TU Delft

Delft University of Technology

Document Version
Final published version

Licence
Dutch Copyright Act (Article 25fa)

Citation (APA)
Gao, Z., Daalman, A., Mubarak, F., Gao, C., De Martino, C., Lehmann, S., & Spirito, M. (2025). Automatic On-Wafer Probe Positioning System Based on Convolutional Neural Network Model. *ARFTG Microwave Measurement Conference, ARFTG*, (2025). https://doi.org/10.1109/ARFTG65332.2025.11168138

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

Automatic On-Wafer Probe Positioning System Based on Convolutional Neural Network Model

Z. Gao*, A. Daalman*, F. Mubarak†, C. Gao*, C. de Martino*§ S. Lehmann‡, M. Spirito*
*Delft University of Technology, Delft, The Netherlands
†Van Swinden Laboratorium (VSL), Delft, The Netherlands
§Vertigo Technologies B.V., Delft, The Netherlands
‡GlobalFoundries, Dresden, Germany

2025 105th ARFTG Microwave Measurement Conference (ARFTG) | 979-8-3315-9529-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/ARFTG65332.2025.11168138

***Abstract*—In this work we present an automatic probe positioning system based on a commercial station environment and employing open loop piezo actuators for probe control. The proposed work utilizes open source convolutional neural networks (CNNs) to identify and track the probe and the landing pad. Benchmarking of the various software modules composing the automatic station is presented in terms of their speed, and absolute position tracking control. Moreover, a procedure to define t he " zero-landing" h eight i s p roposed a nd u sed to evaluate the overall system landing accuracy. The proposed system demonstrates the ability to position the probe on IC aluminum pads with a standard deviation of 1.31 μm and 0.96 μm on the X and Y axes, respectively.**

***Index Terms*—instruments, probes, neural networks, image recognition.**

I. Introduction

An increasing number of commercial applications, such as 6G communication systems and automotive radars are targeting the large available spectra at sub-THz frequencies. In view of this, silicon foundries are pushing to improve their model validation at these frequencies. In order to obtain an accurate indication on the quality of the developed models, accurate S-parameter measurements are required. However, as frequency increases, the wavelength reduction causes an increase in the relative contribution (when compared to other error sources) of the error caused by probe misplacement [1].

Several works have studied (on the impact and minimization) the errors arising from probe misplacement. In [2], a method is proposed to estimate this error and further improve the measurement results by compensating the position of the probe during the measurements. Nevertheless, this method requires monomodal propagation which cannot always be achieved at the pad interface and was only demonstrated during transmission line measurements. Automatic probe systems have been developed in [3] [4] employing S-parameter measurements to enable adjustments on X, Y, Z and also the probe tilt angle. Nevertheless, these works require a vertical displacement of metal artifacts, which is only available on impedance standard substrates (ISS) but not on integrated circuit, where the layers are planarized and the top one is covered by nitride layers for protection. Image processing is also investigated to improve the repeatability of the probe contact position [6]. However, in the work available in the literature, there is a lack of quantified uncertainty of the absolute position that can be achieved using these approaches.

In this contribution we present an automatic probe positioning and landing system based on the combination of computer vision and machine learning models. An open source convolutional neural network (CNN) model is trained and used to identify the location and the centroids of both the pad and the probe objects. This approach allows to iteratively track their centroids relative position and achieve accurate placing of the probe along the X and Y axes. Moreover, we present an approach to employ the ΔS_{11} phase during the Z axis movements to evaluate the $zero - landing$ position and to use it to estimate the overall landing uncertainty.

The implemented CNN models are described in section III and IV. The $zero - landing$ concept and the overall landing uncertainty are presented in Section V.

