Supplementary Information

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

This Supplementary Information describes the Moiré geometry, optical forward model, inverse estimation and validation protocols. Sections S1–S6 distinguish synthetic estimation, target-guided image reconstruction and physical process overlay. Supplementary Tables S1–S8 and Figures S1–S3 provide supporting results.

S1 Moiré geometry and state definition

S2 Broadband transmission optical forward model

S3 CNN initialization and inverse reconstruction

S4 Experimental setup and validation

S5 Compact-mark PDTOM-C dataset provenance and validation limits

S6 Experimental ZDMOR reconstruction and ROI sensitivity

Terminology. Both PDTOM-A and PDTOM-C estimate four state variables $[\Delta x, \Delta y, \Delta\theta, z_gap]^T$. The reported PDTOM-A optimizer refines only Δx and Δy after four-parameter CNN initialization, whereas the separate archived PDTOM-C evaluation records four-parameter refinement. ZDMOR is the experimental image-reconstruction analysis, distinct from either synthetic estimator benchmark.

S1 Moiré geometry and state definition

Measurement Concept

PDTOM combines a calibrated optical forward operator with four-parameter initialization and implementation-specific inverse refinement. The method overview is given in main-text Fig. 3; the equations and qualification limits below support that workflow. [2, 3, 4, 8, 16, 17]

Moiré Formation and Translation Amplification

Let P_1 and P_2 denote the local periods of corresponding lower- and upper-mark grating families, with $P_2 > P_1$. Their spatial-frequency difference produces the Moiré period used in wafer-bond alignment [2,3,4]:

$$P_M = \frac{P_1 P_2}{P_2 - P_1} \tag{S1}$$

Defining the pitch ratio $r = P_2/P_1$ gives $PM = rP_1/(r-1)$. For displacement of the P_1 grating relative to P_2 along the corresponding local grating normal, the translation amplification factor is therefore [2,3]:

$$\boldsymbol{A_M = \frac{P_M}{P_1} = \frac{r}{r-1}, \qquad r = \frac{P_2}{P_1}} \tag{S2}$$

For r = 1.06, motion of P_1 gives $A_1 = 17.67$, so an ideal 100-nm translation shifts the fringes by approximately 1.77 μm. Motion of P_2 gives $A_2 = 16.67$. Rotation is represented by the full two-dimensional rigid transformation rather than a scalar gain.[2,3,4]

Unified Wafer State

The instantaneous state reconstructed by PDTOM is

$$\mathbf{s} = [\Delta x \quad \Delta y \quad \Delta\theta \quad z_{\text{gap}}]^{\mathrm{T}} \tag{S3}$$

Here Δx and Δy are lateral translations, Δθ is in-plane rotation, and z_gap is the nonnegative effective ambient-propagation coordinate. Reported axial displacement is Δz = z_gap − z_ref, with a shared calibration offset z_ref at the reproducible AML working position. Thus, Δz = 0 denotes the selected reference, not zero physical gap. The lateral variables change mark geometry; z_gap changes propagation and image contrast.[6,8]

Geometrical Transformation

For a coordinate $x = [x, y]^T$ expressed in the lower-mark frame and an upper-mark pose defined by translation $t = [\Delta x, \Delta y]^T$ and counterclockwise rotation Δθ, inverse mapping into the upper-mark texture coordinates gives

$$\mathbf{x}' = \mathbf{R}(-\Delta\theta)(\mathbf{x} - \mathbf{t}), \qquad \mathbf{R}(\alpha) = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \tag{S4}$$

The transformed coordinates determine upper-grating transmission. Small rotations produce displacement ρΔθ at radius ρ, giving outer segments greater rotational leverage. Translation shifts phase along local grating normals; rotation changes all eight sides coherently. Axial displacement acts through propagation rather than this rigid transformation.[3,4]

Concentric octagonal large-mark geometry

Supplementary Fig. S1 shows the concentric octagonal mark derived from Deep-learning Optical Moiré Alignment (DLOMA). Two grating sets encode lateral translation and rotation. For N_1 and N_2 cycles, the approximate outer diameter is[2,3,4,1]:

$$D_{\text{mark}} \approx 2\max(N_1 P_1, N_2 P_2) \tag{S5}$$

The experimental mark uses $P_1 = 8.00$ μm, $P_2 = 8.48$ μm, $N_1 = 37$, $N_2 = 35$ and 50% duty cycle. This gives r = 1.06, a Moiré period of 141.33 μm, moving-P_1 gain of 17.67 and outer extent of approximately 593.6 μm. Future sub-200-μm designs require separate optical and fabrication qualification.[3,4,1]

Mark-size-dependent cycle design and finite-overlap criterion

Define the concentric-octagonal radial coordinate $\rho(x,y) = \max(|x|, |y|, (|x| + |y|)/\sqrt{2})$. With nominal flat-to-flat mark width L, the permitted radial extent is R = L/2. If the i-th grating contains N_i complete radial periods P_i, its occupied radius is $R_i = N_i P_i \leq L/2$, and the maximum integer cycle count is $N_i = \lfloor L/(2P_i) \rfloor$. (S6)

For locally parallel radial gratings, the product of $\cos(2\pi\rho/P_1)$ and $\cos(2\pi\rho/P_2)$ contains a difference-frequency term with $1/P_m = |1/P_1 - 1/P_2|$. Consequently, $P_m = P_1P_2/|P_2 - P_1| = rP_1/(r - 1)$, with $r = P_2/P_1 > 1$. (S7) This is a local geometrical approximation, not an exact prediction of the propagated finite-

octagon image.

Because the gratings may have unequal radii after integer truncation, their shared radial region is $R_o = \min(N_1P_1, N_2P_2)$. The number of nominal Moiré beat periods inside that overlap is $N_{m,overlap} = R_o/P_m$. (S8) The inequality $N_{m,overlap} \geq 1$ is a preliminary geometric screening guideline. (S9) It is not a necessary condition for pose estimation and is not a guarantee of optical visibility. Fringe phase, finite aperture, propagation, contrast and camera sampling require separate qualification.

