



High-availability displacement sensing with multi-channel self mixing interferometry

ROBIN MATHA,^{1,2,3} STÉPHANE BARLAND,^{2,*} AND FRANÇOIS GUSTAVE¹

¹*DOTA, ONERA, Université Paris-Saclay, F-91123, Palaiseau, France*

²*Université Côte d'Azur - CNRS, Institut de Physique de Nice, 1361 route des Lucioles, F-06560, Valbonne, France*

³*Safran Aerosystems, 61 rue Pierre Curie, F-78373 Plaisir, France*

*stephane.barland@univ-cotedazur.fr

Abstract: Laser self-mixing is in principle a simple and robust general purpose interferometric method, with the additional expressivity which results from nonlinearity. However, it is rather sensitive to unwanted changes in target reflectivity, which often hinders applications with non-cooperative targets. Here we analyze experimentally a multi-channel sensor based on three independent self-mixing signals processed by a small neural network. We show that it provides high-availability motion sensing, robust not only to measurement noise but also to complete loss of signal in some channels. As a form of hybrid sensing based on nonlinear photonics and neural networks, it also opens perspectives for fully multimodal complex photonics sensing.

© 2023 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Complex (nonlinear or disordered) systems have long been considered as high-capacity "information processors" [1]. For the specific task of processing environmental changes (*ie* "sensing"), complex waves and photonic systems are particularly useful thanks to their capability to process information from a distance. Recently, hybrid solutions leveraging complex wave physics and computer neural networks have enabled outstanding achievements [2,3], at the intersection between neural networks and optics [4]. Thanks to its non-linearity, self-mixing interferometry can carry much more information than its linear counterpart, which makes it potentially useful for many sensing tasks [5–11].

Self mixing interferometry is based on a delayed feedback effect, where a laser beam re-enters the emitting laser itself (in most cases in a semiconductor laser) after being reflected on a target. As is well known, this feedback alters the operation point of the laser and by monitoring this operation point (either via an integrated photodiode or by simply measuring the voltage across the diode), information about the displacement of the target can be retrieved. Thanks to this simplicity and compactness, this scheme is expected to provide a robust and reliable sensing technique, for instance for the displacement of a target along the light propagation axis. However, the operation range in which information can be reliably retrieved is in fact an important constraint, notably in terms of the amount of light which re-enters the emitting laser. This is often characterized via the coupling parameter C which also depends on the target distance. We refer the reader to for instance [5] for a very complete discussion of the different feedback regimes depending on C (including more parameters in [12]) but here we only underline that in general, the shape of the signal strongly depends on C , which makes signal processing difficult. More importantly, when the amount of light re-entering the device is too small ($C \ll 1$), the self mixing signal becomes harmonic and the information about the direction of the displacement of the target is lost. At the other edge of the operation regime (when $C > 4.6$), the dynamics of the laser with re-injection becomes multistable or even unstable towards chaotic regimes. Again, in this range, the self-mixing signal does not carry reliable information about the displacement of the target.

For these reasons, great care must be taken to keep the system within the useful operation range. Different approaches have been taken to mitigate unwanted variation of C for instance due to speckle effect, including dedicated algorithms [13], on the fly parameter estimation [14,15] or training a neural network in different alignment conditions [16,17]. However, the hardest limitation is the progressive loss (eventually up to complete lack) of significance of the signal when $C \ll 1$ or $C > 4.6$. Since this limitation is of physical origin, tracking hardware has been proposed for the case of speckle [18] and adapted beam focusing has been analyzed for large displacements [19].

Here, we propose that high-availability motion sensing can be achieved with a multi-channel self-mixing interferometer equipped with a simple, embeddable neural network. Multichannel self mixing has already been envisioned for complex measurements (multi-dimensional motion in [20,21], imaging in [22] and flow sensing in [23]) but only seldomly considered as a potential enhancement [24,25] for the acquisition of a single measurement. On the other hand, machine learning has been increasingly used in self-mixing applications, including for fringe detection [26–28], parameter estimation [15], signal enhancement [29,30], vibration measurement [31] and displacement inference [16]. Here we show that, thanks to the intrinsic capacity of neural networks to process high-dimensional data, a multichannel self-mixing sensor can provide high availability displacement measurements, robust against signal loss and with enhanced resolution.

In section 2 we present the experimental arrangement (2.1), the neural network design (2.2.1) and its training procedure (2.2.2). We assess the performance of the system in section 3 in terms of accuracy (3.1), robustness against noise (3.2) and measurement availability (3.3). We present our conclusions in section 4.

2. Experimental set up, model and training

2.1. Experiment

The principle of the experimental arrangement is presented on Fig. 1. It consists of three independent self-mixing measurement channels with a (calibrated) speaker which acts as a target, a moving surface. Each channel is composed of a power supply, a laser diode and a signal amplification stage. The lasers of channels 1 and 2 emit at $\lambda = 1310$ nm (ML725B8F) and their coherent emission threshold is 6.5 mA. The laser of channel 3 emits at $\lambda = 1550$ nm (ML925B45F) and its threshold is around 10 mA. For all the acquisitions that will be described later, the lasers were respectively driven with DC currents of 7.7mA (1 and 2, driven by battery-powered laser driver) and 41.7mA (channel 3, commercial driver). Each laser is inserted in a passive mount with an adjustable collimation tube featuring an antireflection coated aspheric lens (3.1 mm focal length and 0.68 numerical aperture). The laser beams are focused onto the speaker surface, which is located about 15cm away from each laser. The lasers are placed very close to each other to minimize the angle between the beams and the normal to the speaker surface, minimizing the error between the actual displacement and its projection along each beam propagation direction.

