Journal of Instrumentation

PAPER • OPEN ACCESS

Machine learning for precise hit position reconstruction in Resistive Silicon Detectors

Published by IOP Publishing for Sissa Medialab

Received: *November 14, 2023*
Accepted: *December 11, 2023*
Published: *January 18, 2024*

16th Topical Seminar on Innovative Particle and Radiation Detectors
September 2023
Siena

Machine learning for precise hit position reconstruction in Resistive Silicon Detectors

F. Siviero,[a,*] **R. Arcidiacono,**[a,c] **N. Cartiglia,**[a] **M. Costa,**[a,b] **M. Ferrero,**[a,c] **L. Lanteri,**[b] **C. Madrid,**[d] **L. Menzio,**[a] **R. Mulargia**[a,b] **and V. Sola**[a,b]

[a] *INFN Sezione di Torino, Via Giuria 1, Torino, Italy*
[b] *Università degli Studi di Torino, Via Giuria 1, Torino, Italy*
[c] *Università degli Studi del Piemonte Orientale, Via del Duomo 6, Vercelli, Italy*
[d] *Fermi National Accelerator Laboratory, Batavia, IL, U.S.A.*

E-mail: federico.siviero@to.infn.it

Abstract: RSDs are LGAD silicon sensors with 100% fill factor, based on the principle of AC-coupled resistive read-out. Signal sharing and internal charge multiplication are the RSD key features to achieve picosecond-level time resolution and micron-level spatial resolution, thus making these sensors promising candidates as 4D-trackers for future experiments. This paper describes the use of a neural network to reconstruct the hit position of ionizing particles, an approach that can boost the performance of the RSD with respect to analytical models. The neural network has been trained in the laboratory and then validated on test beam data. The device-under-test in this work is a 450 µm-pitch matrix from the FBK RSD2 production, which achieved a resolution of about 65 µm at the DESY Test Beam Facility, a 50% improvement compared to a simple analytical reconstruction method, and a factor two better than the resolution of a standard pixel sensor of equal pitch size with binary read-out. The test beam result is compatible with the laboratory ones obtained during the neural network training, confirming the ability of the machine learning model to provide accurate predictions even in environments very different from the training one. Prospects for future improvements are also discussed.

Keywords: Analysis and statistical methods; Data reduction methods; Si microstrip and pad detectors; Solid state detectors

*Corresponding author.

https://doi.org/10.1088/1748-0221/19/01/C01028

2024 JINST 19 C01028

Contents

1 Introduction

Resistive Silicon Detectors (RSDs, also known as AC-LGAD) are ~50 µm-thick silicon sensors based on the LGAD (Low-Gain Avalanche Diode) technology [1], implementing the principle of AC-coupled resistive read-out [2, 3].

In the RSD design, sketched in figure 1, metal read-out pads are coupled to the sensor through an oxide layer, while a resistive n^+ layer implanted underneath allows the sharing of charge among multiple read-out pads. The gain implant, typical of LGAD sensors, spreads across the whole active area, allowing for the multiplication of primary electrons. The key ingredients of the RSD design are thus the internal signal sharing combined with the internal charge multiplication. A detailed description of the RSD design can be found in [4, 5].

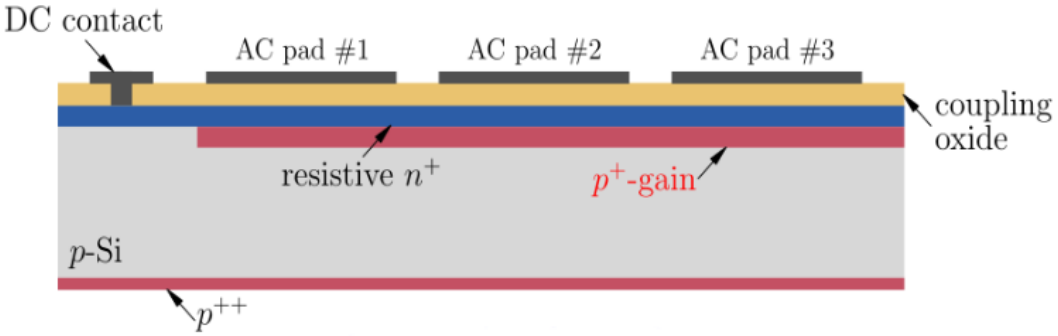

Figure 1. Sketch of the RSD design.

RSDs can provide a spatial resolution as low as 3% of their pitch [6], to be compared to about 30% of the pitch in standard silicon pixel sensors with binary read-out: combining the information brought by multiple read-out channels, exploiting in this way signal sharing, is the key to achieving such accurate position reconstruction. Moreover, the RSDs are LGADs, hence capable of providing a time resolution of 30–40 ps [6], whereas standard pixel sensors are known to have rather poor timing performance. Finally, the RSD design has 100% fill-factor. Because of such features, RSDs are particularly suited as 4D-trackers for future high-energy physics experiments [7].