Ignoring integer truncation and assuming full use of the footprint, a target $N_{m,target}$ gives $L/(2P_m) \geq N_{m,target}$, hence $r \geq L/(L - 2N_{m,target}P_1)$, provided $L > 2N_{m,target}P_1$. (S10) The actual integer N_1 and N_2 must then be recalculated and checked against (S8)–(S9); the inverse bound is not sufficient for finite masks.

The analytical 45 × 45 μm² candidate in Table S3 uses $P_1 = 1.50$ μm, $P_2 = 1.65$ μm, $N_1 = 15$ and $N_2 = 13$, yielding $P_m = 16.50$ μm and $N_{m,overlap} = 1.30$. It is distinct from the actual PDTOM-C benchmark: $P_1 = 1.50$ μm, $P_2 = 1.59$ μm, $N_1 = 15$ and $N_2 = 14$, $P_m = 26.50$ μm, $R_o = 22.26$ μm and $N_{m,overlap} \approx 0.84$. For the experimental large mark, the 8.00/8.48-μm, 37/35-cycle geometry gives $R_o = 296$ μm and $N_{m,overlap} \approx 2.09$; its outer extent is approximately 593.6 μm. The 592-μm-constrained candidate instead has $N_2 = 34$.

Talbot scaling and optical qualification

For a monochromatic periodic grating in a homogeneous medium and in the paraxial limit, $z_t \approx 2P^2/\lambda$. (S11) At $\lambda = 1.00$ μm, $P = 8.00$ μm gives $z_t \approx 128$ μm, whereas $P = 1.50$ μm gives $z_t \approx 4.50$ μm. These are characteristic diffraction-phase scales, not Z-gap measurement ranges. For wavelength-scale pitches, nonparaxial propagation and the number of collected diffraction orders can substantially change recurrence behavior; the existing full broadband angular-spectrum treatment in S2 is required.

Proposed optical qualification comprises visibility and four-parameter identifiability tests using the actual wafer stack, source, NA, camera integration and calibrated noise. Existing AML sampling is 1.2 μm/pixel, so 1.50–1.59 μm periods occupy only about 1.25–1.33 samples. Adequate resolution cannot be assumed; magnification, aperture transfer and sampling must be qualified on hardware. These checks are future work, not completed compact-mark experiments.

Supplementary Table S3 lists the geometrical candidates and makes the distinction between the existing experimental mark and footprint-constrained design proposals explicit.

S2 Broadband transmission optical forward model

Transmission imaging and grating representation

The modeled path passes through the lower 725-μm Si wafer, lower mark, inter-wafer layers, upper mark and upper 725-μm Si wafer. The source, objective and camera specifications are consolidated in Table S2. Each imaging channel requires its own calibration. [9, 10, 20, 21, 22]

At wavelength λ and illumination angle α, tL and tU denote patterned complex-amplitude transmissions, u0 is the unit-power incident field, and $\tau_{Si,L}$ and $\tau_{Si,U}$ are the silicon field-transmission coefficients. Source radiance enters only through the detection weight.

$$U_L(x, y; \lambda, \boldsymbol{\alpha}) = \tau_{\mathrm{Si},L}(\lambda) u_0(x, y; \lambda, \boldsymbol{\alpha}) t_L(x, y; \lambda) \tag{S12}$$

Gap Propagation and Upper-Wafer Interaction

The band-limited multilayer angular-spectrum method (ASM) propagates the lower patterned field

through the SiO_2 caps and ambient segment at each wavelength and source angle.[8,12,11,13]

$$k_{z,j}(f_x, f_y; \lambda) = 2\pi\sqrt{\left(\frac{n_j}{\lambda}\right)^2 - f_x^2 - f_y^2} \tag{S13}$$

Sampling is checked for every wavelength and axial coordinate. Unsupported spatial frequencies are suppressed or assigned evanescent decay.[8,12]

$$H_j(f_x, f_y; \lambda, d_j) = \exp\left[i\, k_{z,j}(f_x, f_y; \lambda) d_j\right] \tag{S14}$$

In (S14), longitudinal spatial frequency and distance are evaluated per layer. Propagating components accumulate phase with unit magnitude; evanescent components decay. Equation (S15) multiplies the lower-cap, ambient and upper-cap transfer functions. Multiple reflections are neglected and require separate ablation.[8,12,11]

$$H_{\text{stack}} = H_{\text{ox}}(d_{\text{ox}})\, H_{\text{gap}}(z_{\text{gap}})\, H_{\text{ox}}(d_{\text{ox}}) \tag{S15}$$

Let FT_2 and IFT_2 denote the two-dimensional Fourier transform and its inverse. The field arriving at the upper W-mark plane is

$$U_z = \mathcal{F}_{2\text{D}}^{-1}\{\mathcal{F}_{2\text{D}}\{U_L\}\, H_{\text{stack}}\} \tag{S16}$$

After multilayer propagation, the upper-wafer interaction is

$$U_U(x, y; \lambda, \boldsymbol{\alpha}, z) = \tau_{\text{Si},U}(\lambda) t_U(x', y'; \lambda) U_z(x, y; \lambda, \boldsymbol{\alpha}) \tag{S17}$$

The nominal model assumes locally uniform separation. Independently measured tilt or bow may be incorporated as z(x,y) using split-step propagation or a validated local-phase approximation.[8,12,11]

Broadband integration and camera sampling

The objective pupil has cutoff $f_c(\lambda) = NA/\lambda$, where NA is numerical aperture (0.112), and may include calibrated aberration. Equation (S18) gives the monochromatic sensor intensity. Camera formation includes 4× magnification, 4.8-µm pixel-area integration, exposure, gain, offset, quantization and calibrated noise.[7,20,21,12,11,13]

$$I_{\lambda,\boldsymbol{\alpha}}(x, y; \mathbf{s}) = \left|\mathcal{F}_{2\text{D}}^{-1}\{\mathcal{F}_{2\text{D}}\{U_U\}\, P(f_x, f_y; \lambda)\}\right|^2 \tag{S18}$$

Silicon transmission is already included at field level in (S12) and (S17). The detection weight therefore contains only source radiance, objective throughput and camera response:

$$W(\lambda, \boldsymbol{\alpha}) = S_{\text{lamp}}(\lambda, \boldsymbol{\alpha}) H_{\text{opt}}(\lambda, \boldsymbol{\alpha}) R_{\text{cam}}(\lambda) \tag{S19}$$

