Huaichen Zhang, Cyrus E. Tabery, Ruben Maas, Oleksandr Khodko, Victor M. Blanco Carballo, Eren Canga, Filip Schleicher, "Predictive compact model for stress-induced on-product overlay correction," J. Micro/Nanopattern. Mats. Metro. 21(4) 043201 (13 October 2022) https://doi.org/10.1117/1.JMM.21.4.043201

Predictive compact model for stress-induced on-product overlay correction

Huaichen Zhang,[a,*] Cyrus Tabery,[a] Ruben Maas,[a] Oleksandr Khodko,[a] Victor M. Blanco Carballo,[b] Eren Canga,[b] and Filip Schleicher[b]

[a]ASML Netherlands B.V. (Netherlands), Veldhoven, Netherlands
[b]IMEC, Heverlee, Belgium

Abstract. On-product overlay (OPO) is an important indicator of device yield. In this work, we show that stressed thin films used in semiconductor manufacturing can be an important contributor to OPO at multiple length scales. Depending on the stress level, film thickness, and the mask design, the overlay impact can be a few nanometers for the exposure of the next lithography layer. A predictive compact model based on pattern density is developed to accurately predict this overlay impact. The model is then verified using short-loop dual damascene wafers with stress split. The predictive model opens a new opportunity for model-based mask correction during optical proximity correction to increase the overlay margin for subsequent lithography exposures. © *2022 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.JMM.21.4.043201]

Keywords: compact model; overlay; on-product overlay; plasma etching; mask; thin film; stress; optical proximity correction; holistic lithography.

Paper 22030G received May 31, 2022; accepted for publication Sep. 22, 2022; published online Oct. 13, 2022.

1 Introduction

Overlay error between an exposure layer and a reference layer, or simply overlay, is an important indicator for the lithography scanner performance. In the most advanced nodes, the overlay error margin is becoming so small that holistic solutions to minimize overlay error throughout the chip manufacturing process are required. Overlay error can originate from both the lithography scanner and the subsequent manufacturing processes. The overlay error as measured after a plasma etching step is named on-product overlay (OPO), which represents the layer-to-layer connectivity in a geometric sense. Therefore, optimizing OPO is more important than optimizing the scanner overlay error alone, thus calling for a mask-scanner-process co-optimization in the holistic lithography[1] approach.

Understanding the different sources of OPO is essential for advanced overlay process correction.[2] One of the sources that contributes to OPO is thin-film stress. As an example, to increase carrier mobility, the p-channel field-effect transistor and n-channel field-effect transistor are preferably stressed, either by inducing intrinsic stress, or by using thin-film stressor, such as SiN.[3] Figure 1 shows how stressed thin film produces OPO. When a wafer substrate is deposited with a stressed thin film, the wafer shrinks or expands on the wafer table depending on whether the stress is tensile or compressive. Such a shrinkage or expansion is uniform (up to certain radius near the wafer edge where the wafer is no longer supported by the wafer table). The first layer is then exposed on the photo resist of the uniformly shrunk or expanded wafer. During the etching step, the resist pattern is transferred to the stressed thin-film layer. When pattern density non-uniformity is present, the amount of the stress released during the etching step is also not uniform because of the unequal amount of material removed, resulting in local expansion and local contraction. At the next exposure step, if the top layer is exposed without proper correction, the resultant grid mismatch is translated into the layer-to-layer overlay error. Such grid distortion induced overlay error was reported in prior studies.[4,5] Stobert et al.[6] proposed model-based correction during mask tape out flow on limited use cases. However, these studies

*Address all correspondence to Huaichen Zhang, huaichen.zhang@asml.com

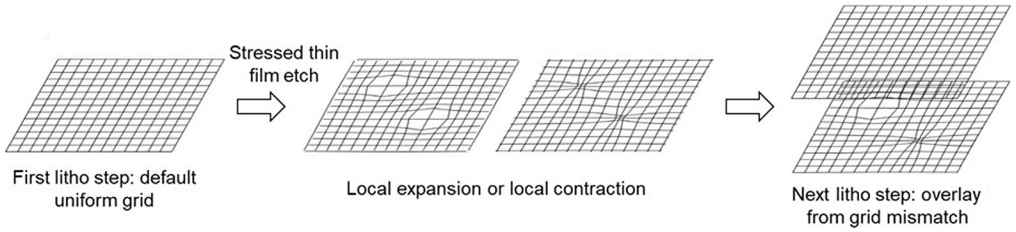

Fig. 1 Mechanism of stress-induced overlay error between two layers. Local distortion is created when a stressed film is etched. The exposure grid of the second top layer mismatches the distorted bottom layer grid.

did not address the mechanism of the short length scale local grid distortions and their link to wafer scale fingerprints.

In this work, we revisit the mechanism of stress-induced grid distortion, and derive a predictive model for stress-induced OPO error using coarse grid pattern density map as input. The compact model is fully predictive and does not require a fitting parameter. In the case of model-based mask correction, the mechanical properties of the thin films on the substrate are used as input; whereas in the case of overlay monitoring and scanner control, the wafer alignment data can be used as input, without knowing the actual mechanical properties. To validate the model, a set of wafers is produced with an in-house mask and a short-loop dual damascene flow. The pattern density of the designed metrology targets is modulated by changing the pattern's critical dimension (CD). The metrology targets are then measured with critical dimension scanning electron microscope (SEM) to extract local overlay error. The obtained results are compared with the predictive model. In addition, we also obtained wafer alignment sensor data to check the wafer and field scale grid distortions. The results are also compared with the model.

2 Grid Distortion Model

2.1 *Compact Model-based on Convolution*

The aim of the work is to derive a compact model for stress-induced grid distortion. Therefore, it makes sense to first neglect the exact patterns and use a mesh grid to represent the local stress on each grid point (k, l). Within each grid cell, the stress $\sigma_{xx}(k, l)$ and $\sigma_{yy}(k, l)$ in x- and y-direction, respectively, are viewed as constants. Assuming a linear-elastic material, the total displacement at cell (i, j) can be expressed as the sum of displacements induced by each grid cell on the two-dimensional (2D) plane, by assuming the superposition principal of elastic material

$$\begin{aligned} U_x(i,j) &= \sum \sum_{k,l=-\infty\to+\infty} u_x(i-k, j-l, \sigma_{xx}(k,l), \sigma_{yy}(k,l)) \\ U_y(i,j) &= \sum \sum_{k,l=-\infty\to+\infty} u_y(i-k, j-l, \sigma_{xx}(k,l), \sigma_{yy}(k,l)), \end{aligned} \tag{1}$$

where u_x and u_y are the displacements at cell (i, j) in response to a stressed cell at (k, l) with stress $\sigma_{xx}(k, l)$, $\sigma_{yy}(k, l)$ in x and y directions, respectively; and $U_x(i, j)$, $U_y(i, j)$ are the total displacements at cell (i, j) in response to a distributed stress on all other cells. We will later show that the displacement function is separable:

$$\begin{aligned} u_x(i,j,\sigma_{xx0},\sigma_{yy0}) &= \sigma_0 u_{x1}(i,j) + \sigma_{\Delta 0} u_{x2}(i,j) \\ u_y(i,j,\sigma_{xx0},\sigma_{yy0}) &= \sigma_0 u_{y1}(i,j) + \sigma_{\Delta 0} u_{y2}(i,j), \end{aligned} \tag{2}$$

where $\sigma_0 = \frac{\sigma_{xx0}+\sigma_{yy0}}{2}$ and $\sigma_{\Delta 0} = \frac{\sigma_{xx0}-\sigma_{yy0}}{2}$ denote isotropic and anisotropic stress, respectively; u_{x1}, u_{x2}, u_{y1}, u_{y2} are kernel functions that describe the displacement responses at (i, j) when the system is subjected to isotropic stress σ_0 and anisotropic stress $\sigma_{\Delta 0}$ at the center cell (0,0), respectively. Now, we convert the cell index to continuous Cartesian coordinates,

$$U_x(x,y) = \iint [\sigma(x_0,y_0)u_{x1}(x-x_0,y-y_0) + \sigma_\Delta(x_0,y_0)u_{x2}(x-x_0,y-y_0)]\mathrm{d}x_0\,\mathrm{d}y_0$$
$$U_y(x,y) = \iint [\sigma(x_0,y_0)u_{y1}(x-x_0,y-y_0) + \sigma_\Delta(x_0,y_0)u_{y2}(x-x_0,y-y_0)]\mathrm{d}x_0\,\mathrm{d}y_0, \quad (3)$$

where σ and σ_Δ denote isotropic and anisotropic stress distribution, respectively. Using convolution notation, the above formulae can be written as

$$U_x(x,y) = \sigma(x,y) \otimes u_{x1}(x,y) + \sigma_\Delta(x,y) \otimes u_{x2}(x,y)$$
$$U_y(x,y) = \sigma(x,y) \otimes u_{y1}(x,y) + \sigma_\Delta(x,y) \otimes u_{y2}(x,y). \quad (4)$$

The convolution operation makes the calculation fast to perform in an algorithm by utilizing the fast Fourier transform algorithm. It is easy to spot the computation will rely on the kernel functions u_{x1}, u_{x2}, u_{y1}, and u_{y2}.

