QUADRA: METROLOGY FOR THE ONGOING EVOLUTION OF MOORE'S LAW

The semiconductor roadmap is currently fuelled by innovations along three trends – hybrid bonding for advanced packaging at system level, the transition from FinFET towards GAAFET 3D design technology for faster switching times, and the move towards EUV lithography and high-NA allowing 3 nm and lower nodes in high-volume manufacturing (HVM). Conventional metrology solutions in semiconductor fabs face the daunting task of enabling process control with these trends. Nearfield Instruments' QUADRA has been proven to provide the necessary metrology solution for these three developments in HVM. This article shows how QUADRA tackles these challenges head-on through a systems engineering approach including measurement mode, probe and image processing development.

EKATERINA CHERNYSHEVA, SAMANTHA MAZZAMUTO, ARNOLD ZONDERVAN, RUDI WILHELM AND NIRANJAN SAIKUMAR

Semiconductor roadmap trends

The semiconductor roadmap features three manufacturing trends for 3D device architecture:

1. Hybrid bonding or direct bond interconnects provide vertical connections on a die-to-wafer or wafer-to-wafer level through copper pads allowing for heterogeneous integration and 3D device stacking. Reliable interconnections between copper-to-copper/dielectric require a highly smooth and planar surface. Insufficient surface quality may lead to the formation of voids along bond lines followed by device failure. This requires wafer-level metrology of the planarised copper pads where height variations need to be captured at the sub-nm level.
2. The 3D design technology for transistor fabrication, which started with Fin Field-Effect Transistor (FinFET) technology, allowed for continued device scaling and has been cemented with Gate All Around Field-Effect Transistor (GAAFET) technology. GAAFET's use of nanosheets or nanowires allows for a gate-enclosing channel design overcoming the leakage current and short channel effect issues faced by FinFET, while further allowing channel width scaling. However, the strict full 3D critical dimension control necessary for such complex epitaxial structures is very challenging.
3. High-NA EUV lithography is essential for continued scaling beyond 3-nm nodes with single patterning methods for improved resolution and reduced edge placement error benefits. This requires thinner photoresist layers, which consequently makes roughness and stochastic errors worse during the lithography step and creates metrology problems due to the low-contrast nature of the materials used.

Metrology challenges

With the development of the semiconductor roadmap towards advanced nodes (Figure 1) and the new innovations introduced in HVM production, a higher and a more precise control over the manufacturing process is required. Moreover, the increased complexity of designs and the typical dimensions used need a re-evaluation of the traditional parameters of interest and the wafer-level sparse sampling of metrology measurements used for process control.

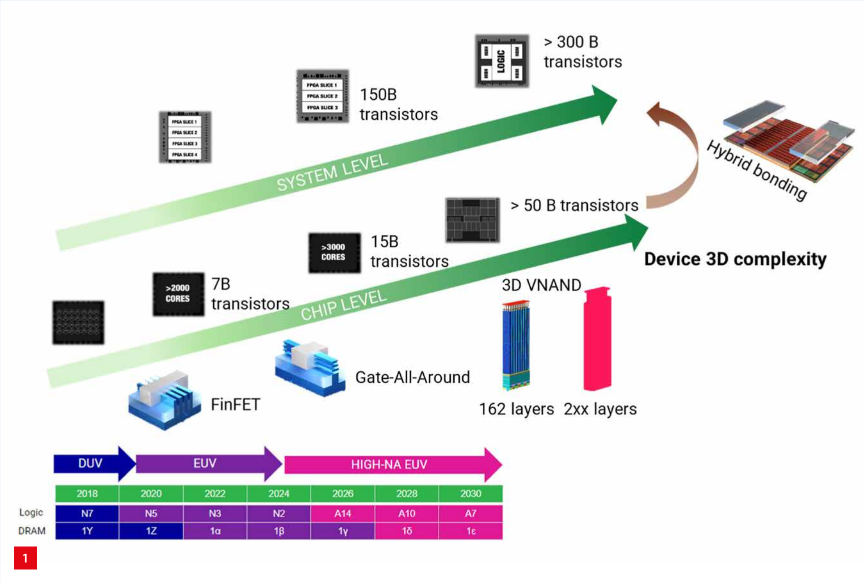
An HVM metrology system needs to satisfy several requirements. First of all, it must be a high-throughput, highly adaptive and robust in-line metrology system with fast time-to-solution. Moreover, measurements must be non-destructive, show high resolution and precision to successfully measure at the sub-nm level, and minimise tool and environment-induced measurement errors. This latter point is critical because of the small dimensions of the measured features. The characterisation techniques and metrology systems available on the market reach physical limits and, thus, cannot be fully adopted as metrology tools in an HVM environment for advanced nodes.

