a higher accuracy.

Supplementary Figures

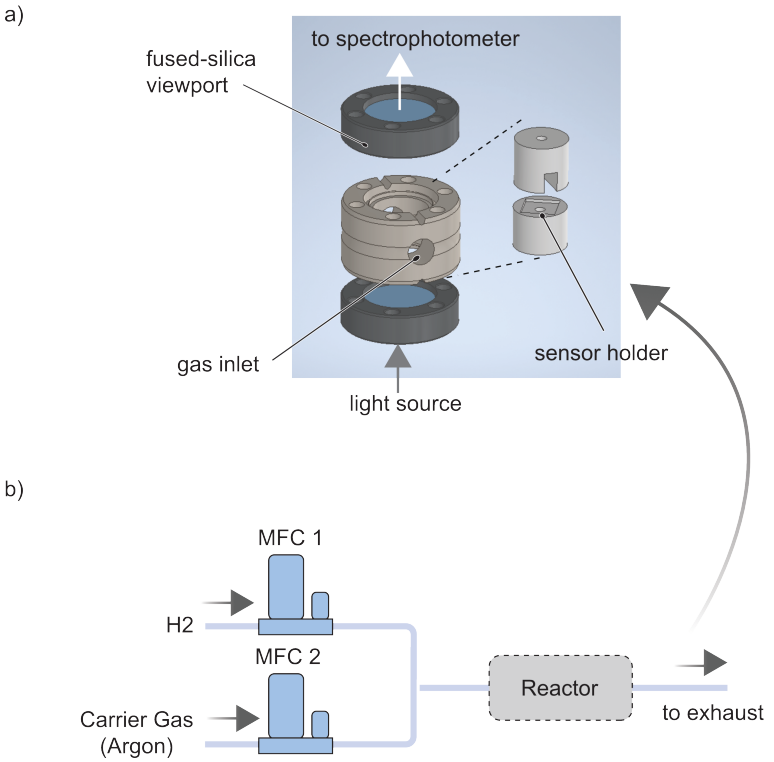


Figure S1: **Experimental setup.** Schematic illustration of the (a) reaction chamber and (b) gas mixing system. Adapted with permission from Ref. 1.

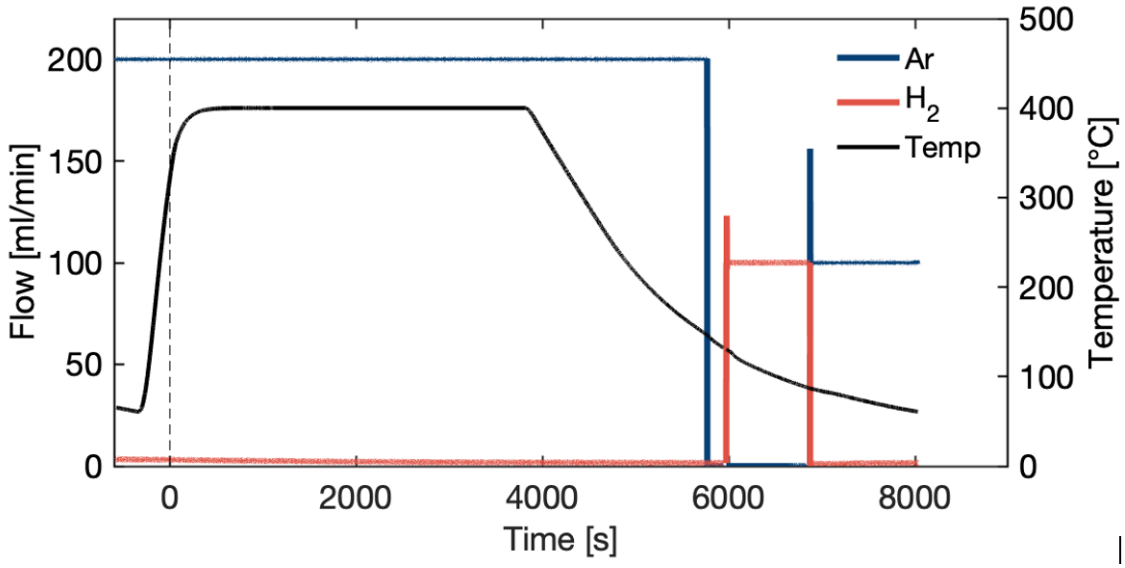


Figure S2: **Gas background quadrupole mass spectrometer measurements.** The gas flows of 100 % Ar (blue, left y-axis) followed by 100 % H_2 (red, left y-axis) and temperature (black, right y-axis) during the quadrupole mass spectrometer (QMS) measurements of the concentrations of possible trace gases in the mixture. The temperature was first increased to 400 °C in 100 % Ar flow, to desorb possible absorbed gases in the flow reactor. This was followed by one 15 minutes 100 % H_2 pulse, and again one pulse of 100 % Ar. The QMS measurement was started at time 0, marked by the dashed line.