The three self-mixing signals are obtained by measuring the voltage across each laser diode for channels 1 and 2 and the current through the internal photodiode for the third one (dictated by available equipment). This last laser is driven further above threshold with respect to the other two because it appeared to be necessary to obtain a usable signal from the internal photodiode. These signals being of low amplitude, each voltage signal is amplified by an AC-coupled amplifier with 10^4 gain factor and several MHz bandwidth and the photodiode current is amplified by a transimpedance amplifier. Since self-mixing interferometry with non-cooperative targets is often plagued with signals of varying quality (mostly due to speckle effect) we purposefully align the lasers slightly differently so that each laser operates in a slightly different regime. An example is shown on Fig. 2. On the top row, the self-mixing signal is not particularly noisy but it corresponds to a rather low re-injection value and therefore the asymmetry of the fringes is not very visible. On the second row, the signal is of excellent quality with low amount of noise and

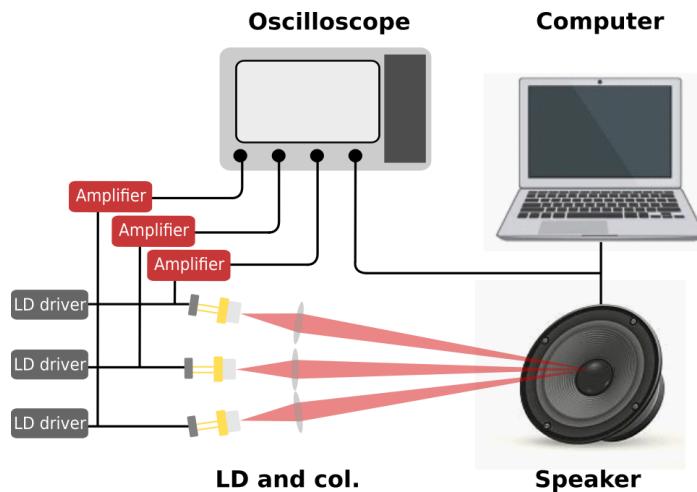


Fig. 1. Scheme of the experimental setup. Three independent self-mixing channels monitor the displacement of a single non-cooperative target and a neural network processes the resulting high dimensional data to estimate the displacement of the target.

well defined asymmetric fringes. Finally, the signal shown on the third row is hardly exploitable at all, with a very poor signal to noise ratio. On the right column of Fig. 2, the difference between the first two channels can be better appreciated, with channel 2 featuring very sharp jumps while channel 1 is much smoother.

The displacement of the speaker has been calibrated and the linearity of the response of the speaker to a harmonic signal in a range of 5 to 100 Hz allows us to have access to the effective displacement of the surface. In this frequency range the response of the loudspeaker does not show any phase shift. These harmonic signals are generated by the sound card of the computer with a normalized amplitude between 0.1 and 1, which means displacements between $3.5\mu\text{m}$ and $7.5\mu\text{m}$. This electrical signal is recorded during the experiment and (after some preprocessing described below) forms the basis of the truth signal. An example of the displacement per time unit signal is shown on the bottom row of Fig. 2. Of course a piezo-electric displacement system could offer a much higher accuracy (*i.e.* deeply subwavelength resolution) of motion, but thanks to its large displacement and frequency ranges, a calibrated speaker offers a convenient trade-off for our purpose.

2.2. Model and training

2.2.1. Model design

In this experiment, the neural network is supposed to infer the displacement per time unit of the target on the basis of three interferometric signals. The problem is therefore in principle a (multi-)sequence to sequence task. However, as was pointed out in [16], defining a spectral band of operation for the displacements to be measured allows one to convert the task into a simpler regression task by down-sampling the displacement signal. In this case, the frequency operation regime we define ranges from 10 to 100 Hz. Within this range of displacement frequency and with $4\ \mu\text{s}$ sampling time for the signal acquisition (both interferometric signals and displacement), we choose to analyze the time series in chunks of 1.024 ms, each of them containing 256 data points per interferometric channel and per displacement measurement. Due to the selected frequency band and the Nyquist theorem, each chunk of 256 measurements of speaker voltage can be replaced by a single average displacement value over the 256 time steps.

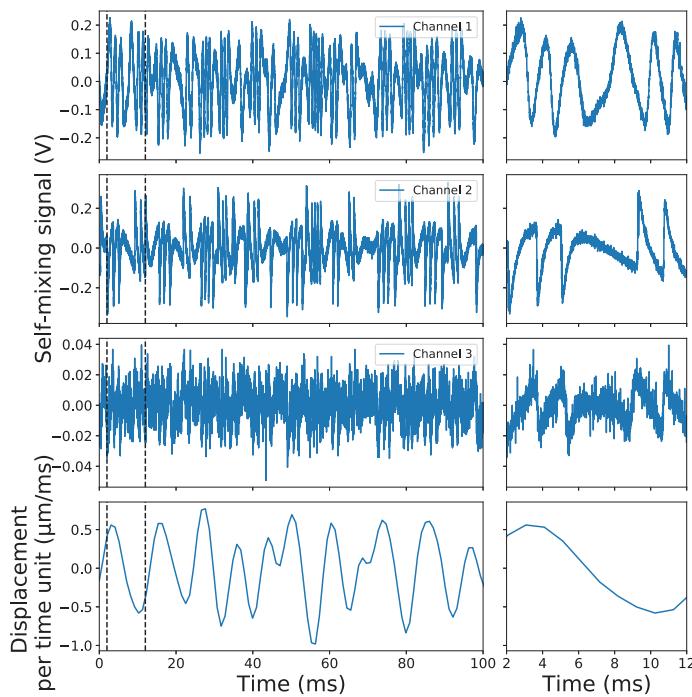


Fig. 2. Experimental data example. The top three rows show the self-mixing signal provided by each laser and the bottom row the corresponding displacement per time unit. The lasers are purposefully set to provide different quality of self-mixing signals (see text). On the right column, a zoomed view on a shorter time interval.