The device under test (DUT) in this work comes from the RSD2 production, manufactured by Fondazione Bruno Kessler (FBK, Italy) [8]; it is a 6×6 matrix with an active area of 2.6×2.6 mm^2 and 450 µm pitch (figure 2). As a defining characteristic of the sensor design, the metal electrodes are

2024 JINST 19 C01028

designed to be a "cross-shape", in order to minimize the area covered by metal and hence improve the reconstruction uniformity: indeed, if the metal pads are large, signal sharing does not occur (or it is reduced) when the particle hits the metal (only the hit pad sees the signal in that case).

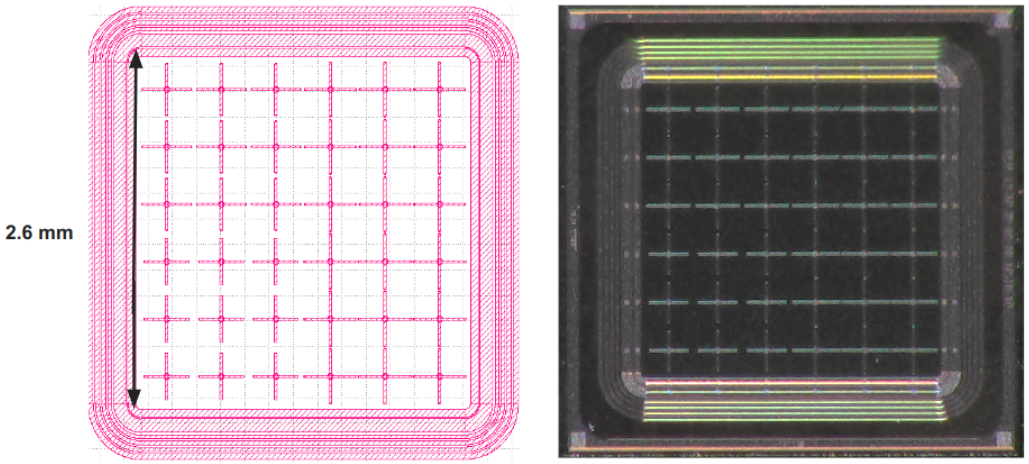

Figure 2. The drawing (left) and photograph (right) of the RSD2 sensor.

In the following section, methods of position reconstruction with RSDs will be described, focusing in particular on machine learning techniques. The subsequent sections will describe the data collection for the DUT, the machine learning training, and experimental results. Finally, a summary and discussion of future improvements will be presented.

2 Position reconstruction with machine learning

Position reconstruction with RSDs is based on the extraction of valuable information from the signals read out by each electrode, which can then be used to infer the $x - y$ coordinates of the particle hit position.

Analytical models that can accurately infer the position do exist [9–11]: they work well in a simple 4-electrode version of the RSD design, but become less and less accurate as the number of read-out pads increase, as would be the case in a realistic scenario.

Machine learning algorithms are a natural alternative choice [10, 12]: they can be trained with input features extracted from each read-out signal, with no need to know the underlying sharing laws; once the algorithm has been trained, it can be used to predict the hit positions.

The amplitudes of the positive and negative lobes (the RSD signal is bipolar, as shown in figure 3) of the signals read out by each electrode are used as input features in this work.

A fully-connected dense neural network has been chosen as reconstruction algorithm. Simpler algorithms have been used in previous works [12, 13]: they work well but have been found to be less performing than neural networks.

The neural network has been developed with the PyTorch framework [14], which offers flexible and tunable features that can be tailored to a specific task.

The neural network used in this work consists of 6 hidden layers with 36 nodes each; the Adam optimizer is used as stochastic gradient descent method [15]. A dropout and a regularization layer