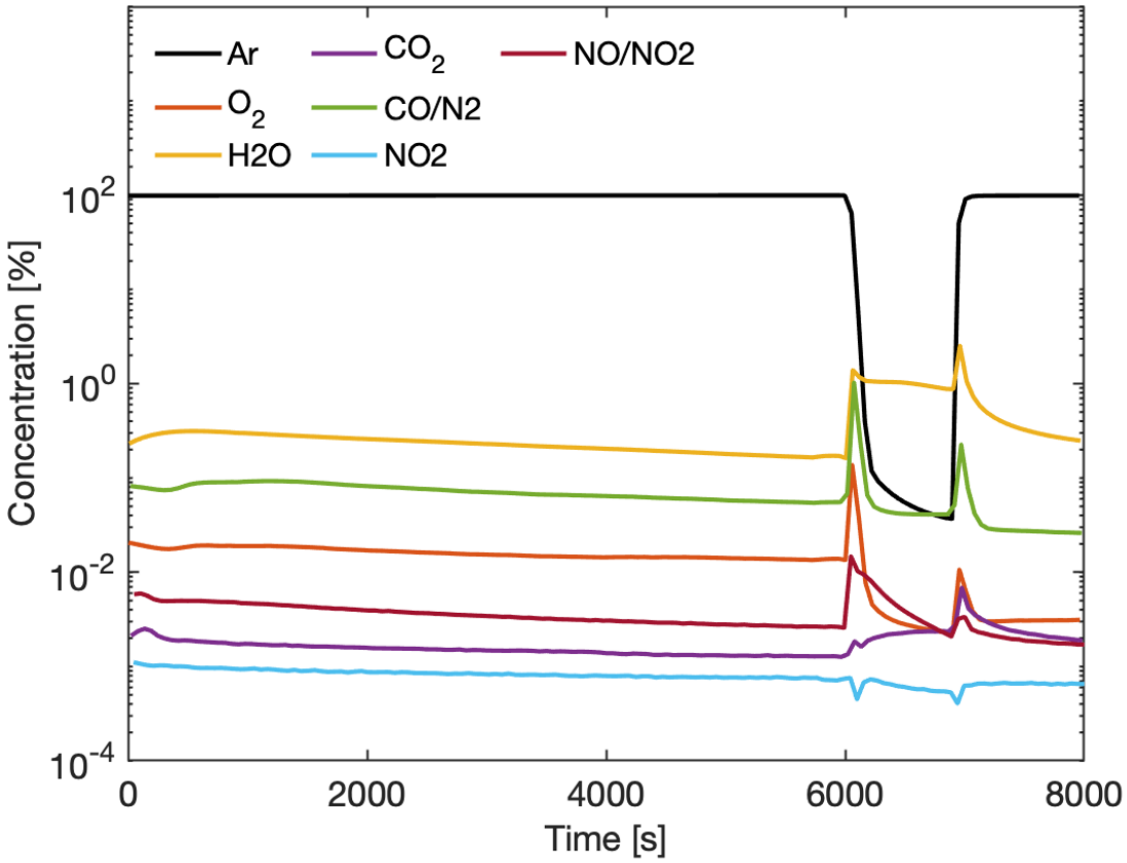


Figure S3: **Gas background quadrupole mass spectrometer measurements.** The concentrations of Ar and the trace gases during the Ar and H2 pulses in Fig. S2. The concentrations of the trace gases at the end of the Ar pulse at 8000 s were 99.7 % Ar, 0.26 % H_2O, 0.027 % CO or N_2, 31 ppm O_2, 18 ppm NO, and 6.6 ppm NO_2.

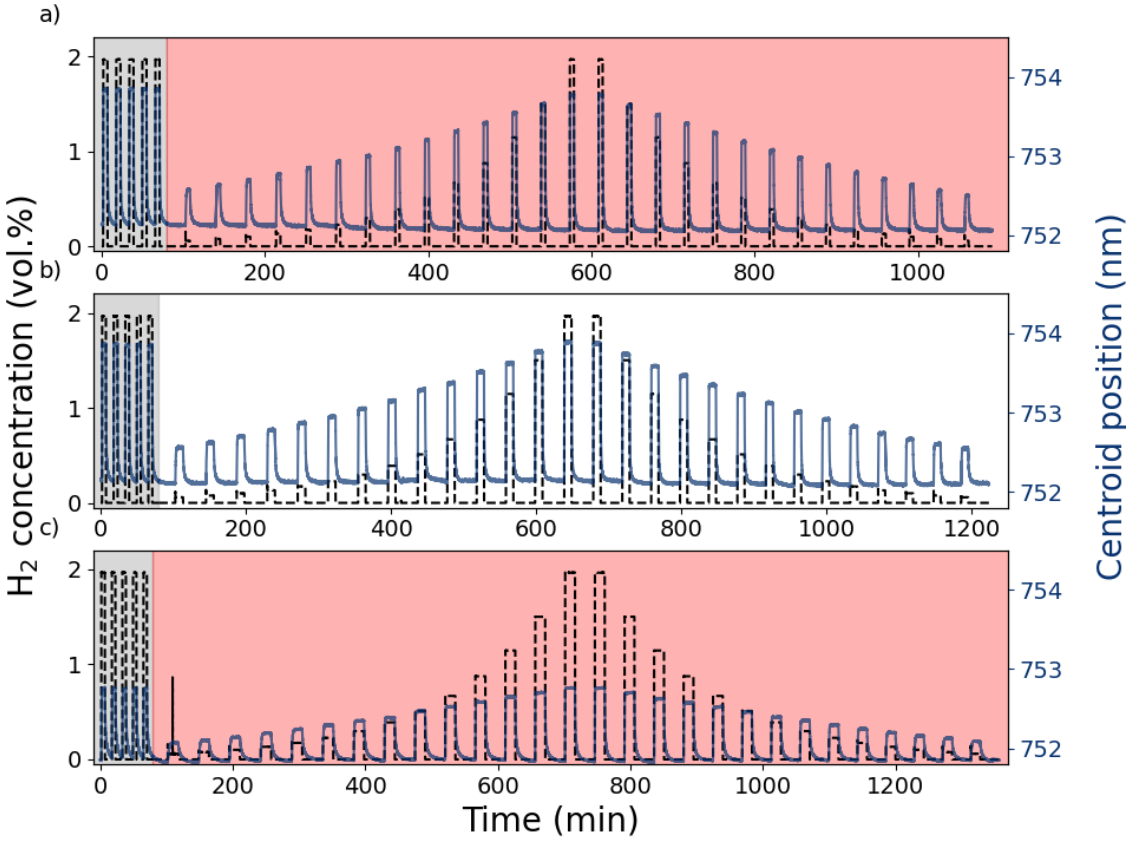


Figure S4: **Step wise hydrogenation measurements.** Hydrogenation protocol for the step wise increase/decrease, alongside the measured centroid shift. (a) 5 min, (b) 10 min, (c) 15 min of pulse duration followed by 30 minutes between each pulse for all measurements. The spectra collected during the first 5 pulses to 1.97 vol.% H_2 (gray shaded area) were used to pre-process the subsequent spectra in each measurement and were not used for training or testing. The measured spectra and set hydrogen concentrations from the red shaded region in measurement (a) and (c) were used for training/validation, with the first 14 pulses being used for training and the last 14 pulses for validation. The measured spectra and set hydrogen concentrations from non-shaded region in measurement (b) were used for testing.