This converts the sequence to sequence task into a much simpler regression task, where one displacement value must be inferred from three interferometric sequences. We underline that the inferred value is a *displacement per time interval*, where the time interval is 1.024 ms, which is dimensionally a *velocity*, which (contrary to an absolute displacement) can be reconstructed with bounded error as discussed in [16]. At this point, we can use the very simple convolutional neural network architecture proposed in [16], since it appeared to carry sufficient information capacity to efficiently process single-channel data. However, each input vector is now composed of three channels, one per interferometric signal. Although advanced use of multiple sensing modalities is a very relevant area of research in itself (see *eg* [32–35]), here we deal with the simpler case of multiple measurement channels which operate on similar modalities. Thus, we stick to a very simple "early fusion" approach: the first convolutional layer operates on the three input channels and fuses the resulting kernels into a single output space, irrespectively of what specific channel activated them. After this stage, the network consists of a contracting stack of down-sampling and 1D, single channel, convolutional layers. At the end of the stack, two fully connected layers perform the final regression towards a single displacement value per time interval. All layers use a Rectified Linear Unit activation function except for the last layer which is linear. Details about the network as well as a schematic view are available in the Appendix (Fig. 6).

2.2.2. Training

Training a neural network requires acquiring and pre-processing data. We describe below with some details first the physical ranges of motions we want to analyze and second how the data is pre-processed.

The training set (available at [36]) consists only of experimental data. We record simultaneously the three self-mixing signals and the speaker voltage, calibrated as a displacement. The sampling rate is $4 \mu\text{s}$ and the record length is 499968 points. The training set features only harmonic displacements with eight evenly spaced frequencies between 53Hz and 93Hz and six evenly spaced amplitudes between 0.4 and 0.9 V applied to the speaker, corresponding to 3.5 and $7.5 \mu\text{m}$. Thus, the training set contains 48 different configurations in terms of frequency and amplitude of periodic displacement. The ranges of frequencies and amplitudes we define here sets the range of displacements per time unit that the sensor will be able to measure accurately.

Before processing by the network, the interferometric signals are normalized and centered around zero by dividing them by their standard deviation and subtracting the mean value of the signal over the full record length (a standard procedure often called "z-scaling" in the machine learning context). In order to reinforce the robustness of the network to noise, we add to each signal a gaussian white noise of standard deviation $\sigma_n = \sigma_s$ where σ_s is the standard deviation of the original self-mixing signal.

As for the truth signal, first we remove the electronic noise (intrinsic oscilloscope noise and digitalization noise) by smoothing the displacement signal using a Savitsky-Golay filter whose parameters are a sliding interval of 1001 points and a polynomial degree of 2. Then we compute the average displacement during a time window of 256 sampling points which gives a signal in units of Volts per time window. We then use the duration of the sampling rate $dt = 4\mu\text{s}$ which gives a time window of 1.024 ms and the speaker calibration of $31.3 \mu\text{m}/\text{V}$ to convert the signal in physical units of $\mu\text{m}/\text{ms}$.

The network can then be trained using as input about $9.5 * 10^4$ arrays of size 3×256 corresponding to three interferometric signals over time windows of $256 \times dt$ and as truth the corresponding average displacement over each time window, with 10% of the samples kept for validation. Of course during training the order of samples is randomized at each epoch but an additional randomization is also operated on the order of the channels themselves, *ie* to a single displacement value can correspond any permutation of the measurement channels. This point implies that the network must be trained in physical units of $\mu\text{m}/\text{ms}$ but we have observed that it leads to a remarkable improvement of the reconstruction performance. We attribute it to an enhancement of training for all weights especially at the first layer instead of focusing training immediately on the most significant channel. In addition to randomization, during training we randomly replace one of the self-mixing signals with white noise (with probability 1/4) to prepare the network to potential channel loss during the inference phase. The whole data preparation procedure is shown in a synthetic view in the [Appendix](#) (Fig. 7).

We train the network by minimizing the mean squared error between the inferred and the true displacement during 18 epochs in batches of 32 samples. On a (consumer grade) GTX1080 GPU the training time is about 20 min.

3. Results

After training, we assess the performance of a multi-channel measurement approach against that of the usual single-channel approach, first in terms of accuracy and second in terms of robustness to measurement noise up to the complete loss of one or the other channel. Since each channel provides an independent measurement, we will refer to a model processing all three channels as a three-dimensional (3D) model, as opposed to a model processing a single dimension, which we will refer to as a one-dimensional (1D) model.