2.2 *Kernel Functions*

The physical system in consideration is a constant thickness thin film attached to a substrate where a square inclusion region of the film is initially stressed, as shown in Fig. 2. The substrate is restricted to not move in the z-direction on the bottom side, representing the wafer in a well clamped state. Similar systems were studied in various literatures,[7–10] however, no direct solution is found for the exact system of interest. Here we construct an approximate solution using analytically solved systems.

2.2.1 *Long-range kernel considering the 2D film-substrate system*

First consider a thin film without substrate. The thin film has a stressed circular inclusion of radius a centered at the origin of the Cartesian coordinate. The initial stress of the inclusion region is σ_1 in x-direction and σ_2 in y-direction. For the region outside the inclusion area ($r \geq a$), the displacement field of the film is[8,11]

$$u_{f,x}(x,y) = \frac{1+\nu_f}{4E_f}x\left[\frac{\sigma_1+\sigma_2}{2}\left[2(1-\nu_f)\frac{a^2}{r^2}\right] + \frac{\sigma_1-\sigma_2}{2}\left[4\frac{x^2-\nu_f y^2}{r^2}\frac{a^2}{r^2} + (1+\nu_f)\frac{3y^2-x^2}{r^2}\frac{a^4}{r^4}\right]\right]$$
$$u_{f,y}(x,y) = \frac{1+\nu_f}{4E_f}y\left[\frac{\sigma_1+\sigma_2}{2}\left[(1-\nu_f)\frac{a^2}{r^2}\right] - \frac{\sigma_1-\sigma_2}{2}\left[4\frac{\nu_f x^2-y^2}{r^2}\frac{a^2}{r^2} + (1+\nu_f)\frac{y^2-3x^2}{r^2}\frac{a^4}{r^4}\right]\right], \quad (5)$$

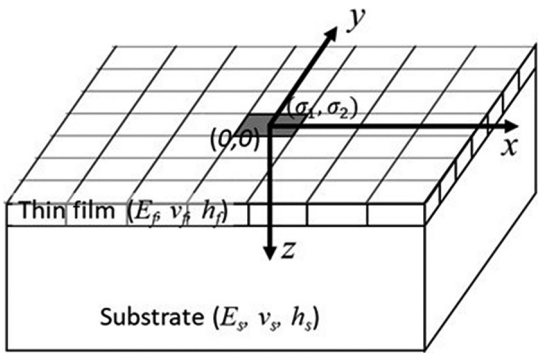

Fig. 2 The physical system used to derive the kernel functions comprises of a thin film attached to a substrate. The bottom surface of the substrate is restricted to not move in z-direction, mimicking a clamped wafer. The film is coarse grained into mesh grid cells. The center cell at (0,0) is initially internally stressed by $(\sigma 1, \sigma 2)$ in x- and y-direction, respectively.

where $r = \sqrt{x^2 + y^2}$ denotes the radial distance to origin; E_f is the elastic modulus of the film; ν_f is the Poisson ratio of the film; and subscript f denotes the thin film.

Now consider the thin film is attached to a "clamped" substrate whose z-direction movement is restricted. Assuming the stress field is equally distributed along z-direction, and assuming the thin film is much thinner than the substrate and no delamination, the force balance requires

$$u_{fs,x} = u_{f,x}\frac{E_f h_f(1-\nu_s)}{E_s h_s(1-\nu_f)}, \quad u_{fs,y} = u_{f,y}\frac{E_f h_f(1-\nu_s)}{E_s h_s(1-\nu_f)} \tag{6}$$

where subscript fs denotes the film-substrate system, E_s is the elastic modulus of the substrate, ν_s is the Poisson ratio of the substrate, h_f and h_s are the film thickness and substrate thickness, respectively. In the far field approximation ($r \gg a$), the circular shape can be viewed as a square, following the area equivalence

$$\pi a^2 = \mathrm{d}x\mathrm{d}y. \tag{7}$$

In addition, the fourth-order term can be neglected for far field approximation ($r \gg a$). Combining Eqs. (5)–(7) and neglect the $(a/r)^4$ term

$$\begin{aligned} u_{fs,x}(x,y) &= \frac{(1+\nu_f)(1-\nu_s)h_f}{2\pi E_s h_s}\frac{x}{x^2+y^2}\left[\frac{\sigma_1+\sigma_2}{2} + \frac{\sigma_1-\sigma_2}{2}\cdot\frac{2(x^2-\nu_f y^2)}{(1-\nu_f)r^2}\right]\mathrm{d}x\,\mathrm{d}y \\ &= F_1(x,y)\frac{\sigma_1+\sigma_2}{2}\mathrm{d}x\,\mathrm{d}y + F_2(x,y)\frac{\sigma_1-\sigma_2}{2}\mathrm{d}x\,\mathrm{d}y, \\ u_{fs,y}(x,y) &= \frac{(1+\nu_f)(1-\nu_s)h_f}{2\pi E_s h_s}\frac{y}{x^2+y^2}\left[\frac{\sigma_1+\sigma_2}{2} - \frac{\sigma_1-\sigma_2}{2}\cdot\frac{2(y^2-\nu_f x^2)}{(1-\nu_f)r^2}\right]\mathrm{d}x\,\mathrm{d}y \\ &= F_3(x,y)\frac{\sigma_1+\sigma_2}{2}\mathrm{d}x\,\mathrm{d}y + F_4(x,y)\frac{\sigma_1-\sigma_2}{2}\mathrm{d}x\,\mathrm{d}y, \end{aligned} \tag{8}$$

where F_1, F_2, F_3, and F_4 are kernel functions

$$\begin{aligned} F_1(x,y) &= \frac{(1+\nu_f)(1-\nu_s)h_f}{2\pi E_s h_s}\frac{x}{x^2+y^2} \\ F_2(x,y) &= \frac{(1+\nu_f)(1-\nu_s)h_f}{2\pi E_s h_s}\frac{x}{x^2+y^2}\frac{2(x^2-\nu_f y^2)}{(1-\nu_f)(x^2+y^2)} \\ F_3(x,y) &= \frac{(1+\nu_f)(1-\nu_s)h_f}{2\pi E_s h_s}\frac{y}{x^2+y^2} \\ F_4(x,y) &= \frac{(1+\nu_f)(1-\nu_s)h_f}{2\pi E_s h_s}\frac{y}{x^2+y^2}\frac{2(\nu_f x^2-y^2)}{(1-\nu_f)(x^2+y^2)}. \end{aligned} \tag{9}$$

If the kernel functions are converted to polar coordinates, it is easy to spot the kernel functions scale with $1/r$ [by taking $x = r\cos(\theta)$ $y = r\sin(\theta)$]. The slowly decaying functions indicates non-local long-range effect on the displacement field when integrated in the convolution calculations. Therefore, the kernels are named long-range kernel functions. Similarly, the following paragraph will derive a set of kernel functions that scale with $1/r^2$. These fast decaying functions indicates local short-range effect and are thus named short-range kernel functions.

2.2.2 *Short-range kernel approximated by 3D semi-infinite elastic body inclusion problem*

In the close proximity of the stressed region (<100 mm), the substrate with 775-μm thickness can be viewed as a semi-infinite elastic body. Assuming no z-displacement on the top of the film, the system can be approximated by a full 3D elastic body with a spherical inclusion considering the z-plane symmetry. The 3D body inclusion problem was solved by.[12,13] For a given inclusion of radius a and internal stress σ_1, σ_2, and σ_3 in x-, y-, and z-direction, respectively, the displacement

field outside the inclusion area ($R \geq a$) is

$$\begin{aligned}\frac{2E}{(1+\nu)}\frac{u_x}{x} &= \frac{2(1-2\nu)}{3(1-\nu)}\left(\frac{a}{R}\right)^3\sigma_1 + \frac{2}{5(1-\nu)}\left(\frac{a}{R}\right)^5(\sigma_1-\overline{\sigma}) \\ &+ \frac{1}{(1-\nu)}\left[\left(\frac{a}{R}\right)^3-\left(\frac{a}{R}\right)^5\right]\left(\frac{x^2}{r^2}\sigma_1+\frac{y^2}{r^2}\sigma_2+\frac{z^2}{r^2}\sigma_3-\overline{\sigma}\right) \\ \times\frac{2E}{(1+\nu)}\frac{u_y}{y} &= \frac{2(1-2\nu)}{3(1-\nu)}\left(\frac{a}{R}\right)^3\sigma_2 + \frac{2}{5(1-\nu)}\left(\frac{a}{R}\right)^5(\sigma_2-\overline{\sigma}) \\ &+ \frac{1}{(1-\nu)}\left[\left(\frac{a}{R}\right)^3-\left(\frac{a}{R}\right)^5\right]\left(\frac{x^2}{r^2}\sigma_1+\frac{y^2}{r^2}\sigma_2+\frac{z^2}{r^2}\sigma_3-\overline{\sigma}\right) \\ \times\frac{2E}{(1+\nu)}\frac{u_z}{z} &= \frac{2(1-2\nu)}{3(1-\nu)}\left(\frac{a}{R}\right)^3\sigma_3 + \frac{2}{5(1-\nu)}\left(\frac{a}{R}\right)^5(\sigma_3-\overline{\sigma}) \\ &+ \frac{1}{(1-\nu)}\left[\left(\frac{a}{R}\right)^3-\left(\frac{a}{R}\right)^5\right]\left(\frac{x^2}{r^2}\sigma_1+\frac{y^2}{r^2}\sigma_2+\frac{z^2}{r^2}\sigma_3-\overline{\sigma}\right), \end{aligned} \tag{10}$$

where $r = \sqrt{x^2+y^2}$ denotes the radial distance to origin within an xy-plane, and $R = \sqrt{x^2+y^2+z^2}$ denotes the radial distance to origin in 3D coordinates. $\overline{\sigma} = (\sigma_1+\sigma_2+\sigma_3)/3$ is the mean initial stress.