It is no longer sufficient to use high-speed 2D inspection techniques, such as Optical Critical Dimension (OCD) or Critical Dimension-Secondary Electron Microscopy (CD-SEM). The most used optical techniques lack resolution, are material dependent and, thus, are not able to measure the critical dimensions of complex structures such as FinFETs or GAAFETs. OCD provides global values and has low sensitivity to local variation of dimensions. With CD-SEM, due to the higher resolution needed with the ever-decreasing

AUTHORS' NOTE

Ekaterina Chernysheva (product marketing manager), Samantha Mazzamuto (metrology application engineer), Arnold Zondervan (team lead Mechatronics), Rudi Wilhelm (director of System Engineering & Program Management) and Niranjan Saikumar (R&D manager) work at Nearfield Instruments in Rotterdam (NL).

ekaterina.chernysheva@nearfieldinstruments.com
www.nearfieldinstruments.com



Semiconductor manufacturing trend for 3D device architecture ("Unleash the future of innovation", ISSCC, Feb 15, 2021).

sizes of semiconductor structures, higher-energy-density electrons are needed for good contrast and this leads to damage, especially with photoresists, referred to as resist loss.

Transmission Electron Microscopy (TEM) can achieve, on one side, the desired precision and resolution, but, on the other side, it is a destructive measurement and the throughput (TPT) is generally very low for being considered an ideal characterisation technique for metrology and inspection not only in an HVM environment but also in the ramp-up stage of R&D. Finally, the currently available Atomic Force Microscopy (AFM) tools are non-destructive and can achieve the required level of precision, but generally have a low TPT.

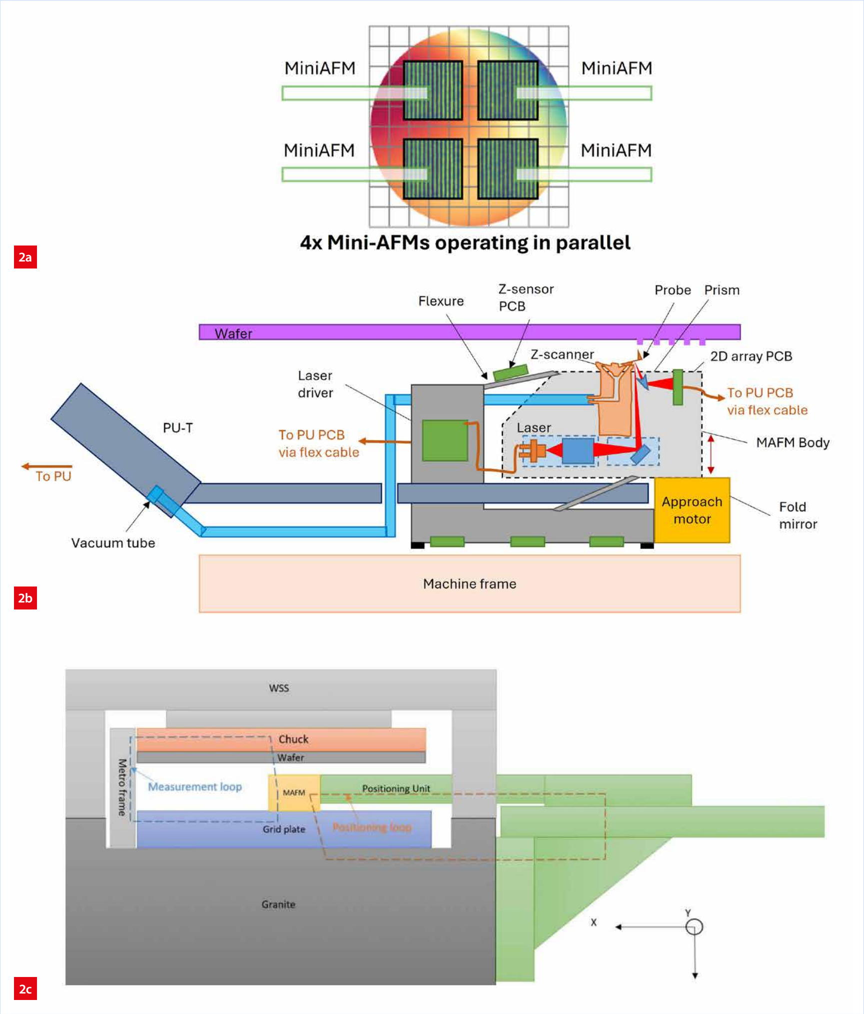
Nearfield Instruments has overcome the throughput limitations of the traditional AFM systems and offers an effective solution for current and future metrology challenges in the semiconductor industry. QUADRA is a high-TPT AFM-based metrology tool providing non-destructive characterisation in all three dimensions with sub-Ångström resolution and very high precision, long-term stability, a long probe lifetime and a versatile data processing toolkit to extract all necessary parameters of interest and relevant statistics.