3.1. Accuracy

On Fig. 3, we show the reconstruction of the target trajectory from each measurement channel and the reconstruction based on the three channels simultaneously. The displacement reconstruction is expected to operate correctly for arbitrarily complex displacements therefore we analyze its

performance (as in [16]) on a random displacement with a prescribed bandwidth (here 10 to 100 Hz). To achieve this, we generate a random δ -correlated signal which we Fourier filter with a fifth order Butterworth filter between 10 and 100 Hz. This signal is sent to the speaker, which then undergoes a random motion. On the top three traces, we use three 1D models trained on the experimental data set described above, each model processing only one of the three self-mixing signals. We quantify the quality of the reconstruction by measuring the Pearson correlation coefficient and the root mean squared error (RMSE) between the true displacement and the reconstruction on a 512-seconds long time trace of this random displacement ($5 * 10^5$ samples of 1.024 ms duration). On the top trace, the reconstruction is rather good with a correlation coefficient between the true displacement and the reconstruction of $r = 0.83$ and an $RMSE < 0.23 \mu\text{m}/\text{ms}$. The second row is based on the best measurement channel, with low detection noise and well defined, clearly asymmetric fringes. Correspondingly, the quality of the target displacement reconstruction is excellent, with a correlation coefficient between the true displacement and the reconstruction of $r = 0.97$ and an $RMSE < 0.11 \mu\text{m}/\text{ms}$. On the third row instead, the reconstruction of the displacement is of again of lower quality, which is not unexpected since the self-mixing signal is really very poor (see Fig. 2, third row). We note however that the reconstruction is essentially equivalent to that of the first channel, with a correlation coefficient $r = 0.83$ and $RMSE < 0.24 \mu\text{m}/\text{ms}$. This reconstruction even with such a poor signal quality is a result of adequately training the network with similar quality signal: During training, the network gradually adapts to capture significant features of the data while discarding the irrelevant ones, which would be very hard to achieve with conventional programming techniques. Finally, on the bottom row, we show the reconstruction based on a 3D model processing the three channels simultaneously. This reconstruction achieves a correlation coefficient of $r = 0.96$ and an $RMSE = 0.13 \mu\text{m}/\text{ms}$. Remarkably, it is almost as good as the one obtained only on the single highest quality measurement (channel 2) and is markedly better than the one obtained on the two poorest channels (1 and 3).

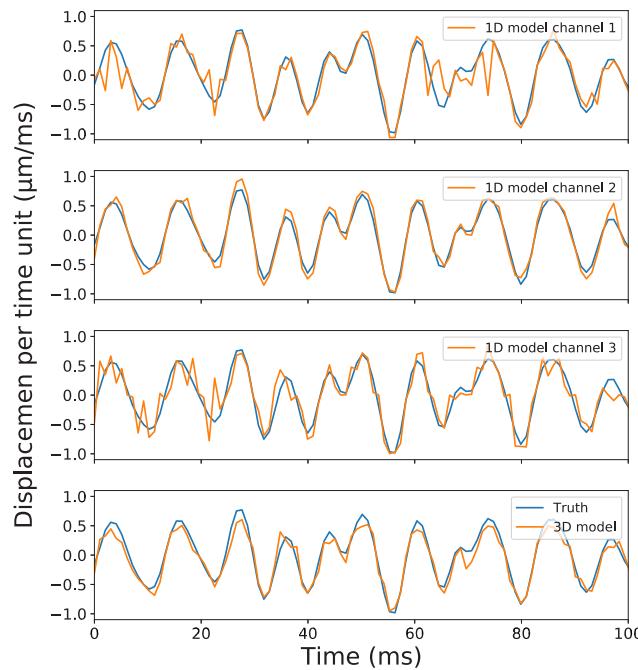


Fig. 3. Reconstruction performance based on the single channels in isolation (the top three rows are three 1D models) and on the three channels simultaneously (bottom row, 3D model).

3.2. Robustness against signal degradation

We have seen above that (at least with the simple early fusion approach we use here in the neural network), the redundancy in measurement channels does not immediately translate into a higher measurement precision. However, as we shall see in the following, it does improve very strongly the resilience of the system to noise present in one or more channels.

To quantify this effect, we check the reconstruction performance on the same data set as above after adding to one or more channel(s) a δ -correlated Gaussian noise with standard deviation σ_n which models the noise-induced degradation of self-mixing measurements. The results of this procedure are shown on Fig. 4, where we show (top row) the correlation coefficient between the true displacement and the reconstruction of the displacement provided by several models and (bottom row) the root mean square error on the reconstruction for each model depending on noise added to one or more channels.

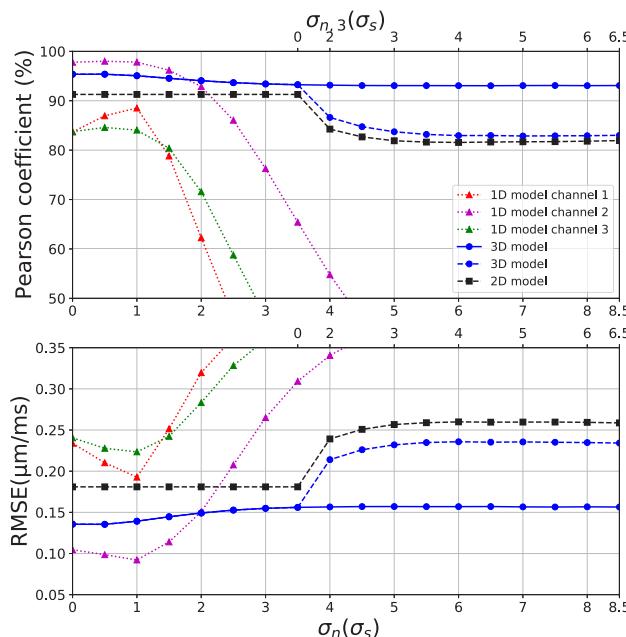


Fig. 4. Robustness of the reconstruction against noise in one or more self-mixing signals. Top: Pearson correlation coefficient between the reconstruction provided by models and the true displacement. Bottom: Root mean squared error on the displacement reconstruction.