II. Overview of the Probe Station

A custom probe station was jointly developed between TU Delft and VSL, see Fig. 1. The system is based on a Cascade Microtech Summit 9000 and employs 4 piezo actuators (for X, Y, Z and tilt control) from Newport, for each manipulator, enabling a step resolution of 20 nm [7]. The employed

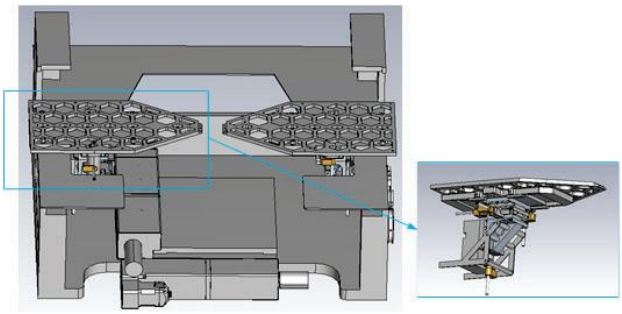


Fig. 1: CAD 3D view of the probe station developed at TU Delft and VSL and used in this work. Inset showing the 3 axis stage plus tilt assembly with the Newport piezo motors.

picomotors are open loop systems which therefore require a vision tracking/control system to implement accurate and repeatable placement. The vision system is realized using a Mitutoyo FS70Z microscope, coupled with a 10x magnification objective. The computer image is acquired via a Basler

979-8-3315-9529-6/25/$31.00 ©2025 IEEE
Authorized licensed use limited to: TU Delft Library. Downloaded on August 06,2026 at 08:11:51 UTC from IEEE Xplore. Restrictions apply.

ace2 usb 3 camera with 24M pixels. The resulting resolution of the vision system is 0.3 μm/pixel.

The data processing hardware is based on a desktop computer with an Intel i9-13900 processor and 16Gb of RAM. The motherboard hosts an Nvidia RTX 3060 GPU to reduce the computational time of the CNN model.

III. Pad Detection Model

The machine learning algorithm to identify the pads relies on collecting and labeling the data for the training. In this setup, the type of pads used is constrained to a single type to establish the most favorable outcome before extending generalization to other pads types. The annotation of the pad images is done through bounding boxes, which are computationally efficient.

This project employs the YOLOv5 [9] model with a configuration that utilizes low data augmentation (for restricted generalization yielding to higher accuracy, in the chosen setup) and large models (76.8 million parameters) to maintain a balance between accuracy and process time. A total of 100 labeled pad images are fed to the model with epochs of 400 iterations. To evaluate the performance of the model, mean Average Precision (mAP) is used as the popular metric for object detection model. The mAP is close to 1 with Intersection over Union Threshold (IoU, intersection between predicted and ground truth bounding boxes) at 0.5, indicating that the model is well trained for practical applications. The steps for both training and model usage, and the estimated processing times are shown in Fig. 2.

To estimate the pad centroid uncertainty that can be achieved with the trained model, 78 different pad photos are used. Each pad is positioned by manually controlling the wafer chuck. This, in combination with the small mechanical vibration of the microscope camera view, results in the pad images not sharing the same coordinate system which then cannot be used for estimating the centroid uncertainty. For this reason, an image registration function is used to align all pads onto the same coordinate system, enabling statistical comparison. The function applies a correction of rotation, scaling and transformation based on the mean squared error of the reference and the test images. Fig. 3 shows an example of the function capability. In the example, an extra 5 degree rotation is applied to the test image for clarity purposes. It is worth noting that, in actual measurements, the rotation angle is less than ±0.04 degrees and the scale factor is within ±0.008, while the major difference lies in the X-Y plane translation. The processed image is limited to a 400x400 pixel size to reduce the processing time while providing enough information for alignment. The σ of the detected pad centroids is 0.69 μm along the X axis and 0.59 μm along the Y axis, see Fig.4.

IV. Probe Centroid Detection Model

The probe identification and model are based on the probe shadow on the substrate. For this reason the probe is kept in close proximity, i.e., circa 5μm, to the flat surface to minimize

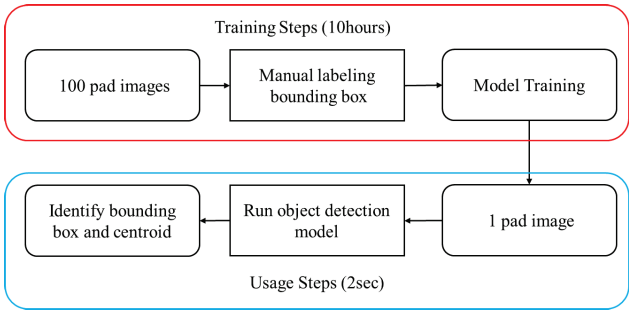


Fig. 2: Flow chart describing the steps and the time duration of the pad model training and the model execution.

