

Residual speckle in a lithographic illumination system

Gregg M. Gallatin*

Applied Math Solutions, LLC
6 Castle Lane
Newtown, Connecticut 06470
E-mail: ggallatin@charter.net

Naonori Kita

Nikon Corporation
Development Headquarters
Optical Design Department
Strategic Technology Development Section
201-9 Miizugahara, Kumagaya
Saitama 360-8559, Japan

Tomoko Ujike

Nikon Corporation
Research and Development Headquarters
Core Technology Center
Optical Research Laboratory
4th Development Section
201-9 Miizugahara, Kumagaya
Saitama 360-8559, Japan

Bill Partlo

Cymer Incorporated
17075 Thornmint Court
San Diego, California 92127

1 Introduction

Partially coherent imaging in a lithographic exposure tool is usually modeled in an idealized fashion. For example, Kohler illumination is treated as a set of plane waves, spread over a given range of angles, with each plane wave taken to be incoherent with respect to every other plane wave. The term “incoherent” here means that when we compute the intensity by squaring the sum of the plane waves, we are to drop all the cross terms and keep only the sum of the squares of each plane wave with itself. Since the intensity of any plane wave is uniform, the sum of the intensities of a set of incoherent plane waves is also uniform. But why do we drop the cross terms? After all, Maxwell’s equations are linear, therefore superposition holds and the cross terms should be there. The answer is the cross terms really are there but they vary with time for any illumination that has a finite bandwidth and they would contribute if we detected the instantaneous intensity. But what counts for exposure is the time integrated intensity or dose. In this case, since the cross terms are time varying, they integrate approximately to zero over times long compared

Abstract. Finite bandwidth and finite exposure time place a fundamental limit on dose uniformity. We evaluate the amplitude and spatial distribution of this residual speckle in a given type of lithographic illumination system. For nominal bandwidths and exposure times, the level of dose nonuniformity is on the order of several percent. We argue that this effect actually makes only a small contribution to line edge roughness. © 2009 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.3256007]

Subject terms: speckle; illumination; lithography; line edge roughness.

Paper 09068R received Apr. 6, 2009; revised manuscript received Aug. 26, 2009; accepted for publication Sep. 13, 2009; published online Nov. 25, 2009.

to the reciprocal of the bandwidth. The diagonal terms on the other hand are static, i.e., not time varying, and hence their integral increases linearly with time. So, although the integrals of the cross terms are on average zero, there will be some residual root-means-square (rms) spatial variation in the dose. It is the statistics of this residual dose nonuniformity that we discuss in this work.

This effect can be understood from a physical point of view by noting that the instantaneous interference pattern formed by multiple plane waves with arbitrary or random phases and frequencies spread over a finite bandwidth is a nonstationary, i.e., time varying, 100% contrast speckle pattern.^{1,2} The pattern is not stationary in time due to the finite bandwidth, which constantly changes the relative phases of the plane waves. The assumption here is that the instantaneous frequency of each plane wave, in Kohler illumination for example, is uncorrelated with the all other plane waves and that the frequency of each plane wave varies more or less randomly over the full bandwidth during the scale of the time integration, i.e., during the exposure time. This assumption is equivalent to requiring spatial incoherence in the source pupil, which is the standard assumption in partially coherent imaging. (If the plane waves all had the same time-varying frequency, then the speckle pattern would be stationary and would not average in time.) The accumulated dose or time integral of this nonstationary

* Current address: Institute of Standards and Technology, Center for Nanoscale Science and Technology, Gaithersburg, Maryland 20899.
1932-5150/2009/\$25.00 © 2009 SPIE

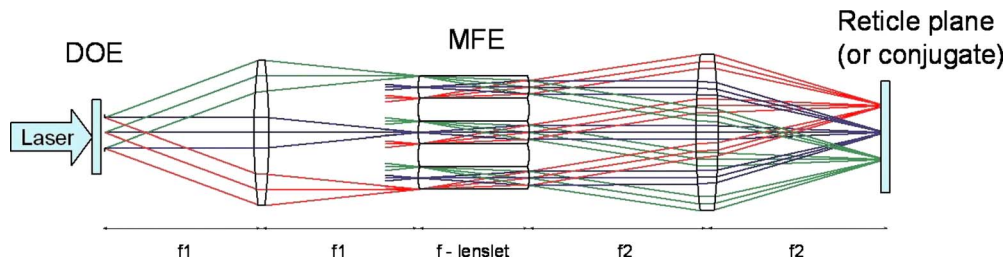


Fig. 1 Schematic layout of an illuminator.