The three 1D models curves (red, green and purple triangles) are shown for reference and display essentially the same behavior. For each case, we use a single channel for the reconstruction and degrade this channel by adding noise to it. At first, each model provides an acceptable level of performance but even the best signal (channel 2) becomes very unreliable when the added noise is larger than $\sigma_n = 3 * \sigma_s$ where σ_s is the standard deviation of the uncontaminated original self-mixing signal. The fact that there is an optimal non-zero level of noise in these curves is a result of our training procedure: for robustness all models have been trained with $\sigma_n = 1$ and therefore have never seen uncontaminated signals.

The two curves based on 3D models (blue disks) show the robustness of a model operating on three channels simultaneously. The continuous blue line is obtained when noise is added only to the second self-mixing channel *ie* the one with the best performance when used in isolation. We observe that when a single channel is fully degraded ($\sigma_n = 8\sigma_s$ for instance) this model still operates at a very high level of performance ($r = 0.93$, $RMSE = 0.16$). This is an

excellent point in terms of measurement quality and availability. About measurement availability, it shows that a 3D measurement system can very well operate under highly degraded conditions *ie* even when the highest performance channel is basically lost. The dashed blue line is obtained when noise is added to the second self-mixing channel and also (starting from $\sigma_n > 3.5\sigma_s$) to the third measurement channel. When this noise term $\sigma_{n,3}$ is added to the third channel, the measurement quality of course decreases since the model in the end operates basically on only a single remaining channel. Accordingly, when both channels 2 and 3 have become unusable due to very high noise (for instance $\sigma_n = 7$, $\sigma_{n,3} = 5$), the performance level of the 3D model is basically the performance level of the 1D model operating on the uncontaminated channel 1 ($r = 0.83$, $RMSE = 0.23$). Again, this is an excellent point in terms of measurement availability since a three-channels measurement system can operate at the performance level of a single channel when the other two are lost.

Finally, for completeness, the performance of a 2D model (trained only on the poorest self-mixing signals 1 and 3) is shown as black squared dashed line. Of course the performance of that model is not affected by noise added to channel 2 which it does not process but it is important to underline that the 2D model performance is better than that of the 1D models based only on channels 1 and 3 in isolation. When noise is added also to channel 3, the performance of this model smoothly degrades down to that of the 1D model based only on an unperturbed channel 1.

Overall, the above observations demonstrate several outstanding properties (in terms of accuracy and robustness) of a 3D approach to displacement measurement:

- In absence of noise, the 3D model's performance almost matches that of the best 1D model and is much better than the other two 1D models
- When noise degrades the most informative channel, the 3D model strongly outperforms all 1D models and also outperforms a 2D model based on unperturbed channels
- Even if noise strongly degrades two of the three available channels, the performance of the 3D model never degrades below that of the 1D model based on the only unperturbed channel
- In presence of noise on all 3 channels, we checked that the 3D model performance is essentially equivalent to that of the best 1D model (channel 2), as was also observed in absence of noise (Fig. 3).

For all the analysis above we use a Gaussian white noise but this is nothing more than a practical choice and the network is expected to be robust to any stochastic function seen during training. For instance the network is remarkably robust also to the intrinsic physical noise (detection, electronics, photonics) heavily present in channel 3.

3.3. Measurement availability

The features described above have a considerable impact in terms of measurement availability, as we discuss below. Our goal here is to demonstrate that a displacement measurement system based on three simultaneous measurement channels processed by an adequately trained neural network provides a considerable step towards a high-availability self mixing displacement sensor. In fact, it is well known that with uncooperative targets, speckle can strongly modify the self-mixing signal shape, including leading to an effectively vanishing feedback rate. Several possibly complementary ways to mitigate this issue exist, including dedicated tracking hardware [18], dedicated algorithms [13], on-the-fly parameter estimation [14] or training a neural network in several feedback conditions to enable self-mixing signal processing in many operation ranges [16]. Here we focus on the case in which speckle (or in fact any other unwanted perturbation) leads to the full degradation of a self-mixing measurement channel.

To simulate this phenomenon, we replace one or the other channel by a Gaussian white noise during a certain time interval (simulating the case of $C = 0$ due to speckle or any other disturbance

for this channel) and feed that modified interferometric data to the 3D model. The top three rows of Fig. 5 show the three self-mixing signals and the bottom row shows the true displacement and the reconstruction. This analysis is performed on the same segment of displacement as in Fig. 2. During the interval $60 < t < 110$ ms, channel 1 is degraded and during the interval $25 < t < 75$ ms channel 2 is degraded, again simulating $C = 0$. As we see, the 3D model always provides a meaningful reconstruction, even in the worst conditions such as the central region, during the interval $60 < t < 75$ ms where two channels out of three are unusable. Most importantly for a high-availability measurement system, the 3D model transparently makes optimal use of the available information but also the optimal quality of the reconstruction is recovered as soon as the quality of the self-mixing data itself is restored.

