Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

PDTOM: Physics-guided Moiré metrology for precision wafer-level bonding alignment in Thin-3D integrated circuits

Shie-Ping Chang
National Yang Ming Chiao Tung University
Chen-Chia Chang
National Yang Ming Chiao Tung University
Kuan-Neng Chen
knchen@nycu.edu.tw

National Yang Ming Chiao Tung University

Article

Keywords:

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11609588/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

PDTOM: Physics-guided Moiré metrology for precision wafer-level bonding alignment in Thin-3D integrated circuits

Shie-Ping Chang[1] , Chen-Chia Chang[2] and Kuan-Neng Chen[1,2]*

[1] Institute of Electronics, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan.

[2] International College of Semiconductor Technology, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan.

*Correspondence: knchen@nycu.edu.tw

Abstract

Wafer-to-wafer alignment is challenged by axial changes in optical Moiré contrast. We present Physics-guided Digital Twin Optical Moiré Metrology (PDTOM), integrating broadband optical image synthesis, synthetic supervision, four-parameter CNN initialization and inverse refinement. PDTOM-A and PDTOM-C estimate lateral translations (Δx and Δy), rotation (θ), and axial separation using large and compact octagonal marks, respectively. On 500 independent synthetic images, PDTOM-A achieved lateral mean absolute errors of 3.295 and 3.711 nm after translation-only refinement. An archived 1,000-record PDTOM-C evaluation yielded refined errors of 0.634 nm (Δx), 0.663 nm (Δy), 0.003510° (θ), and 1.801 nm (simulated axial separation); checkpoint provenance remains unverified. Four AML axial images revealed Moiré contrast redistribution and local polarity reversal. Background-excluded ZDMOR reconstruction achieved mean image NCC of 0.998026 with target-guided axial matching. Independent Vernier measurements (n = 90) yielded mean wafer overlay of (0.08, 0.22) μm and empirical 3σ spread of 0.51 μm. These findings distinguish synthetic estimation, experimental image reconstruction, and physical overlay.

Keywords: optical Moiré metrology; digital twin; wafer bonding; broadband transmission imaging; deep learning; heterogeneous integration

Introduction

Wafer-to-wafer (W2W) bonding enables dense vertical interconnects and heterogeneous device integration, but its yield depends on reliable control of in-plane translation and rotation. Thin-3D integration provides the application context for the present study; characterization of the full electrical, thermal and mechanical platform lies beyond the scope of this optical-metrology investigation. [1]

Conventional optical alignment methods—including cross targets, box-in-box structures, Vernier scales, template matching and edge localization—are affected by illumination, focus, mark quality and imaging resolution. Moiré metrology offers a complementary route: superposition of gratings with slightly different periods converts small relative translations into larger changes in the observed fringe pattern. A two-dimensional concentric octagonal geometry samples several orientations and can encode both lateral translation and rotation within a single target. [2, 3, 4, 5, 6]

A previously reported deep-learning optical Moiré alignment (DLOMA) approach provided an image-to-pose baseline under the represented optical conditions. During W2W approach, however, the inter-wafer separation changes continuously, modifying the recorded optical field even when lateral pose remains unchanged. Fringe contrast can diminish, intensity extrema can change polarity and energy can redistribute among rings. Consequently, a model trained at one axial state may confuse a separation-dependent optical change with in-plane displacement. Supplementary Table S1

compares the scope and validation evidence of the AML baseline, DLOMA and PDTOM. [1]

In broadband transmission measurements through opposing silicon wafers, the recorded intensity also depends on spectral weighting, optical propagation, spatial coherence, objective aperture and detector sampling. The resulting axial contrast evolution must therefore be accounted for when interpreting concentric Moiré patterns. [7, 8, 9, 10, 11, 12, 13, 14, 15]

We address this issue with a physics-guided representation of the optical measurement chain. The forward model maps a candidate wafer state to a camera image, while the inverse method combines a CNN-derived initial estimate with constrained refinement. The axial coordinate is reported relative to a mechanical-stage reference because an independently calibrated absolute wafer-to-wafer surface gap was not available. The digital-twin and inverse-optimization concepts motivate model-based interpretation, without supplying missing physical reference labels. [16, 17]

The framework comprises two related four-parameter estimation configurations: PDTOM-A, using the larger experimental-mark geometry, and PDTOM-C, using a compact 45 × 45 μm² synthetic mark. Both represent the state as $[\Delta x, \Delta y, \Delta\theta, z_{gap}]^T$, but their reported inverse-refinement scopes differ. In the frozen PDTOM-A evaluation, the CNN initializes all four parameters and only the lateral coordinates are locally optimized; PDTOM-C evaluates four-parameter refinement using archived initial-state predictions. Figure 1 introduces the Thin-3D and AML platform, Figure 2 describes the broadband forward model, and Figure 3 summarizes the method workflow. ZDMOR experimental image reconstruction is assessed separately in Figure 4, while synthetic benchmarks and wafer-level Vernier overlay are presented in Figures 5–7. Supporting protocols are given in Supplementary Sections S1–S6.

Results

Because independently traceable nanometre-scale labels for translation, rotation and wafer separation are difficult to obtain in large numbers, synthetic supervision is central to PDTOM. We begin with a calibrated broadband forward model and generate Moiré images under known parameter perturbations. A CNN predicts four alignment-state components in each configuration, and physics-guided inverse refinement updates the implemented parameter subset for the respective benchmark (Fig. 3). For the measured large mark, the concentric octagonal gratings have pitches of 8.00 and 8.48 μm and radial cycle counts of 37 and 35, respectively. The nominal geometric beat period is 141.33 μm, corresponding to an ideal translation amplification of 17.67 for motion of the P_1 grating. Rotation is treated through a two-dimensional rigid transformation rather than the same scalar amplification factor. Supplementary Fig. S2 provides model-consistency examples of directional and axial image variation.

Mark-size-dependent grating-cycle design

For a concentric octagonal grating with nominal mark width L and radial pitches P_1 and P_2 ($P_2 > P_1$), the integer numbers of radial grating cycles are $N_i = \lfloor L/(2P_i) \rfloor$. Their local geometric beat period is $P_m = P_1P_2/(P_2 - P_1)$. The usable common radial extent is $R_o = \min(N_1P_1, N_1P_2)$, giving N_m, overlap = R_o/P_m. A value N_m, overlap ≥ 1 is a preliminary geometric screening guideline for accommodating a nominal beat period. It is not a necessary condition for useful pose information and does not guarantee a camera-resolved bright–dark cycle. Finite aperture, fringe phase, broadband propagation, and sampling require separate visibility and identifiability checks (Supplementary Section S1).

Supplementary Table S3 compares candidate mark geometries from approximately 30 to 592 μm. These values represent design calculations and should not be interpreted as fabricated or optically

qualified marks. The measured 8.00/8.48-μm configuration has an outer extent of approximately 593.6 μm; the 592-μm footprint-constrained row in Supplementary Table S3 represents a distinct candidate. Smaller periods also shorten the approximate paraxial Talbot scale, $z_T \approx 2P^2/\lambda$, but this estimate alone does not specify an unambiguous axial measurement range under broadband, finite-aperture imaging.

The actual PDTOM-C benchmark uses $P_1 = 1.50$ μm, $P_2 = 1.59$ μm, $N_1 = 15$ and $N_2 = 14$ within a 45 × 45 μm² footprint. Its beat period is $P_m = 26.50$ μm and common radius $R_o = 22.26$ μm, yielding N_m, overlap ≈ 0.84. The $P_2 = 1.65$ μm, 45-μm candidate in Supplementary Table S3 instead gives P_m = 16.50 μm and N_m, overlap = 1.30. These are different designs; the screening candidate must not be substituted for the archived compact-mark configuration.