For the semi-infinite elastic body problem, we take the above solution for $z \geq 0$ region (note the positive direction is defined downward in Fig. 2). Because of the free surface at $z = 0$, the initial z-stress can be assumed zero, i.e., $\sigma_3 = 0$, rearranging the terms and neglect the $(a/R)^5$ term

$$\begin{aligned} u_{s,x}(x,y) &= \frac{1+\nu_s}{2E_s(1-\nu_s)}x\frac{a^3}{(r^2+z^2)^{3/2}}\left[\frac{\sigma_1+\sigma_2}{2}\left(\frac{3-4\nu_s}{3}-\frac{z^2}{r^2+z^2}\right)+\frac{\sigma_1-\sigma_2}{2}\left(\frac{x^2-y^2}{r^2+z^2}+\frac{2-4\nu_s}{3}\right)\right] \\ u_{s,y}(x,y) &= \frac{1+\nu_s}{2E_s(1-\nu_s)}y\frac{a^3}{(r^2+z^2)^{3/2}}\left[\frac{\sigma_1+\sigma_2}{2}\left(\frac{3-4\nu_s}{3}-\frac{z^2}{r^2+z^2}\right)+\frac{\sigma_1-\sigma_2}{2}\left(\frac{x^2-y^2}{r^2+z^2}-\frac{2-4\nu_s}{3}\right)\right]. \end{aligned} \tag{11}$$

In far field approximation ($r \gg a$), the semi-spherical shape can be viewed as a cuboid, following the area equivalence

$$\frac{2\pi a^3}{3} = dxdyh_f. \tag{12}$$

Substituting into Eq. (11) and rearranging the terms,

$$\begin{aligned} u_{s,x}(x,y) &= \frac{(1+\nu_s)(3-4\nu_s)h_f}{4\pi E_s(1-\nu_s)}\frac{x}{(r^2+z^2)^{3/2}}\left[\frac{\sigma_1+\sigma_2}{2}\left(1-\frac{3}{3-4\nu_s}\frac{z^2}{r^2+z^2}\right)\right. \\ &\left.+\frac{\sigma_1-\sigma_2}{2}\left(\frac{2-4\nu_s}{3-4\nu_s}+\frac{3}{3-4\nu_s}\frac{x^2-y^2}{r^2+z^2}\right)\right]\mathrm{d}x\,\mathrm{d}y \\ &= G_1(x,y)\frac{\sigma_1+\sigma_2}{2}\mathrm{d}x\,\mathrm{d}y + G_2(x,y)\frac{\sigma_1-\sigma_2}{2}\mathrm{d}x\,\mathrm{d}y \\ u_{s,y}(x,y) &= \frac{(1+\nu_s)(3-4\nu_s)h_f}{4\pi E_s(1-\nu_s)}\frac{y}{(r^2+z^2)^{3/2}}\left[\frac{\sigma_1+\sigma_2}{2}\left(1-\frac{3}{3-4\nu_s}\frac{z^2}{r^2+z^2}\right)\right. \\ &\left.+\frac{\sigma_1-\sigma_2}{2}\left(-\frac{2-4\nu_s}{3-4\nu_s}+\frac{3}{3-4\nu_s}\frac{x^2-y^2}{r^2+z^2}\right)\right]\mathrm{d}x\,\mathrm{d}y \\ &= G_3(x,y)\frac{\sigma_1+\sigma_2}{2}\mathrm{d}x\,\mathrm{d}y + G_4(x,y)\frac{\sigma_1-\sigma_2}{2}\mathrm{d}x\,\mathrm{d}y, \end{aligned} \tag{13}$$

where G_1, G_2, G_3, and G_4 are kernel functions

$$G_1(x,y)=\frac{(1+\nu_s)(3-4\nu_s)h_f}{4\pi E_s(1-\nu_s)}\left(1-\frac{3}{3-4\nu_s}\frac{z^2}{r^2+z^2}\right)\frac{x}{(x^2+y^2+z^2)^{3/2}}$$
$$G_2(x,y)=\frac{(1+\nu_s)(3-4\nu_s)h_f}{4\pi E_s(1-\nu_s)}\left(\frac{2-4\nu_s}{3-4\nu_s}+\frac{3}{3-4\nu_s}\frac{x^2-y^2}{r^2+z^2}\right)\frac{x}{(x^2+y^2+z^2)^{3/2}}$$
$$G_3(x,y)=\frac{(1+\nu_s)(3-4\nu_s)h_f}{4\pi E_s(1-\nu_s)}\left(1-\frac{3}{3-4\nu_s}\frac{z^2}{r^2+z^2}\right)\frac{y}{(x^2+y^2+z^2)^{3/2}}$$
$$G_4(x,y)=\frac{(1+\nu_s)(3-4\nu_s)h_f}{4\pi E_s(1-\nu_s)}\left(-\frac{2-4\nu_s}{3-4\nu_s}+\frac{3}{3-4\nu_s}\frac{x^2-y^2}{r^2+z^2}\right)\frac{y}{(x^2+y^2+z^2)^{3/2}}. \quad (14)$$

2.2.3 *Combining the kernel functions of the 2D system and the 3D system*

In the last step, the kernel functions are assembled. Because the 2D system already considers the mean x and y-displacement of the film-substrate system, the redundant mean displacement field calculated from the 3D system should be removed. For convenience, the displacement as calculated for the 3D system at half the substrate thickness, i.e., at $z = h_s/2$, is taken out. Plugging the kernel functions back to Eq. (4) yields the final result

$$\begin{aligned}U_x(x,y)=\sigma(x,y)\otimes\left[F_1+G_1(z=0)-G_1\left(z=\frac{h_x}{2}\right)\right]+\sigma_\Delta(x,y)\\ \otimes\left[F_2+G_2(z=0)-G_2\left(z=\frac{h_x}{2}\right)\right]\\ U_y(x,y)=\sigma(x,y)\otimes\left[F_3+G_3(z=0)-G_3\left(z=\frac{h_x}{2}\right)\right]+\sigma_\Delta(x,y)\\ \otimes\left[F_4+G_4(z=0)-G_4\left(z=\frac{h_x}{2}\right)\right].\end{aligned} \quad (15)$$

2.3 *Stress Distribution*

The grid displacement as calculated in Eq. (15) shows a critical dependence on the spatial distribution of stress, both the isotropic stress $\sigma(x,y)=(\sigma_{xx}+\sigma_{yy})/2$ and the anisotropic stress $\sigma_\Delta(x,y)=(\sigma_{xx}-\sigma_{yy})/2$. The simplest form of the stress equation is to assume the remaining stress is proportional to the remaining stressed material. The fraction of patterned material is commonly named as pattern density, denoted as $\rho(x,y)$, which can be extracted from various electronic design automation software using the detailed chip layout as input. In this case

$$\sigma_{xx}(x,y)=C_x[1-\rho(x,y)]\sigma_0 \quad \sigma_{yy}(x,y)=C_y[1-\rho(x,y)]\sigma_0, \quad (16)$$

where σ_0 is the initial film stress, C_x and C_y are correction factors. Practically, the remaining stress percentage is smaller than the remaining material percentage. Therefore, C_x and C_y are smaller than unity. The exact values are further dependent on the pitch and CD or the actual shapes and the anisotropy of the printed patterns. These complications can be further solved by finite element methods if needed. In this work, we take $C_x = C_y = 1$.