speckle pattern will be smoother than a snapshot of the speckle pattern, with the uniformity improving with increasing integration time and/or bandwidth of the illumination. But, for any finite integration time and/or finite bandwidth, there will always be some residual illumination nonuniformity or residual speckle. The mathematical origin of this residual nonuniformity can be understood almost trivially by considering the integration of $\sin(\theta)$ and $\sin(\theta)^2$ over a range of θ much greater than 2π , where $\sin(\theta)$ and $\sin(\theta)^2$ act like the cross terms and diagonal terms between plane waves, respectively. For each 2π segment of the range, $\sin(\theta)$ integrates to zero and hence the net value of its integral comes solely from the last fraction of 2π in the integration range. On the other hand, $\sin(\theta)^2 = 1/2[1 - \cos(2\theta)]$. The constant term $1/2$ integrates to the range/2 and the $\cos(2\theta)$ term leaves only a small residual, just like the $\sin(\theta)$ integration. Normalizing by the range/2, we see that the contribution from the $\sin(\theta)^2$ term is fixed at unity but the cross terms decrease as $2/\text{range}$ and are not generally zero for any finite integration range. The practical implication of all this is that for a short enough integration time and/or a narrow enough bandwidth, the illumination nonuniformity can be large compared to the specification on dose uniformity. But this is precisely where lithographic exposure tools are designed to operate. High throughput requires short integration times, and minimizing chromatic aberrations in refractive optical systems requires narrow bandwidth.

We present a detailed analysis of the residual speckle that occurs in a generic type of lithographic illumination system. The work is organized as follows. Section 2 presents a simple derivation of the expected level of residual speckle under general conditions of illumination and relates this value to dose nonuniformity. Section 3 presents an overview of the illumination system. Section 4 describes how we represent an excimer pulse. Section 5 analyzes, step by step, the propagation of an excimer pulse through the illuminator. In Sec. 6 the mean dose and the expected residual speckle or rms variation in the dose is calculated.

2 Simple Derivation of Residual Speckle

Let T be the integration time and let τ be the correlation time for the speckle pattern, i.e., how long on average it takes the speckle pattern to become uncorrelated with itself. The correlation time is related to the bandwidth $\Delta\nu$, in Hertz, via $\tau = 1/(2\pi\Delta\nu)$. It follows from this that the number of statistically independent speckle patterns N that are summed during the exposure time is given by

$$N \approx T/\tau = 2\pi T\Delta\nu = 2\pi cT \frac{\Delta\lambda}{\lambda^2},$$

where λ is the wavelength and c is the speed of light. For uncorrelated processes we expect the root mean square or 1σ variation to grow as \sqrt{N} , and hence the relative nonuniformity varies as $\sqrt{N}/N = 1/\sqrt{N} \approx \sqrt{\tau}/T$. Using the fact that each speckle pattern by itself has 100% contrast, we expect the 1σ dose nonuniformity to be on the order of

$$\sigma_{\% \text{dose}} \approx \sqrt{\frac{1}{2}} \sqrt{\frac{\lambda^2}{2\pi c T \Delta\lambda}} \times 100\%,$$

where the $1/\sqrt{2}$ comes from assuming two independent polarizations of the field. Taking the nominal value of 30 pulses to expose with each pulse being ~ 20 nsec long and assuming a bandwidth $\Delta\lambda \sim 0.2$ pm at $\lambda = 193$ nm, we find $\sigma_{\% \text{dose}} \sim 1\%$, which is essentially the entire dose budget.

The spatial distribution of the speckle pattern depends on the pupil fill geometry with the autocorrelation function being the Fourier transform of the pupil fill geometry. This is shown in detail next. See also the work by Rydberg et al.,^{3,4} Kritsun et al.,⁵ and Noordman et al.⁶ For now, note that the characteristic length scale or spatial period for the residual nonuniformity is the length scale of the speckle pattern itself, which is roughly $\lambda/(2\sigma\text{NA})$, where σ refers to top-hat pupil fill and NA is the numerical aperture of the optics measured at the wafer. For top-hat pupil fills on the order of 0.6 to 0.8, this length scale is on the order of the minimum printable feature size, and therefore we would expect this dose nonuniformity to manifest itself as line edge roughness with a period on the order of twice the resolution.

3 Illuminator Design

A generic design for an illuminator based on a diffractive optical element (DOE) is shown in Fig. 1. After beam expanding, and possibly pulse stretching, the beam coming from the excimer passes through the DOE, which is designed so that the diffraction pattern it generates has the geometry of the desired pupil fill. The lens with focal length f_1 would actually be a zoom system, but we treat it as a single lens and adjust the value of f_1 as needed to incorporate the zoom aspect into the model. This lens places the desired diffraction pattern on the entrance surface of the micro-flies eye (MFE). The field entering each lenslet of the MFE is magnified and reimaged onto the illumination area at the reticle conjugate plane with the angles of incidence corresponding to the diffraction angles

from the DOE. The pattern can be further propagated in standard fashion to the reticle plane itself or to the wafer plane.

4 Comment on the Modeling Approach

In general, at the reticle plane an illumination system will mix both transverse and longitudinally shifted regions of the incident excimer pulse. As long as the longitudinal shift is much less than the longitudinal coherence length, the longitudinal shift can be ignored and standard physical optics methods for monochromatic fields that have an infinite longitudinal coherence length can be used to compute the propagated fields. The longitudinal coherence length is approximately $\lambda^2/\Delta\lambda$, where $\Delta\lambda$ is the full-width-half-maximum bandwidth of the laser. The longitudinal shift caused by the illuminator can be estimated as the product of the maximum NA in the illuminator and the maximum radius of the intensity distribution inside the illuminator. For the illuminator considered here, the maximum angle is roughly the illumination NA, and the maximum transverse radius is the NA times the largest focal length, and so the longitudinal shift is on the order of $NA^2 \times f_{\max}$. We assume that this shift is much less than the longitudinal coherence length, in which case it is the transverse spatial coherence length that controls the mixing of the speckle patterns at the reticle plane. It is certainly possible to do the analysis that follows to include the effect of the longitudinal coherence length, but this nominally leads to a much more complex calculation.