2024 JINST 19 C01028

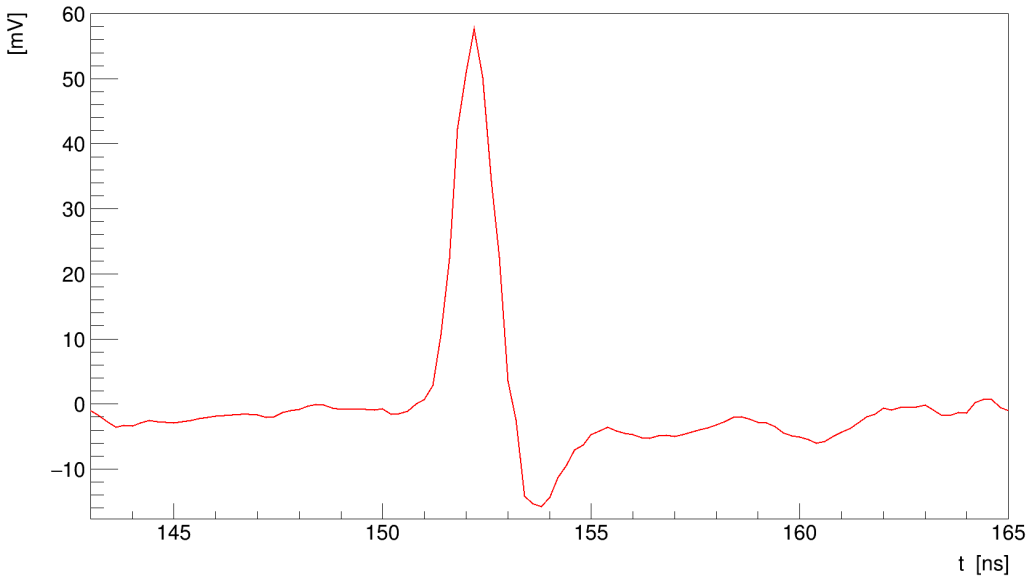

Figure 3. A typical RSD signal.

have been introduced, mitigating in this way overfitting and ruling out most of the outliers in the final distribution of the predicted positions.

The neural network has been trained with laboratory data, taken with a precise TCT (Transient Current Technique) setup [16], where a near-IR laser simulates the passage of an ionizing particle through the device and a movable $x - y$ translation stage provides the laser reference position with a resolution that was measured to be $\sim 2\,\mu$m. The measurement procedure and data acquisition are the same as those described in [13].

Once the model has been trained, it has been tested on a different dataset, taken at the DESY Test Beam Facility using 4 GeV/c electrons. The facility, described in [17], is equipped with an EUDET-type pixel telescope [18] that provides the reference positions to be compared with the algorithm predictions. During the measurements presented in this work, the telescope spatial resolution was measured to be of about 15 µm.

Figure 4 presents the loss curves[1] of the training and test datasets as a function of the training epochs (one epoch represents one passage of the entire dataset through the algorithm): given the low values of the loss on both training and test, this plot proves both the goodness of the chosen model (training loss minimum is low) and its ability to generalize the predictions to different scenarios (test loss minimum is low, too).

Several reasons led to the decision to train the reconstruction algorithm with a dataset very different from the test one: (i) laboratory datasets may have millions of events, whereas test beam ones available in this work were limited to a few tens of thousands at most; (ii) laboratory data are taken in ideal conditions: the uncertainty on the reference positions provided by the TCT setup is $\sim 2\,\mu$m, to be compared with the 15 µm of the tracker at DESY; (iii) training with a different set of data is a way to prove the generalization power of the model: the goal is to develop an algorithm

[1]The loss is a parameter describing how well the model fits the data; the lower it is the better the agreement between model's predictions and data.

2024 JINST 19 C01028

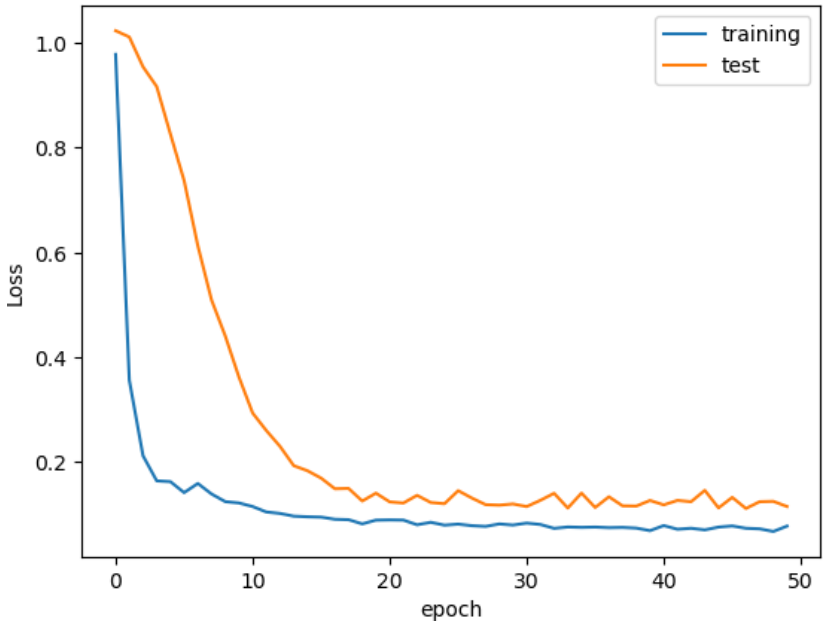

Figure 4. Train and test loss curves as a function of the training epochs.

that can be trained once in the laboratory and then used in very different scenarios, without having to re-train it every time. Proving such generalization power (i.e. obtaining test beam results as good as the laboratory ones) is one of the main goals of this work.