3 Experiment

To test the compact model, an in-house mask is used with custom designed patterns. The patterns consist of two layers: An M-layer and a V-layer (representing metal-like and via-like layers, respectively). Both layers are squares with the center of the squares perfectly aligned. The V-layer square has a CD (side length) of fixed 50 nm placed at a fixed pitch of 400 nm. The M-layer square has an adjustable CD to facilitate the pattern density requirements. An example is given in Fig. 3. The test patterns are filled within a 200 μm × 200 μm area (except for the locations of optical targets). The circular pink stripe regions are filled with squares corresponding to

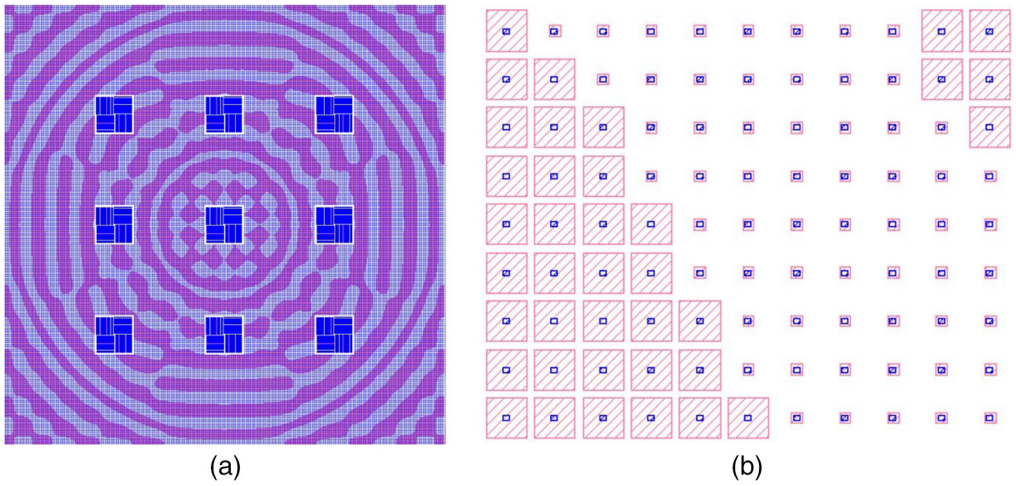

Fig. 3 A snapshot of a layout of a test structure to study the stress-induced substrate distortion. (a) 200 μm $\times$ 200 μm area with binary M-layer pattern density at either 70% or 5%. (b) An inset region of the left figure showing how the M-layer pattern density is modulated. Red: M-layer and blue: V-layer.

70% pattern density (CD = 334.6 nm) while the other regions are filled 5% (CD = 89.4 nm). It should be noted the actual pattern density as printed on wafer and etched would be different from the designed nominal pattern density.

The stressed layer used in this experiment is SiN. The deposition condition is tuned to achieve different stress levels. In the experiment, the tensile SiN film stress is measured to be 1.79 GPa; the neutral SiN film stress is measured to be −0.05 GPa; the compressive SiN film stress is measured to be −2.82 GPa, based on in-line metrology tools. These numbers are calculated from wafer warpage measurements. The SiN film is further split into different thickness for different wafers. Three values used are 30, 50, and 90 nm.

A short-loop dual damascene flow is used to process the test wafers, as shown in Fig. 4. First, the *M*-layer is exposed and etched into the stressed SiN. During the etching step, the stress is partially released because of the removed SiN. At this step, the substrate is already expected to distort. The wafers are sent to a different scanner for after-etch grid distortion measurements using the built-in scanner alignment sensor (ASML ORION Alignment Sensor). Afterward, the wafers are coated with hard mask and photo resist for V-layer patterning. A 6-parameter wafer scale linear correction is applied during exposure (wafer shift, rotation, and scaling, two degrees of freedom each), correcting the wafer scaling that is originated from the SiN film coating. However, any substrate distortion beyond the linear wafer scaling is not corrected. Therefore, the V-layer is expected to have a local overlay error with respect to the bottom M-layer. The wafers are then etched in several steps to have the V-layer patterns transferred to the inter layer

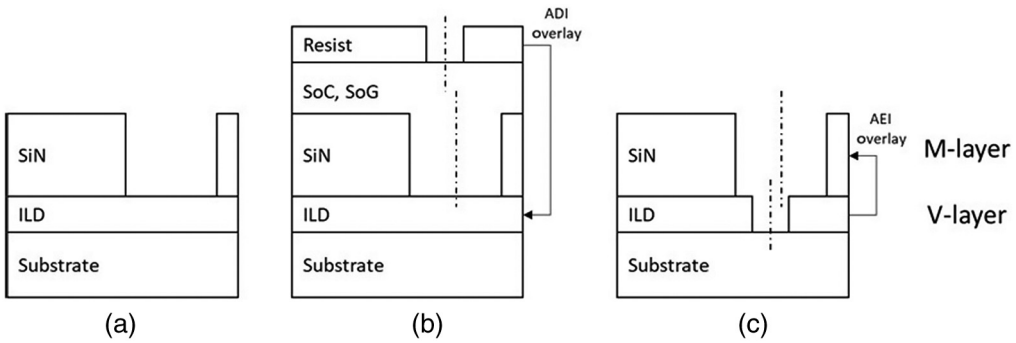

Fig. 4 Process flow of the test structure. (a) M-layer is patterned into the SiN layer; (b) after V-layer is exposed and developed; and (c) V-layer pattern is transferred to the ILD layer.

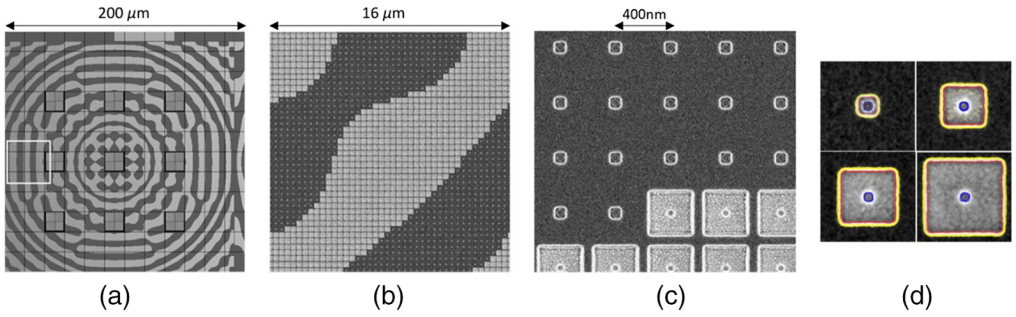

Fig. 5 (a) Stitched SEM image of the designed test structure; (b) one SEM image as captured by ep5 metrology system using 16 μm × 16 μm field-of-view; (c) a portion of the acquired SEM image; and (d) individual features and extracted contours.

dielectric (ILD) layer. After this step, the test patterns are inspected with ASML's HMI eP5 SEM metrology system. Both M-layer and V-layer are visible, enabling local overlay extraction directly from the SEM images.

4 Results

4.1 *SEM Image Data and Treatment*

SEM images of 16 μm × 16 μm field-of-view are acquired for the region of interest. During data treatment, the single SEM images are sliced into imagelets as shown in Fig. 5(d) for contour extraction. The center of mass delta between the inner feature (V-layer) contour and the outer feature (M-layer) contour is used as the overlay error. If the M-layer CD is close to the V-layer CD, the edges are no longer separable and we cannot extract the overlay error; for larger M-layer CD, this method provides a clear image based overlay measurement.

4.2 *Extracted Overlay Data and Model Prediction*

In the model prediction, the pattern density is calculated using the pre- optical proximity correction layout, at a 1-μm mesh grid size. We use 90-nm SiN film thickness and −2.82 GPa initial compressive stress as input values to the model. The stress distribution after the M-layer etch is modeled using Eq. (16). The film's mechanical properties are assumed to be $\nu_f = 0.28$ (note the elastic modulus E_f is not used in the calculations). The wafer substrate's mechanical properties are assumed to be $E_s = 179$ GPa, $\nu_s = 0.278$, and $h_s = 775$ μm, in-line with various literature sources.[14,15] The extracted overlay data is plotted in Figs. 6(a) and 6(b). To be noted, the overlay data extracted from the small M-layer squares (low pattern density features) are excluded because of their large measurement noise mentioned in the previous section. All the arrows displayed in the figure comes from the large M-layer squares (high pattern density features). It can be seen that on the edge of the high pattern density features, the overlay arrows are pointing into the low pattern density region. This is because of the relatively larger compressive stress remaining in the low pattern density regions pushing the large pattern density regions outward, resulting in local expansion of the low pattern density region and local contraction of the high pattern density region. As shown in Fig. 4(b), the overlay is defined as V-to-M; therefore, the measured overlay fingerprint for the locally contracted high density region points outward. Figure 6(b) plots the modeled overlay of the high pattern density features.

To test the model's performance, a correlation plot is shown Fig. 7(b). The modeled and measured results show good correlation for overlay-X. Overlay-Y does not correlate very well because the overlay signal in Y is much lower than the metrology uncertainty. The model residue, defined as the difference between measurement and model, is plotted in Fig. 7(a). Compared with the original overlay fingerprint, the model residue reduces overlay 3-sigma (3 times standard deviation) from 2.60 to 1.18 nm for overlay-X, and from 1.12 to 1.10 nm for overlay-Y.