System architecture and mechatronics design

The exceptionally high throughput, compared to existing AFM systems, is achieved by a revolutionary mechatronic design. Its innovative architecture based on four high-speed miniaturised AFM heads (MAFMs) allows to measure four areas of the wafer in parallel (Figure 2a). Each MAFM has its own positioning unit (PU) arm to be able to reach the required location of the wafer.

Compared to the conventional AFM systems, QUADRA has adapted an inverted architecture: the wafer is vacuum clamped on a chuck upside down and the tip approaches its surface from underneath (Figures 2b and 2c). This architecture combined with the unique design of the MAFM allows for shortening of the mechanical loop and reduces the noise contribution from mechanical resonances by removing excitation points. The shortening is achieved by releasing the kinematic coupling of the MAFMs with respect to the positioning units after moving to the desired location, where instead they make a local kinematic connection to the 'grid plate' in a process called 'landing' (Figure 2c). This grid plate is the local reference in the larger measurement loop between wafer and cantilever.

The MAFM scan head consists of:
- the Z-stage, which is the principal actuation mechanism in Z for measurements, also referred to as Z-scanner;
- the optical beam deflection system to sense the cantilever deflection;



QUADRA core overview.
(a) Top view, showing a configuration of four scan heads.
(b) Miniaturised AFM (MAFM) decoupled from the positioning unit (PU), which allows low noise even at high speeds.
(c) Global schematics.

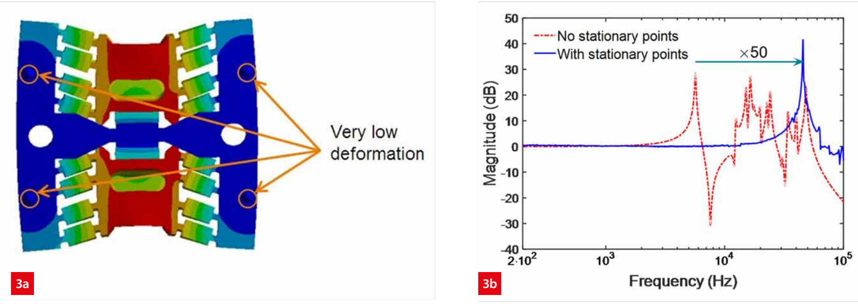
- an approach motor for larger Z-range motion;
- optical encoders used in combination with the positioning units for the scan-head positioning in XY;
- surrounding structure and frame to enable integration with other components of the system.

A decoupled scanning architecture is utilised to improve the Z-scanner response and eliminate the scanner-induced bow that is generally introduced by piezo-tube scanners in conventional AFM systems, which provide combined XYZ-motion. A wafer-scanning stage (WSS, figure 2c), which the wafer is clamped to through the chuck, is used for XY-scanning, allowing for a single module to provide this functionality to all the scan heads, while the Z-scanner of each MAFM is used to independently track the topography of the surface area.