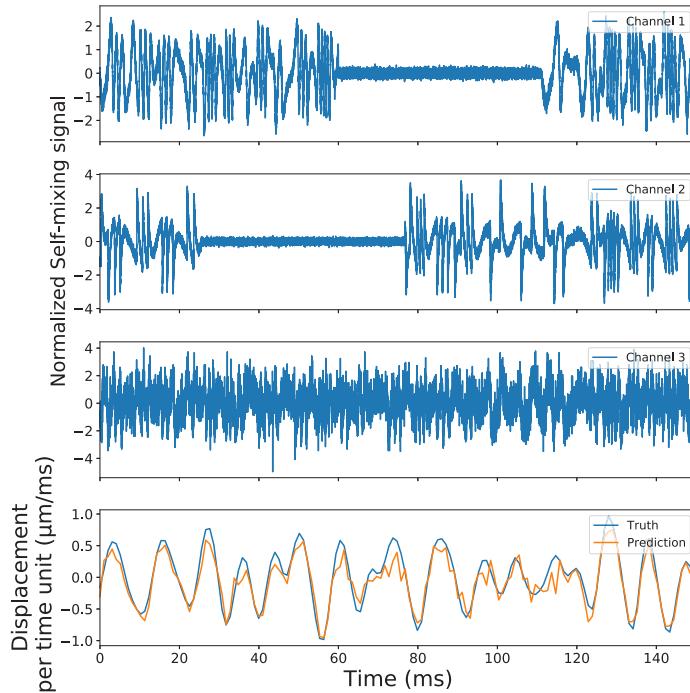


Fig. 5. Channel loss and measurement availability. At different times, one or more self-mixing signals (top three rows) are replaced by white noise, simulating $C = 0$ as could be caused for instance by speckle. The 3D model transparently makes use of the available data and provides (bottom row) a meaningful reconstruction of the displacement even in the worst case scenario of two channels becoming unavailable simultaneously.

4. Conclusion

In conclusion, we have analyzed the performance of a high-availability displacement sensor based on three independent self-mixing interferometry channels processed by a simple (in terms of architecture, number of layers and trainable weights, see Appendix) neural network. Thanks to the inherent capacity of neural networks to process high-dimensional input, the multichannel sensor can transparently make optimal use of the self-mixing signal, using one or more input channels depending on their availability. The multichannel system performance nearly matches that of the best quality channel in absence of any disruption and as soon as one channel is degraded the multichannel sensor outperforms any single-channel sensor. Even when two channels are entirely degraded, the multichannel sensor still provides reliable displacement inference based on the only

remaining channel. The network is trained to infer a displacement without relying on physical modeling, therefore the approach does not require any parameter estimation. Together with the ability of a single neural network to process many shapes of self-mixing signals (including those obtained from different experimental set-ups) [16], we believe that the multi-channel approach constitutes a key element in solving the long-standing issue of speckle-affected self-mixing interferometry with non-cooperative targets.

Since the neural network is purposefully designed to process self-mixing data with near-minimal number of parameters, the network is amenable to embedding on tiny computing devices [37,38], which opens the way towards small-footprint and low-power smart sensors leveraging the intrinsic simplicity of self-mixing interferometry.

For future work, we underline that the approach outlined here is only meant as a proof of concept. First about the neural network itself: as in [16], the network is extremely basic and can certainly be improved. In particular, we expect that more advanced channel fusion could provide better reconstruction, in particular when all channels are degraded. Also about the photonic stages, the present work opens perspectives along the lines of more advanced and multimodal sensing [11,39], perhaps including compressed sensing [40] for onboarding on low-footprint components.

Appendix

A.1 Network details

The structure of the network is identical to that of [16] albeit with a larger number of convolutional kernels so as to accommodate three measurement channels instead of one. The key elements of the network are shown on Table 1. The network is implemented thanks to the Keras library [41] and we refer the reader to deep learning fundamentals [42] and implementations [41] for background information. The total number of trainable parameters is 207 905. Networks of identical architecture with more cells per layer did not lead to significant improvements. The training of the network takes about twenty minutes on a consumer-grade GPU (NVIDIA GeForce GTX 1080).

A.2 Data pre-processing

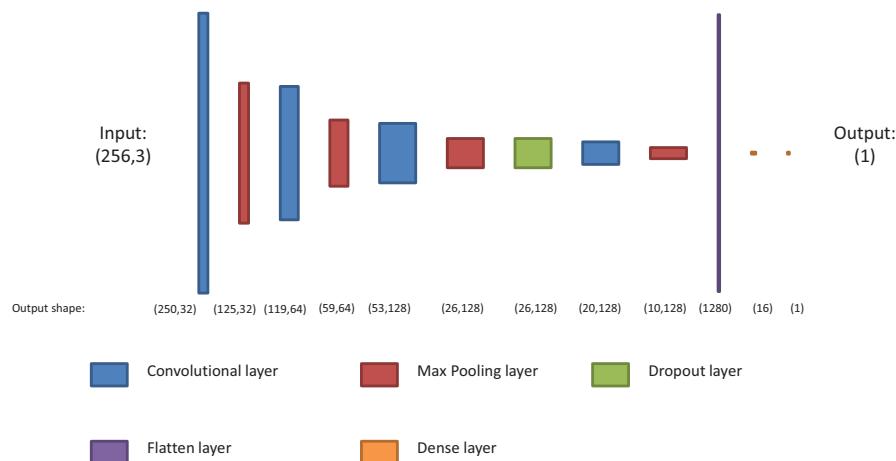


Fig. 6. Schematic view of the model structure, see main text for details.

Disclosures. The authors declare no conflicts of interest.

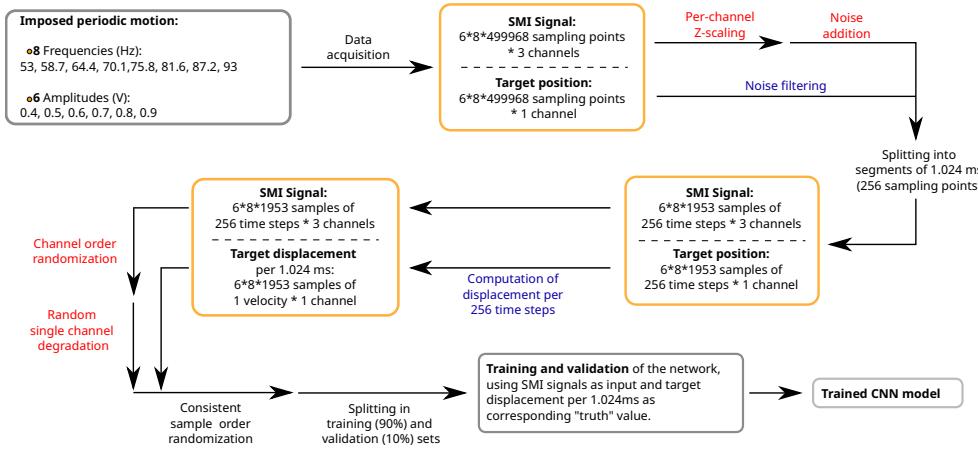


Fig. 7. Synthetic view of the preparation of data for training. Operations marked in red, blue and black act respectively on the self mixing signal, the displacement and both, see main text for a complete description.