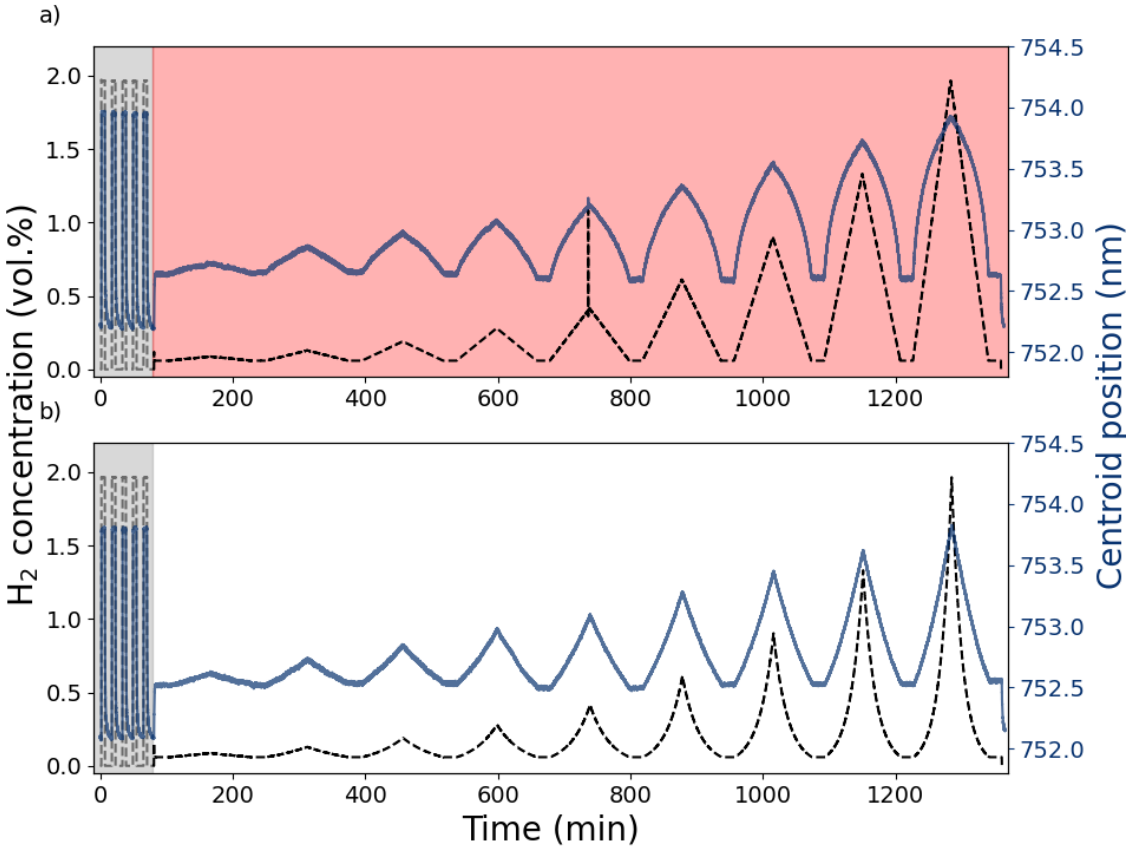


Figure S5: **Linear and exponential hydrogenation measurements.** Hydrogenation protocol for the (a) linear and (b) exponential increase/decrease alongside the measured centroid shift. The ramp-up/down duration for each cycle was 1 h. The spectra collected during the first 5 pulses to 1.97 vol.% H_2 (gray shaded area) were used to pre-process the subsequent spectra in each measurement and were not used for training or testing. The measured spectra and set hydrogen concentrations from the red shaded region in measurement (a) were used for training, while the measured spectra and set hydrogen concentrations from non-shaded region in measurement (b) were used for testing.

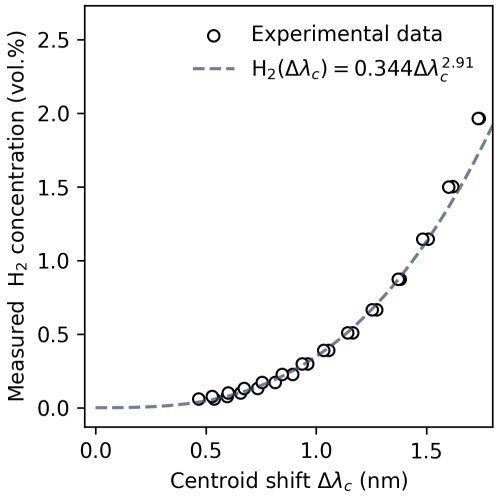


Figure S6: **Calibration function for step-wise increases/decreases.** Experimental data used from measurement Fig. S4a to fit the calibration function for the step-wise increases and the obtained calibration function.

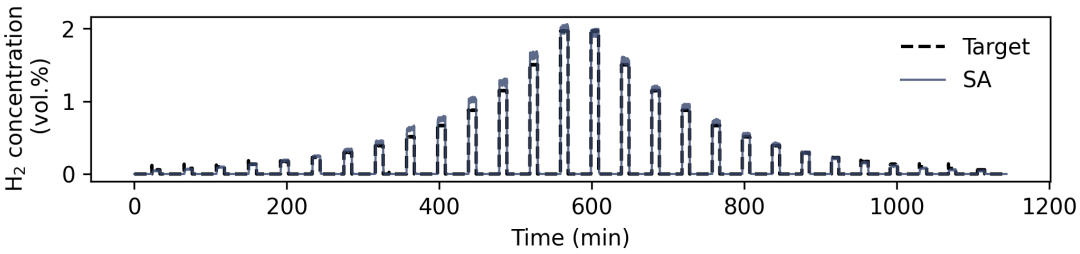


Figure S7: **Calibration function applied to the test data.** Calibration function obtained from Fig. S6 used on the test data, measurement Fig. S4b.

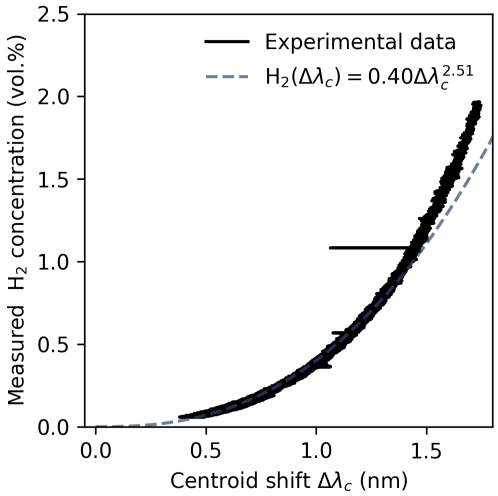


Figure S8: **Calibration function for linear/exponential increases/decreases .** Experimental data used from measurement Fig. S5a to fit the calibration function for the linear/exponential increases and the obtained calibration function.