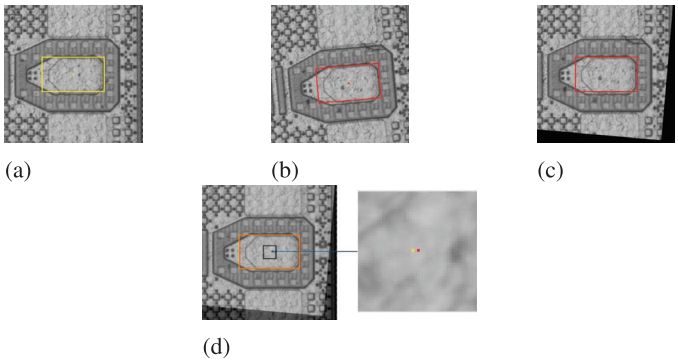


Fig. 3: Example showing the process of pad image registration: a) reference pad, b) test pad with increased rotation of 5 degrees for example purpose, c) alignment after registration, d) overlay of the two pad images with zoom in showing the centroid distance.

the shadow blur. The detection model for the probe is similar to the one used for the pad, which starts from identifying a bounding box that contains the probe shadows for further processing. To reduce the time spent on manual labeling, a set of probe images is taken where the steps only vary along the X axis or Y axis. While the first and last images are labeled manually, all other images can be labeled automatically, based on the knowledge of the motor movements.

The same open source model (YOLOv5) is employed to detect the shadow of the probe. A total number of 100 pictures are used to train the model. The probe detection model also has achieved a performance with mAP(IoU at 0.5) close to 1.0. The steps and estimated processing time are illustrated in Fig. 5.

To further identify a consistent centroid of the signal probe tip over various images (i.e., arising from the probe moving

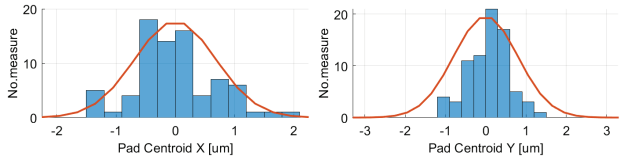


Fig. 4: Pad centroid histogram along a) X and b) Y axis.

Authorized licensed use limited to: TU Delft Library. Downloaded on August 06,2026 at 08:11:51 UTC from IEEE Xplore. Restrictions apply.

towards the pad target) the segment anything model (SAM) [8] is used to recognize the contour of the probe shadow. To increase the accuracy, the probes are restricted to a type of probe that share the same sahpes. Nevertheless, the model is capable of recognizing probes with different pitches without requiring a new training procedure.

To consistently identify a centroid over the segmented probe contour, a circle can be defined around the probe center that best matches the contour of the middle tip using Hough transform algorithms. The probe centroid can be defined as the point on the contour that has the same Y axis coordinates as the circle's centroid, see in Fig. 6.

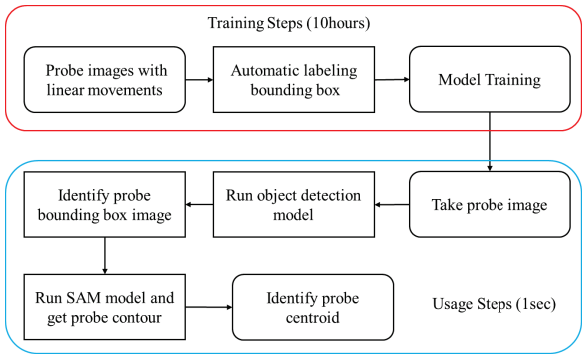


Fig. 5: Flow chart for pad detection model training ans using.