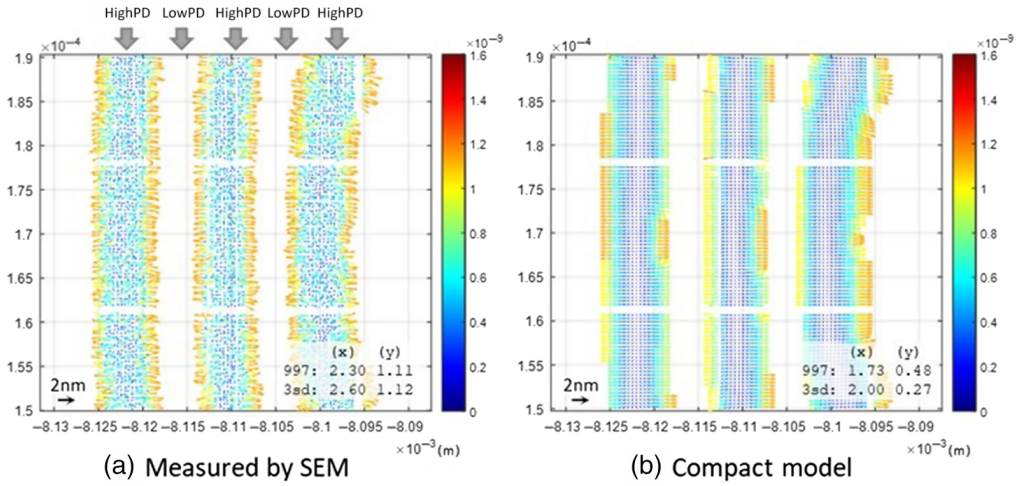

Fig. 6 (a) Quiver plot of the extracted overlay using SEM images of a 40 μm by 40 μm area of interest, as labeled by the white square in Fig. 5(a). The stripes with higher and lower pattern density are labeled on top of the figure with PD denoting pattern density (b) quiver plot of modeled overlay at area of interest.

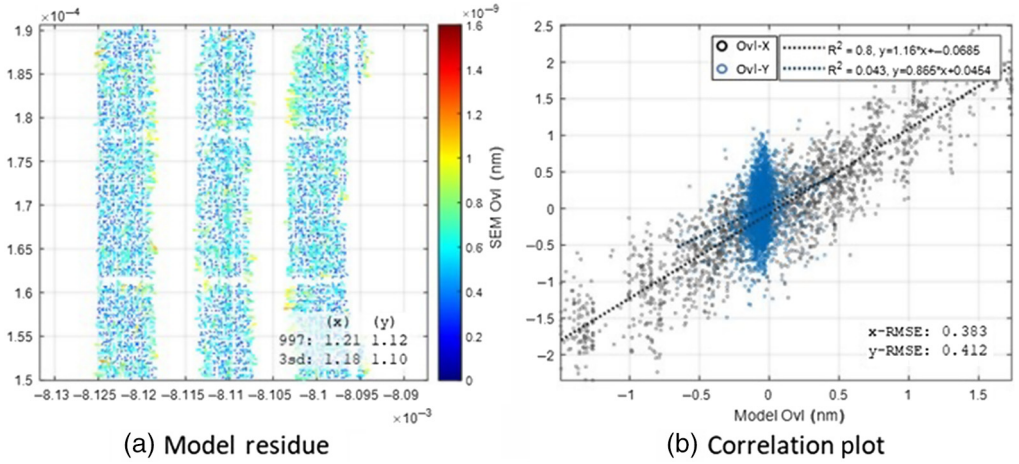

Fig. 7 (a) Quiver plot of the model residue defined as the difference between measurement and model; the inset numbers are the 99.7% percentile and 3-sigma of the residue in nanometers and (b) correlation plot between measured overlay and modeled overlay.

The residue overlay variations are expected to be dominated by the metrology uncertainty, which is approximately isotropic.

Some discrepancy between the model and the measurement is expected because of the coarse grained pattern density map at a finite resolution of 1 um. Decreasing pattern density mesh size can certainly improve accuracy at the cost of computational speed. Practically, we consider 1 to 5 μm mesh size sufficient to use.

A few other pattern density designs of 200 μm $\times$ 200 μm area are also measured. These designs have different sized blocks of uniform pattern density ranging from 5% to 70%. The results are summarized in Fig. 8. In all cases, the model residue significantly reduces the overlay fingerprint. Compared with the original overlay plots, the overlay distribution in the residue plot is much more uniform, and the large overlay error at pattern density jump locations is not seen any more. To be noted, the "noise," as indicated by the standard deviation of the model residue, contains both random placement error and metrology noise. Smaller M-layer patterns have more measurement uncertainty because of the lack of secondary electrons generated from the shaded V-layer patterns. This effect is especially visible in the third design in Fig. 8, where the low pattern density blocks have a much higher measurement noise.

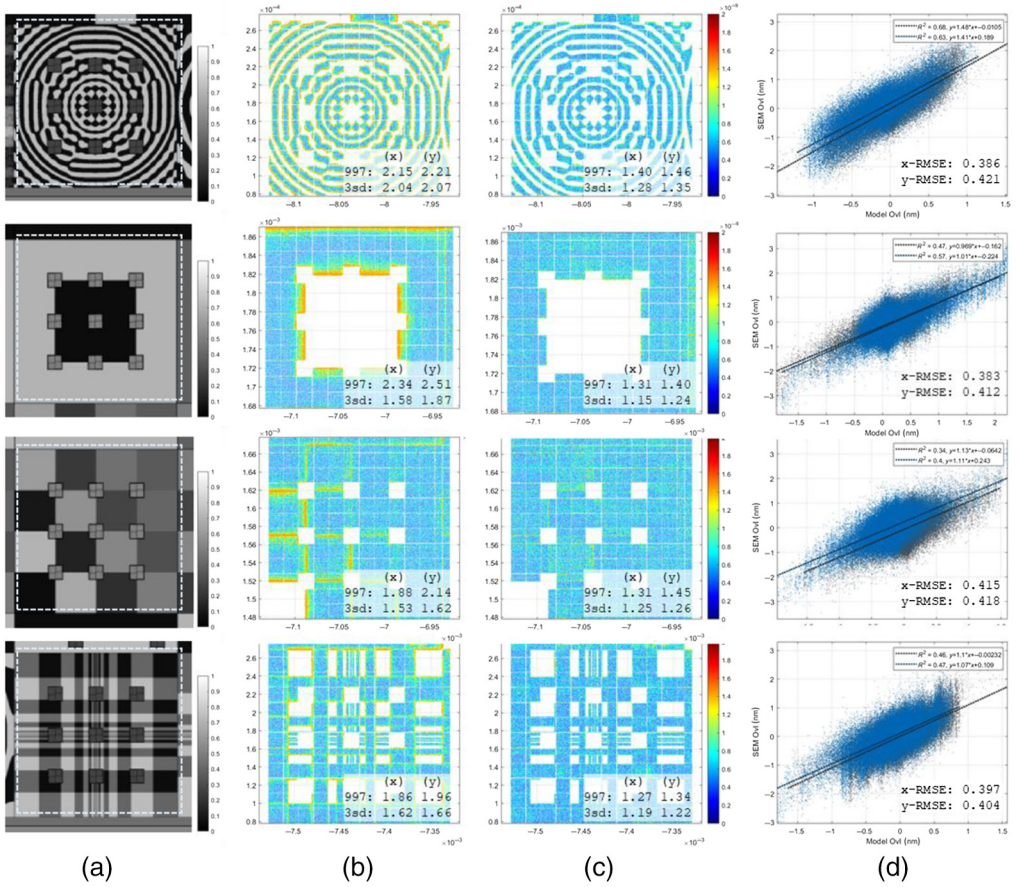

Fig. 8 (a) Pattern density maps and inspected area; (b) quiver plot of overlay extracted from the SEM images; (c) quiver plot of model residue; inset numbers are the 99.7% percentile and 3-sigma values; and (d) correlation plots.

4.3 *Correlation to Thin-Film Stress and Thickness*

Based on the model, the stress-induced local grid distortion and local overlay should be proportional to the initial thin-film stress and thin-film thickness. This is tested by extracting the overlay fingerprints of different wafers. Figure 9 compares the overlay fingerprint of a 40 μm by 40 μm area of interest, corresponding to the white box area labeled in Fig. 5(a). Clearly, the measured overlay scales with the product of SiN thickness and SiN stress. When the SiN changes from tensile to compressive, the measured overlay fingerprint also flips direction.