Distinct wavelengths and independently sampled illumination angles are combined in intensity. Figure 2(b) shows an illustrative effective detected spectrum including transmission through both Si wafers; it is not W(λ,α), the numerical integration weight. No additional Si-transmission factor is applied to W. No band-pass filter or assumed 1.00–1.05-µm passband is introduced. [7, 9, 10, 11]

$$I^{\text{BB}}(x, y; \mathbf{s}, \boldsymbol{\psi}) = \int_{\Lambda} \int_{\mathcal{A}} W(\lambda, \boldsymbol{\alpha}) I_{\lambda,\boldsymbol{\alpha}}(x, y; \mathbf{s}, \boldsymbol{\psi})\, \mathrm{d}\boldsymbol{\alpha}\, \mathrm{d}\lambda \tag{S20}$$

Here Λ and A are the calibrated wavelength and illumination-angle domains, $\mathcal{D}$ is the camera-formation operator, and ψ contains frozen instrument parameters. Silicon transmission must not be counted again in W(λ,α).[7,12,11]

$$I_{\text{DT}}(x, y; \mathbf{s}, \boldsymbol{\psi}) = \mathcal{D}\{I^{\text{BB}}(x, y; \mathbf{s}, \boldsymbol{\psi})\} + n(x, y) \tag{S21}$$

The resulting image averages spectral, angular, finite-NA and pixel responses. Blur, background and

radiometric parameters are shared across axial frames.[21,12,13]

Near-Field Diffraction and Optical-State Evolution

Talbot distance provides an axial phase scale, while the full broadband, multilayer, finite-NA model determines the measured response.[12,11,14,15]

$$z_T(p,\lambda_0) = \frac{\lambda_0}{1-\sqrt{1-(\lambda_0/p)^2}} \tag{S22}$$

For p ≫ λ₀, (S22) reduces to the familiar paraxial approximation [14]:

$$z_T(p,\lambda_0) \approx \frac{2p^2}{\lambda_0}, \qquad p \gg \lambda_0 \tag{S23}$$

The 8.00/8.48-μm grating pair has wavelength-dependent recurrence scales. Measured recurrence and ambiguity must therefore be obtained from the calibrated axial sweep.[14,15]

$$\Delta\phi_1(z,\lambda_0) = \frac{2\pi z}{z_T(p,\lambda_0)} \tag{S24}$$

In an ideal monochromatic two-order interpretation, a half-Talbot-scale displacement can produce a phase difference near π. This conditional model relation does not establish a measured π shift in the broadband AML images. [14]

$$I_{\lambda_0}\left(x,\frac{z_T}{2}\right) \approx I_{\lambda_0}\left(x+\frac{p}{2},0\right) \tag{S25}$$

The idealized relation explains how propagation can reverse contrast, but the actual intensity follows the collected orders and spectral-angular average. Main-text Results discusses the local polarity reversal in Fig. 2(e); it is not direct phase measurement. [7, 11, 12, 14, 15]

Process state

Equation (S26) allows the stack to depend on process state. Before contact, z_gap represents ambient propagation; a bonded-state extension would replace that segment with a qualified interface model. Absolute local gap requires independent offset, bow and tilt calibration.[9,10,8,11]

$$I_{\mathrm{DT}} = \mathcal{F}(\mathbf{s}, c; \boldsymbol{\psi}), \qquad c \in \mathcal{C} \tag{S26}$$

main-text Figs. 2 and 3 show the optical model and inverse workflow, respectively.

Generator implementation and consistency checks

The generator evaluates s = [Δx, Δy, Δθ, z_gap]ᵀ. Its broadband ASM component is combined with weak geometrical carrier texture, an effective point-spread function (PSF) and a frozen radiometric transform:

$$I(\mathbf{s};\boldsymbol{\psi}) = \mathcal{T}_{\boldsymbol{\psi}}\left\{(1-w_c)[I_{\mathrm{ASM}}(\mathbf{s}) * h_b(z)] + w_c\left[I_{\mathrm{geo}}(\mathbf{s}) * h_s(z)\right]\right\} \tag{S27}$$

$$h_b(z) = \eta G(\sigma_b) + (1-\eta)A[r_A(z)],\ h_s(z) = G[\sigma_s(z)] \tag{S28}$$

$$\sigma_s(z) = \sigma_0 + k_\sigma \frac{\Delta z}{Z_{\max}}, \quad \sum_{\lambda,\boldsymbol{\alpha}} W(\lambda,\boldsymbol{\alpha}) = 1 \tag{S29}$$

IASM is the angle-averaged broadband result; Igeo uses the same lateral pose; hG and hA are Gaussian and Airy kernels; and Tψ is the radiometric mapping. Carrier weight, Airy radius, axial sharpness slope and radiometric coefficients are shared empirical calibration parameters.

Consistency checks cover analytical Moiré gain, rigid transformations, H(fx,fy;0) = 1, unit-modulus

propagating transfer, normalized spectral-angular weights, sampling convergence and a frozen calibration across the axial sequence. Supplementary Fig. S2 shows directional translation and rotation responses and axial contrast evolution with fixed grating cycles. These checks assess generator consistency, not independent measurement accuracy. 8,12,11,19

S3 CNN initialization and inverse reconstruction

Main-text Fig. 3 provides the workflow overview. This section specifies the objective, CNN initialization and implemented refinement scopes; it does not report experimental stage feedback.