3 Setup and results

The DUT is an RSD2 6×6 matrix, however, only 9 pads, arranged in a 3×3 matrix, have been read out (figure 5). Since position reconstruction in RSD works using those pads that see a signal above the noise level (those closer to the hit position), the whole sensor active area could not be used: only positions within the region framed in blue in figure 5 have been considered, as positions outside would require using a different set of read-out pads.

In order to actually limit the reconstruction to the area framed in blue, one has to require that the central pad in the 3×3 matrix (in pink in figure 5) sees the highest signal among all read-out channels, and such signal has to be larger than 6 mV, to reject most of the noise events effectively, given that the RMS noise is 1–1.5 mV (figure 6). In the end, the study was limited to a region of interest of about 500×500 μm^2, framed in pink in figure 5.

The results presented in the following will thus concern a fraction of the whole active area, requiring a specific selection of the events: the final results, those obtained by the full 6×6 matrix, may be different.

The laser training of the neural network has been performed in the laboratory, shooting the laser of the TCT setup in different positions, arranged on a grid within the region of interest previously defined. The laser has been shot 100 times in each position, to let the model better grasp the signal fluctuations due to the electronics noise; that is a way of adding variance to the model, an important aspect that can help generalize it [19]. Once the training is done, the model is asked to predict the very same positions used for its training, in order to check its consistency.

2024 JINST 19 C01028

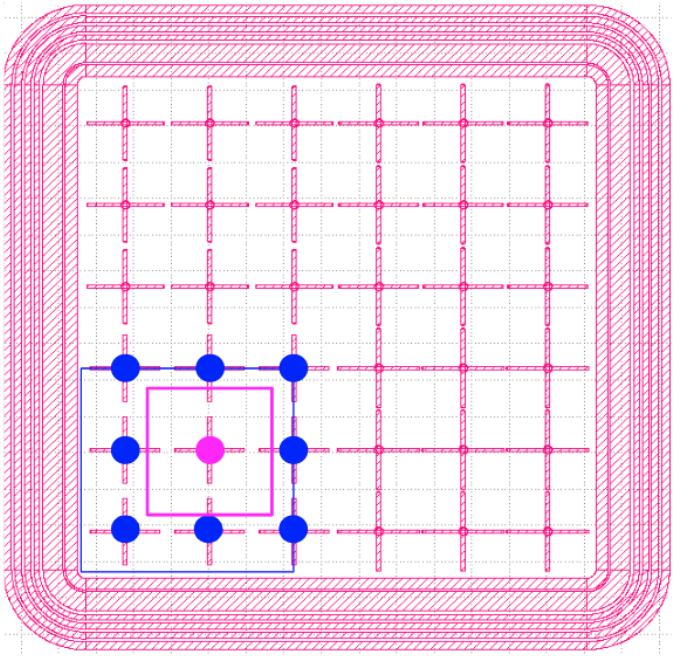

Figure 5. The RSD2 sensor tested in this work, with the 9 read-out pads highlighted. The pink electrode is required to see the highest signal among all pads, limiting the reconstruction to the area framed in blue. The pink electrode is also required to see a signal above 6 mV to reject noise, therefore eventually the region of interest is limited to the 500 µm x 500 µm area framed in pink.

As shown in figure 7, the predicted positions (orange) are not arranged on the same grid as the reference ones (blue), rather they make a different pattern that, in its central part, resembles that of a cross. This can be explained by the electrode design, based on squares: indeed, pads in the corner of the 3×3 matrix are further away by a factor $\sqrt{2}$ from the central pad than the others, and consequently are less important in the reconstruction, because, on average, they record lower signals. The reconstructed positions thus tend to be pulled towards the center of the region of interest, closer to those pads seeing higher signals, resulting in a different pattern than that of the reference positions. In order to improve the reconstruction, future RSD productions are being designed with electrodes arranged in triangles, so that they are all equidistant.

The trained neural network has been eventually tested on the DESY dataset, finding, as shown in figure 8, a total resolution of ~ 65 µm; this is the squared sum of the RSD and tracker resolution, but the tracker contribution (~ 15 µm) can be neglected. Also in this case, the positions in the corner tend to be pulled towards the center of the region of interest, reflecting the pattern already observed in the training data.