Table 1. Main parameters of the network used in this work. The network is a sequence of 1-dimensional convolutional and dropout layers followed by two fully connected layers for the final regression. The total number of trainable parameters is 207 905.

Network structure: sequential				
Layer type	Main hyperparameters	Trainable Parameters	Output shape	Activation function
1D convolutional	kernel size: 7 filters: 32	704	(250, 32)	relu
Max Pooling	pool size: 2	0	(125, 32)	
1D convolutional	kernel size: 7 filters: 64	14400	(119, 64)	relu
Max Pooling	pool size: 2	0	(59, 64)	
1D convolutional	kernel size: 7 filters: 128	57472	(53, 128)	relu
Max Pooling	pool size: 2	0	(26, 128)	
Dropout	10%	0	(26, 128)	
1D convolutional	kernel size: 7 filters: 128	114816	(20, 128)	relu
Max Pooling	pool size: 2	0	(10, 128)	
Flatten		0	1280	
Fully connected	units: 16	20495	16	relu
Fully connected	units: 1	17	1	linear

Data availability. The neural network training and test data are available in [36].

References

1. G. Nicolis and C. Nicolis, *Foundations of complex systems: emergence, information and prediction* (World Scientific, 2012).
2. P. Caramazza, A. Boccolini, D. Buschek, M. Hullin, C. F. Higham, R. Henderson, R. Murray-Smith, and D. Faccio, “Neural network identification of people hidden from view with a single-pixel, single-photon detector,” *Sci. Rep.* **8**(1), 11945 (2018).
3. M. Del Hougne, S. Gigan, and P. Del Hougne, “Deeply subwavelength localization with reverberation-coded aperture,” *Phys. Rev. Lett.* **127**(4), 043903 (2021).
4. D. Mengu, M. S. S. Rahman, Y. Luo, J. Li, O. Kulce, and A. Ozcan, “At the intersection of optics and deep learning: statistical inference, computing, and inverse design,” *Adv. Opt. Photonics* **14**(2), 209–290 (2022).
5. G. Giuliani, M. Norgia, S. Donati, and T. Bosch, “Laser diode self-mixing technique for sensing applications,” *J. Opt. A: Pure Appl. Opt.* **4**(6), S283–S294 (2002).
6. D. M. Kane and K. A. Shore, *Unlocking dynamical diversity: optical feedback effects on semiconductor lasers* (John Wiley & Sons, 2005).
7. S. Donati, “Developing self-mixing interferometry for instrumentation and measurements,” *Laser Photonics Rev.* **6**(3), 393–417 (2012).
8. T. Taimre, M. Nikolić, K. Bertling, Y. L. Lim, T. Bosch, and A. D. Rakić, “Laser feedback interferometry: a tutorial on the self-mixing effect for coherent sensing,” *Adv. Opt. Photonics* **7**(3), 570–631 (2015).
9. J. Li, H. Niu, and Y. X. Niu, “Laser feedback interferometry and applications: a review,” *Opt. Eng.* **56**(5), 050901 (2017).
10. A. Rakić, T. Taimre, K. Bertling, Y. Lim, P. Dean, A. Valavanis, and D. Indjin, “Sensing and imaging using laser feedback interferometry with quantum cascade lasers,” *Appl. Phys. Rev.* **6**(2), 021320 (2019).
11. M. Brambilla, L. L. Columbo, M. Dabbicco, F. De Lucia, F. P. Mezzapesa, and G. Scamarcio, “Versatile multimodality imaging system based on detectorless and scanless optical feedback interferometry—a retrospective overview for a prospective vision,” *Sensors* **20**(20), 5930 (2020).
12. K. Bertling, X. Qi, T. Taimre, Y. L. Lim, and A. D. Rakić, “Feedback regimes of lfi sensors: Experimental investigations,” *Sensors* **22**(22), 9001 (2022).
13. A. A. Siddiqui, U. Zabit, O. D. Bernal, G. Raja, and T. Bosch, “All Analog Processing of Speckle Affected Self-Mixing Interferometric Signals,” *IEEE Sens. J.* **17**(18), 5892–5899 (2017).
14. O. D. Bernal, U. Zabit, F. Jayat, and T. Bosch, “Toward an estimation of the optical feedback factor c on the fly for displacement sensing,” *Sensors* **21**(10), 3528 (2021).
15. L. An and B. Liu, “Measuring parameters of laser self-mixing interferometry sensor based on back propagation neural network,” *Opt. Express* **30**(11), 19134–19144 (2022).
16. S. Barland and F. Gustave, “Convolutional neural network for self-mixing interferometric displacement sensing,” *Opt. Express* **29**(8), 11433–11444 (2021).
17. S. Barland and F. Gustave, “Displacement measurement via self mixing interferometry and neural network training set,” 10.5281/zenodo.7303745 (2022).
18. O. D. Bernal, U. Zabit, and T. M. Bosch, “Robust method of stabilization of optical feedback regime by using adaptive optics for a self-mixing micro-interferometer laser displacement sensor,” *IEEE J. Sel. Top. Quantum Electron.* **21**(4), 336–343 (2015).
19. F. De Lucia, M. Putignano, S. Ottonelli, M. Di Vietro, M. Dabbicco, and G. Scamarcio, “Laser-self-mixing interferometry in the gaussian beam approximation: experiments and theory,” *Opt. Express* **18**(10), 10323–10333 (2010).
20. S. Ottonelli, M. Dabbicco, F. De Lucia, and G. Scamarcio, “Simultaneous measurement of linear and transverse displacements by laser self-mixing,” *Appl. Opt.* **48**(9), 1784–1789 (2009).
21. S. Ottonelli, M. Dabbicco, F. De Lucia, M. Di Vietro, and G. Scamarcio, “Laser-self-mixing interferometry for mechatronics applications,” *Sensors* **9**(5), 3527–3548 (2009).
22. Y. L. Lim, M. Nikolic, K. Bertling, R. Kliese, and A. D. Rakić, “Self-mixing imaging sensor using a monolithic vcsel array with parallel readout,” *Opt. Express* **17**(7), 5517–5525 (2009).
23. Y. L. Lim, R. Kliese, K. Bertling, K. Tanimizu, P. Jacobs, and A. D. Rakić, “Self-mixing flow sensor using a monolithic vcsel array with parallel readout,” *Opt. Express* **18**(11), 11720–11727 (2010).
24. R. Atashkhooei, S. Royo, and F. J. Azcona, “Dealing with speckle effects in self-mixing interferometry measurements,” *IEEE Sens. J.* **13**(5), 1641–1647 (2013).
25. J. R. Tucker, A. Mowla, J. Herbert, M. A. Fuentes, C. S. Freakley, K. Bertling, Y. L. Lim, R. S. Matharu, J. Perchoux, T. Taimre, S. J. Wilson, and A. D. Rakic, “Self-mixing sensing system based on uncooled vertical-cavity surface-emitting laser array: linking multichannel operation and enhanced performance,” *Opt. Lett.* **39**(2), 394–397 (2014).
26. H. Li, C. Zhang, N. Song, and H. Li, “Deep learning-based interference fringes detection using convolutional neural network,” *IEEE Photonics J.* **11**(4), 1–14 (2019).
27. K. Kou, C. Wang, T. Lian, and J. Weng, “Fringe slope discrimination in laser self-mixing interferometry using artificial neural network,” *Optics Laser Technol.* **132**, 106499 (2020).