Inverse Problem and Contrast-Robust Objective

Measured and simulated images IE and IDT(s) are standardized to zero mean and unit standard deviation. State s = [Δx, Δy, Δθ, z_gap]^T and contrast polarity are estimated by

$$(\hat{\mathbf{s}}, \hat{\kappa}) = \underset{\mathbf{s}\in\Omega,\, \kappa\in\{-1,+1\}}{\arg\min} \; J(\mathbf{s}, \kappa) \tag{S30}$$

Here Ω bounds the calibrated state domain. Contrast polarity κ is +1 for direct contrast and −1 for bright–dark reversal. The objective is

$$\begin{aligned} J(\mathbf{s},\kappa) = {} & w_{\mathrm{NCC}} d_{\mathrm{NCC}}\left(\tilde{I}_E, \kappa \tilde{I}_{\mathrm{DT}}(\mathbf{s})\right) + w_g d_g\left(\tilde{I}_E, \kappa \tilde{I}_{\mathrm{DT}}(\mathbf{s})\right) \\ & + w_F d_F(|F(\tilde{\mathrm{I}}_E)|, |F(\tilde{\mathrm{I}}_{DT}(s))|) \; + \; w_\phi d_\phi(s,\kappa) \\ & + w_r d_r(s,\kappa) \; + \; w_{ring} d_{ring}(s,\kappa) \; + \; \lambda R(s) \end{aligned} \tag{S31}$$

The discrepancies compare NCC (dNCC = 1 − NCC), image gradients (dg), Fourier magnitudes (dF), radial profiles (dr) and ring energies (dring). R(s) represents state-domain and continuity constraints. The optional polarity variable introduces a π correction to the selected image-sideband phase in the objective; this mathematical contrast convention is not a measurement of an optical diffraction-order phase shift. [17, 18, 19]

$$d_\phi(\mathbf{s},\kappa) = 1 - \cos\left[\phi_E - \phi_{\mathrm{DT}}(\mathbf{s}) - \frac{1-\kappa}{2}\pi\right] \tag{S32}$$

where ϕ_E and $\phi_{\mathrm{DT}}(\mathbf{s})$ are extracted from the same selected Moiré sideband. For κ = +1, the correction in (S32) is zero; for κ = −1, it is π. The correlation and gradient terms retain spatial registration, the Fourier-magnitude term constrains period and diffraction-envelope agreement independently of contrast polarity, and the wrapped-phase term preserves lateral-displacement sensitivity. Radial-profile and ring-energy residuals are retained only when calibration and held-out tests demonstrate improved reference-relative axial observability. All objective weights are selected using calibration data and frozen before independent evaluation. Geometric registration precedes optical fitting, and the direct and inverted contrast hypotheses are evaluated explicitly rather than corrected independently for each frame [17,18,14,15,19].

Coarse-to-Fine State Estimation

Both configurations output four state components. In the evaluated PDTOM-A implementation, CNN initialization supplies all four coordinates, but the frozen local search updates only Δx and Δy. The archived PDTOM-C evaluation records four-parameter refinement. The generic state-update expression below describes bounded optimization; it does not establish identical executed solvers or verified checkpoint lineage for both archives. [17, 18]

After normalization by the calibrated parameter ranges, the candidate state is updated according to

$$\mathbf{s}_{k+1} = \Pi_{\Omega}[\mathbf{s}_k - \eta_k \nabla_{\mathbf{s}} J(\mathbf{s}_k, \kappa_k)] \tag{S33}$$

Each update reevaluates the forward model and objective, then projects into the calibrated domain. The general convergence criterion combines objective change with state-update magnitude.[17,18]

The following directional constraint describes a possible approach-sequence extension. It is not evidence of experimentally completed closed-loop control. [25, 26]

$$\left|z_{\mathrm{gap},t+1} - z_{\mathrm{gap},t} - \Delta z_{\mathrm{cmd},t}\right| \le \varepsilon_z, \qquad z_{\mathrm{gap},t} \ge 0. \tag{S34}$$

Stage history, the finite mark boundary and nonperiodic features reduce modulo-pitch ambiguity. PDTOM-A searches laterally with 80-, 20- and 5-nm steps and at most one accepted move per level. A confidence gate rejects score gains below 10^{-5} or excessive displacement. Boundary continuation uses 80-, 40-, 20- and 5-nm steps, at most two moves per level and a maximum 640-nm excursion. All schedules and thresholds are frozen before evaluation. 2,3,4,18

Future uncertainty extension

A prospective uncertainty extension would estimate repeated-acquisition feature covariance Σg and local sensitivity as follows. [25]

$$\mathbf{G} = \left.\frac{\partial\, \mathbf{g}[I_{\mathrm{DT}}(\mathbf{s})]}{\partial \mathbf{s}}\right|_{\mathbf{s}=\hat{\mathbf{s}}} \tag{S35}$$

If the local model is identifiable and $\mathbf{G}^{\mathsf{T}}\boldsymbol{\Sigma}_g^{-1}\mathbf{G}$ is nonsingular, the first-order state covariance is

$$\boldsymbol{\Sigma}_{\mathbf{s}} \approx \left(\mathbf{G}^{\mathsf{T}}\boldsymbol{\Sigma}_{\mathbf{g}}^{-1}\mathbf{G}\right)^{-1} \tag{S36}$$

These equations are retained as a future extension only. Identifiability, covariance calibration, reference uncertainty and failure coverage have not been validated; uncertainty-gated stage control is not an achieved result. [25, 26]

CNN training

A learned regressor fθ supplies the initial state for (S33), reducing repeated evaluations of the broadband operator F(s,c;ψ).

The frozen objective and state-domain projection are retained. The four-variable PDTOM-A estimator locally refines lateral translation only, preserving its CNN rotation and axial outputs. This implementation-specific refinement scope does not change the four-parameter definition of the predicted state.

The reported PDTOM-A CNN initializer uses a 7,000-image synthetic pool from the frozen operator. A grayscale 18-layer residual network (ResNet-18) predicts normalized $[\Delta x, \Delta y, \Delta\theta, \Delta z]^{\mathsf{T}}$. Training uses adaptive moment estimation with decoupled weight decay (AdamW; learning rate 10^{-4}; weight decay 10^{-4}), batch size 8, learning-rate reduction on a performance plateau (ReduceLROnPlateau), and early stopping. Weak gain, offset, gamma and additive-noise augmentation is limited to training to preserve axial sharpness cues. The main-text Fig. 5 test set uses an unused random seed and is excluded from model selection and refinement tuning.

$$\mathbf{s}^{(0)} = f_{\boldsymbol{\theta}}(I_{\mathrm{meas}}) \tag{S37}$$

Synthetic training images follow I = F(s,c;ψ) + n(x,y); the optical, process and state variables are sampled from their specified distributions:

$$\boldsymbol{\psi} \sim p_{\boldsymbol{\psi}}, \qquad c \sim p_c, \qquad \mathbf{s} \sim p_{\mathbf{s}|c} \tag{S38}$$