As a comparison, it is possible to assess the results achieved by the same sensor using a simple analytic model based on the centroid, i.e. a model where the predicted position can be computed as:

$$x_{\text{RSD-centroid}} = \frac{\sum_{i=1}^{9} A_i \cdot x_i}{\sum_{i=1}^{9} A_i} \tag{3.1}$$

with A_i, x_i being the amplitude and position of the $i-th$ electrode. The reference and reconstructed positions are shown in figure 9 left: the reconstructed positions cluster in the center of the region

2024 JINST 19 C01028

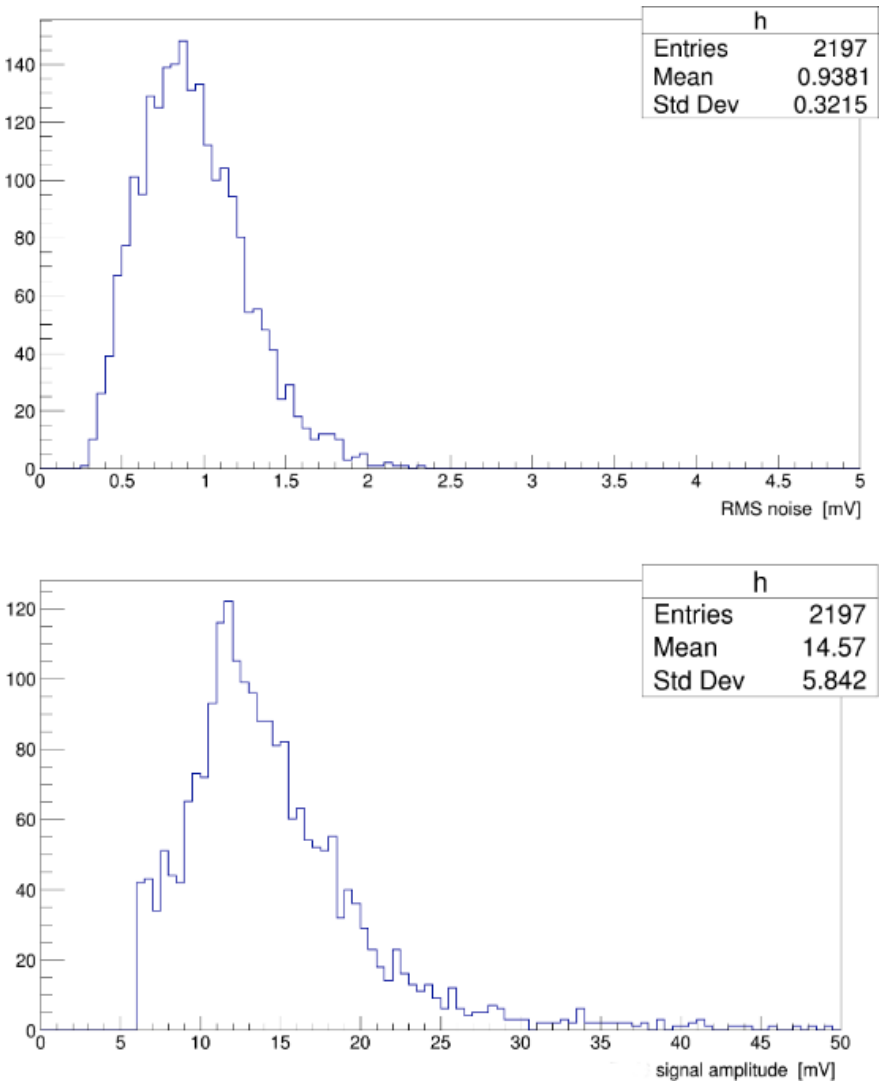

Figure 6. The RMS noise (top) and signal amplitude (bottom) distributions of the central pad in the 3×3 matrix of the DUT.

of interest, given the requirement that the central pad sees the highest signal. The resolution (see figure 9 right) is about 95 µm, ~ 50% larger than the machine learning result.[2]

A standard pixel sensor with binary read-out and the same pitch size as the DUT would instead provide a spatial resolution of ~ 130 µm, twice as large as the RSD resolution achieved using the neural network.

Figure 10 compares the results (dots) obtained with the DUT in the laboratory (using the TCT laser setup, see [9]) to the test beam result (green cross), as a function of the signal amplitude: the resolution achieved at the test beam is similar to that obtained in the laboratory for an equivalent signal amplitude, despite the very different conditions found at the test beam. The result demonstrates the validity of our approach and the generalization power of the model, which can provide accurate predictions beyond a specific training scenario.

[2]The centroid is the simplest reconstruction method that can be used with a 3×3 matrix: more sophisticated and performing analytical methods may be developed in the future.