To achieve a high tracking bandwidth for the Z-scanner, it is necessary to reduce the impact of the vibrations caused by reaction forces from the Z-scanner excitations. While the latest conventional AFM-system scan heads have been able to improve their mechanical designs to achieve a higher first resonance frequency and hence potentially higher bandwidth in Z for operation in isolation, the excitation of other resonances in the mechanical loop becomes the limiting factor.

On QUADRA, to minimise the parasitic excitation of elements in the measurement loop, Nearfield Instruments have developed an IP termed “Suspension on dynamically determined points” to isolate the Z-scanner from the body (Figure 3a). The dynamically determined connection points between the Z-scanner and the rest of the system allow the flat frequency response up to 53 kHz with no parasitic resonances (Figure 3b), keeping the noise low even when scanning at high speeds.

Another unique feature of QUADRA is the presence of a 6-DoF metro frame (DoF stands for degrees of freedom – three rotational and three translational), which allows for the correction of any remnant errors generated by reaction forces due to the acceleration of the wafer-scanning stage.



Suspension on dynamically determined points.
(a) Schematic layout of the mechanical Z-scanner (https://doi.org/10.1063/1.4935584).
(b) Bandwidth with (blue line) and without (red line) dynamically determined points.

This solution comes together with a 6-DoF active vibration isolation system, pushing the measurements into the realm of Ångström-level precision and below, while enabling, for example, high-aspect-ratio (HAR) measurements (see below).

A challenge unrelated to the measurement capability, but crucial for metrology with existing AFM systems, is related to navigating to the right location of the wafer where wafer alignment, automated pattern recognition and overall positioning accuracy are critical.
With a unique and novel optical imaging system consisting of autofocusing elements to accommodate different wafer thicknesses and a reference light path for drift correction, QUADRA has been tested for all of the necessary navigation functions, showing robust pattern recognition against various layers like metals, photoresist, dielectrics, polysilicon, etc., with thicknesses all the way down to 10 nm. More than 50,000 wafer cycles have been tested in the fab.

There are two options for wafer alignment depending on the application of interest. Global alignment is needed to determine the wafer centre and correct for eventual rotation errors. Fine alignment is done by pattern recognition training on the selected shotkeys and/or unique features on wafer/die level.

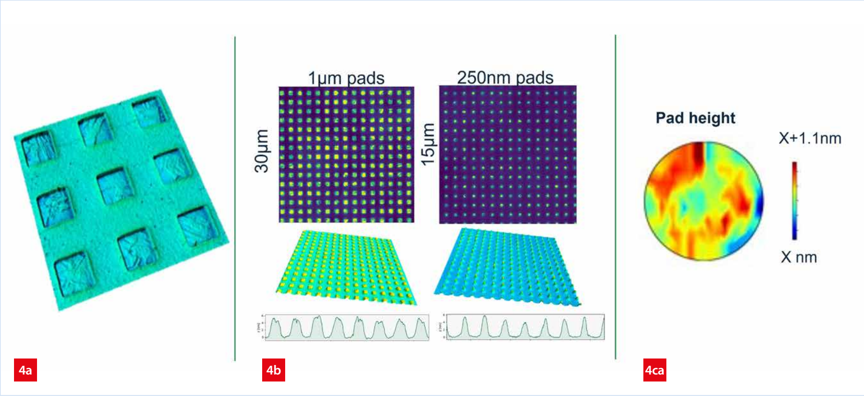
Optimised measurement modes

Nearfield Instruments has introduced new scanning modes that can be chosen depending on the application and the desired metrology parameters of interest to be extracted.

High-speed measurement for polished surfaces
Amplitude Modulation (AM) is the most widely used mode in most AFM systems, also known as intermittent contact or tapping mode. It works by using a dither piezo that excites the probe at its resonance frequency. The amplitude of oscillation of the cantilever, whose tip is in contact with the sample, will change whenever the topography changes and is used as the feedback parameter. The closed-loop control is used to maintain the setpoint amplitude constant by moving the Z-scanner appropriately enabling the latter to follow the surface topography when the scanning stage moves.