Gap-Dependent Optical-State Evolution

Four AML transmission images were acquired at relative mechanical-stage displacements of $\Delta Z = 0$, +30, +60, and +80 μm, corresponding to stage readings of −0.190, −0.160, −0.130, and −0.110 mm (Fig. 2e). The measured octagonal Moiré patterns exhibited pronounced changes in radial intensity and fringe contrast. Smoothed radial profiles revealed local bright–dark contrast-polarity reversal while the underlying grating geometry remained unchanged. These observations demonstrate that axial measurement conditions can alter optical appearance without corresponding lateral pose changes, motivating the inclusion of an axial coordinate in the four-parameter forward and inverse models. The reported ΔZ represents relative mechanical-stage displacement, not an independently calibrated physical wafer gap.

Axial propagation changes the relative phases of spatial-frequency components generated by the lower grating. In angular-spectrum propagation, each collected component accumulates a longitudinal phase determined by wavelength, transverse frequency and propagation distance. The upper grating mixes this propagated field with its own spatial structure. Their interference therefore changes ring intensity and the slowly varying Moiré envelope, even when lateral registration is held fixed. Near-field recurrence provides a useful physical scale, but the finite octagonal aperture and multilayer stack prevent interpretation through a single infinite-grating Talbot distance. [8, 12, 14, 15]

The observed axial contrast is also an ensemble response of the instrument. Broadband spectral averaging combines intensities whose diffraction-phase evolution occurs on different axial scales and can smooth monochromatic recurrences. The illumination-angle distribution averages shifted angular spectra and modifies spatial coherence. Finite numerical aperture filters the transmitted spatial frequencies and collected diffraction orders, changing which interference terms survive. Finally, pixel-area integration and camera sampling reduce high-frequency modulation and can obscure or alias unresolved grating structure. These operations affect contrast amplitude, apparent ring shape and axial distinctiveness; none alone establishes a unique absolute-gap label. [7, 11, 12, 13]

Figure 2(e) shows local bright–dark contrast-polarity reversal in the measured intensity profiles: a radial region that is relatively bright at one stage position becomes relatively dark at another. Such reversal is compatible with propagation-dependent interference, but the broadband camera records intensity rather than the complex optical field. The observation therefore does not directly demonstrate an exact π phase shift of a specified diffraction-order pair. The $\Delta Z = 0$, +30, +60 and +80 μm labels denote relative mechanical-stage displacement throughout this comparison, not absolute wafer gap.

The spectral conventions in Figure 2 and Supplementary Section S2 are identical. Silicon attenuation is applied through the lower- and upper-wafer field-transmission coefficients before intensity formation. Numerical spectral integration weights contain source radiance, optical throughput and

camera response, with the angular distribution included consistently, but no second silicon-transmission factor. Figure 2(b) instead illustrates an effective detected spectrum that includes transmission through both Si wafers. It is an illustrative spectrum for interpreting the detection band, not the numerical integration weight. This distinction prevents double counting of silicon attenuation.

Experimental validation of Z-dependent Moiré optical reconstruction

The four AML images were compared with Z-dependent Moiré optical reconstruction (ZDMOR) using leave-one-Z-out (LOZO) reference selection and target-guided optimization of the reconstructed axial coordinate. To assess agreement while reducing contributions from peripheral non-Moiré features, the registered 512 × 512 AML and reconstructed images were evaluated with an identical background-exclusion mask. The mask removed predefined upper-left text and lower-right shadow regions and retained 89.11% of the pixels in each image. The background-excluded Pearson NCC values were 0.998328, 0.999455, 0.999507 and 0.994814 at AML stage positions −0.190, −0.160, −0.130 and −0.110 mm, respectively (mean 0.998026). Horizontal and vertical normalized-intensity profiles were obtained by averaging the central seven image rows or columns (indices 253–259). The corresponding horizontal/vertical profile RMSEs were 0.020494/0.013773, 0.010826/0.008238, 0.006550/0.011672 and 0.031666/0.032937 (mean 0.017384/0.016655). The corner exclusions do not intersect the central profile bands, so all 512 samples per profile were retained. Figure 4 shows mark-centred image crops for visualization, whereas the stated NCC values were computed from the background-excluded registered images, not from the displayed octagonal crops. Because the best-matching axial coordinates were selected by comparison with the target frames, these results characterize target-guided reconstruction consistency rather than blind Z-gap estimation accuracy (Supplementary Section S6).

The correlation value depends on the evaluation region. The full-image mean NCC was 0.998323, while an octagonal-only ROI gave 0.960148. The official background-excluded mean is 0.998026. These values are not interchangeable: retaining or excluding broad common intensity structure changes the pixel population and covariance entering Pearson NCC. The restricted ROI is reported as a sensitivity analysis, not a replacement primary score. The complete horizontal and vertical profiles complement NCC by displaying where local modulation agrees or departs, with mean RMSEs of 0.017384 and 0.016655, respectively. High reconstruction correlation establishes image agreement under the stated mask and target-guided matching protocol; it does not independently measure axial prediction error.

PDTOM-A: Physics-guided lateral alignment reconstruction

On the independent 500-sample held-out synthetic test set, the CNN coarse estimator achieved Δx and Δy mean absolute errors (MAEs) of 21.483 and 24.674 nm. The frozen PDTOM-A refinement reduced them to 3.295 and 3.711 nm. The empirical per-axis 3σ errors decreased from 90.263 to 13.884 nm for Δx and from 102.281 to 15.682 nm for Δy. The final radial mean, 95th percentile (P95), and mean plus three standard deviations (mean + 3SD) errors were 5.665, 14.713, and 17.900 nm. Boundary continuation was activated for 17 of 500 samples, while the confidence gate retained the coarse estimate in 6 cases. Because the refinement acts only on Δx and Δy, the CNN $\Delta\theta$ and reference-relative z MAEs remain 0.00961° and 0.299 μm. Fig. 5 shows both signed-error stability across the sampled displacement range and the reduction of the absolute-error tails.

Near-Zero Displacement Qualification

A second synthetic dataset probed the behavior of the frozen PDTOM-A estimator near zero translation. Across 500 images spanning ±1 μm, refined Δx and Δy MAEs were 3.383 and 3.376 nm,

and the corresponding signed-error 3σ values were 14.095 and 14.562 nm. The radial mean, P95 and mean + 3SD errors were 5.463, 13.380 and 17.422 nm, respectively. These values were calculated from a dataset distinct from the 500-image primary held-out benchmark (Supplementary Fig. S3 and Table S5).

For the 313 samples satisfying $|\Delta x|, |\Delta y| \leq 500$ nm, the lateral MAEs remained 3.400 and 3.252 nm, with a radial mean of 5.379 nm and radial P95 of 13.482 nm. Within the 102-sample ±100-nm subset, the corresponding MAEs were 3.575 and 2.960 nm and the radial P95 was 13.207 nm. Exact zero-valued coordinate cases showed residual biases of +2.064 nm in Δx and +0.691 nm in Δy. Thus, no observable loss of synthetic-domain estimation accuracy occurs near zero.

Empirical sign recovery exceeded 95% for nonzero displacements within 100 nm but decreased for magnitudes below 20 nm. The result therefore supports an approximate 50-nm minimum discernible displacement under the present synthetic conditions; it does not establish guaranteed 10-nm resolution. Because MAE is an accuracy statistic rather than a resolution metric, the nanometer-scale numerical errors must not be interpreted as traceable wafer-level resolution.

Physical wafer-level overlay assessed with Vernier marks

To provide an experimental wafer-overlay benchmark, we analyzed two Vernier mark pairs with 45 measurements per pair. Pair 1 yielded a mean offset of (0.02, 0.24) μm and an empirical 3σ spread of 0.50 μm, while Pair 2 yielded (0.14, 0.19) μm and 0.51 μm. Across all 90 measurements, the combined mean offset was (0.08, 0.22) μm, with an empirical 3σ spread of 0.51 μm (Fig. 6). The observed distribution reflects wafer-level overlay rather than a direct test of the nanometre-scale synthetic pose predictions.

These measurements demonstrate submicrometre overlay in the investigated W2W process. They do not independently verify a 3–5-nm physical estimator accuracy because the Vernier readout resolution and its associated uncertainty have not been established at that scale. The measured distribution also includes contributions beyond optical pose reconstruction, such as bonding and processing variability.