2024 JINST 19 C01028

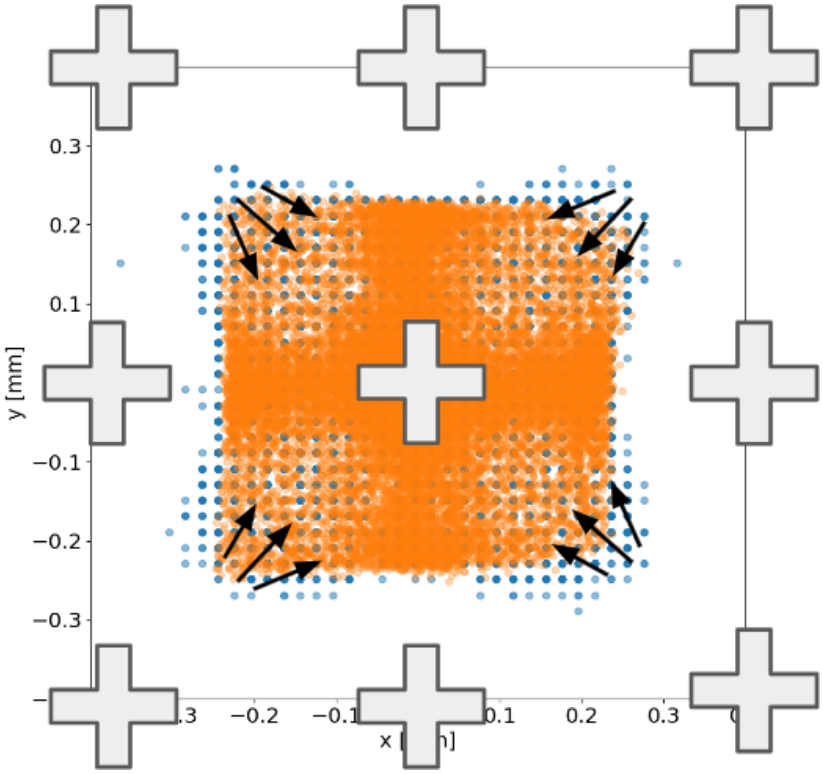

Figure 7. Reference (blue) and predicted (orange) positions of the training dataset, taken with the TCT laser setup. The nine electrodes used for the reconstruction are superimposed on the 2D map representing the region of interest. The black arrows in the corners represent the shift of the positions in the corners, caused by the electrodes pattern.

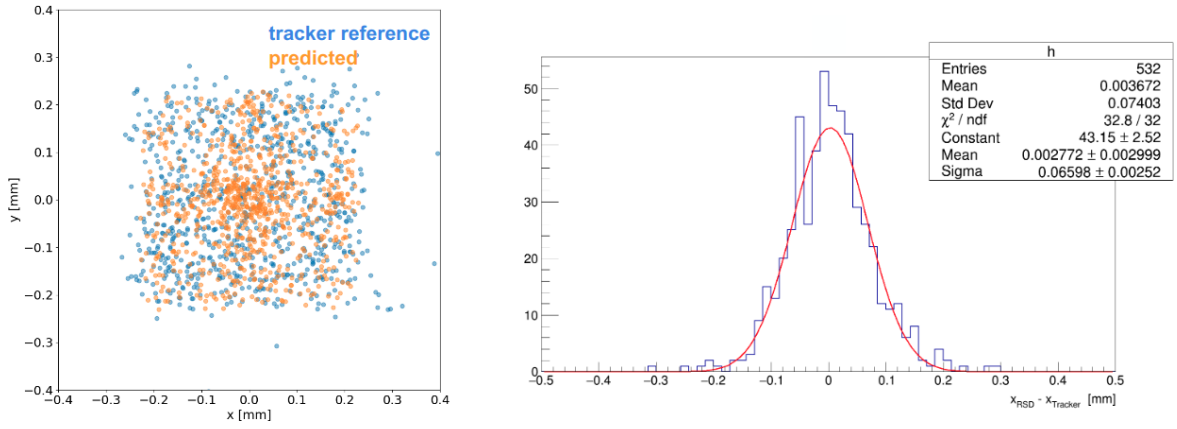

Figure 8. Left: 2D map comparing the reference (blue) and predicted (orange) position of the test dataset acquired at DESY. Right: difference between the positions predicted (x_{RSD}) by the machine learning model and the reference (x_{Tracker}) positions; the standard deviation of the distribution represents the squared sum of the RSD and tracker (negligible) resolutions.