Depending on the selected pixel resolution, image size and number of shots per wafer, the throughput can range from 10 to 100 wafers per hour. This mode has been successfully tested for hybrid bonding applications, in particular on copper-to-copper bonding (C2C); see Figure 4. QUADRA is now the Process of Record metrology tool for this application at the launching customer.

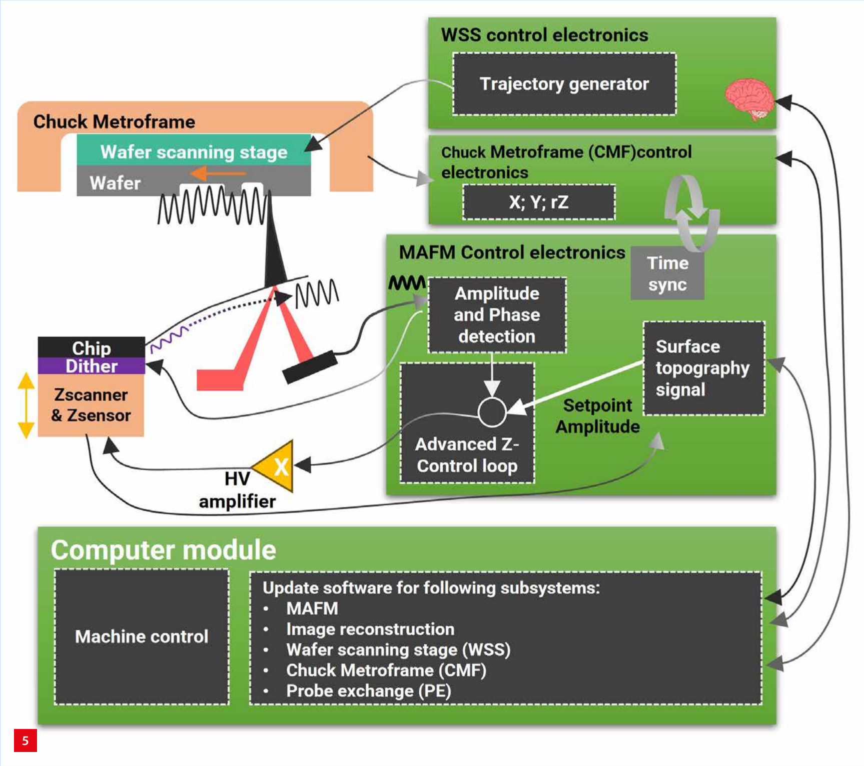
Lightning mode, the latest innovation on QUADRA, allows high scan speeds with pixel rates up to 1 MPx/s (Figure 5), showcasing the capability of this AFM technology to match



Copper-to-copper bonding.
(a) 3D image of copper pads.
(b) Example of pad profiles with different pad and pitch sizes, and with pad heights down to < 0.1 nm.
(c) Wafer maps can be obtained for every parameter of interest for better visualisation of its uniformity across the whole wafer.

imaging speeds of other established semiconductor metrology and inspection technologies. This innovation is enabled on the MAFM side through:

- advanced Z-scanner control algorithms allowing for exploiting the full mechatronic bandwidth of the Z-scanner for tracking topography;
- active control of fast-reacting cantilevers allowing the cantilever acting as the sensor to react to topography changes quickly;
- smart data-fusion algorithms where multiple signals from the MAFM are combined to reconstruct the topography image.



Schematic representation of the Lightning mode.

Combining this with advanced wafer-scanning-stage trajectory shapes and control allows the highest line rate possible for a given mechanical design while keeping the same vibrational floor.