Since the pulse is finite and propagates at a finite speed, the time at which a given portion of the pulse arrives at each element in the illuminator is different. This time lag is roughly $2f_1/c$ to get from the DOE to the MFE, and roughly $2f_2/c$ to get from the MFE to the reticle (or reticle conjugate) plane. We effectively ignore this time lag and define $t=0$ at each element to be the time the front edge of the pulse arrives at that element. We also define $z=0$ separately for each element to be the (entrance or exit) plane of the element. With this notation and ignoring the longitudinal shearing of the pulse, the parameter z just floats through the calculation and is not shown explicitly until the end when, to compute the statistics, we do an integral over z to account for the contribution of the entire pulse.

For the low NA at the reticle plane or at any conjugate plane with close to unit magnification, polarization effects can nominally be ignored. But at the wafer where the NA is on the order of unity or greater, polarization effects must be properly accounted for to accurately predict the illumination intensity distribution. To add in polarization effects, we need to include the electric field vector in the prior considerations. This can be done by “attaching” the appropriate electric field vector to each plane wave incident on the wafer, which is the standard procedure used in modern models of partially coherent imaging. Since our purpose here is to understand the fundamental origin of residual speckle, we do not consider polarization effects any further.

5 Representing the Excimer Laser Pulse

Excimer laser beams are neither spatially nor temporally coherent. The lack of temporal coherence follows from the finite bandwidth $\Delta\lambda$ of the beam. The finite spatial coherence follows from the fact that at any point in time, differ-

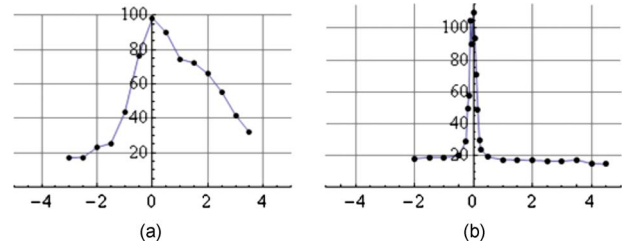


Fig. 2 The graphs show the average fringe contrast, measured using two-pinhole interference, as a function of (a) horizontal pinhole position and (b) vertical pinhole after a $4\times$ magnification of the beam. The contrast has been corrected by dividing the raw contrast by the theoretical contrast predicted based on the relative energy passing through each pinhole. The approximate 15 to 20% floor in the fringe contrast is clearly visible.

ent spatial positions inside the beam have (slightly) different wavelengths λ , and hence their phase difference changes with time. Clearly if the beam had zero bandwidth $\Delta\lambda=0$, then it would be perfectly temporally coherent. This also implies perfect spatial coherence, since all positions within the beam would then also have exactly the same wavelength and hence fixed relative phases.

Data from at least one excimer (see Fig. 2) shows that the autocorrelation or coherence of a pulse with itself as measured by the interference contrast does not necessarily go to zero with large spatial separation, but rather to a constant nonzero value. This nonzero “floor” in the contrast occurs in both the longitudinal and transverse directions with the same nonzero value in all directions. The model described here automatically allows for the existence of a contrast “floor,” but can also handle the zero floor case depending on the parameter values chosen.

To model the excimer pulse, let the field of the pulse (at say time $t=0$ just as it begins to impinge on the DOE) be given by

$$\phi(\vec{r}) = A(r)\exp[ikz + i\theta(\vec{r})]. \quad (1)$$

Here $A(\vec{r})$ is a real valued function giving the overall shape of the pulse and $k=2\pi/\lambda$ with λ the nominal wavelength. The values of z correspond to positions just upstream of the DOE, which we take to be at $z=0$. The variation in the phase $\theta(\vec{r})$ is treated as a random function of position $\vec{r}=(x,y,z)$ within the pulse. The statistics of $\theta(\vec{r})$ are taken to be Gaussian (in function space). This can be written as

$$P[\theta] = \mathcal{N} \exp\left[-\frac{1}{2} \int d^3r d^3r' \theta(\vec{r}) C^{-1}(\vec{r}, \vec{r}') \theta(\vec{r}')\right], \quad (2)$$

where \mathcal{N} is a normalization factor and the notation $P[\theta]$ is meant indicate that P depends only on the functional form of $\theta(\vec{r})$ and not on its value at any particular position \vec{r} . Since the overall average phase of the pulse is irrelevant, we have defined $P[\theta]$ so that $\theta(\vec{r})$ has zero mean, $\langle\theta(\vec{r})\rangle=0$. The function $C^{-1}(\vec{r}, \vec{r}')$ is meant to indicate the reciprocal or inverse of the function $C(\vec{r}, \vec{r}')$, i.e.,

$$\int d^3s C^{-1}(\vec{r}, \vec{s}) C(\vec{s}, \vec{r}') = \delta(\vec{r} - \vec{r}'). \quad (3)$$