2024 JINST 19 C01028

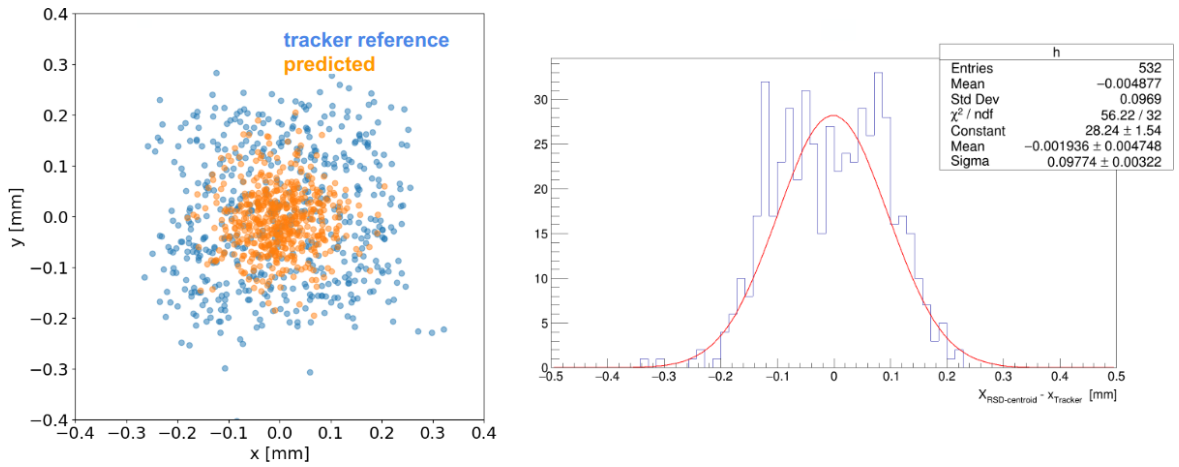

Figure 9. Left: 2D map comparing the reference (blue) and predicted (orange) position of the test dataset acquired at DESY. Right: difference between the positions predicted ($x_{\text{RSD-centroid}}$) by the centroid analytical model and the reference (x_{Tracker}) positions; the standard deviation of the distribution represents the squared sum of the RSD and tracker (negligible) resolutions.

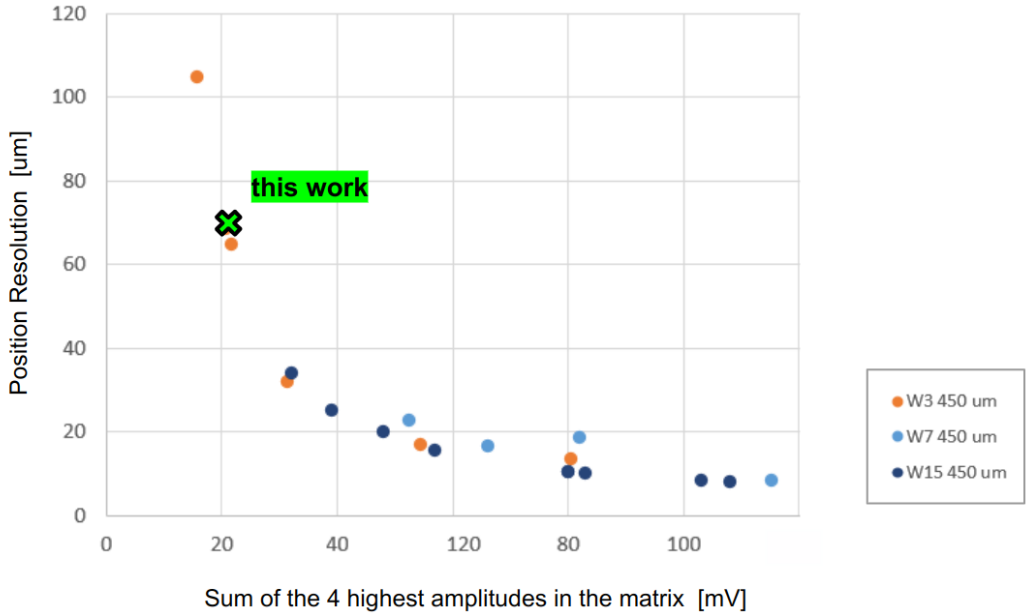

Figure 10. The DUT resolution as a function of the total amplitude of the 4 pads seeing the highest signals, as measured with the TCT laser in the laboratory (dots) and at the DESY test beam (cross). Figure taken from [9].

Figure 10 also shows that by increasing the signal amplitudes (i.e. increasing gain), the resolution can improve, plateauing at 10–15 µm. Different sensor designs and new read-out boards will be used in future test beams to achieve higher signals and further push the performance of the RSDs.

2024 JINST 19 C01028

4 Conclusions

RSDs are innovative LGAD silicon sensors with 100% fill factor, implementing the principle of the AC-coupled resistive read-out. RSDs are suited as 4D-trackers for future experiments, able to provide picosecond-level time resolution and micron-level spatial resolution.

Although accurate analytical models do exist, the reconstruction of the hit position of ionizing particles with RSD is enhanced by using machine learning techniques. In particular, those techniques are particularly helpful when many pads are involved in the reconstruction process, as is the case in a realistic detector.

The sensor presented in this work comes from the FBK RSD2 production, and features cross-shaped metal electrodes (needed to minimize the fraction of active area covered by metal) and 450 µm pitch size. It achieved a resolution, measured during a test beam at the DESY Test Beam Facility using an EUDET-type pixel telescope as reference, of about 65 µm. Such resolution is 50% better than the resolution achieved by the same sensor using a simple analytical model for the reconstruction, and a factor two better than what would be provided by a standard silicon pixel sensor with the same pitch and binary read-out. This result was achieved considering only a fraction of the sensor active area, requiring specific event selection: once a dedicated read-out board is produced (work in progress), it will be possible to move to a more realistic scenario where all channels are read out, possibly leading to different results.