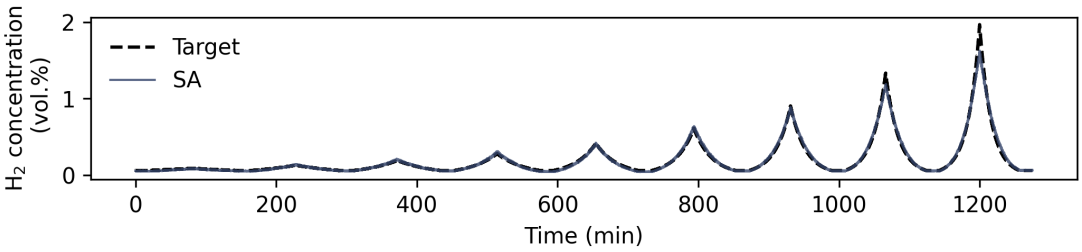


Figure S9: **Calibration function applied to the test data.** Calibration function obtained from Fig. S8 used on the test data, measurement Fig. S5b.

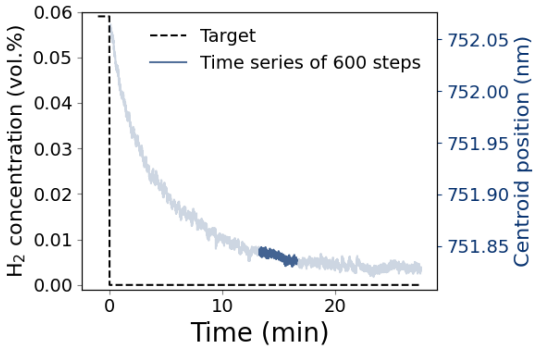


Figure S10: **Time series length for models optimized for accelerating sensor response.** Change in centroid position during the slowest occurring process in the training data, namely the desorption from 0.06 vol.% H_2 in measurement Fig. S4c. A time series comprising 600 time steps is chosen here to allow the model to estimate the rate of change in the spectrum, which is a prerequisite for accurately describing this process.

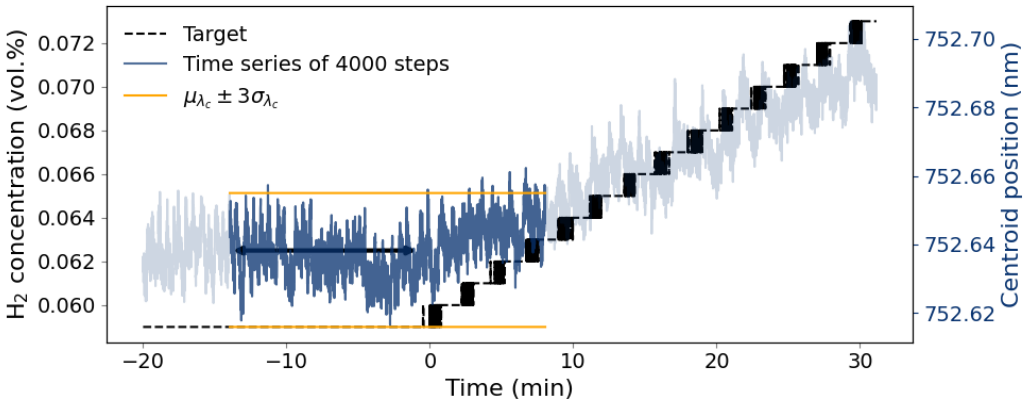


Figure S11: **Time series length for models optimized for leak detection.** The change in centroid position in the beginning of the leak with the smallest leak rate in Fig. S5a, compared to the mean and standard deviation of the centroid position before the leak has started (indicated by the double-headed arrow). Through using a time series comprising 4000 time steps the model should, in the worst case, be able to differentiate the leak occurring with the smallest slope from the noise in the sensor output data approximately 10 min after the leak has started.

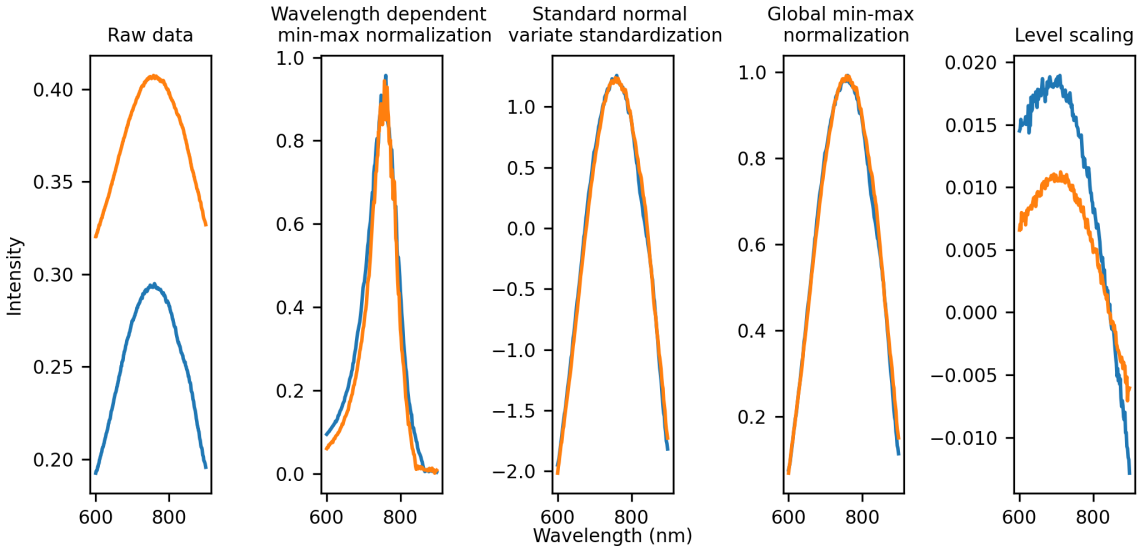


Figure S12: **Different pre-processing methods.** Illustration of the effect of different pre-processing methods on model prediction as well as raw data for spectra from measurement Fig. S4a (blue) and Fig. S4c (orange) at 0 vol.% H_2.