The implemented normalized weighted Smooth-L1 loss uses s = [Δx, Δy, Δθ, Δz]^T, weights w = [2, 2, 1.25, 1]^T and transition β = 0.10:

$$\mathcal{L}_{\boldsymbol{\theta}} = \mathbb{E}\left[\frac{1}{\sum_{j=1}^{4} w_j}\sum_{j=1}^{4} w_j \,\mathrm{SmoothL1}_{\beta}\left(\hat{s}_j - s_j\right)\right], \quad \mathbf{w} = [2,2,1.25,1]^{\mathsf{T}},\ \beta = 0.10 \quad \text{(S39)}$$

Main-text Fig. 5 and Supplementary Table S4 evaluate the complete CNN-to-PDTOM-A pipeline, under the frozen evaluation configuration. Learned estimators in fringe metrology provide context for this initialization approach. 23,24

S4 Experimental setup and validation

Instrument and acquisition

The AML AWB08 provides two transmission-imaging channels. Instrument specifications and source records are given in Table S2; main-text Fig. 1 shows the hardware arrangement. [20, 21, 22]

Test vehicle and calibration

The mark geometry is defined in S1. Both channels acquired images at stage readouts −0.11, −0.13, −0.16 and −0.19 mm. These are repeatable relative coordinates rather than traceable local surface gaps.[1,27]

The process context comprises activation, alignment, approach, pre-bonding, inspection and annealing. The reported Vernier distribution describes the resulting process overlay. It is not an isolated estimator-accuracy test or a demonstrated closed-loop correction experiment.

Dataset Partitioning and Metrics

Calibration covers magnification, camera rotation, distortion, pupil/PSF, background, gain, spectral-angular weighting and axial scale. Controlled translations and rotations establish geometry. Parameters are frozen before held-out evaluation and must remain frozen for future blind-wafer tests.[1,25,19]

For N samples, signed error is eq,i = q̂i − qi. Bias is its mean; σq is the sample standard deviation about that mean (denominator N−1). Reported metrics are mean absolute error (MAE), root mean square error (RMSE), 95th percentile (P95) of |eq| and empirical 3σq. Nonnegative radial error uses mean, P95 and mean plus three standard deviations (mean + 3SD):

$$\mathrm{MAE}_q = \frac{1}{N}\sum_{i=1}^{N}\left|e_{q,i}\right|, \quad \mathrm{RMSE}_q = \sqrt{\frac{1}{N}\sum_{i=1}^{N} e_{q,i}^2}, \qquad e_r = \sqrt{e_x^2 + e_y^2} \quad \text{(S40)}$$

The reported experimental reconstruction metrics are background-excluded NCC and complete horizontal/vertical intensity-profile RMSE (Section S6). Synthetic pose-error statistics and physical Vernier overlay are assessed separately.

The independent near-zero set contains 500 synthetic samples with Δx and Δy within ±1 μm, densely sampled within ±500 and ±100 nm and including zero-axis cases. Rotation, axial, optical and nuisance ranges follow the primary evaluation. It was generated after freezing CNN and refinement parameters and excluded from training, validation and tuning.

Wafer-Level Vernier Overlay Evaluation

A separate automated Vernier program evaluated two physical mark pairs at 45 measurements each. Horizontal and vertical keys with 0.10-, 0.25-, 0.50- and 1.00-μm increments were detected in

dedicated regions of interest and converted to Δx and Δy. These peripheral features are independent of the octagonal inverse objective.

The 0.10-μm finest increment, without calibrated sub-tick interpolation uncertainty, supports submicrometer overlay assessment rather than validation of 3–5-nm synthetic MAEs. Empirical 3σ is an overlay spread, not a traceable expanded uncertainty.

Comparison and validation limits

The experimental AML images, synthetic PDTOM benchmarks and physical Vernier measurements provide complementary but distinct evidence. Figure 4 reports background-excluded target-guided ZDMOR reconstruction agreement, not blind axial-gap estimation. Figure 5 and Supplementary Tables S4–S5 report PDTOM-A synthetic evaluation; Figure 6 and Supplementary Table S6 report physical overlay; Figure 7 and Supplementary Table S7 report archived PDTOM-C synthetic performance. Compact-mark hardware qualification and independent closed-loop stage-feedback validation remain future work.

S5 Compact-mark PDTOM-C dataset provenance and validation limits

The actual PDTOM-C configuration is P_1 = 1.50 μm, P_2 = 1.59 μm, N_1 = 15 and N_2 = 14 within a 45 × 45 μm² mark, giving P_m = 26.50 μm and N_m,overlap ≈ 0.84. Table S3 instead lists analytical design-screening candidates; its 45-μm, P_2 = 1.65 μm row is not the evaluated compact-mark geometry.

The PDTOM-C research archive contains 7,000 synthetic label records, partitioned into 5,000 training, 1,000 validation and 1,000 test records, together with initial and refined estimates from a separate 1,000-record evaluation. The sample-level linkage between these datasets and their correspondence to a specific executed CNN checkpoint have not been established.

Main-text Fig. 7 and Supplementary Table S7 summarize the 1,000-record compact-mark evaluation. The recorded initial state is treated as the optimization starting point, rather than as the output of a separately verified CNN checkpoint. These results therefore characterize recorded initialization and refinement performance; exact end-to-end reproducibility remains unverified.

Main-text Fig. 7(b) and Table S7 document the remaining tails. Small P95 values do not exclude isolated rotational or axial outliers, including maxima of 0.139486° and 99.466 nm in the recorded synthetic evaluation. Physical AML compact-mark validation remains outstanding; nanometre-scale physical four-parameter accuracy is not claimed.

S6 Experimental ZDMOR reconstruction: background-excluded image NCC and profile verification

Z-dependent Moiré optical reconstruction (ZDMOR) was evaluated against four AML transmission images acquired at mechanical stage coordinates −0.190, −0.160, −0.130 and −0.110 mm. The four stage positions define a reference-relative axial series and are not independently calibrated physical wafer gaps. Each AML frame and its reconstructed counterpart were compared using the stored 512 × 512 registered grayscale arrays and their existing intensity normalization; the ZDMOR best-match axial coordinate was selected by target-image comparison within the LOZO reconstruction protocol.