PDTOM-C: Compact-mark four-parameter reconstruction

The compact PDTOM-C configuration was assessed using a 45 × 45 μm² octagonal mark with P_1 = 1.50 μm, P_2 = 1.59 μm, N_1= 15, and N_2 = 14. The archived 1,000-record synthetic evaluation contains initial and refined estimates for Δx, Δy, θ, and simulated z_{gap}. Mean absolute errors decreased from 13.202 to 0.634 nm (Δx), 12.251 to 0.663 nm (Δy), 0.032052° to 0.003510° (θ), and 27.926 to 1.801 nm (simulated z_gap), corresponding to reductions of 95.2%, 94.6%, 89.1%, and 93.6%, respectively (Fig. 7a). The tail-error analysis in Fig. 7b further examines recorded initial-versus-refined absolute errors. The refined 95th-percentile errors were 2.729 nm, 2.808 nm, 0.008109°, and 6.764 nm, while the largest residuals were 13.827 nm, 11.470 nm, 0.139486°, and 99.466 nm, respectively. The isolated rotational and axial outliers demonstrate that low average errors do not guarantee uniformly small errors across the synthetic evaluation. Because the archived predictions have not been conclusively linked to a specific executed CNN checkpoint, these results describe recorded initialization and refinement performance rather than verified checkpoint-level accuracy. Experimental compact-mark validation remains necessary.

The present AML object-side sampling is 1.2 μm/pixel. A 1.50–1.59 μm grating period therefore spans only about 1.25–1.33 detector samples under that optical configuration. The compact grating cannot be assumed to be resolved by the existing AML optics from a synthetic Moiré envelope alone. Optical and hardware qualification must establish magnification, NA, wavelength-dependent transfer,

camera sampling and usable contrast for the actual compact target. Figure 7(a) retains the four-parameter initial/refined distributions, and Figure 7(b) retains the tail-error analysis; both describe the synthetic evaluation only.

Discussion

Axial-dependent Moiré imaging makes wafer separation part of the optical measurement state. The AML sequence shows that unchanged grating geometry can produce substantially different ring brightness and local contrast polarity as the stage moves. Propagation-dependent diffraction phases, followed by spectral and angular averaging and aperture filtering, explain why a purely geometric overlay image is insufficient to represent the acquisition chain. The measured reversal is an intensity observation, rather than direct phase metrology; neither an exact π shift nor an absolute wafer gap can be inferred from it alone. The calibrated forward representation therefore provides a physically interpretable way to accommodate changing appearance during approach.

The two PDTOM configurations share a four-component state definition but implement different refinement scopes. PDTOM-A uses the CNN for Δx, Δy, rotation and reference-relative axial initialization, followed by lateral refinement only in the frozen benchmark. Its reported rotation and axial errors remain those of the initializer. PDTOM-C records refinement of all four coordinates. Comparing their final MAEs as though the marks, optical sampling, parameter domains and optimizers were identical would confound these differences. The compact archive further lacks a fully verified lineage between sample records, initial predictions and a specific executed CNN checkpoint. Its improvement is consequently attributed to recorded initialization followed by refinement, with checkpoint-level reproducibility still to be established.

Synthetic MAE measures agreement with imposed simulator labels within a specified data population. It does not include all errors encountered in physical wafer metrology, such as imperfect material and illumination models, mark fabrication variations, wafer bow, drift, stage calibration or reference-readout uncertainty. Very small synthetic residuals can therefore demonstrate numerical consistency without establishing corresponding physical accuracy. The present results also lack a controlled, fully matched ablation against DLOMA; claims of superiority would require common data partitions, optical conditions, initialization and evaluation constraints. The independent held-out and near-zero PDTOM-A sets support the reported synthetic behavior, while the compact benchmark remains a separate population.

Compact marks are attractive for reducing alignment-target area, but the 45×45 μm² geometry changes the optical problem as well as the footprint. Its $N_{m,\mathrm{overlap}} \approx 0.84$ illustrates why a one-beat screening guideline cannot serve as a necessary estimator condition. Conversely, useful synthetic information below one nominal beat does not prove practical visibility. At 1.2 μm/pixel, the existing AML sampling is insufficient to assume faithful recording of 1.50–1.59 μm grating structure. A qualified compact-mark optical configuration must demonstrate appropriate sampling, aperture transfer, contrast and parameter identifiability before physical four-parameter accuracy can be assessed.

ZDMOR NCC measures reconstruction fidelity under a declared image-processing and selection protocol. The background-excluded mask removes annotation and shadow while retaining 89.11% of registered pixels; its mean NCC of 0.998026 is the primary comparison. The full-image and octagonal-only means show that ROI choice changes the score. The central profiles add spatially resolved information that a single correlation statistic cannot convey. Although LOZO excludes the target from the reference-image set, that target still participates in best-match Z selection. Consequently, the experiment tests target-guided reconstruction, not blind prediction of a withheld

axial label. A future blind test must freeze calibration and selection rules before evaluating independently referenced target states.

The Vernier measurements provide a complementary physical process-overlay result. Across n = 90 observations, the combined mean offset of (0.08, 0.22) μm and empirical 3σ spread of 0.51 μm reflect alignment, bonding and processing contributions together with readout limitations. They do not identify the isolated optical-estimator error, and the 0.10-μm nominal key increment does not establish nanometre-scale uncertainty. These measurements should therefore remain separate from both synthetic pose-error statistics and target-guided image-correlation results.

Closed-loop alignment has not been experimentally demonstrated. Future work should qualify compact-mark optics, verify checkpoint and dataset lineage, acquire independently referenced pose labels and test randomized stage perturbations with repeated correction cycles. Convergence, capture range, throughput and pre-bond/post-bond overlay should be evaluated alongside failure tails and calibrated uncertainty. Such experiments would connect the current offline reconstruction framework to operational wafer control while preserving the distinction among image fidelity, synthetic estimation and physical overlay.

Methods

Thin-3D fabrication and AML optical alignment architecture

The Thin-3D process integrates thinned device tiers through temporary carrier bonding, backside silicon removal, wafer-level oxide bonding and interconnect formation (Fig. 1). This process context motivates precise pre-bond alignment; the complete three-tier process is not independently validated by the present optical-metrology experiments. The AML W2W bonder uses dual transmission-imaging channels to observe opposing octagonal Moiré marks (Fig. 1).

Gap-dependent broadband optical forward model

The nominal mark-plane separation is $d = z_0 + Z_{gap}$, with $z_0 = t_{ox,U} + t_{ox,L}$, where $t_{ox,U}$ and $t_{ox,L}$ are process-dependent upper and lower SiO_2 thicknesses. Z_gap is the nonnegative ambient separation between opposing oxide surfaces; the reported Δz remains reference-relative and does not establish an absolute local gap. In the optical model, the SiO_2 and ambient segments have distinct refractive indices and are propagated with layer-specific angular-spectrum transfer functions rather than one homogeneous transfer distance (Fig. 2; Supplementary Section S2).