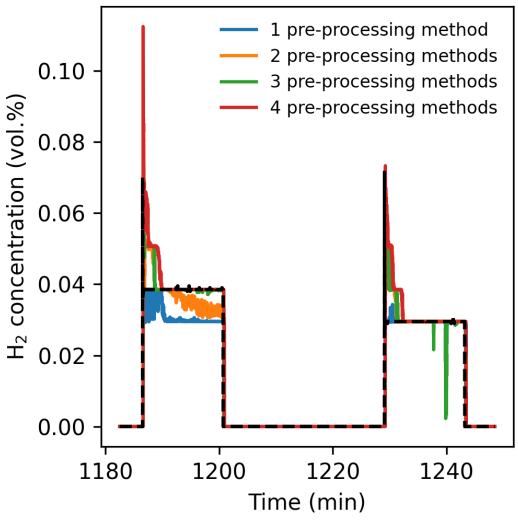


Figure S13: **Impact of using different pre-processing methods.** Predicted H_2 concentration on the two last step-wise increases in measurement Fig. S4c for models trained on the first half of measurement Fig. S4a and Fig. S4c, using different combinations of pre-processing methods.

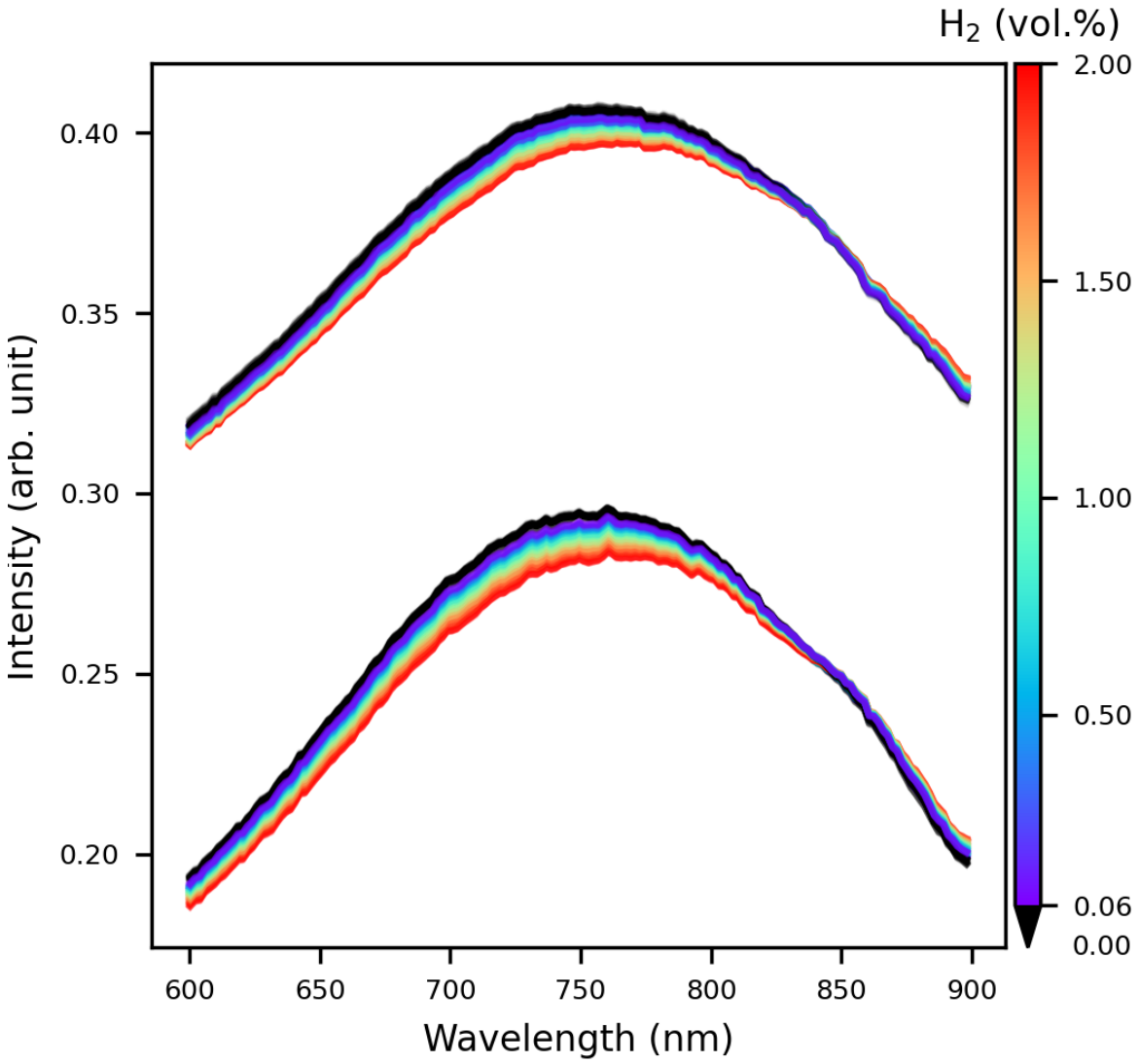


Figure S14: **Difference in spectra from different measurements.** Illustration of spectra from measurement Fig. S4a and Fig. S4c as function of the hydrogen concentration. The spectra differ in total intensity, where spectra from measurement Fig. S4c higher intensity, while the spectra from measurement Fig. S4a has lower intensity. The spectra also differ slightly in shape, most prominent around 850 nm.