28. A. A. Siddiqui, U. Zabit, and O. D. Bernal, "Fringe detection and displacement sensing for variable optical feedback-based self-mixing interferometry by using deep neural networks," *Sensors* **22**(24), 9831 (2022).
29. L. Wei, J. Chicharo, Y. Yu, and J. Xi, "Pre-processing of signals observed from laser diode self-mixing interreferometries using neural networks," in *2007 IEEE International Symposium on Intelligent Signal Processing*, (IEEE, 2007), pp. 1–5.
30. I. Ahmed, U. Zabit, and A. Salman, "Self-mixing interferometric signal enhancement using generative adversarial network for laser metric sensing applications," *IEEE Access* **7**, 174641–174650 (2019).
31. H. Liang, M. Chen, C. Jiang, L. Kan, and K. Shao, "Combined feature extraction and random forest for laser self-mixing vibration measurement without determining feedback intensity," *Sensors* **22**(16), 6171 (2022).
32. K. Gadzicki, R. Khamsehashari, and C. Zetzsche, "Early vs late fusion in multimodal convolutional neural networks," in *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, (IEEE, 2020), pp. 1–6.
33. L. Khacef, L. Rodriguez, and B. Miramond, "Brain-inspired self-organization with cellular neuromorphic computing for multimodal unsupervised learning," *Electronics* **9**(10), 1605 (2020).
34. G. M. Dimitri, "A short survey on deep learning for multimodal integration: Applications, future perspectives and challenges," *Computers* **11**(11), 163 (2022).
35. P. P. Liang, A. Zadeh, and L.-P. Morency, "Foundations and recent trends in multimodal machine learning: Principles, challenges, and open questions," *arXiv* arXiv:2209.03430 (2022).
36. R. Matha, S. Barland, and F. Gustave, "Multichannel displacement measurement via self mixing interferometry and neural network : training and test datasets," Zenodo (2023) <https://10.5281/zenodo.7554008>.
37. P.-E. Novac, G. Boukli Hacene, A. Pegatoquet, B. Miramond, and V. Gripon, "Quantization and deployment of deep neural networks on microcontrollers," *Sensors* **21**(9), 2984 (2021).
38. S. S. Saha, S. S. Sandha, and M. Srivastava, "Machine learning for microcontroller-class hardware—a review," *arXiv* arXiv:2205.14550 (2022).
39. L. Columbo, M. Brambilla, M. Dabbicco, and G. Scamarcio, "Self-mixing in multi-transverse mode semiconductor lasers: model and potential application to multi-parametric sensing," *Opt. Express* **20**(6), 6286–6305 (2012).
40. L. Li, Y. Zhang, Y. Zhu, Y. Dai, X. Zhang, and X. Liang, "Absolute distance measurement based on self-mixing interferometry using compressed sensing," *Appl. Sci.* **12**(17), 8635 (2022).
41. F. Chollet and others., "Keras," <https://github.com/fchollet/keras> (2015).
42. I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, vol. 1 (MIT Press Cambridge, 2016).