For the primary image-correlation evaluation, a common background-exclusion mask removed two predefined non-Moiré regions in the zero-indexed registered image coordinate system: rows 0–91 with columns 0–152 (upper-left annotation), and rows 399–511 with columns 384–511 (lower-right shadow). The same mask was applied to every measured/reconstructed pair, retaining 233,604 of 262,144 pixels (89.112854%) per frame. Image NCC was calculated as the Pearson correlation coefficient over these retained pixels. Horizontal and vertical intensity profiles were formed by averaging central rows or columns 253–259, respectively. Since the excluded corner regions do not intersect the central seven-pixel bands, all 512 profile positions remained valid. RMSEs were calculated from the paired profiles on the original stored normalized-intensity scale, without renormalization after masking.

At AML stage positions −0.190, −0.160, −0.130 and −0.110 mm, background-excluded NCC values were 0.998328, 0.999455, 0.999507 and 0.994814, respectively (mean 0.998026). The corresponding horizontal RMSEs were 0.020494, 0.010826, 0.006550 and 0.031666 (mean 0.017384), whereas vertical RMSEs were 0.013773, 0.008238, 0.011672 and 0.032937 (mean 0.016655; Supplementary Table S8). For context, full-image NCC averaged 0.998323. An alternative octagonal-only ROI sensitivity analysis produced a mean NCC of 0.960148, but that restricted-region result is not the primary metric used in Figure 4. The image crops in the figure illustrate the mark morphology; they are distinct from the quantitative background-exclusion mask. Under LOZO, the target frame was excluded from the reference-image set but used to select the best-matching model Z; thus, reported correlations assess target-guided image consistency rather than independently predicted gap accuracy. Independent reconstruction-code and reference-image audits remain necessary to assess the effects of shared registration and background information.

Supplementary Tables

Supplementary Table S1. Comparison of alignment methods implemented on the AML AWB08 platform

Method	Mark / footprint	Principle / estimated state	Gap awareness / performance
Native AML	Cross-shaped; tool-specific	Automatic image alignment; Δx, Δy	No explicit gap estimate; specified accuracy <2 μm
DLOMA [1]	Concentric octagonal Moiré; prior reported design	Deep-learning regression; Δx, Δy, Δθ	Represented conditions only; reported accuracy <1 μm
Proposed PDTOM	Concentric octagonal Moiré; ≈594 × 594 μm²	Calibrated forward/inverse reconstruction; Δx, Δy, Δθ, z	Reference-relative z; target <0.5 μm; instrument validation not reported

The PDTOM footprint refers to the present 8.00/8.48-μm gap-evolution vehicle; compact sub-200-μm designs require separate optical requalification.

Supplementary Table S2. Experimental and optical parameters

Parameter	Nominal value / setting	Traceability / source
Bonding platform	AML AWB08, 8-in W2W bonder	Equipment record
Illumination	2 × Philips Type 7023 projection-halogen lamps; no band-pass filter	Equipment record / optical inspection
Wafer pair	725-μm double-polished Si / 725-μm double-polished Si	Process traveler
Gap-evolution mark	P_1 = 8.00 μm; P_2 = 8.48 μm; 50% duty cycle	Mask layout / critical-dimension (CD) metrology
Objective	Moritex MML4-HR65D; 4×; numerical aperture (NA) = 0.112; nominal working distance (WD) = 65 mm	Manufacturer specification ref. 20
Camera	Basler ace U acA1300-200um; 1280 × 1024; 4.8-μm pixel	Manufacturer specification ref. 21
Object-side imaging	1.2 μm/pixel; field of view (FOV) ≈ 1.536 × 1.229 mm²	Magnification calibration
Relative stage-Z readouts	−0.11, −0.13, −0.16, and −0.19 mm	AML acquisition log; not absolute gap

Supplementary Table S3. Analytical candidate geometry for design screening only; marks are not fabricated or optically qualified.

L (μm)	P_1 (μm)	P_2 (μm)	N_1	N_2	P_m (μm)	N_m,overlap
30	1.50	1.72	10	8	11.73	1.17
45	1.50	1.65	15	13	16.50	1.30
60	2.00	2.20	15	13	22.00	1.30
100	3.00	3.30	16	15	33.00	1.45
150	4.00	4.40	18	17	44.00	1.64
200	5.00	5.50	20	18	55.00	1.80
300	6.00	6.48	25	23	81.00	1.84
592	8.00	8.48	37	34	141.33	2.04

The 45-μm candidate uses P_2 = 1.65 μm and is distinct from actual PDTOM-C (P_2 = 1.59 μm; N_2 = 14; P_m = 26.50 μm; N_m,overlap ≈ 0.84). The 592-μm row is footprint constrained; the measured 37/35-cycle large mark extends to approximately 593.6 μm. Counts are geometrical beat-period estimates, not camera-visible fringe counts.

Supplementary Table S4. PDTOM-A held-out synthetic performance (N = 500).

Output	Estimator	MAE/mean	P95	3σ metric
Δx	CNN	21.483 nm	62.088 nm	90.263 nm
Δx	PDTOM-A	3.295 nm	11.194 nm	13.884 nm
Δy	CNN	24.674 nm	74.152 nm	102.281 nm
Δy	PDTOM-A	3.711 nm	12.716 nm	15.682 nm
Δθ	CNN = PDTOM-A	0.00961°	0.02628°	0.03905°
Δz	CNN = PDTOM-A	0.2992 μm	0.7693 μm	1.1440 μm
Radial x–y	CNN	36.290 nm	92.926 nm	118.462 nm*
Radial x–y	PDTOM-A	5.665 nm	14.713 nm	17.900 nm*

* Per-axis 3σ is three sample standard deviations of signed error; radial rows marked * report mean + 3SD of nonnegative radial error.

Supplementary Table S5. PDTOM-A near-zero synthetic qualification.