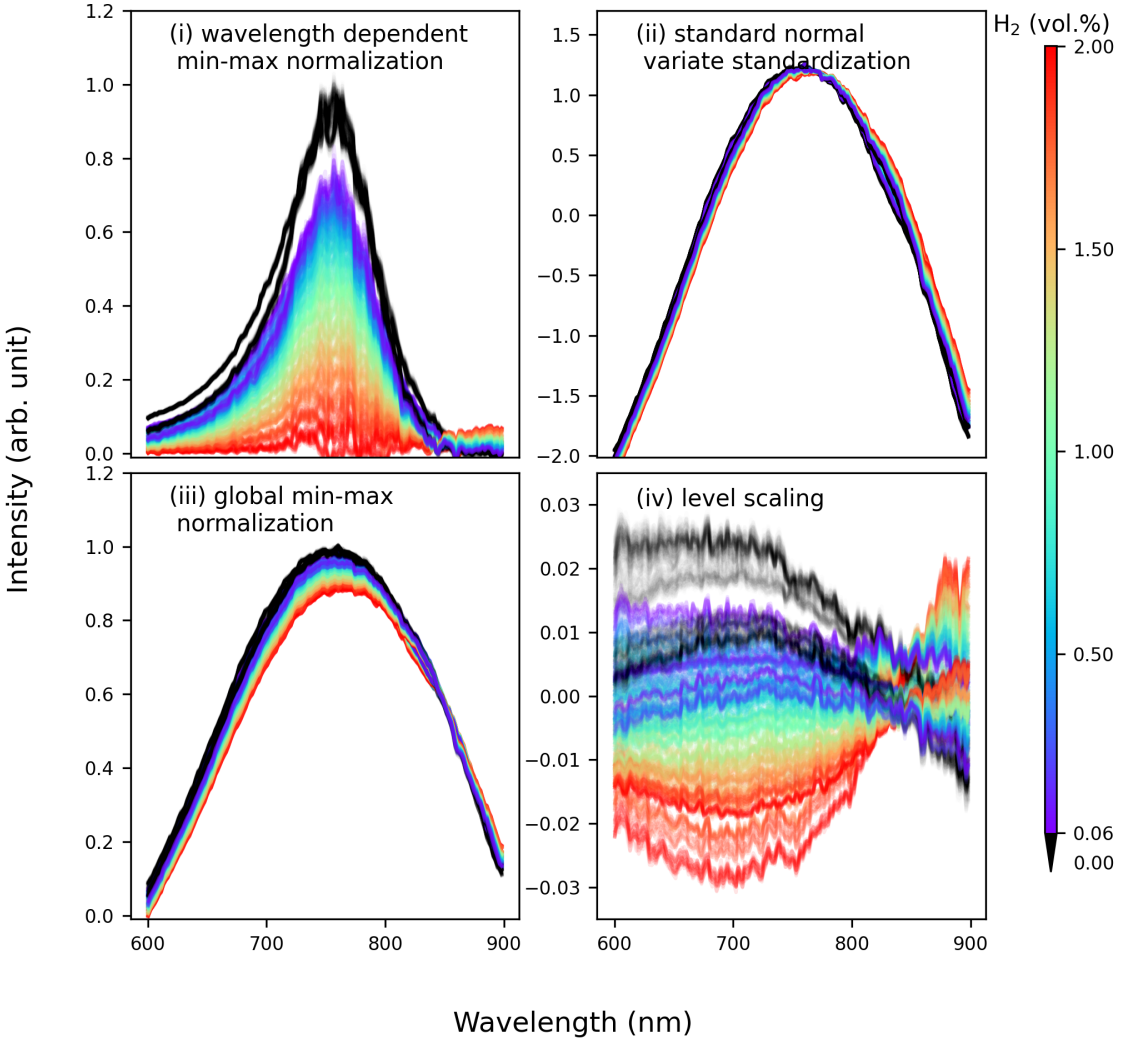


Figure S15: **Difference in spectra from different measurements after pre-processing.** Illustration of spectra from measurement Fig. S4a and Fig. S4c as function of the hydrogen concentration after they have been pre-processed. Comparing with Fig. S14 the difference between the two measurements are smaller, however some disparity still exists, which can be seen by e.g, noticing the two distinct black lines in (i) or (iv) corresponding to spectra at 0 vol.% H_2 in measurement Fig. S4a and Fig. S4c respectively.

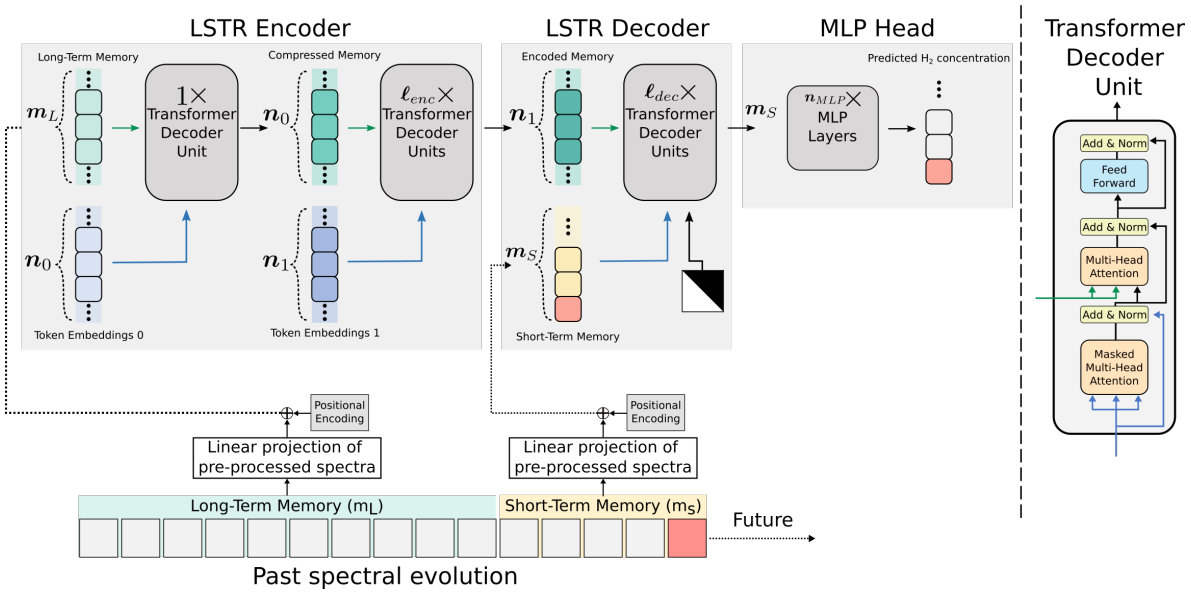


Figure S16: **Schematic illustration of the deep learning architecture used in this work.** Adapted from Ref. 2.