Fig. 6: Probe centroid detection process, images left to right: probe shadow over pad, B/W image, segment anything result, centroid identified using Hough transform, position of the centroid on the probe contour.

To estimate the probe centroid error, a methodology has been developed based on stacking various contours of the probe images and calculating the distance between each centroid, as shown in Fig. 7. The alignment of various contours is achieved using template-matching algorithms [10]. In this measurement, the probe is randomly placed in an area of 120 μm$\times$180μm and a total of 70 photos are taken. Fig. 8 presents the probe centroid error, a standard deviation of 0.86 μm along the X axis and 1.36 μm along the Y axis is obtained.

V. Landing Mark Detection and Zero-Landing Concept

The landing mark can be identified by taking the photo of the pad before and after landing and comparing the difference between those two pictures. However, when considering a mark including a probe skating length the initial landing

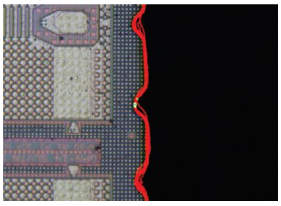
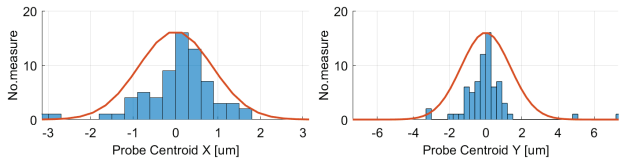


Fig. 7: Template Matching algorithm used on probe contours acquired from different locations.



(a) (b)

Fig. 8: Probe centroid histogram along a) X and b) Y axis.

position of the probe is difficult to identify given the different shapes of the soft aluminum that is moved and accumulated (on the sides) by the probe.

To address this challenge, the concept of zero-landing is introduced. The concept is to produce a mark which is small enough to represent the start point of the probe landing, while being consistently recognized by the image processing steps.

As described in [4] and [5], the measured S11 can be used to estimate the distance between the probe in air and on pad. Fig. 9 reports the uncalibrated phase of ΔS_{11} when landing a probe on pads for different devices(i.e., device, open and short). The VNA is set to 40 GHz with an IF BW of 30 Hz. In this case, we set up 3 thresholds with ΔS_{11} at 0.05, 0.3 and 1 degree (after the peak value) to check the landing marks on the pads.

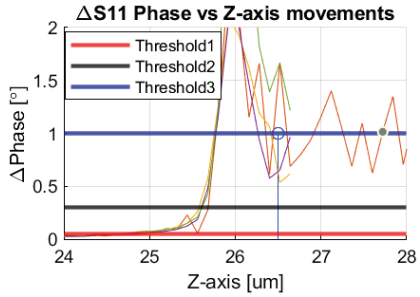


Fig. 9: Delta phase of S_{11} for different devices (i.e., short, open and dut) versus the Z axis (increasing from the start point) highlighting the three threshold levels.

Fig. 10 shows the mark size at different thresholds to identify the initial contact mark of the probe. The gray scale difference of 0.05 used for binary detection allows to identify the marker while filtering out background noises. In addition, marks that are less than 1.8 μm^2 size (20 pixels) are filtered to exclude the noise of dust on different pads. While the mark in threshold 1 is almost invisible, the mark in threshold 3 leads to higher uncertainty in the zero-landing definition due

Authorized licensed use limited to: TU Delft Library. Downloaded on August 06,2026 at 08:11:51 UTC from IEEE Xplore. Restrictions apply.

to its size. To maintain a consistent mark of the pad while keeping the size as small as possible, a ΔS_{11} threshold of 0.3 is chosen when the motor is moving with a step of 120 nm. Once the landing mark is detected, a rectangle bounding box is drawn around the mark, assuming the pad landing is along the X axis. The centroid of the box is recognized as the initial landing position.