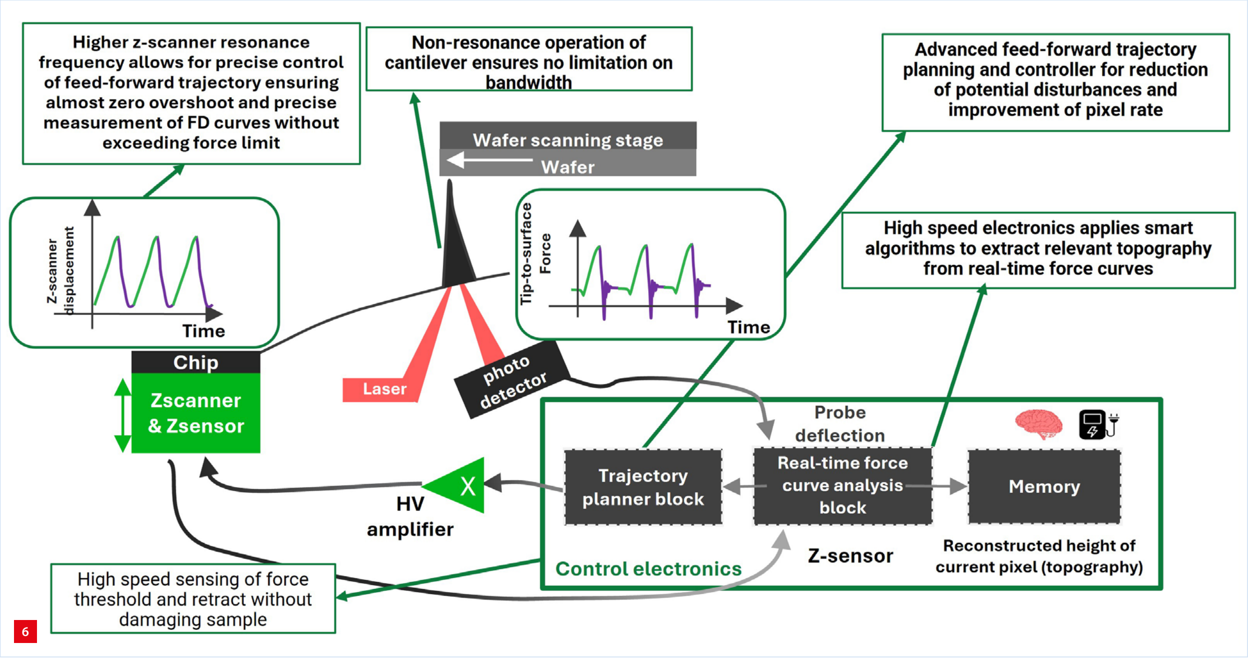
Novel mode for high-aspect-ratio structures
Nearfield Instruments has a proprietary scan mode called Feed Forward Trajectory Planning (FFTP). The working principle of this mode is shown in Figure 6. With AM and Lightning mode, the cantilever is excited at its resonance frequency. However, a resonating cantilever cannot measure HAR structures where the tip experiences damping in non-vacuum conditions.

FFTP is a non-resonant AFM mode (the cantilever is not excited by the dither piezo) developed to ensure that the tip can reach the bottom of a trench or hole. Cantilevers with lower stiffness are used for this mode, allowing for highly precise control of the tip-wafer interaction forces and true non-destructive measurements.

This mode enables imaging with individual force-distance curve (FD-curve, also known as deflection-distance curve) measurements per pixel. The Z-scanner moves the probe towards the sample from a predefined distance, the tip then engages with the surface with a continuous motion along the Z-direction based on certain predetermined measurement settings. After this stage of FD-curve measurement, the tip is moved away from the surface to the same initial point at a very fast rate allowing the tip to overcome any adhesion with the sample.

Although other AFM systems provide non-resonant measurement modes based on the analysis of FD-curves, the FFTP mode has some unique features through the use of a feed-forward mechanism instead of a closed-loop feedback operation. A defined tip trajectory (Figure 6) designed to avoid excitation of any parasitic resonances, combined with the exceptional linear motion of the vertical Z-scanner, ensures negligible motion along other DoFs, reducing the interaction with the sidewalls. With the unique feed-forward approach, the higher first-resonance-frequency-based fast reaction of the Z-scanner allows for precise control of the trajectory, ensuring almost zero overshoot and precise measurement of an FD-curve, without excessive force experienced by the tip.

High-speed sensing allows for negligible delays, reducing the tip-wafer interaction forces, achieving excellent measurement quality for after-development inspection on EUV photoresist structures and after-etch inspection on HAR structures (Figure 7). This also contributes to the tip-wear reduction and increase of tip lifetime.



Schematics of the FFTP scanning mode.