Next we see that $C(\vec{r}, \vec{r}') = \langle \theta(\vec{r}) \theta(\vec{r}') \rangle$. The generating functional $Z[J]$ for the autocorrelation functions of $\theta(\vec{r})$ is defined as the (functional) Fourier transform of $P[\theta]$,

$$Z[J] = \int \mathcal{D}\theta \exp\left[i \int d^3r J(\vec{r}) \theta(\vec{r})\right] P[\theta]. \quad (4)$$

It can be evaluated exactly by completing the square in the exponent to get

$$Z[J] = \exp\left[-\frac{1}{2} \int d^3r d^3r' J(\vec{r}) C(\vec{r}, \vec{r}') J(\vec{r}')\right]. \quad (5)$$

Here $\int \mathcal{D}\theta$ indicates functional integration. Functional techniques such as these are described in many books on quantum field theory. See for example the excellent book by Zee.⁷

The definition of $Z[J]$ implies

$$\langle 1 \rangle = \int \mathcal{D}\theta P[\theta] = Z[0] = 1,$$

$$\langle \theta(\vec{r}) \rangle = \int \mathcal{D}\theta \theta(\vec{r}) P[\theta] = \left(-i \frac{\mathcal{D}}{\mathcal{D}J(\vec{r})} Z[J]\right)_{J=0} = 0,$$

$$\begin{aligned} \langle \theta(\vec{r}) \theta(\vec{r}') \rangle &= \int \mathcal{D}\theta \theta(\vec{r}) \theta(\vec{r}') P[\theta] \\ &= \left(-\frac{\mathcal{D}^2}{\mathcal{D}J(\vec{r}) \mathcal{D}J(\vec{r}')} Z[J]\right)_{J=0} = C(\vec{r}, \vec{r}'), \end{aligned} \quad (6)$$

where $\mathcal{D}/\mathcal{D}J(\vec{r})$ indicates functional differentiation. The first result shows that $Z[J]$ as given properly accounts for the normalization of $P[\theta]$, the second that the mean value of $\theta(\vec{r})$ vanishes as desired, and finally that C is the autocorrelation function of $\theta(\vec{r})$. Next we show that this approach for representing the excimer pulse automatically allows for a nonzero contrast floor with an amplitude that depends on the root-mean-square (rms) variation of $\theta(\vec{r})$. Note that to determine the speckle we need the expectation value of products of $\exp[i\theta]$, e.g., $\langle \exp[i\theta(\vec{r}) - i\theta(\vec{r}')] \rangle$. In terms of these expectation values, the Gaussian statistics used here smoothly transition in the limit of $\langle \theta(\vec{r})^2 \rangle = \sigma^2$ approaching infinity to the results obtained using the standard uniform 0 to 2π distribution for θ . This is shown briefly in the Appendix.

Figure 3 provides an illustration of this way of representing a pulse. The upper graph illustrates schematically an “ideal” pulse $A(\vec{r}) \exp[ikz]$ propagating in the horizontal or $+z$ direction. The center wavelength corresponds to the value of k and the bandwidth is controlled by the length of $A(\vec{r})$ in the propagation direction. The decrease in amplitude around the edges is meant to correspond to the overall length and width of the pulse as given by the amplitude factor $A(\vec{r})$. The lower graph illustrates the case where $\theta(\vec{r})$

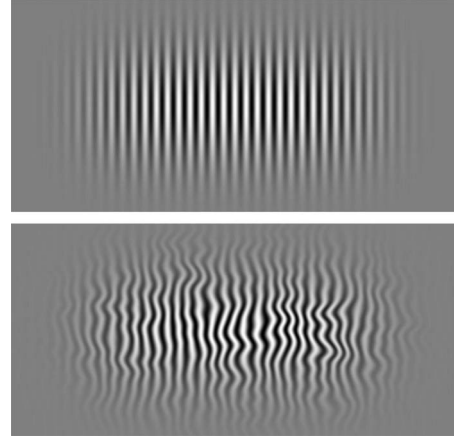


Fig. 3 The upper graph is an illustration of the field of an ideal [$\theta(\vec{r})=0$] perfectly collimated excimer pulse propagating horizontally. The lower graph shows how we model a real excimer laser pulse, i.e., one with $\theta(\vec{r}) \neq 0$. An example of what $\theta(\vec{r})$ itself looks like is given in Fig. 4. In this case, the wavefronts are not flat nor equally spaced. The nonflatness leads to the added beam divergence and the nonequal spacing to the bandwidth. The pulses shown are only about 40 wavelengths long. Real pulses are on the order of 10^8 wavelengths long.

is not zero. The wavefronts are no longer flat or uniformly spaced. The nonflatness leads to the extra beam divergence and the nonequal spacing to the increase in bandwidth beyond that corresponding to the length of the pulse. The bandwidth of the ideal pulse in the upper graph would be on the order of $10^{-8}\lambda$, since the width of its longitudinal Fourier transform is controlled by the length of the pulse in the z direction, which for a real pulse is about 10^8 wavelengths. The phase variations within a real pulse increases the bandwidth significantly so that it is actually on the order of $0.1 \text{ pm} \sim 10^{-6}\lambda$. Detailed evaluation of the bandwidth and contrast of this model for the pulse is given next.