At each wavelength λ, the incident field U_0 is modulated by the lower mark and lower-wafer silicon field transmission. Multilayer propagation to the upper mark is evaluated as $U_2 = \mathscr{F}^{-1}\{H_{stack}\,\mathscr{F}[U_1]\}$, with $H_{stack} = H_{ox,L}\,H_{gap}\,H_{ox,U}$ and $Hj = \exp(i\,k_{z,j}\,d_j)$. The propagated field then interacts with the upper mark and upper-wafer field transmission before objective-pupil filtering. Distinct wavelengths and illumination angles are combined in intensity. Because silicon attenuation is already included in these field transmissions, the numerical detection weight excludes silicon transmission and contains source, optical-throughput and camera-response terms only. The two-wafer effective detected spectrum in Fig. 2(b) is illustrative and is not reused as that weight. The full editable equations are retained in Supplementary Section S2. [8, 9, 10, 11, 12]

AML transmission imaging and axial reference definition

Experiments used an AML AWB08 8-in. W2W bonder with two independent transmission-imaging channels. Each channel combines a Moritex MML4-HR65D 4× telecentric objective (NA = 0.112, nominal working distance 65 mm) and a Basler acA1300-200um monochrome camera (1280 × 1024 pixels, 4.8-μm pixel pitch). The object-side sampling is 1.2 μm/pixel. Unfiltered Philips Type 7023

projection-halogen illumination traverses two opposing 725-μm double-polished silicon wafers. [20, 21, 22]

Mechanical stage readouts of −0.190, −0.160, −0.130 and −0.110 mm correspond to relative axial displacements of 0, +30, +60 and +80 μm. We define $\Delta z = z_{gap} - z_{ref}$, where z_{ref} denotes the shared reference coordinate; thus, $\Delta z = 0$ is a stage-reference convention rather than physical wafer contact. The forward calculation retains an effective nonnegative propagation gap, while reported axial motion is reference-relative. The instrument configuration and additional optical assumptions are summarized in Supplementary Table S2 and main-text Fig. 1.

Physics-guided synthetic image generation and inverse estimation

In the absence of large experimentally labeled four-parameter datasets, supervised CNN initialization was based on synthetic images generated from octagonal grating geometry and the calibrated optical forward model. This representation includes multilayer angular-spectrum propagation, broadband and angular averaging, finite-aperture imaging and camera sampling. An effective point-spread function and a fixed radiometric mapping account for residual instrument response. The procedure should therefore be described as calibrated and physics-guided, rather than as an entirely independent first-principles prediction. Figure 3 separates the synthetic training workflow from ZDMOR image reconstruction and Vernier process-overlay evaluation. The four-state representation is common to PDTOM-A and PDTOM-C, but the optimization variables reported for the respective benchmarks are not identical.

A common set of optical calibration parameters was fixed across the four registered axial frames. The inverse objective incorporates complementary image descriptors, including intensity correlation, gradients, Fourier content, phase and ring-resolved profiles, and considers both direct and contrast-inverted image hypotheses. Supplementary Sections S1–S3 provide the objective and implementation details. The separate ZDMOR LOZO experiment additionally uses reference images and target-guided axial selection and should not be confused with a blind four-parameter prediction test.

The current PDTOM-A CNN initializer was trained on a 7000-image domain-randomized synthetic pool generated by the frozen forward operator. The grayscale 18-layer residual network (ResNet-18) regressor outputs four normalized components $[\Delta x, \Delta y, \Delta\theta, \Delta z]^T$. Training used the adaptive moment estimation optimizer with decoupled weight decay (AdamW; initial learning rate 10^{-4}, weight decay 10^{-4}), batch size 8, learning-rate reduction on a performance plateau (implemented using ReduceLROnPlateau), and early stopping. Weak gain, offset, gamma, and additive-noise perturbations were applied only to the training subset to preserve the reference-relative z sharpness cue. The independent held-out synthetic test set used for Fig. 5 and Supplementary Table S4 was generated with a previously unused random seed and was not used for training, model selection, or refinement-parameter tuning.

The training objective was normalized weighted Smooth-L1 with output weights [2, 2, 1.25, 1] and transition $\beta = 0.10$. Frozen lateral refinement used 80-, 20- and 5-nm steps with confidence gating. Boundary continuation used 80-, 40-, 20- and 5-nm steps and a maximum excursion of 640 nm. The complete PDTOM-A estimate still contains rotation and reference-relative z, but these two initialized coordinates were not locally refined in the archived benchmark. Detailed search limits and the proposed, unvalidated uncertainty extension are retained in Supplementary Section S3.

For PDTOM-C, the 45 × 45 μm² octagonal configuration was assessed through a recorded initial estimate followed by constrained four-parameter refinement. The available per-image output comprises an independent 1,000-image synthetic evaluation, distinct from both PDTOM-A test populations. Figure 7 presents the resulting four-parameter residual distributions. Because the

underlying checkpoint correspondence and some optical implementation records remain unverified, the reported results are appropriately qualified in Supplementary Section S5.

Evaluation and physical overlay

PDTOM-A was tested on 500 independent synthetic images generated with a previously unused random seed. A separate 500-image near-zero set, including nested ±500- and ±100-nm subsets and zero-axis cases, was also evaluated. Neither set was used to train the CNN or select inverse-refinement settings. Errors are summarized using per-axis MAE and signed-error 3σ; radial results are reported by the mean, 95th percentile and mean plus three standard deviations. Confidence-gate rejections remained part of each evaluated population.

Physical overlay was analyzed separately using an automated Vernier-key recognition workflow applied to two mark pairs (45 measurements each). Dedicated horizontal and vertical Vernier regions contained nominal increments of 0.10, 0.25, 0.50 and 1.00 μm. Detected tick offsets were converted into Δx and Δy. Because the peripheral Vernier structures and the octagonal Moiré marks rely on different image features, the Vernier measurement constitutes an independent process-overlay benchmark, not a reuse of the PDTOM reconstruction objective.

The smallest nominal Vernier increment was 0.10 μm, and a calibrated sub-tick interpolation uncertainty was unavailable. We therefore interpret the experimental result as evidence of submicrometre wafer overlay, not as a verification of nanometre-scale synthetic-model accuracy. The reported 3σ statistic describes empirical spread and is not assigned a traceable expanded-uncertainty interpretation.

Use of artificial intelligence

ChatGPT (OpenAI) was used to assist with manuscript reformatting and language editing during preparation of the Scientific Reports version. This assistance did not generate the experimental measurements or replace scientific validation.

Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Code availability

The custom code used for optical simulation, image reconstruction and quantitative analysis is available from the corresponding author upon reasonable request.

Acknowledgements

This work was financially supported by Center for Advanced Semiconductor Technology Research under Higher Education Sprout Project of Ministry of Education, Taiwan, the Co-creation Platform of the Industry Academia Innovation School, NYCU, under the framework of the National Key Fields Industry-University Cooperation and Skilled Personnel Training Act, and the National Science and Technology Council, Taiwan, under NSTC 115-2221-E-A49-159-MY3, and NSTC 113-2221-E-A49-090-MY3, and NSTC T-Star Center Project: Future Semiconductor Technology Research Center under NSTC 115-2634-F-A49-009-.

Author contributions

Shie-Ping Chang and Kuan-Neng Chen conceived the study. Shie-Ping Chang developed the optical model and reconstruction framework, performed the simulations and quantitative analyses, and prepared the figures and original manuscript draft. Chen-Chia Chang contributed to the wafer-alignment experiments, experimental data acquisition and interpretation. Kuan-Neng Chen supervised the research and provided scientific guidance. All authors discussed the results, revised the manuscript and approved the final version.

Competing interests

[Competing-interest statement to be confirmed by the authors.]

Additional information

Supplementary Information accompanies this article. Correspondence and requests for materials should be addressed to Kuan-Neng Chen.