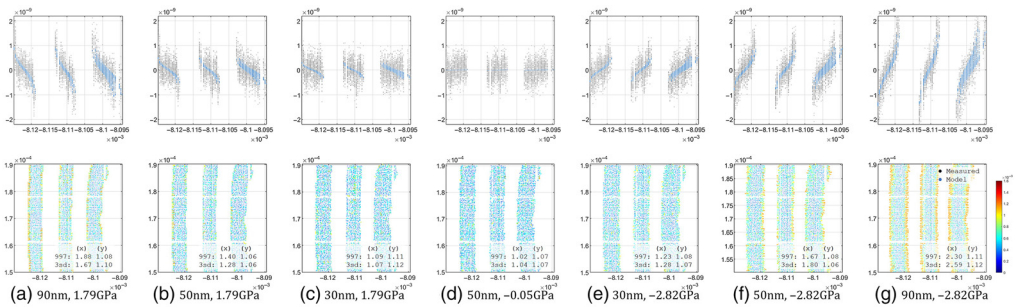

Fig. 9 Overlay information of the same area measured on different wafers with different SiN film thickness and stress levels. (Top) Scatter plot of overlay-*x* versus *x*-coordinate. (Bottom) Quiver plot of overlay. (a)–(g) corresponds to different film thickness and stress levels: (a) 90 nm, 1.79 GPa (b) 50 nm, 1.79 GPa (c) 30 nm, 1.79 GPa (d) 50nm, -0.05 GPa (e) 30 nm, –2.82 GPa (f) 50 nm, –2.82 GPa (g) 90 nm, –2.82 GPa.

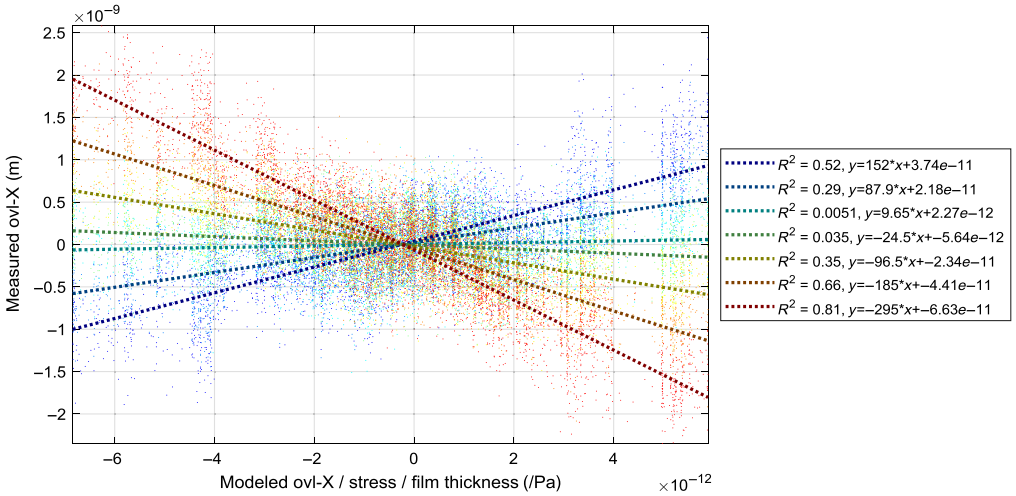

Fig. 10 Measured overlay versus normalized modeled overlay defined as overlay per GPa of film stress per nanometer of film thickness. The thick dashed lines are fitted results of the seven wafers with different film thickness and stress.

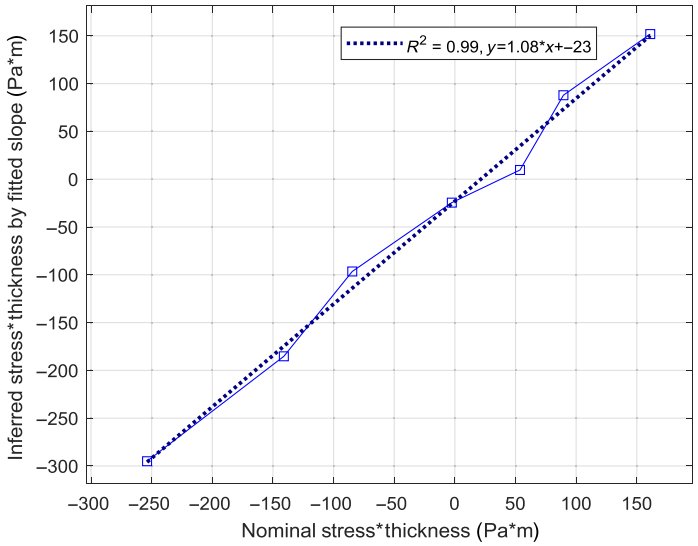

Fig. 11 Inferred SiN stress × SiN thickness from the fitted slopes of each wafer and the nominal product of the thin-film stress and its thickness.

To further quantify the relation between local overlay fingerprint and stress level, the measured overlay data of the seven wafers are compared with the normalized model result defined as overlay per GPa of film stress per nanometer of film thickness. The result is plotted in Fig. 10. The root mean square error of all the fits are below 0.387 nm. The fitted slopes are expected to be the product of the film stress and thickness of each wafer. The numbers are compared in Fig. 11 and they match relatively well, despite the expected discrepancies from the mismatch between the actual stress and the nominal stress; the pattern density drift from litho bias and litho-etch bias, and the linear stress model assumption used in Eq. (16).

4.4 *Grid Distortion Measured by Alignment Sensor*

Apart from the feature scale test patterns analyzed in the previous sections, the longer length scale wafer distortions are directly measured using the built-in alignment sensor in the lithography scanner. The measurement is done after the SiN etching step, as shown is Fig. 4(a). At the M-layer exposure step, the wafer is exposed on the default grid, without correcting the uniform

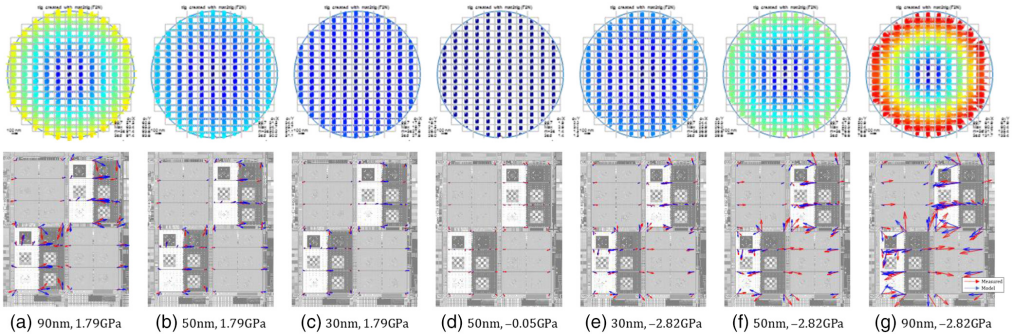

Fig. 12 Wafer alignment measurement result after SiN etching step of seven wafers. Top row: wafer scale data showing mainly symmetric wafer magnification; bottom row: intra-field result with wafer scaling fingerprint removed (red: measured; blue: model) displayed as quiver plots superposed on the pattern density map of an 8 mm × 10 mm area. (a)–(g) Film thickness and stress levels: (a) 90 nm, 1.79 GPa (b) 50 nm, 1.79 GPa (c) 30 nm, 1.79 GPa (d) 50 nm, −0.05 GPa (e) 30 nm, −2.82 GPa (f) 50 nm, −2.82 GPa (g) 90 nm, −2.82 GPa.

contraction or expansion induced by the SiN film. However, after the etching step, part of the stressed SiN is removed, resulting in a partial stress release. When the etched wafers are measured again with the alignment sensor, delta wafer contraction or expansion is expected. The alignment readout results, which measure the wafer distortion grid versus the scanner's default grid, are given in Fig. 12 for seven wafers with different SiN stress and thickness values. To be noted, the average scanner printing fingerprint is removed in the figure. Also shown in Fig. 12 is the intra-field grid distortion as measured by the alignment sensor. This is calculated by first removing the wafer scale linear magnification fingerprint, then taking a field average. In the selected region of interest of roughly 8 mm × 10 mm area, the wafer grid distorts by a few nanometers.

To test the model accuracy, the measured grid distortion is compared with the model of all seven wafers in Fig. 13. Because of the linearity of the model, the measurement is compared with the normalized model. The fitted slopes are the expected product of SiN stress and SiN thickness. In all seven cases, the model predicts very well the wafer grid distortion, with the root mean square error ranging between 0.05 and 0.28 nm. The fitted slopes are compared with the product of the nominal stress and nominal thickness in Fig. 14. The inferred product of thickness and stress matches very well with the nominal values. Some discrepancies are expected, such as the deviation from nominal values, pattern density drift from CD biases, and oversimplified stress equations. Nevertheless, the discrepancies are very similar to those shown in Fig. 11, proving the model's consistency while hinting at common sources of error.