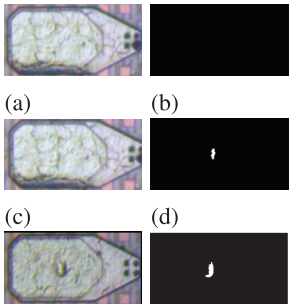


Fig. 10: Landing marks(optical figures) and detected binary marks (binary figures) with threshold 1 (a-b) , 2 (c-d), and 3 (e-f)

To estimate the overall landing uncertainty, the probe is placed at the locations of the red dots shown in Fig. 11, with an initial height around 5 μm. A total of 38 new pads are used to perform zero-landing to derive the statistical data of the final landing position. The mean width of the detected landing bounding boxes is 2 μm and the mean length is 4.5 μm, with standard deviation of 0.87 μm and 0.86 μm respectively. The same image registration function is used to align different pad photos as described in Section III. The final landing mark uncertainty results in a standard deviation of 1.31 μm along the X axis and 0.97 μm along the Y axis.

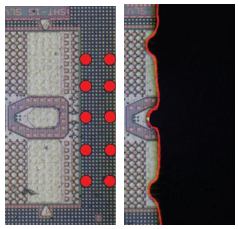


Fig. 11: a) Initial probe tip position, b) Final positioning.

VI. Conclusion

In this contribution, we presented an approach based on CNN model to identify the location of the probe and the pad. Moreover, the derivative of the phase of S_{11} is used to estimate the distance between the pad and probe along the Z axis. A zero-landing concept is proposed to estimate the initial landing location while not heavily modifying the pad aluminum surface. This approach has been tested on 38 unused pads. The overall landing uncertainty has a σ of 1.31 μm along the X axis and 0.97 μm along the Y axis.

Acknowledgment

This work is funded by the European Union within "NextGeneration EU", by the Federal Ministry for Economic Affairs and Climate Action (BMWK) on the basis of a decision by the German Bundestag and by the State of Saxony with tax revenues based on the budget approved by the members of the Saxon State Parliament in the framework of "Important Project of Common European Interest - Microelectronics and Communication Technologies", under the project name "EUROFOUNDRY".

References

[1] C. Esposito, C. De Martino, S. Lehmann, Z. Zhao, M. Schroter and M. Spirito, "Pre-Silicon direct Calibration/De-embedding Evaluation and Device Parameters Uncertainty Estimation, " *2021 97th ARFTG Microwave Measurement Conference (ARFTG)*, 2021, pp. 1-4, 2021.

[2] R. Schmidt et al., "Compensating Probe Misplacements in On-Wafer S-Parameters Measurements," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 11, pp. 5213-5223, Nov. 2022.

[3] R. Sakamaki and M. Horibe, "Automatic probing system with machine learning algorithm", *2021 96th ARFTG Microwave Measurement Conference (ARFTG)*, pp. 1-4, 2021.

[4] R. Sakamaki and M. Horibe, "Realization of Accurate On-Wafer Measurement Using Precision Probing Technique at Millimeter-Wave Frequency", *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 8, pp. 1940-1945, 2022.

[5] F. Mubarak, C. D. Martino, R. Toskovic, G. Rietveld and M. Spirito, "Automated Contacting of On-Wafer Devices for RF Testing", *2020 Conference on Precision Electromagnetic Measurements (CPEM)*, pp. 1-2, 2020.

[6] H. Li, F. T. von Kleist-Retzow, O. C. Haenssler, S. Fatikow and X. Zhang, "Multi-target tracking for automated RF on-wafer probing based on template matching," *2019 International Conference on Manipulation, Automation and Robotics at Small Scales (MARSS)*, pp. 1-6, 2019.

[7] Newport, 9-March-2025, "Picomotor™ Piezo Linear Actuators" Source. [Online]. Available: https://www.newport.com/f/picomotor-piezo-linear-actuators.

[8] Segment Anything Model, *Meta AI Inc.*

[9] YOLOv5, *Ultralytics Inc.*

[10] R. Brunelli, Template Matching Techniques in Computer Vision: Theory and Practice, Wiley, ISBN 978-0-470-51706-2, 2009.

Authorized licensed use limited to: TU Delft Library. Downloaded on August 06,2026 at 08:11:51 UTC from IEEE Xplore. Restrictions apply.