Subset	N	Δx MAE (nm)	Δy MAE (nm)	Radial mean (nm)	Radial P95 (nm)
≤1 μm	500	3.383	3.376	5.463	13.380
≤500 nm	313	3.400	3.252	5.379	13.482
≤100 nm	102	3.575	2.960	5.218	13.207
Zero axis	20	2.765	2.574	4.270	11.474

Supplementary Table S6. Physical wafer-level Vernier overlay.

Group	Mean (Δx, Δy), μm	3σ, μm
Pair #1 (N = 45)	(0.02, 0.24)	0.50
Pair #2 (N = 45)	(0.14, 0.19)	0.51
Combined (N = 90)	(0.08, 0.22)	0.51

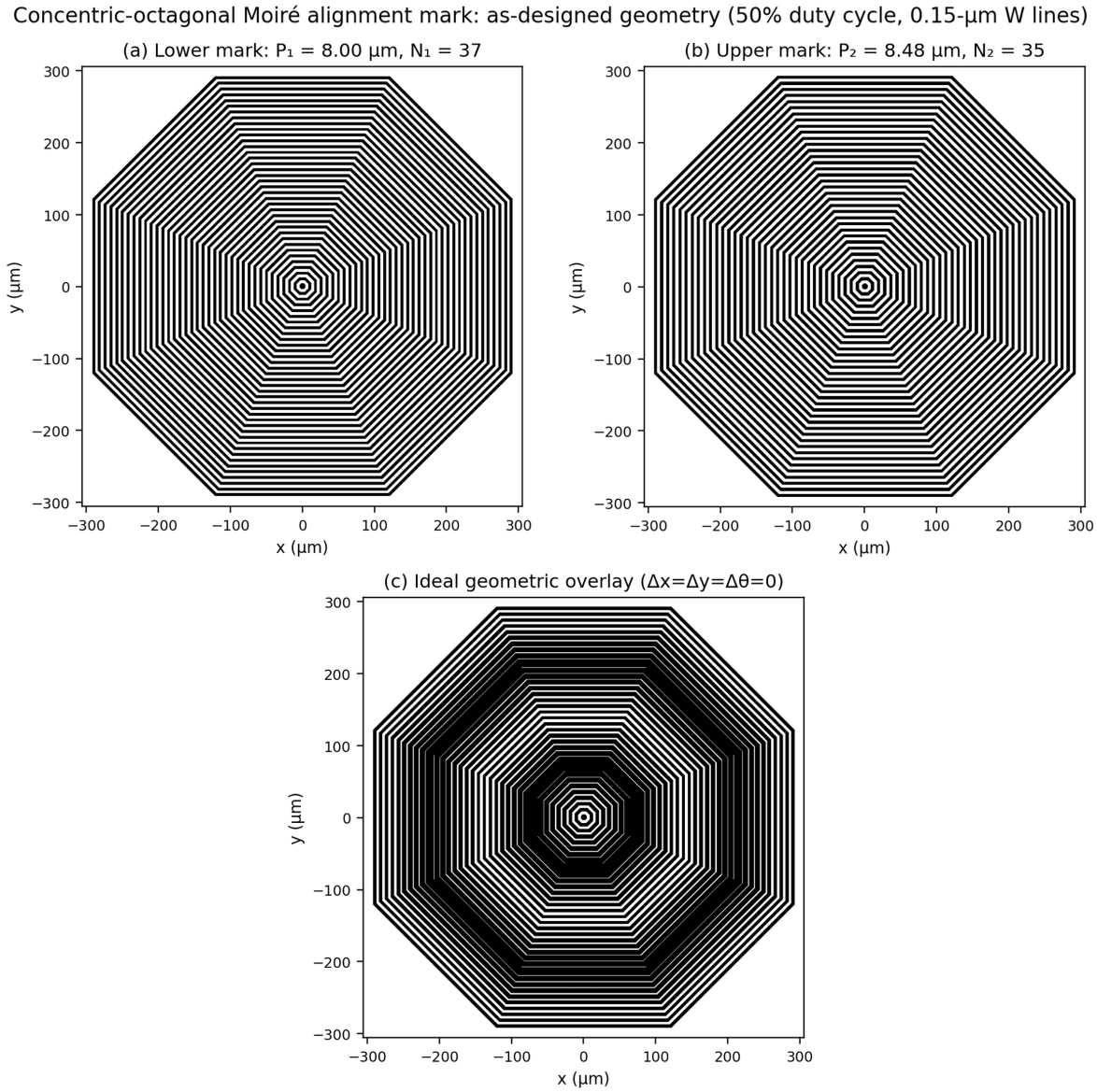
Supplementary Table S7. PDTOM-C compact-mark synthetic performance (N = 1,000).

Parameter	Initial-state MAE	PDTOM-C MAE	PDTOM-C P95	PDTOM-C maximum
Δx	13.202 nm	0.634 nm	2.729 nm	13.827 nm
Δy	12.251 nm	0.663 nm	2.808 nm	11.470 nm
θ	≈0.032°	0.003510°	0.008109°	0.139486°
Z-gap	27.926 nm	1.801 nm	6.764 nm	99.466 nm

Supplementary Table S8 | Background-excluded ZDMOR image reconstruction and central-profile verification.

AML stage Z (mm)	**Best-match Z (mm)**	**Offset (μm)**	**BG-excluded NCC**	**Horizontal RMSE**	**Vertical RMSE**
-0.190	-0.179	+11	0.998328	0.020494	0.013773
-0.160	-0.166	-6	0.999455	0.010826	0.008238
-0.130	-0.141	-11	0.999507	0.006550	0.011672
-0.110	-0.091	+19	0.994814	0.031666	0.032937
Mean	—	—	0.998026	0.017384	0.016655