An innovative extension allowing for the measurement of extremely deep and narrow structures in HAR and EUV structures is the FFTP-fusion mode. Higher tip-wafer interaction force and consequent cantilever deflection is needed for clear imaging of these challenging structures. However, this change in force can lead to tip slip, poor image quality and inaccurate parameter extraction (Figure 8b). Hence, for this mode, multiple data points are collected along the FD-curve and smart data-fusion techniques are applied to simultaneously measure the top and bottom of HAR structures without any loss in throughput and avoiding any tip-slip effects (Figure 8).

FFTP-fusion mode, in combination with specifically designed probe-tip shapes, enables the measurement of holes and trenches with more than 300 nm depth and less than 25 nm top opening width. Some examples of measurements are shown in Figure 7.

Data processing

The on-tool data-processing algorithms are fully automated and robust, and have been tested on various structures (Figure 9). Raw images are flattened and aligned for target features. Structures of interest are identified, and a mask is applied to select areas/features for further analysis.

For copper pads, various pad sizes and shapes (square or circular), patterns (horizontal or diagonal), pitches and pad heights (down to 0.1 nm), and both protruded and recessed pads, even in the same image, can be characterised (Figure 4). Dielectric erosion can be computed, along with dielectric surface roughness and pad roughness.

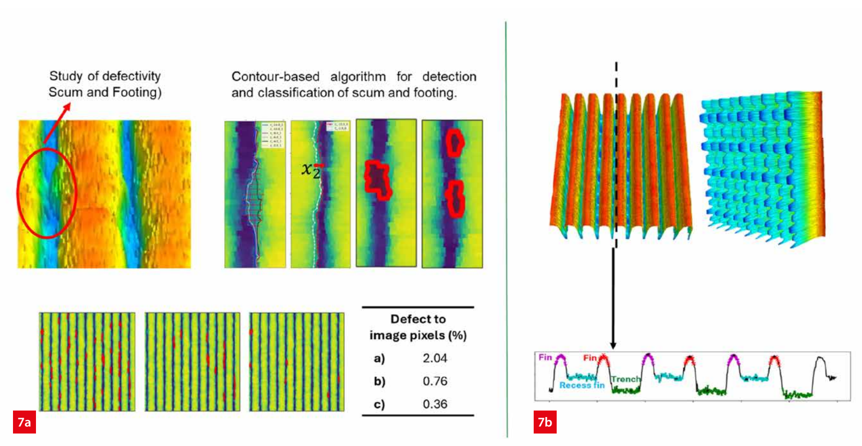
Full wafer maps can be extracted for different parameters of interest to visualise trends along the wafer (Figure 4c), providing a valuable insight for process control at different stages of manufacturing.

Application-critical parameters can include but are not limited to width, depth/height, line-edge and line-width roughness (LER/LWR) and top/bottom roughness.

Complex structures are analysed by using advanced pattern recognition algorithms. For example, for GAAFETs, all parameters can be extracted separately for fins and gates, and at different process steps: fin patterning, dummy gate, spacer and inner spacer generation, and, finally, epi-growth.

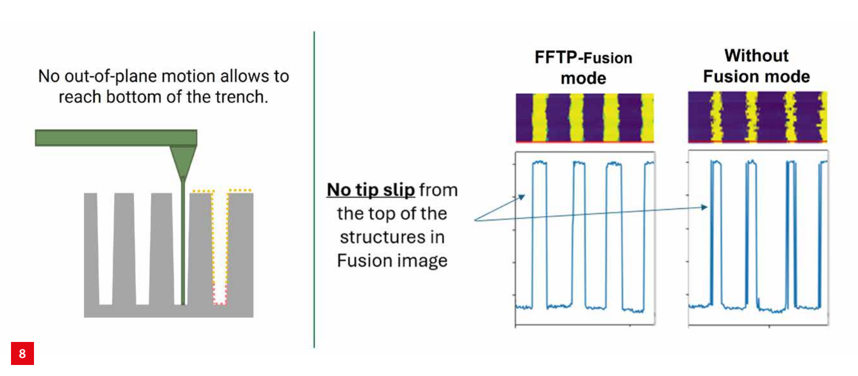
Surface roughness can be also accessed through power spectral density analyses to evaluate extremely flat surfaces.