References

1. Chang, C.-C. et al. High-precision wafer-level bonding in Thin-3D: A Moiré pattern and deep-learning alignment approach. in Proc. 76th IEEE Electron. Compon. Technol. Conf. (ECTC), Orlando, FL, USA, 2026, 1245–1251. https://doi.org/10.1109/ECTC51846.2026.00207. (2026)

2. Wang, C. & Suga, T. A novel Moiré fringe assisted method for nanoprecision alignment in wafer bonding. in Proc. 59th IEEE Electron. Compon. Technol. Conf. (ECTC), 2009, 872–878. https://doi.org/10.1109/ECTC.2009.5074115. (2009)

3. Huang, B., Wang, C., Fang, H., Zhou, S. & Suga, T. Moiré-based alignment using centrosymmetric grating marks for high-precision wafer bonding. Micromachines 10, 339. https://doi.org/10.3390/mi10050339. (2019)

4. Xu, F., Chang, S., Zhao, Y., Li, W. & Xia, H. High-precision Moiré alignment method using centrosymmetric grating for 3D interlayer interconnect. Opt. Lasers Eng. 171, 107815. https://doi.org/10.1016/j.optlaseng.2023.107815. (2023)

5. Wang, N., Jiang, W. & Zhang, Y. Deep learning-based Moiré-fringe alignment with circular gratings for lithography. Opt. Lett. 46, 1113–1116. https://doi.org/10.1364/OL.414617. (2021)

6. Zhu, J. et al. Influence of tilt Moiré fringe on alignment accuracy in proximity lithography. Opt. Lasers Eng. 51, 371–381. https://doi.org/10.1016/j.optlaseng.2012.12.001. (2013)

7. Wolf, E. New theory of partial coherence in the space–frequency domain. Part I: Spectra and cross spectra of steady-state sources. J. Opt. Soc. Am. 72, 343–351. https://doi.org/10.1364/JOSA.72.000343. (1982)

8. Matsushima, K. & Shimobaba, T. Band-limited angular spectrum method for numerical simulation of free-space propagation in far and near fields. Opt. Express 17, 19662–19673. https://doi.org/10.1364/OE.17.019662. (2009)

9. Green, M. A. Self-consistent optical parameters of intrinsic silicon at 300 K including temperature coefficients. Sol. Energy Mater. Sol. Cells 92, 1305–1310. https://doi.org/10.1016/j.solmat.2008.06.009. (2008)

10. Mendeleyev, V. Y., Skovorodko, S. N., Lubnin, E. N. & Prosvirikov, V. M. Optical constants of silicon in near infrared region. Appl. Phys. Lett. 93, 131916. https://doi.org/10.1063/1.2994669. (2008)

11. Born, M. & Wolf, E. Principles of Optics Electromagnetic Theory of Propagation Interference and Diffraction of Light 7th expanded edn (Cambridge University Press, Cambridge). (1999)

12. Goodman, J. W. Introduction to Fourier Optics 4th edn (W. H. Freeman, New York). (2017)

13. Hopkins, H. H. On the diffraction theory of optical images. Proc. Roy. Soc. London A 217, 408–432. https://doi.org/10.1098/rspa.1953.0071. (1953)

14. Kim, M.-S., Scharf, T., Menzel, C., Rockstuhl, C. & Herzig, H. P. Talbot images of wavelength-scale amplitude gratings. Opt. Express 20, 4903–4920. https://doi.org/10.1364/OE.20.004903. (2012)

15. Hua, Y., Suh, J. Y., Zhou, W., Huntington, M. D. & Odom, T. W. Talbot effect beyond the paraxial limit at optical frequencies. Opt. Express 20, 14284–14291. https://doi.org/10.1364/OE.20.014284. (2012)

16. Shao, G. & Helu, M. Framework for a digital twin in manufacturing: Scope and requirements. Manuf. Lett. 24, 105–107. https://doi.org/10.1016/j.mfglet.2020.04.004. (2020)

17. Guizar-Sicairos, M. & Fienup, J. R. Phase retrieval with transverse translation diversity: A nonlinear optimization approach. Opt. Express 16, 7264–7278. https://doi.org/10.1364/OE.16.007264. (2008)

18. Guizar-Sicairos, M., Thurman, S. T. & Fienup, J. R. Efficient subpixel image registration algorithms. Opt. Lett. 33, 156–158. https://doi.org/10.1364/OL.33.000156. (2008)

19. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process. 13, 600–612. https://doi.org/10.1109/TIP.2003.819861. (2004)

20. MORITEX Corp. MML4-HR65D telecentric lens. product specification. https://www.moritex.com/Lens/195_9580.html. accessed Jul. 27, 2026.

21. Basler AG. ace acA1300-200um USB 3.0 camera. product specification. https://www.baslerweb.com/en/shop/aca1300-200um/. accessed Jul. 27, 2026.

22. Signify Holding. 7023 100 W GY6.35 12 V halogen non-reflector lamp. product data sheet. https://www.signify.com/api/assets/v1/file/Signify/content/923870017103_EU.en_AA.PROF.FP/7023_100 W_GY6_35_12V_1CT_10X10F.pdf. accessed Jul. 27, 2026.

23. Zuo, C. et al. Deep learning in optical metrology: a review. Light Sci. Appl. 11, 39. https://doi.org/10.1038/s41377-022-00714-x. (2022)

24. Zhu, X. et al. Hformer: Hybrid convolutional neural network transformer network for fringe order prediction in phase unwrapping of fringe projection. Opt. Eng. 61, 093107. https://doi.org/10.1117/1.OE.61.9.093107. (2022)

25. Joint Committee for Guides in Metrology. Evaluation of Measurement Data Guide to the Expression of Uncertainty in Measurement. JCGM 100:2008(E). https://doi.org/10.59161/JCGM100-2008E. (2008)

26. Hutchinson, S., Hager, G. D. & Corke, P. I. A tutorial on visual servo control. IEEE Trans. Robot. Autom. 12, 651–670. https://doi.org/10.1109/70.538972. (1996)

27. Wang, C. & Suga, T. Measurement of alignment accuracy for wafer bonding by Moiré method. Jpn. J. Appl. Phys. 46, 1989–1993. https://doi.org/10.1143/JJAP.46.1989. (2007)

Tables

Table 1 | Summary of synthetic reconstruction benchmarks and measured wafer-level overlay. Synthetic MAEs should not be interpreted as physical alignment accuracy, and the Vernier spread is not a traceable nanometre-scale uncertainty.

Evidence / dataset	Δx MAE	Δy MAE	Additional result
PDTOM-A held-out synthetic (N = 500)	3.295 nm	3.711 nm	Radial P95 = 14.713 nm
PDTOM-A near-zero synthetic (N = 500)	3.383 nm	3.376 nm	Radial P95 = 13.380 nm
PDTOM-C compact synthetic (N = 1,000)	0.634 nm	0.663 nm	θ MAE = 0.003510°; simulated Z-gap MAE = 1.801 nm
Physical Vernier overlay (N = 90)	—	—	Mean (0.08, 0.22) μm; reported 3σ spread = 0.51 μm

Additional metrics, population definitions and qualifications are provided in Supplementary Tables S4–S7. Provenance limitations of the PDTOM-C benchmark are discussed in Supplementary Section S5. Both configurations estimate four state variables; Table 1 emphasizes the independently reported performance statistics and does not imply four-coordinate local refinement in the PDTOM-A benchmark.