A goal of this work was to prove that the machine learning model can be trained in the laboratory (using a laser) and then used to make reliable predictions in very different scenarios: this has been achieved, as the resolution measured during the test beam is equal to that obtained in ideal conditions in the laboratory for the same signal amplitude.

Laboratory results also show that by increasing signals the RSD resolution can drop significantly, reaching its minimum at 10–15 µm (the 3% of the sensor pitch): future studies will be done at higher signal amplitudes, testing the limit of validity of the models developed in this paper. Intense R&D is being carried out on the RSD project to achieve that: new RSD productions are being manufactured, as well as dedicated read-out boards, both to be tested in future test beams.

Acknowledgments

The measurements leading to these results have been performed at the Test Beam Facility at DESY Hamburg (Germany), a member of the Helmholtz Association (HGF). The research leading to these results has received funding from the European Union's Horizon Europe research and innovation program under grant agreement no. 101057511. We kindly acknowledge the following funding agencies and collaborations: INFN-CSN5, RSD Project, FBK-INFN collaboration framework, MUR PRIN project 4DInSiDe, Dipartimenti di Eccellenza, Torino University (ex L.232/2016, art. 1, cc. 314, 337).

2024 JINST 19 C01028

References

[1] G. Pellegrini et al., *Technology developments and first measurements of Low Gain Avalanche Detectors (LGAD) for high energy physics applications*, *Nucl. Instrum. Meth. A* **765** (2014) 12.

[2] N. Cartiglia et al., *LGAD design for Future Particle Trackers*, *Nucl. Instrum. Meth. A* **979** (2020) 164383.

[3] N. Cartiglia and M. Mandurrino, *Innovative Silicon Sensors for Future Trackers*, in *CERN Detector Seminar* (2020) [https://indico.cern.ch/event/915984/].

[4] M. Mandurrino et al., *Analysis and numerical design of Resistive AC-Coupled Silicon Detectors (RSD) for 4D particle tracking*, *Nucl. Instrum. Meth. A* **959** (2020) 163479.

[5] M. Mandurrino et al., *Demonstration of 200, 100, and 50 µm pitch Resistive AC-Coupled Silicon Detectors (RSD) with 100% fill-factor for 4D particle tracking*, *IEEE Electron Device Lett.* **40** (2019) 1780.

[6] M. Tornago et al., *Resistive AC-Coupled Silicon Detectors: principles of operation and first results from a combined analysis of beam test and laser data*, *Nucl. Instrum. Meth. A* **1003** (2021) 165319 [arXiv:2007.09528].

[7] ECFA Detector R&D Roadmap Process Group, *The 2021 ECFA detector research and development roadmap*, CERN-ESU-017 (2020) [DOI:10.17181/CERN.XDPL.W2EX].

[8] M. Mandurrino et al., *The second production of RSD (AC-LGAD) at FBK*, 2022 *JINST* **17** C08001 [arXiv:2111.14235].

[9] L. Menzio et al., *Latest results on RSD2 performances, a lab update*, in the proceedings of the *18th TREDI workshop*, Trento, Italy (2023) [https://indico.cern.ch/event/1223972/contributions/5262032/attachments/2602158/4493587/Menzio_TREDI_2023.pdf].

[10] R. Heller et al., *Characterization of BNL and HPK AC-LGAD sensors with a 120 GeV proton beam*, 2022 *JINST* **17** P05001 [arXiv:2201.07772].

[11] C. Madrid et al., *First survey of centimeter-scale AC-LGAD strip sensors with a 120 GeV proton beam*, 2023 *JINST* **18** P06013 [arXiv:2211.09698].

[12] F. Siviero et al., *First application of machine learning algorithms to the position reconstruction in Resistive Silicon Detectors*, 2021 *JINST* **16** P03019 [arXiv:2011.02410].

[13] F. Siviero et al., *First experimental results of the spatial resolution of RSD pad arrays read out with a 16-ch board*, *Nucl. Instrum. Meth. A* **1041** (2022) 167313 [arXiv:2204.06388].

[14] https://pytorch.org.

[15] D.P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv:1412.6980.

[16] http://particulars.si.

[17] R. Diener et al., *The DESY II Test Beam Facility*, *Nucl. Instrum. Meth. A* **922** (2019) 265 [arXiv:1807.09328].

[18] H. Jansen et al., *Performance of the EUDET-type beam telescopes*, *EPJ Tech. Instrum.* **3** (2016) 7 [arXiv:1603.09669].

[19] P. Mehta et al., *A high-bias, low-variance introduction to Machine Learning for physicists*, *Phys. Rept.* **810** (2019) 1 [arXiv:1803.08823].

2024 JINST 19 C01028