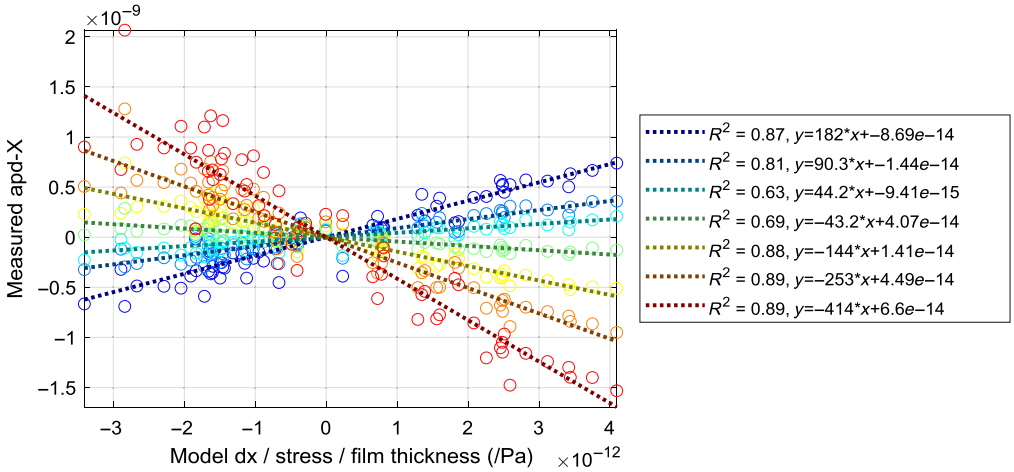

Fig. 13 Measured grid distortion (alignment position deviation) versus modeled grid distortion per GPa stress per nanometer film thickness. The fitted slopes are the expected SiN stress × SiN thickness per wafer.

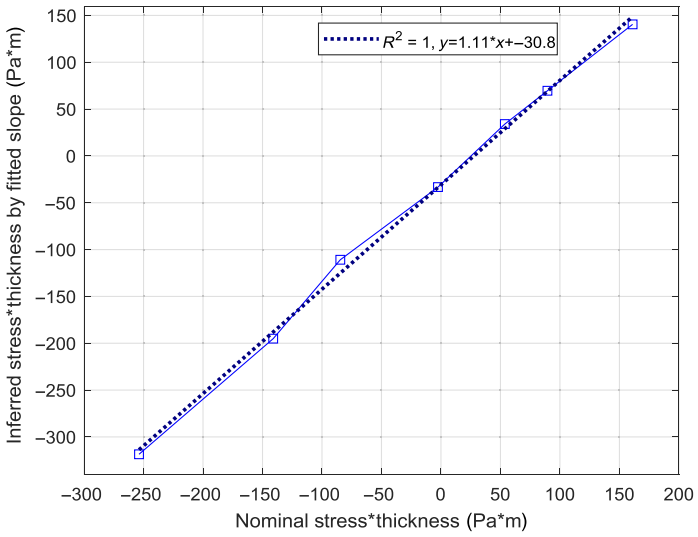

Fig. 14 Inferred SiN stress × SiN thickness from the fitted slopes versus the nominal stress × thickness.

4.5 *Wafer Magnification Term Measured by Alignment Sensor*

In addition to the above local distortion measurements, the alignment sensor also reports the full wafer grid distortion map, as shown in the first row of Fig. 12. At wafer scale, the scaling fingerprints can be fitted with linear wafer models to extract the wafer magnification term, which gives the relative strain (release after dry etch) of the wafers in the clamped state. In these measurements, the wafer magnification ranges from −0.6 to +0.4 parts per million (ppm). These numbers are linked to how much stress is released during the etching process, by the equation as follows:[16]

$$\varepsilon_s = -(1-\upsilon_s)\Delta\sigma_0 \frac{h_f}{E_s h_s}, \tag{17}$$

where ε_s is the strain of the substrate and $\Delta\sigma_0$ is the stress release. According to Eq. (16), assuming remaining stress fraction is equal to remaining material fraction, i.e., $C = 1$, the stress release is

$$\Delta\sigma_0 = \sigma_0 - C(1-\overline{\rho})\sigma_0 = \overline{\rho}\sigma_0. \tag{18}$$

Taking the average mask pattern density $\overline{\rho} = 0.3006$, the initial stress of the wafers can be inferred from the measured wafer strain, as given in Fig. 15.

4.6 *Multiscale Consistency and Discussions*

Comparing Figs. 11, 14, and 15, the model's consistency can be checked across different length scales. It is easy to notice the stress inferred by the measured wafer strain is 57% higher (slope = 1.57) than the nominal values; however, the stress inferred by local grid distortion (both at 9-mm die scale and 40-μm feature scale) are only ~10% higher than the nominal values. Such discrepancy is expected because of the non-linear stress release behavior, as opposed to the assumptions used in Eq. (18). To verify the non-linear stress release behavior, simplified finite element modeling (FEM) simulations are performed with results shown in Fig. 16. The non-linear stress release can be approximately fitted as

$$\frac{\Delta\sigma_0}{\sigma_0} = \rho^{0.68}, \tag{19}$$

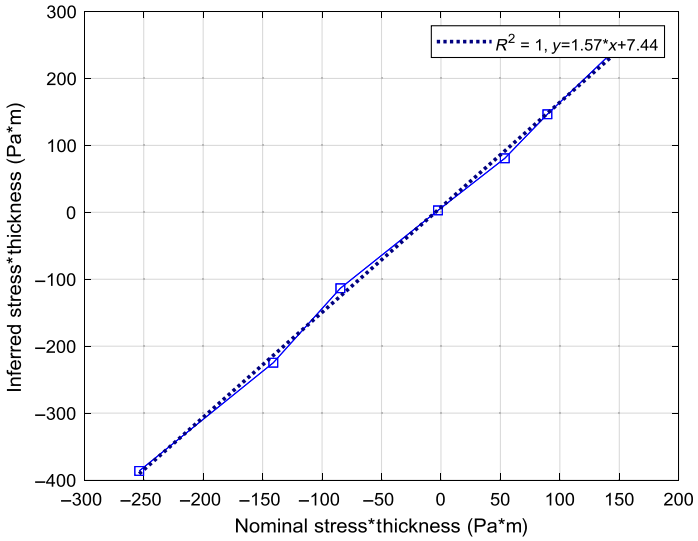

Fig. 15 Inferred SiN stress × SiN thickness from the measured wafer strain release (by alignment sensor).

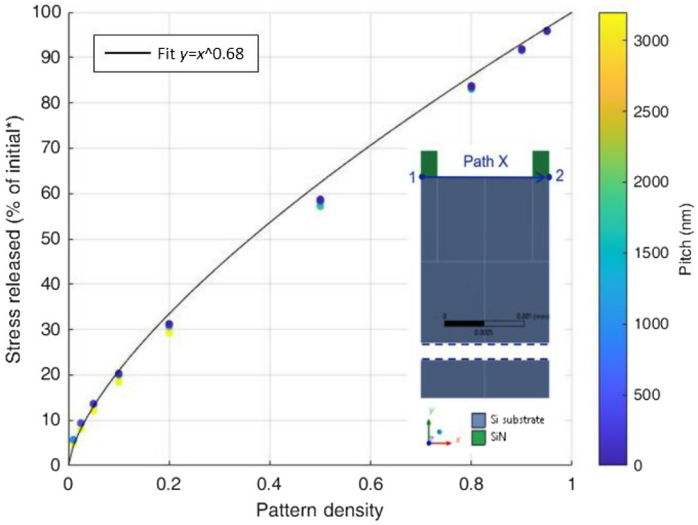

Fig. 16 Simplified FEM simulation to demonstrate the non-linear stress release (versus pattern density) behavior and pitch dependency. The inset shows the configurations of the toy FEM model.

as opposed to the original linear assumption with $\frac{\Delta\sigma_0}{\sigma_0} = \rho$. According to the simulation result, about 44% of stress is released during the etching step when 30% material is removed. Therefore, the correction factor in Eq. (18) is $C(\rho = 30\%) = \frac{1-44\%}{1-30\%} = 0.80$. Plugging this correction factor back to Eqs. (17) and (18) modifies the slope in Fig. 15 to 1.07, much closer to the slopes reported in Figs. 11 and 14. To be noted, if Eq. (19) can be assumed to work locally, we could express the correction factor as

$$\mathrm{C}(\rho) = \frac{1-\rho^{0.68}}{1-\rho}. \tag{20}$$

Such a correction factor also modifies the inferred initial stress from the local grid distortion data. However, the modification is not as big. The inferred initial stress considering the non-unity correction factor versus nominal values are given in Fig. 17 for all three cases at different length scales. The data are very consistent, meaning the model is consistent across the length scales.

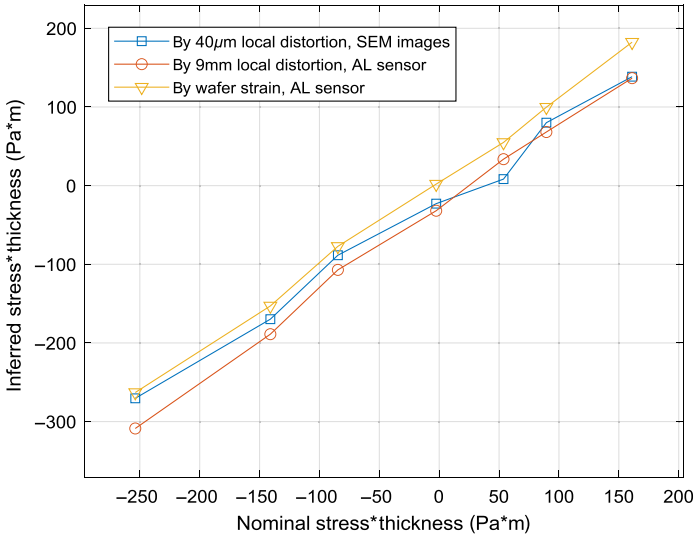

Fig. 17 Inferred SiN stress × SiN thickness by local grid distortion and global wafer strain after considering the non-linear stress release behavior.