The pulses shown in Fig. 3 are meant as illustrations only, since they are only about 40 wavelengths long. An example of what $\theta(\vec{r})$ itself looks like is shown in Fig. 4. The statistics of $\theta(\vec{r})$ are described by two parameters, its rms fluctuation throughout the pulse $\sigma = \langle \theta(\vec{r})^2 \rangle$ and its correlation length ξ . Essentially σ controls the size of the fluctuations and ξ controls how fast they vary with position in the pulse.

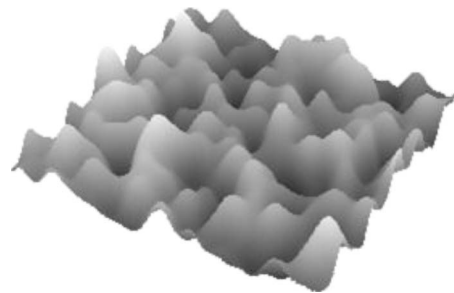


Fig. 4 Illustration of a random local phase variation $\theta(\vec{r})$ throughout the volume of the pulse. The vertical axis is the phase variation and the horizontal axes correspond to the x and z directions in Fig. 3.

Finally, it should be noted that this way of representing the pulse obscures some aspects of the physics. In particular, the fact that $\langle \theta(\vec{r}) \theta(\vec{r}') \rangle$ goes to zero for large transverse spatial separations can occur only after sufficient time integration to allow for the temporal incoherence to produce transverse spatial incoherence. Since the pulse is propagating predominantly in the z direction, the time integration can be replaced by a z integration and so we should consider $\langle \cdots \rangle$ to be equivalent to an integral in z over a distance L , which is much longer than the longitudinal coherence length ξ_z , i.e.,

$$\langle \theta(\vec{r}) \theta(\vec{r}') \rangle = \int_{L \gg \xi_z} dZ \theta(\vec{r} + Z\hat{z}) \theta(\vec{r}' + Z\hat{z}),$$

where \hat{z} is the unit vector in the z direction.

The Gaussian probability distribution we use for the function $\theta(\vec{r})$ should not be confused with the shape of the autocorrelation function of $\theta(\vec{r})$, i.e., with $\langle \theta(\vec{r}) \theta(\vec{r}') \rangle$, which can in principle be anything but commonly is taken to be either a Gaussian itself or an exponential or whatever function best fits the data. Often the autocorrelation function is isotropic and translation invariant, in which case we have

$$\langle \theta(\vec{r}) \theta(\vec{r}') \rangle = \sigma^2 f(|\vec{r} - \vec{r}'|/\xi) = \sigma^2 f(R/\xi), \quad (7)$$

where

$$R = |\vec{r} - \vec{r}'| = [(x - x')^2 + (y - y')^2 + (z - z')^2]^{1/2}$$

is the distance between positions \vec{r} and \vec{r}' , and ξ is the autocorrelation or coherence length of $\theta(\vec{r})$. The function $f(R)$ is normalized with $f(0) = 1$, so that $\langle \theta(\vec{r})^2 \rangle = \sigma^2$ with σ as the rms fluctuation in $\theta(\vec{r})$. An ideal pulse with $\theta(\vec{r}) = 0$ would have $\sigma = \sqrt{\langle \theta(\vec{r})^2 \rangle} = 0$ and thus perfect coherence. The functional form of f is usually either Gaussian $f(R/\xi) = \exp[-R^2/\xi^2]$ or exponential $f(R/\xi) = \exp[-R/\xi]$, where in both cases ξ is the correlation length of $\theta(\vec{r})$.

Now, to compute the speckle we need $\langle \exp[i\theta(\vec{r}) - i\theta(\vec{r}')] \rangle$. This can be evaluated directly by noting that

$$\langle \exp[i\theta(\vec{r}) - i\theta(\vec{r}')] \rangle = \left\langle \exp \left[i \int d^3s J(\vec{s}) \theta(\vec{s}) \right] \right\rangle, \quad (8)$$

with

$$J(\vec{s}) = \delta(\vec{r} - \vec{s}) - \delta(\vec{r}' - \vec{s}). \quad (9)$$

Then, substituting this into the result for $Z[J]$ given before, we get

$$\begin{aligned} \langle \exp[i\theta(\vec{r}) - i\theta(\vec{r}')] \rangle &= \exp \left[-\frac{1}{2} \langle \theta(\vec{r})^2 \rangle - \frac{1}{2} \langle \theta(\vec{r}')^2 \rangle + \langle \theta(\vec{r}) \theta(\vec{r}') \rangle \right] \\ &= \exp[-\sigma^2(1 - f(R/\xi))]. \end{aligned} \quad (10)$$

In the limit of large spatial separations, the autocorrelation function $f(R/\xi)$ generally approaches zero and we have $\langle \exp[i\theta(\vec{r}) - i\theta(\vec{r}')] \rangle = \exp(-\sigma^2) > 0$ for $R \gg \xi$. Thus, as claimed this approach automatically allows for a nonzero

floor in the contrast with the floor depending on the rms variation of the phase, $\sigma = \sqrt{\langle \theta(\vec{r})^2 \rangle}$. The implications that $\langle \exp[i\theta(\vec{r}) - i\theta(\vec{r}')] \rangle$ does not strictly go to zero at large spatial separations are discussed in more detail later.