Figure 1

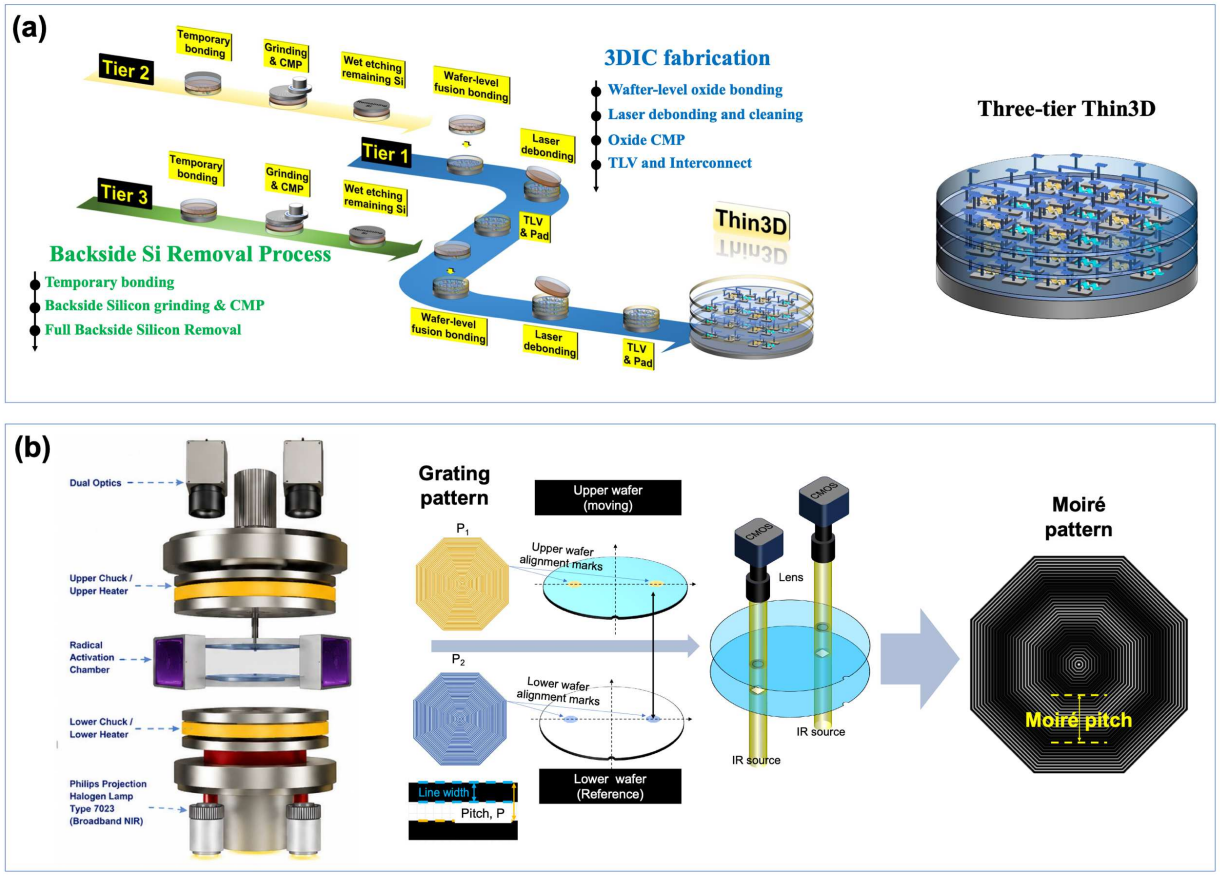

Fig. 1. | Thin-3D process context and AML wafer-to-wafer optical alignment platform. (a) Intended Thin-3D fabrication sequence. (b) Dual-channel AML imaging architecture. The fabrication sequence is illustrative; the present optical-metrology study does not validate all process steps.

Figure 2

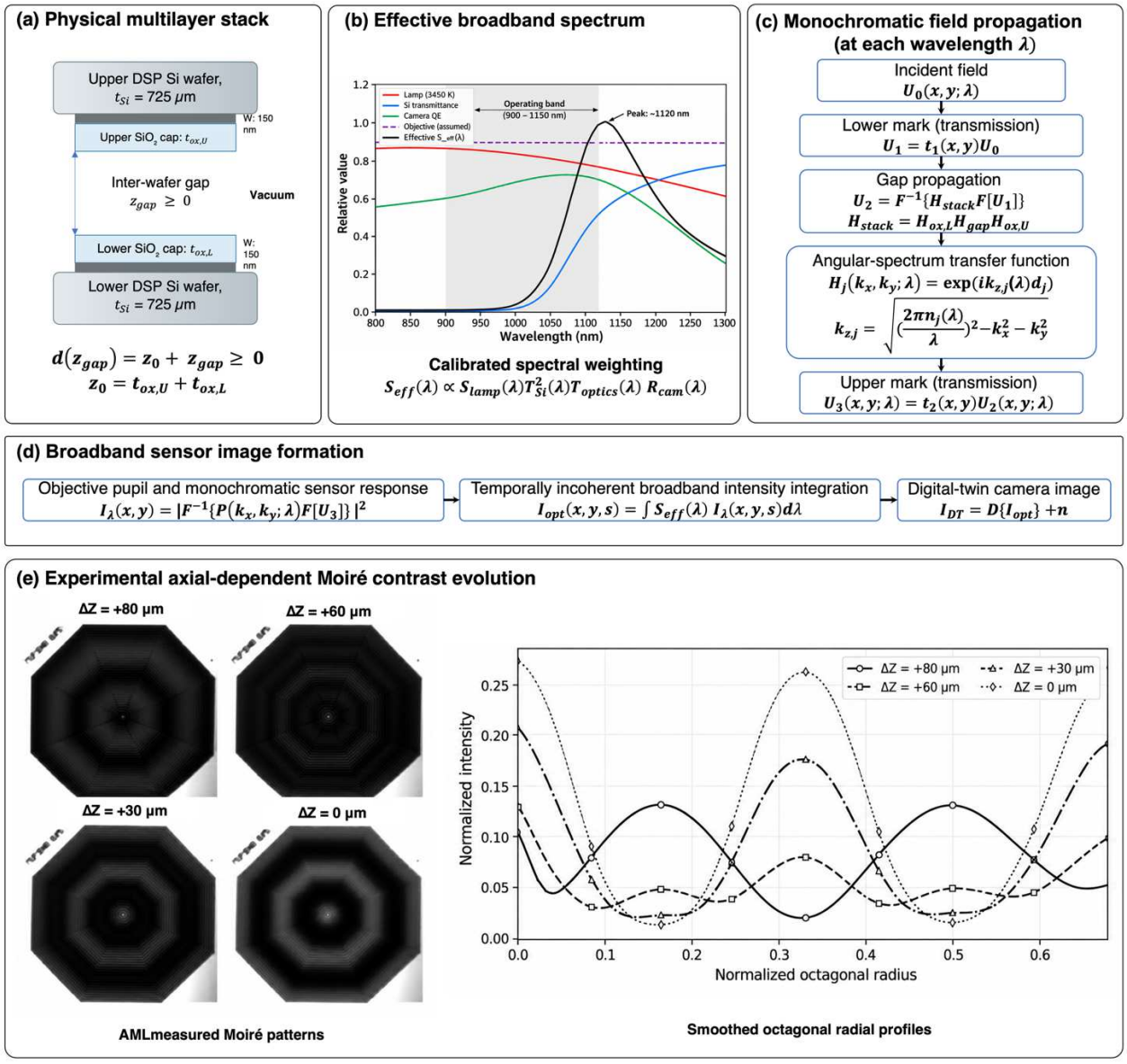

Figure 2 | Gap-aware broadband optical model and experimental axial-dependent Moiré contrast evolution. (a) Physical multilayer stack comprising opposing Si wafers, patterned marks, oxide layers and the inter-wafer gap. (b) Illustrative effective broadband spectrum including source emission, wafer transmission, optical throughput and camera response. Silicon attenuation is included at field level in the numerical model and must not be counted again in the spectral integration weight. The S_eff notation in the preserved schematic denotes the illustrative detected-spectrum representation; the numerical field-level formulation uses W in Supplementary equations (S19)–(S20). (c) Wavelength-dependent multilayer angular-spectrum propagation. (d) Objective-pupil filtering, incoherent spectral integration and sensor image formation. (e) AML-measured octagonal Moiré patterns at relative stage displacements of +80, +60, +30, and 0 μm, with corresponding smoothed radial intensity profiles. The patterns exhibit axial-dependent contrast redistribution and local polarity reversal. The stage displacements are not absolute physical wafer gaps.