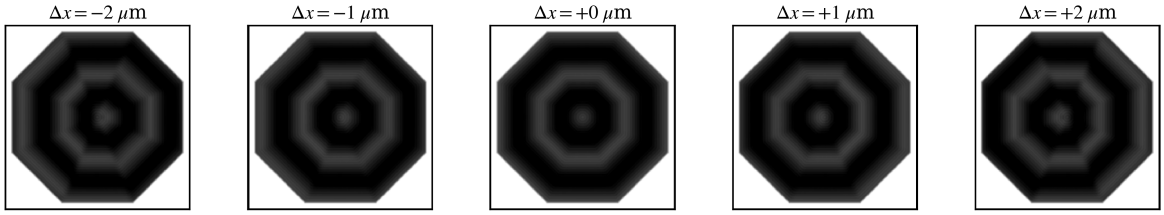
Supplementary Fig. S1



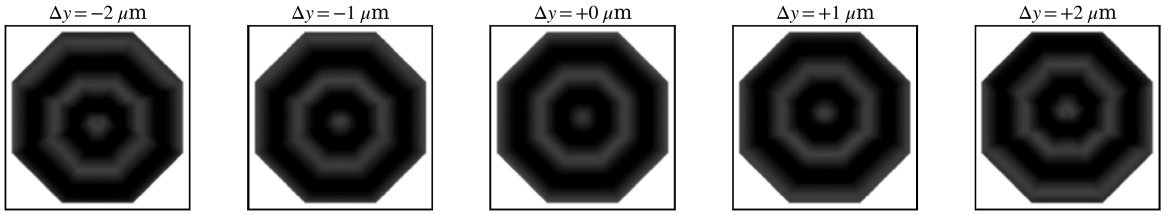
Supplementary Fig. S1. Concentric-octagonal Moiré alignment-mark geometry, rendered directly from the as-designed grating parameters used for the PDTOM gap-evolution experiments (P_1 = 8.00 µm, N_1 = 37; P_2 = 8.48 µm, N_2 = 35; 150-nm-thick W lines, 50% duty cycle). (a) Lower mark. (b) Upper mark. (c) Ideal geometric overlay ($\Delta x = \Delta y = \Delta\theta = 0$) showing the resulting P_m = 141.33 µm Moiré period and A_m = 17.67 translation amplification per (S1)–(S2).

Supplementary Fig. S2

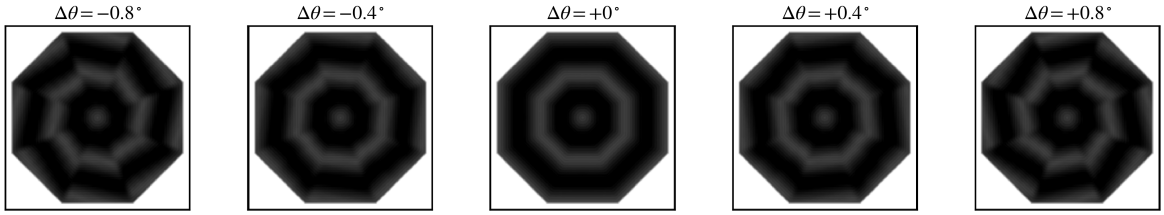
(a) Response to lateral displacement Δx



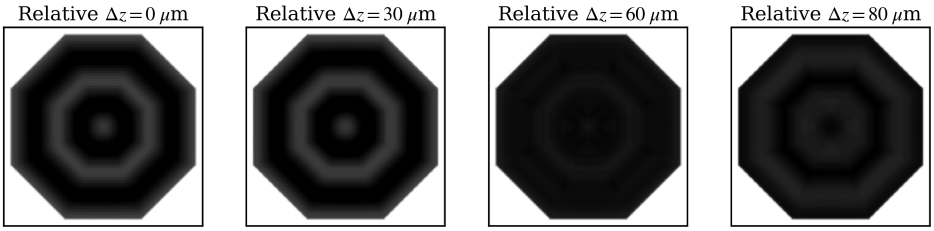
(b) Response to lateral displacement Δy



(c) Response to relative rotation $\Delta\theta$

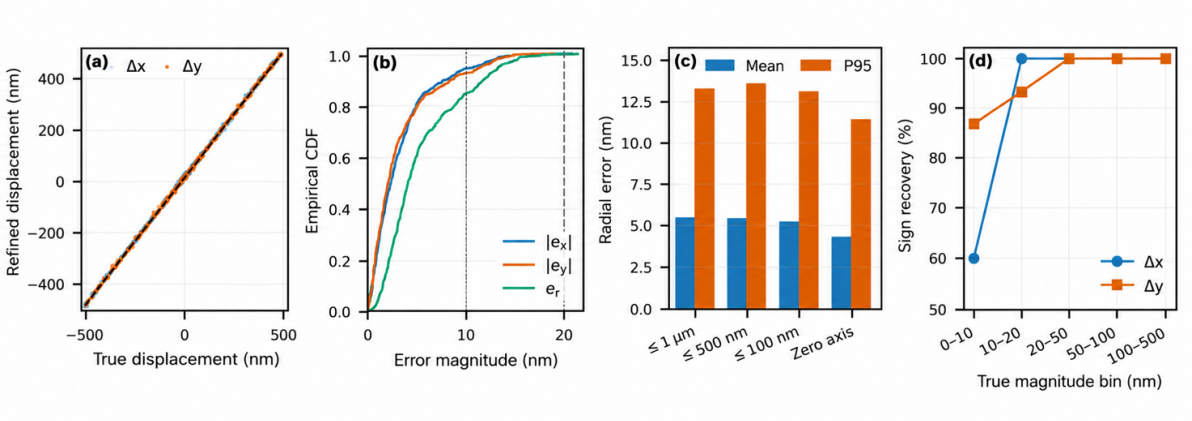


(d) Response to reference-relative axial separation Δz



Supplementary Fig. S2. Controlled generator response to imposed wafer state. (a) Δx sweep, (b) Δy sweep, (c) relative-rotation sweep, and (d) reference-relative Δz sweep. The prescribed grating cycles are unchanged; the figure is a physical-consistency check and not an independent measurement-accuracy result. The ΔZ labels denote the reference-relative axial displacement Δz; Δz = 0 is the selected reference position, not zero physical gap.

Supplementary Fig. S3



Supplementary Fig. S3. Near-zero synthetic qualification of the frozen PDTOM-A pipeline. (a) Predicted versus true lateral displacement; (b) empirical error CDFs; (c) radial error statistics for nested displacement subsets; and (d) sign-recovery rate. The 500-image near-zero test is separate from the primary held-out test and does not establish traceable physical resolution.

Supplementary 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_100W_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)