5.1 Spatial Coherence and Contrast

The spatial coherence function comes from interfering ϕ at two different positions \vec{r}_1 and \vec{r}_2 .

$$|\phi(\vec{r}_1) + \phi(\vec{r}_2)|^2. \quad (11)$$

Since $\theta(\vec{r})$ is a random function, we need to take the expectation value of this to determine what the coherence function looks like ‘‘on average.’’ Thus we need to evaluate

$$\begin{aligned} \langle |\phi(\vec{r}_1) + \phi(\vec{r}_2)|^2 \rangle &= \langle |\phi(\vec{r}_1)|^2 + |\phi(\vec{r}_2)|^2 + \phi(\vec{r}_1) \phi(\vec{r}_2)^* + \phi(\vec{r}_1)^* \phi(\vec{r}_2) \rangle \\ &= \langle |\phi(\vec{r}_1)|^2 \rangle + \langle |\phi(\vec{r}_2)|^2 \rangle + \langle \phi(\vec{r}_1) \phi(\vec{r}_2)^* \rangle \\ &\quad + \langle \phi(\vec{r}_1)^* \phi(\vec{r}_2) \rangle, \end{aligned} \quad (12)$$

where $*$ indicates the complex conjugate and $|\phi|^2 = \phi \phi^*$. Evaluating each term using Eq. (10) and the definitions given before, then substituting back into $\langle |\phi(\vec{r}_1) + \phi(\vec{r}_2)|^2 \rangle$ yields

$$\begin{aligned} \langle |\phi(\vec{r}_1) + \phi(\vec{r}_2)|^2 \rangle &= A(\vec{r}_1)^2 + A(\vec{r}_2)^2 + 2 \cos[k(z_1 - z_2)] \\ &\quad \times A(\vec{r}_1) A(\vec{r}_2) \exp\{-\sigma^2[1 - f(R/\xi)]\}. \end{aligned} \quad (13)$$

The overall factor of $A(\vec{r}_1) A(\vec{r}_2)$ is just the overlap of the pulse shape with itself. This is usually not relevant when looking at contrast, since relevant values of r are generally much smaller than the size of the pulse. We eliminate it by assuming $A(\vec{r}_1) \approx A(\vec{r}_2) \approx A = \text{constant}$ for relevant positions \vec{r}_1 and \vec{r}_2 , which gives

$$\begin{aligned} \langle |\phi(\vec{r}_1) + \phi(\vec{r}_2)|^2 \rangle &= 2A^2(1 + \cos[k(z_1 - z_2)]) \\ &\quad \times \exp\{-\sigma^2[1 - f(R/\xi)]\}. \end{aligned} \quad (14)$$

The cosine term oscillates rapidly between a maximum of +1 and a minimum of -1, so the contrast Γ that is $(\max - \min)/(\max + \min)$ reduces to

$$\Gamma(R) = \exp\{-\sigma^2[1 - f(R/\xi)]\}. \quad (15)$$

For $\vec{r} = \vec{r}'$ so that $R = 0$, we get $\Gamma(0) = 1$ as expected. But, and here is the important point, in general we expect that $f(R/\xi) \rightarrow 0$ for $r \gg \xi$ and that $\Gamma(R)$ would also go to zero, but it does not. Rather $\Gamma(R)$ goes to the fixed nonzero value $\exp[-\sigma^2]$. Of course for σ values on the order of 2π or larger, $\exp[-\sigma^2] \approx 10^{-17}$, which, for all practical purposes is zero. But for σ values much less than 2π , $\exp[-\sigma^2]$ is distinctly and measurably not zero. Thus this model automatically predicts that the contrast does not go to zero but rather to a fixed nonzero value, or floor, for large R as long as σ is much smaller than 2π . This makes sense, since for phase fluctuations that are much less than 2π the contrast

cannot be zero. To match up with an $\sim 20\%$ floor, as seen in the excimer contrast data in Fig. 2, we need $\sigma \approx 1.27 \approx 0.2 \times 2\pi$, which yields $\exp[-\sigma^2] \approx 0.2 = 20\%$.

Finally, to describe a real excimer pulse we need to adjust the form of $f(R/\xi)$ so that it has different coherence lengths in the x , y , and z directions. This can be done by replacing $f(R/\xi)$ with

$$f\left\{\left[\frac{(x_1-x_2)^2}{\xi_x^2} + \frac{(y_1-y_2)^2}{\xi_y^2} + \frac{(z_1-z_2)^2}{\xi_z^2}\right]^{1/2}\right\} \equiv f(\vec{R}, \vec{\xi})$$

and $\Gamma(R)$ with

$$\Gamma(\vec{R}, \vec{\xi}) = \exp\{-\sigma^2[1 - f(\vec{R}, \vec{\xi})]\}, \quad (16)$$

to account for different coherence lengths ξ_x, ξ_y, ξ_z in the x , y , and z directions, respectively.