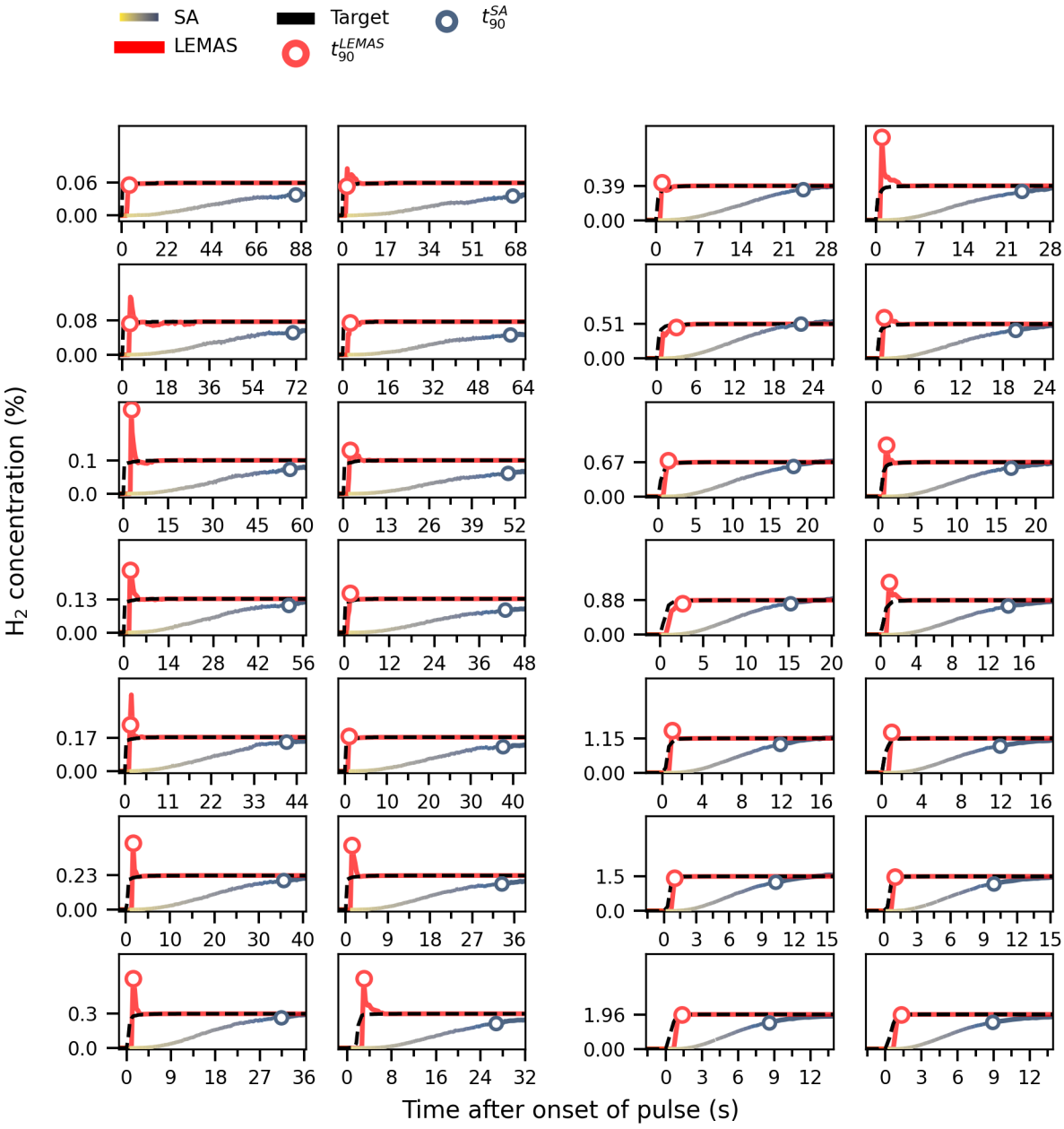


Figure S17: **Comparison of LEMAS and SA for step-wise increases in test data.**

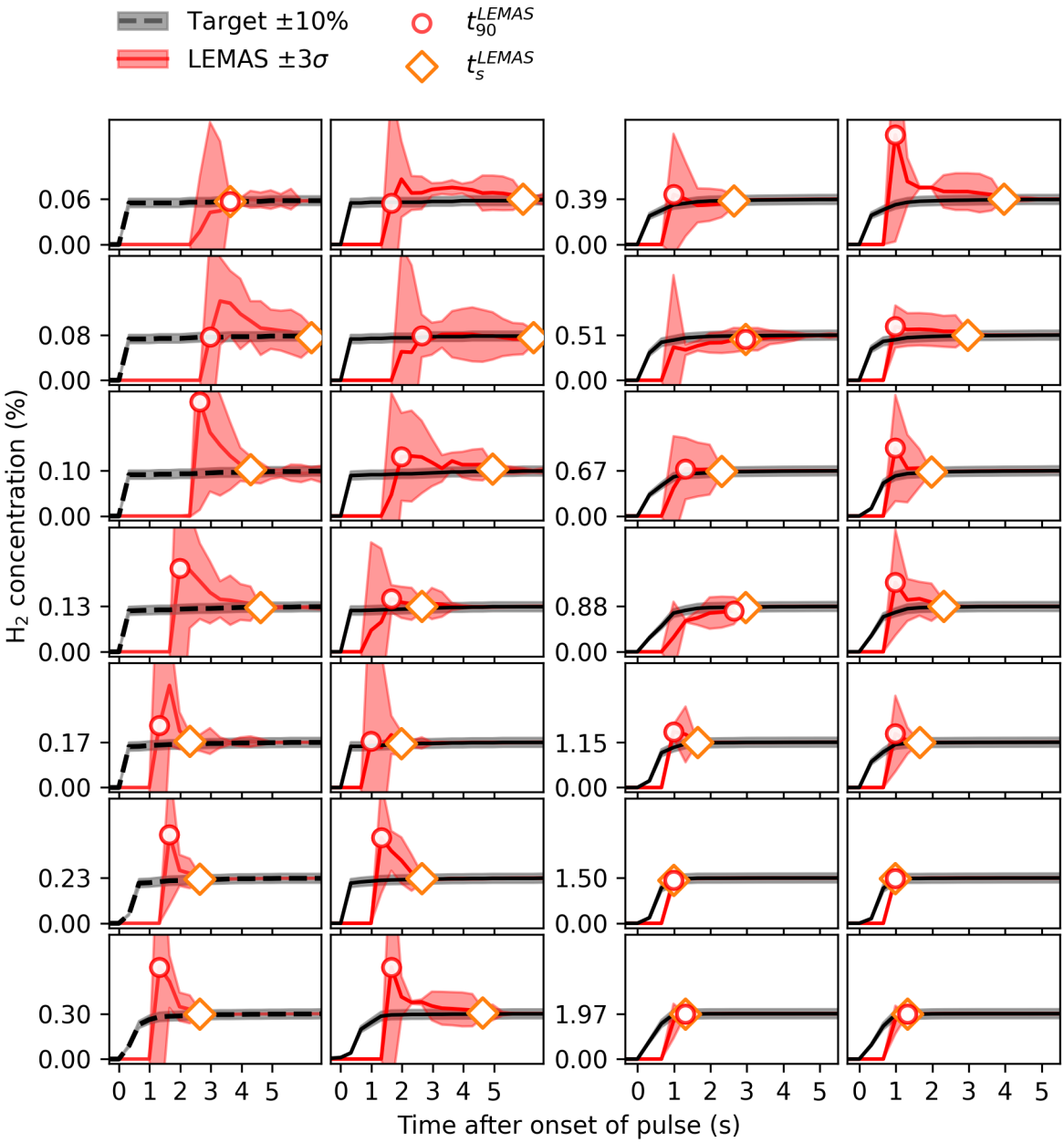


Figure S18: **Initial prediction of mean and standard deviation by LEMAS for step-wise increases in test data.**

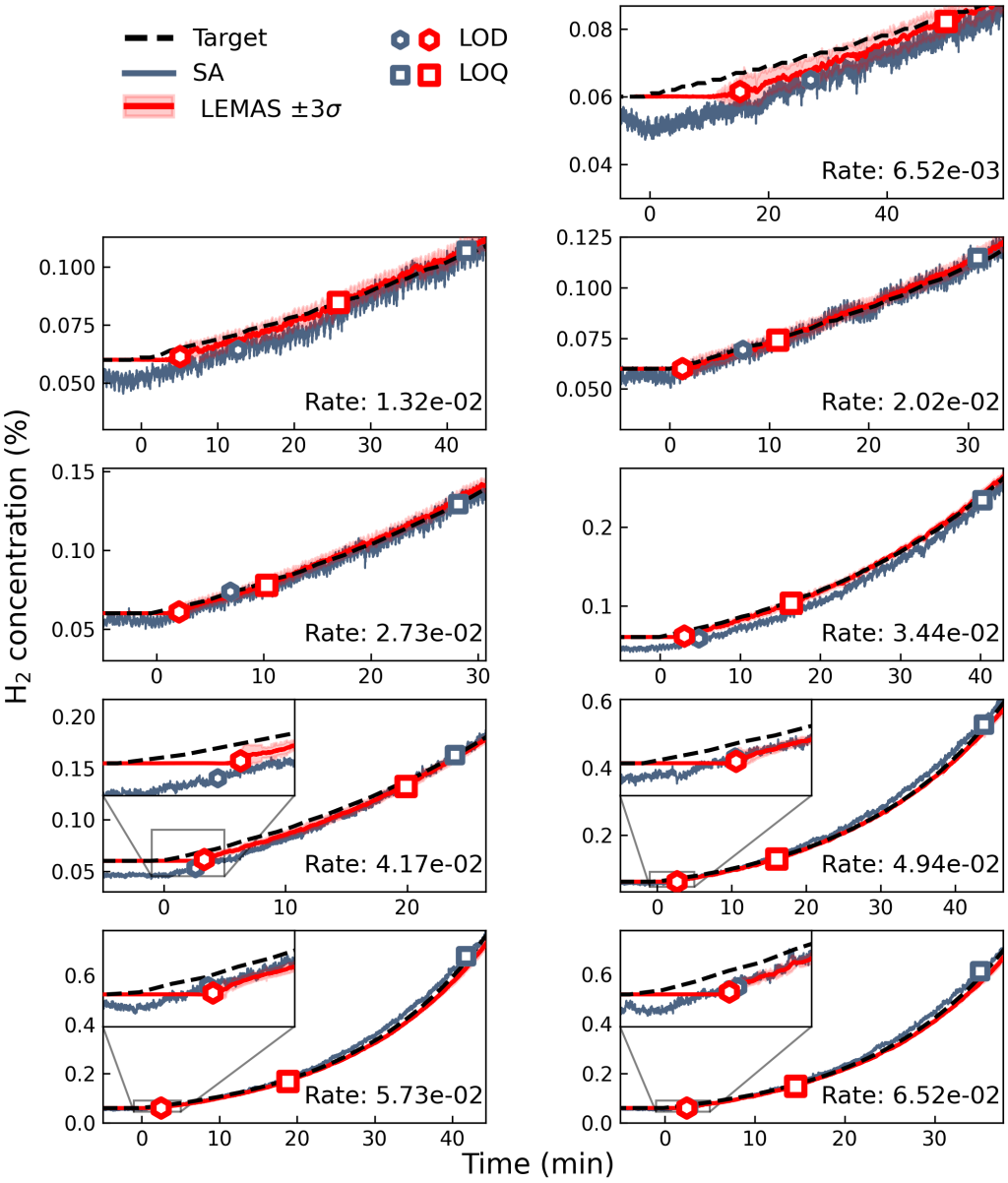


Figure S19: **Comparison of LEMAS and SA for exponential leak rates in test data.**

Supplementary Tables

Table S1: **Hyperparameters used for all models.**

Parameter	Value
n_{enc}	4
n_{dec}	8
n_0	8
n_1	4
d_{model}	256
d_{ff}	512
num heads	8
head size	32
Temporal stride on long term memory	4
n_{mlp}	8
d_{mlp}	512
Dropout rate	0.1

Supplementary References

[1] Iwan Darmadi. *Polymer-Nanoparticle Hybrid Materials for Plasmonic Hydrogen Detection*. PhD thesis, Chalmers University of Technology, Gothenburg, Sweden, January 2021. Available at `https://research.chalmers.se/en/publication/521833`.

[2] Mingze Xu, Yuanjun Xiong, Hao Chen, Xinyu Li, Wei Xia, Zhuowen Tu, and Stefano Soatto. Long short-term transformer for online action detection. In *NeurIPS 2021*, 2021. URL `https://www.amazon.science/publications/long-short-term-transformer-for-online-action-detection`.