For extraction of lateral parameters of interest on the wafer, where the tip shape and dimension can impact the measured images and hence extracted values, Nearfield Instruments has been developing algorithms to remove



Defect to image pixels (%)	
a)	2.04
b)	0.76
c)	0.36

Example of structures on which measurements have been taken with the FFTP mode.
(a) EUV structures - defectivity detection algorithm (i.e. scum and footing); defect classification based on thresholds on X- and Y-axis.
(b) HAR structures - top and bottom view, respectively, (above) and a line profile (below).



8

Smart data fusion and measurement enables removal of tip slip and other artifacts to enable imaging of challenging HAR structures.

these effects, which, along with the highly precise metroframe information, enable excellent accuracy.

Considering the growing demand for defect inspection, Nearfield Instruments has been developing algorithms for the detection of footing and scum (seen on structures due to residue left from the process), bridged and broken lines in high-density metal interconnect layers, and incompletely formed or merged contact holes. For various lithographical steps, residuals of photoresist or failure in mask opening can be characterised by the data-processing toolkit directly. Figure 7a shows an example of detecting footing and scum.

Probe management

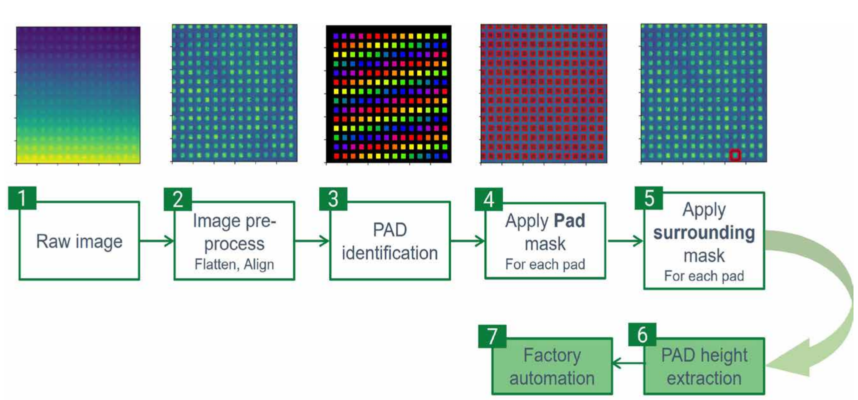
Probe lifetime is a very important parameter to monitor. For QUADRA, the cost of ownership depends mainly on the probe lifetime. The longer it is, the lower the cost per scanned wafer will be. Probe lifetime is maximised thanks to the innovations introduced in the mechatronics design, described above. A low system noise and precise control of the tip-sample interaction ensure minimal interaction between the wafer and tip and, thus, prolong the probe lifetime. All probes are calibrated with a laser Doppler vibrometer at Nearfield Instruments, and a passport for each probe is created.

QUADRA is equipped with two probe cassettes, each one containing 195 probes. The probes (shape, stiffness, material, etc.) are chosen depending on the application of interest and customer requests, and they are filled and shipped together with the passport files. The automatic probe-wear detection ensures a smooth and constant workflow without compromising the image quality. It is achieved by constantly checking specific KPIs (key performance indicators) per application along with a dedicated tip-wear monitoring structure integrated into the system.

Conclusions

The main semiconductor roadmap trends such as hybrid bonding (copper pads), high aspect ratio and EUV lithography structures can be successfully measured by QUADRA (Figure 10) with high throughput. In particular, QUADRA has been qualified for HVM copper-pad metrology. It is the ideal tool for in-line metrology in an HVM production environment because of its unique design and scanning modes tailored to the application to meet the customer needs with its fully automated and robust data-processing algorithms and low time-to-solution.

Nearfield Instruments continuously works on improving the performance of QUADRA in terms of throughput and precision, and broadening the application portfolio, and on new product introduction. For example, recently it has developed a non-destructive, subsurface scanning probe metrology where acoustic excitations are transmitted into the wafer through the tip and the interaction of the acoustic wave with the subsurface features is sensed again through the tip, allowing for high-resolution detection of buried voids and other subsurface features, such as overlay markers.



9

Flow chart of the on-tool data-analysis process of QUADRA; example of C2C application.

10

QUADRA tool overview