5.2 Spatial Frequency and Pulse Bandwidth

The bandwidth of the pulse follows from its Fourier transform, given by

$$\begin{aligned} \tilde{\phi}(\vec{p}) &= \int d^3r \exp(-i\vec{p} \cdot \vec{r}) \phi(\vec{r}) \\ &= \int d^3r \exp[-i(\vec{p} - \vec{k}) \cdot \vec{r} + i\theta(\vec{r})] A(\vec{r}). \end{aligned} \quad (17)$$

Since we are interested only in the deviation in frequency from the nominal value \vec{k} , let $\vec{p} - \vec{k} \rightarrow \vec{p}$ so that now \vec{p} measures the deviation from \vec{k} rather than the absolute frequency. We now have

$$\begin{aligned} \tilde{\phi}(\vec{p}) &= \int d^3r \exp(-i\vec{p} \cdot \vec{r}) \phi(\vec{r}) \\ &= \int d^3r \exp[-i\vec{p} \cdot \vec{r} + i\theta(\vec{r})] A(\vec{r}). \end{aligned} \quad (18)$$

The expectation value of the square of $\tilde{\phi}(\vec{p})$ is given by

$$\begin{aligned} \langle |\tilde{\phi}(\vec{p})|^2 \rangle &= \int d^3r_1 d^3r_2 \exp[-i\vec{p} \cdot (\vec{r}_2 - \vec{r}_1)] A(\vec{r}_1) A(\vec{r}_2) \\ &\quad \times \langle \exp[i\theta(\vec{r}_2) - i\theta(\vec{r}_1)] \rangle. \end{aligned} \quad (19)$$

Substituting results from these and using

$$A(\vec{r}) = \int \frac{d^3p}{(2\pi)^3} \tilde{A}(\vec{p}) \exp(i\vec{p} \cdot \vec{r}),$$

we get

$$\begin{aligned} \langle |\tilde{\phi}(\vec{p})|^2 \rangle &= \int \frac{d^3p_1 d^3p_2}{(2\pi)^6} \tilde{A}(\vec{p}_1) \tilde{A}(\vec{p}_2) \int d^3r_1 d^3r_2 \exp[i(\vec{p}_1 \\ &\quad + \vec{p}) \cdot \vec{r}_1 + i(\vec{p}_2 - \vec{p}) \cdot \vec{r}_2] \exp[-\sigma^2[1 - f(\vec{R}, \vec{\xi})]]. \end{aligned} \quad (20)$$

Since $A(\vec{r})$ is defined to be purely real, its Fourier transform $\tilde{A}(\vec{p})$ satisfies the identity $\tilde{A}(-\vec{p}) = \tilde{A}(\vec{p})^*$.

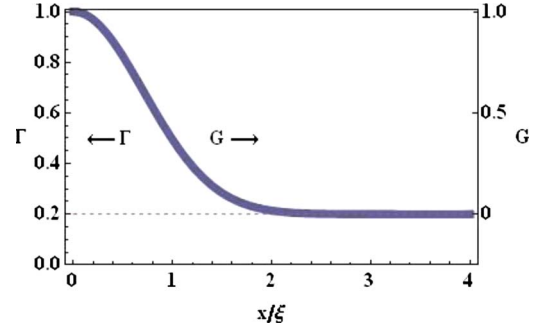


Fig. 5 The graph shows the difference between the contrast Γ and the function G . Γ ranges from 0 to 1 and includes the floor, which in this case is set at 0.2. The values of Γ are read off of the left vertical axis. The values of G are read off of the right vertical axis. G has the floor removed so that it goes to zero for large spatial separations, and it has been rescaled so that it ranges from 0 to 1. Other than that, the shape of G is exactly the same as the shape of Γ .

To separate out the constant background and normalize the nonconstant part of $\exp\{-\sigma^2[1 - f(\vec{R}, \vec{\xi})]\}$, first add and subtract $\exp[-\sigma^2]$, then multiply and divide $\exp\{-\sigma^2[1 - f(\vec{R}, \vec{\xi})] - \exp(-\sigma^2)\}$ by $\exp(\sigma^2) - 1$ to get

$$\begin{aligned} \exp\{-\sigma^2[1 - f(\vec{r}, \vec{\xi})]\} &= [1 - \exp(-\sigma^2)] \\ &\quad \times \left(\frac{\exp[\sigma^2 f(\vec{R}, \vec{\xi})] - 1}{\exp(\sigma^2) - 1} \right) + \exp(-\sigma^2) \\ &\equiv (1 - \Gamma_0)G(\vec{R}, \vec{\xi}) + \Gamma_0, \end{aligned} \quad (21)$$

where $\Gamma_0 = \exp[-\sigma^2]$ and $G(\vec{R}, \vec{\xi}) \equiv \{\exp[\sigma^2 f(\vec{R}, \vec{\xi})] - 1\} / [\exp(\sigma^2) - 1]$. $G(\vec{R}, \vec{\xi})$ is 1 at $r=0$ and goes to 0 for $|\vec{R}| \gg |\vec{\xi}|$. It is just the experimentally measured contrast shape with the floor subtracted and remaining variation scaled to go from 1 at $r=0$ to 0 for r large in any direction. This is shown in Fig. 5.