Figure 3

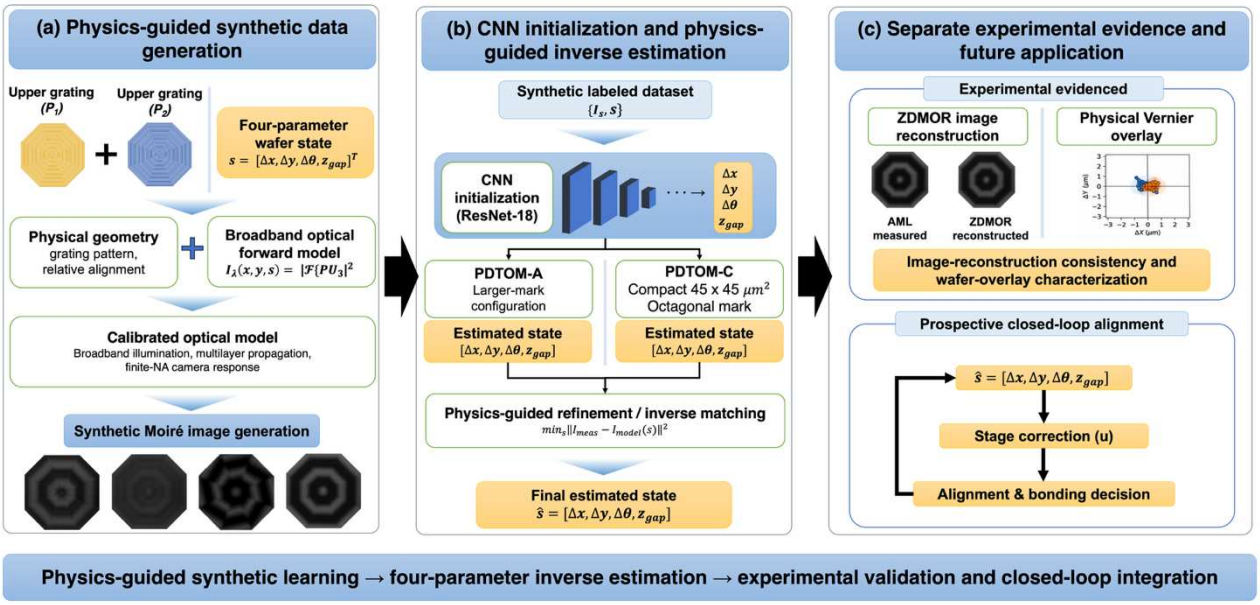

Fig. 3 | Physics-guided PDTOM workflow for four-parameter Moiré alignment metrology. (a) Octagonal geometry and a calibrated broadband forward model produce synthetic images with state labels (Δx, Δy, Δθ, z_{gap}). (b) A ResNet-18 initializer estimates all four components for the larger-mark PDTOM-A and compact-mark PDTOM-C configurations. In the reported PDTOM-A benchmark, local physics refinement updates Δx and Δy while retaining CNN-derived rotation and reference-relative axial estimates. The archived PDTOM-C evaluation records four-parameter refinement, but the associated CNN checkpoint has not been independently verified. (c) ZDMOR image reconstruction and physical Vernier overlay provide separate experimental evidence (Figs. 4 and 6). Thumbnails are illustrative; Fig. 4 shows quantitative AML–ZDMOR image comparisons. Closed-loop wafer alignment is prospective rather than experimentally demonstrated.

Figure 4

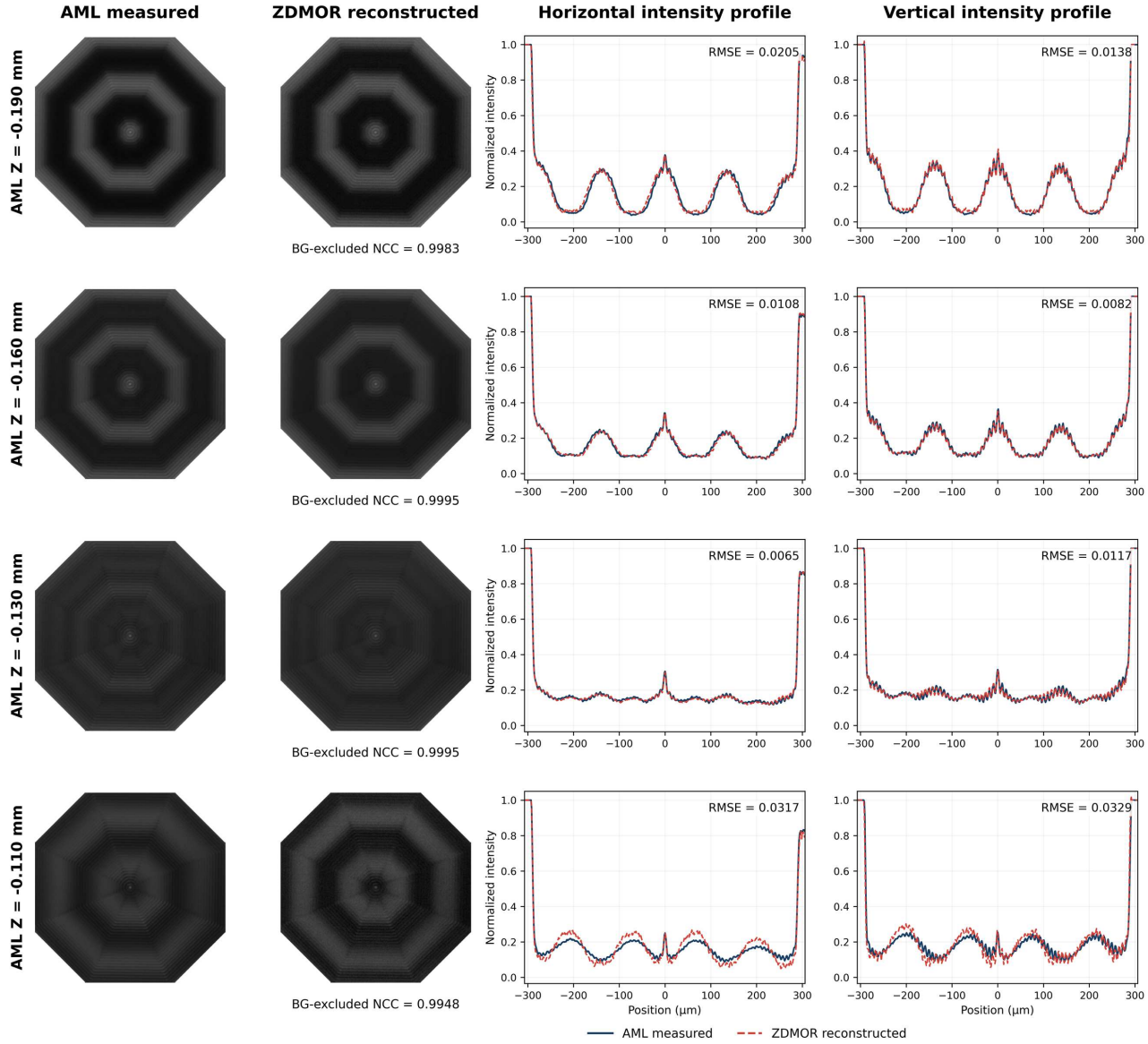

Fig. 4. | Experimental validation of Z-dependent Moiré optical reconstruction (ZDMOR) using a shared background-exclusion mask. Four AML stage coordinates are shown in rows (−0.190, −0.160, −0.130 and −0.110 mm). In each row, panels show the AML measured octagonal Moiré mark, the corresponding ZDMOR reconstruction, and the horizontal and vertical normalized-intensity profiles. The displayed image panels are cropped around the octagonal marks for visual clarity, while quantitative NCC was calculated separately on the registered 512 × 512 images after identically excluding the upper-left annotation and lower-right non-Moiré shadow regions. The common mask retained 89.11% of registered pixels; background-excluded NCC values were 0.998328, 0.999455, 0.999507 and 0.994814 (mean 0.998026). The horizontal/vertical profile RMSEs were calculated from central seven-pixel averaging bands and use all 512 profile samples. Model Z coordinates were optimized against the target images; the stage readings are relative mechanical positions, not absolute wafer gaps.