It should be pointed out that the above analysis was performed to explain the apparent inconsistency in wafer scale strain data. However, such an analysis is not necessary for local grid distortion prediction, as presented in Secs. 4.3 and 4.4. In practice, the pattern density variation is smaller (<20%) and the stress release can be assumed linear in the first-order approximation. Therefore, it is easy to add a constant correction factor C if necessary. Such a correction factor can be obtained from a look-up table or figure as shown in Fig. 16, or obtained by running a feature scale FEM simulation, or by simply experimental calibration. Nevertheless, even without the correction factor, the model already works great for local grid distortions.

As a point of improvement, in the case of extreme anisotropic and arbitrary non-orthogonal patterns, local shear stress variations may occur. This is currently not considered in the kernel function derivations for simplicity.

5 Conclusions

In this work, the idea of using a pattern density-based compact model to predict wafer grid distortion and overlay error is tested. The long-range nature of stress impact makes it practically impossible to simulate a full chip with all the devices explicitly modeled. Based on the results, even with only coarse grid pattern density as input, the grid distortion is already relatively well predicted. The fast and compact model could enable multiple applications, ranging from active and feedforward correction, to monitoring and smart sampling. The accuracy of the model can be further improved by combining device scale FEM simulation results and include the pitch dependency in the correction factors. Nevertheless, even without such calculations, the model can already give good predictions to the intra-field and local grid distortions, with root mean square error <0.42 nm (SEM) and <0.28 nm (alignment sensor), close to the baseline metrology uncertainty.

With the predictive model, the actual stress impact on overlay can be estimated providing an estimated chip layout. Based on the results above and realistic chip designs and design non-uniformity, we expect about ∼1 nm of local grid distortion to happen at below 1-mm length scale for realistic logic chip layout at middle-of-line layers. The predictive model as presented in this paper enables process-aware design optimization and corrections.

Acknowledgments

The authors acknowledge support from the ASML-IMEC APC program.

References

1. M. van den Brink, "Holistic lithography in the age of the artificial intelligence of things," *Proc. SPIE* **11517**, 1151702 (2020).
2. L. Zhang et al., "Process context based wafer level grouping control: an advanced overlay process correction designed for DRAM 1z nm node in high volume manufacturing," *Proc. SPIE* **11325**, 113251O (2020).
3. S. Reboh, R. Coquand, and S. Barraud, "Strain, stress, and mechanical relaxation in fin-patterned Si/SiGe multilayers for sub-7 nm nanosheet gate-all-around device technology," *Appl. Phys. Lett.* **112**(5), 051901 (2018).
4. T. A. Brunner et al., "Characterization of wafer geometry and overlay error on silicon wafers with nonuniform stress," *J. Micro/Nanolithography MEMS MOEMS* **12**(4), 043002 (2013).
5. R. van Haren et al., "On-product overlay characterization after stressed layer etch," *Proc. SPIE* **11615**, 116150N (2021).
6. I. Stobert et al., "Model-based correction for local stress-induced overlay errors," *Proc. SPIE* **10587**, 105870D (2018).
7. L. Lanzoni, "Analysis of stress singularities in thin coatings bonded to a semi-infinite elastic substrate," *Int. J. Solids Struct.* **48**, 1915–1926 (2011).
8. S. Venkataraman, "Elasticity analysis and optimization of a functionally graded plate with hole," in *44th AIAA/ASME/ASCE/AHS Struct., Struct. Dyn., and Mater. Conf.*, Virginia, p. 1466 (2003).
9. Y. Huang, D. Ngo, and A. J. Rosakis, "Non-uniform, axisymmetric misfit strain: in thin films bonded on plate substrates/substrate systems: the relation between non-uniform film stresses and system curvatures," *Acta Mech. Sin.* **21**, 362–370 (2005).
10. D. Ngo et al., "Spatially non-uniform, isotropic misfit strain in thin films bonded on plate substrates: the relation between non-uniform film stresses and system curvatures," *Thin Solid Films* **515**, 2220–2229 (2006).
11. A. K. Rana, S. K. Paul, and P. P. Dey, "Stress field in an isotropic elastic solid containing a circular hard or soft inclusion under uniaxial tensile stress," *Mater. Today: Proc.* **11**, 657–666 (2019).
12. Z. Li, C. Lima, and L. He, "Stress concentration around a nano-scale spherical cavity in elastic media: effect of surface stress," *Eur. J. Mech. A/Solids* **25**, 260–270 (2006).
13. C. W. Lim, Z. R. Li, and L. H. He, "Size dependent, non-uniform elastic field inside a nano-scale spherical inclusion due to interface stress," *Int. J. Solids Struct.* **43**, 5055–5065 (2006).
14. M. A. Hopcroft, W. D. Nix, and T. W. Kenny, "What is the Young's modulus of silicon?" *IEEE J. Microelectromech. Syst.* **19**(2), 229–238 (2010).
15. K.-B. Kim, Y. Oh, and Y.-H. Song, "Simulation of residual stress and its impact on a poly-silicon channel for three-dimensional, stacked, vertical-NAND flash memories," *J. Korean Phys. Soc.* **70**, 1041–1048 (2017).
16. U. o. Cambridge, "11.3: force and moment balances," (2020). https://eng.libretexts.org/Bookshelves/Materials_Science/TLP_Library_I/11%3A_Coating_mechanics/11.03%3A_Section_3-.

Huaichen Zhang is specialized in multiscale multiphysics modeling of process and metrology. He received his PhD in 2017 from Eindhoven University of Technology in the field of mechanical engineering and computational material science, and has been working at ASML Research, Veldhoven, Netherlands, on computational lithography products. He is an architect for the research computational litho+ project, where computational models are developed to capture process impact on the performance of scanners and metrology sensors.

Cyrus Tabery is an expert in yield, DFT, statistics, design technology, lithography, metrology, reticle technology, and OPC. In 2000, he was with AMD, Silicon Valley, California, where he researched lithography, mask making, and OPC. He worked with Globalfoundries, Santa Clara, California, USA, until August 2011, and Intel Custom Foundry from 2011 to 2016 on process design kit development with focus on physical verification and fill and DFM technology. In 2016, he joined ASML, San Jose, California, as a technical staff member working on design

and patterning technology in Brion ATD. He has published 100+ articles and patents in the semiconductor industry and is a member of SPIE.

Ruben Maas received his PhD in nanophotonics from the University of Amsterdam in 2015, working on the fabrication and characterization of optical metamaterials. Since then, he has joined ASML and has worked on EUV lithography, focusing on the role of resist and CD-SEM metrology in pattern variability and stochastic defectivity. Since 2019, he has been project leader for the research computational litho+ project, where computational models are developed to capture process impact on scanner performance.

Oleksandr Khodko is Professional Simulation Engineer certified by NAFEMS. He received his PhD in the field of Mechanical Engineering from the National Aerospace University "KhAI" in 2016. He has diverse research and engineering experience in aerospace, automotive, renewable energy, and semiconductor industries. After joining ASML in 2019, he has been working on various projects, including the research computational litho+ project, where computational models are developed to capture process impact on scanner performance.

Victor M. Blanco Carballo received his PhD in 2009 from the University of Twente, working on CMOS post-processing. In 2010, he was with NIKHEF where he researched radiation imaging detectors for high energy physics, and with ASML from 2011 to 2016 working on immersion and EUV defectivity. Currently, he is with IMEC where his main research interests are EUV lithography, e-beam metrology, and BEOL process integration.

Eren Canga received his MSc degree from the Middle East Technical University in 2009, working on the material characterization for the uncooled infrared detectors. From 2006 to 2018, he worked with METU-MEMS Research and Application Center focusing on the design and process development for the uncooled infrared detectors. In 2018, he joined metrology team at IMEC where his main activities are optical metrology, EUV lithography, and e-beam metrology.

Filip Schleicher received his PhD from the University of Strasbourg in 2012 with the thesis titled "Impact of structural defects on spin-polarized transport across magnetic tunnel junctions." His work was heavily fabrication oriented, with strong focus on mask design, thin-film growth, lithography, etch, and electrical characterization of final devices. After a cooperative post-doctoral position shared between IPCMS in Strasbourg and IJL in Nancy, he joined the dry etch team at IMEC in 2019. His expertise encompasses N3 BEOL applications, lines/spaces, contact holes, and various 3D integration schemes.