Figure 5

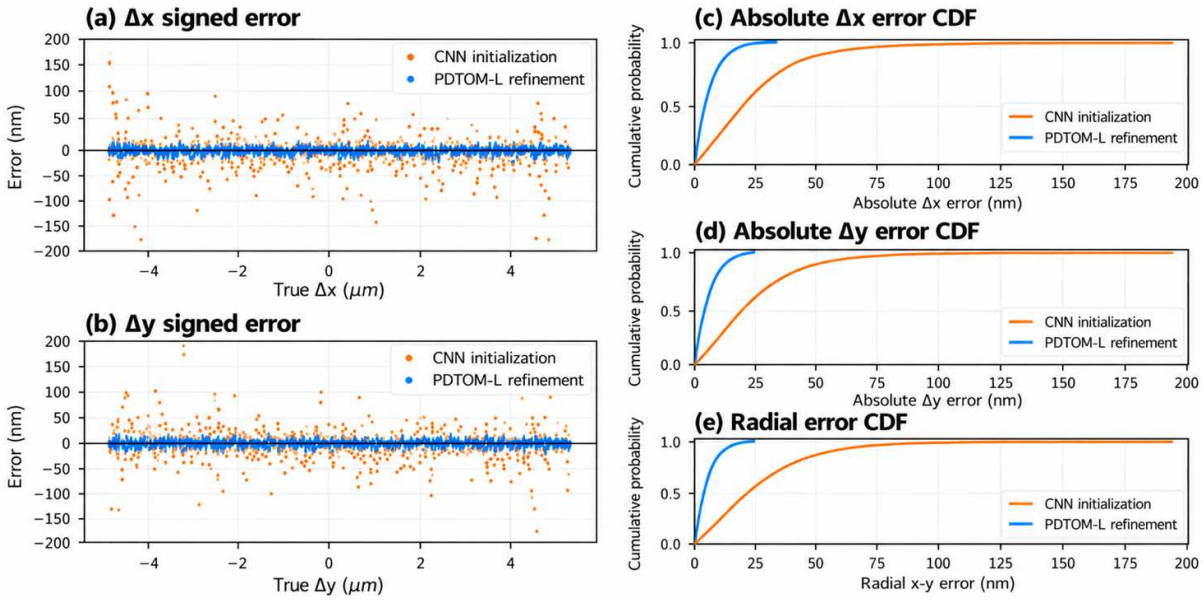

Fig. 5. | PDTOM-A independent held-out synthetic evaluation (N = 500). Signed lateral errors and absolute-error cumulative distributions compare the recorded CNN initializer with physics-guided lateral refinement. The synthetic result is not a physical wafer-overlay accuracy measurement.

Figure 6

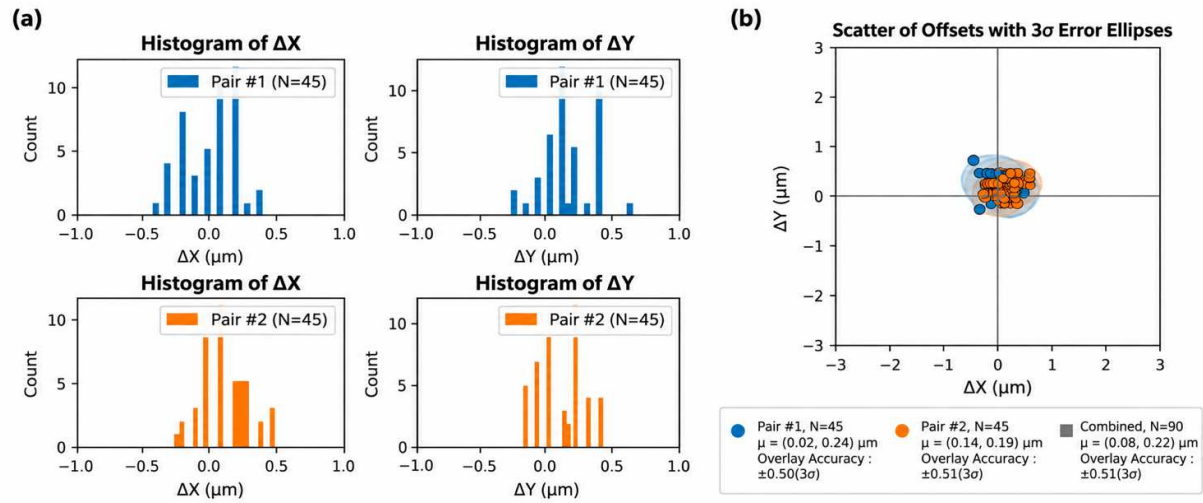

Fig. 6. | Physical wafer-level Vernier overlay. (a) Lateral offset histograms for two-mark pairs (45 measurements each) and (b) two-dimensional distribution. Combined mean offset is (0.08, 0.22) μm; reported 3σ spread is 0.51 μm.

Figure 7

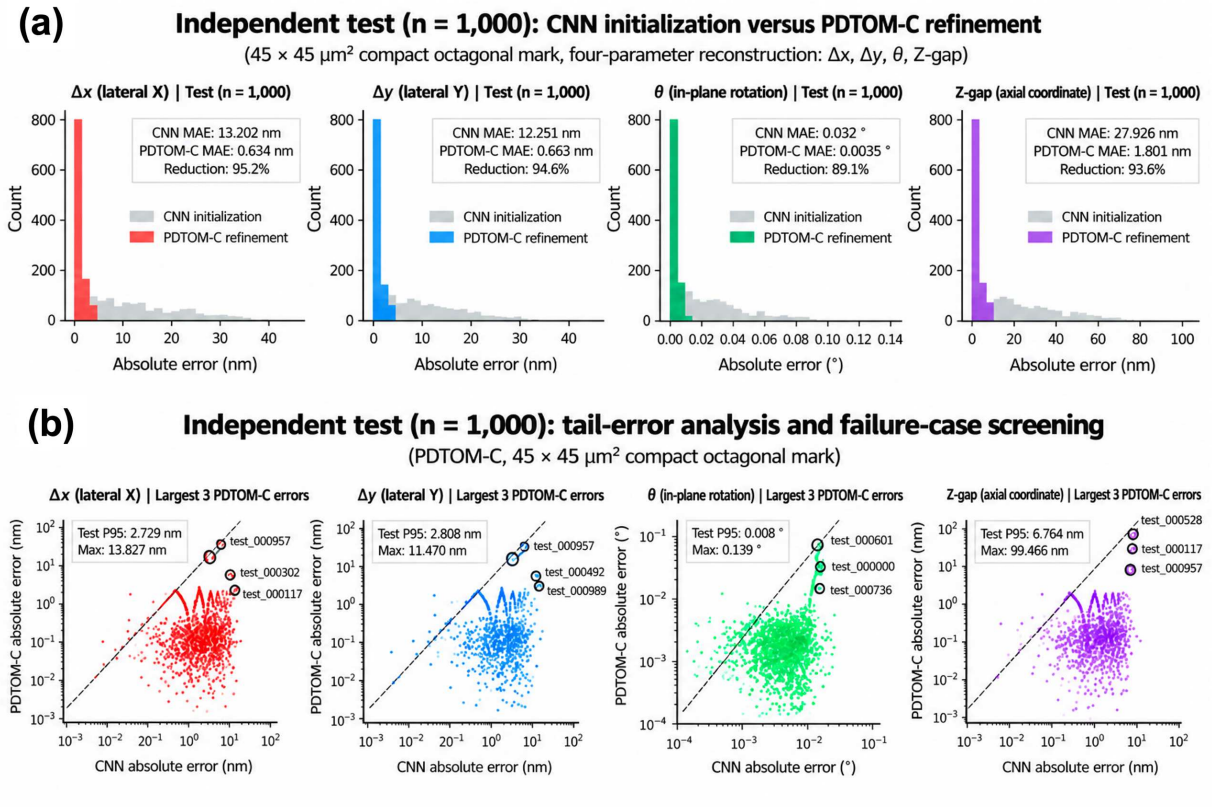

Fig. 7 | Four-parameter synthetic reconstruction and tail-error analysis of PDTOM-C (N = 1,000). (a) Absolute-error distributions comparing recorded initial estimates with refined results for Δx, Δy, θ, and simulated z_{gap}. (b) Initial-versus-refined absolute errors on logarithmic axes. Dashed diagonals indicate equal initial and refined errors; highlighted points identify the three largest final residuals for each parameter. Percentile and maximum errors characterize remaining error tails. The results are synthetic-domain estimates; checkpoint lineage and experimental compact-mark accuracy have not been independently verified.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- PDTOMSupplementaryInformationV1.pdf