Putting this all together, we have

$$\begin{aligned} \langle |\tilde{\phi}(\vec{p})|^2 \rangle &= (1 - \Gamma_0) \int \frac{d^3p_1 d^3p_2}{(2\pi)^6} \tilde{A}(\vec{p}_1) \tilde{A}(\vec{p}_2) \int d^3r_1 d^3r_2 \\ &\quad \times \exp[i(\vec{p}_1 + \vec{p}) \cdot \vec{r}_1 + i(\vec{p}_2 - \vec{p}) \cdot \vec{r}_2] G(\vec{R}, \vec{\xi}) \\ &\quad + \Gamma_0 \int \frac{d^3p_1 d^3p_2}{(2\pi)^6} \tilde{A}(\vec{p}_1) \tilde{A}(\vec{p}_2) \int d^3r_1 d^3r_2 \\ &\quad \times \exp[i(\vec{p}_1 + \vec{p}) \cdot \vec{r}_1 + i(\vec{p}_2 - \vec{p}) \cdot \vec{r}_2]. \end{aligned} \quad (22)$$

The second term is simply $\Gamma_0 |\tilde{A}(\vec{p})|^2$, which is the contribution to the bandwidth coming from the pulse shape itself scaled by the factor $\Gamma_0 = \exp[-\sigma^2]$.

To evaluate the first term, make the change of variables $\vec{t} = 1/2(\vec{r}_1 + \vec{r}_2)$ and $\vec{R} = \vec{r}_2 - \vec{r}_1$, then using $\int d^3t \exp[i(\vec{\beta}_1 + \vec{\beta}_2) \cdot \vec{t}] = (2\pi)^3 \delta(\vec{\beta}_1 + \vec{\beta}_2)$, the first term becomes

$$\begin{aligned}
 (1 - \Gamma_0) \int \frac{d^3 p'}{(2\pi)^3} |\tilde{A}(\vec{p}')|^2 \int d^3 R \exp[-i(\vec{p} - \vec{p}') \cdot \vec{R}] G(\vec{R}, \vec{\xi}) \\
 = (1 - \Gamma_0) \int \frac{d^3 p'}{(2\pi)^3} |\tilde{A}(\vec{p}')|^2 \tilde{G}(\vec{p} - \vec{p}', \vec{\xi}), \quad (23)
 \end{aligned}$$

after defining $\tilde{G}(\vec{p}, \vec{\xi}) = \int d^3 R \exp(-i\vec{p} \cdot \vec{R}) G(\vec{R}, \vec{\xi})$. The final result is just the convolution of \tilde{G} with $|\tilde{A}|^2$. Nominally the pulse size is much larger than the correlation length in any direction, and so its Fourier transform \tilde{A} is generally much narrower than \tilde{G} . For a real excimer pulse, this is certainly true in the longitudinal and vertical directions, but only approximately true in the horizontal direction. But just to see how things scale, we make this assumption in which case the first term becomes

$$\begin{aligned}
 (1 - \Gamma_0) \int \frac{d^3 p'}{(2\pi)^3} |\tilde{A}(\vec{p}')|^2 \tilde{G}(\vec{p} - \vec{p}', \vec{\xi}) \\
 \approx [1 - \exp(-\sigma^2)] \tilde{G}(\vec{p}, \vec{\xi}) V_A, \quad (24)
 \end{aligned}$$

where $V_A \equiv \int d^3 p' / (2\pi)^3 |\tilde{A}(\vec{p}')|^2$ is the total volume of the pulse shape in frequency space. Combining this with the second term and normalizing by V_A yields

$$\frac{\langle |\tilde{\phi}(\vec{p})|^2 \rangle}{V_A} \approx (1 - \Gamma_0) \tilde{G}(\vec{p}, \vec{\xi}) + \Gamma_0 \frac{|\tilde{A}(\vec{p})|^2}{V_A}. \quad (25)$$

Thus we have that the bandwidth is the sum of two terms. The first term is essentially the bandwidth corresponding to the measured coherence or contrast of the pulse, since as shown before F has the same shape as Γ , and it is weighted by $1 - \Gamma_0 = 1 - \exp[-\sigma^2]$. The second term is the bandwidth of the pulse shape itself multiplied by $\Gamma_0 = \exp[-\sigma^2]$. Hence for small enough Γ_0 or equivalently large enough σ , the spatial frequency distribution and therefore the bandwidth in this model is controlled predominantly by the width of the coherence or autocorrelation function as desired.

6 Representing the Diffractive Optical Element

Since we are interested only in the fractional fluctuation of the intensity and not its absolute value, there is no need to keep track of overall constant factors such as the $1/2\pi$ in the definition of the Fourier transform. Hence from now on, we drop all such overall constant factors.

The DOE will be taken to be a quasirandom binary (0 and π) phase grating. It can be represented as a binary transmission phase factor, i.e.,

$$T_{\text{DOE}}(\vec{\rho}) = +1 \text{ or } -1, \quad (26)$$

depending on position x, y . Here $\vec{\rho} = (x, y)$ represents a position in the various xy planes of the optical elements. Which plane, DOE, MFE, reticle, etc., should be clear from the context. The function $T_{\text{DOE}}(\vec{\rho})$ is designed to produce a particular pupil fill, and so the transmission factor can be represented directly in terms of this desired pupil fill as

$$|\tilde{T}_{\text{DOE}}(\vec{\beta})|^2 = \left| \int d^2 \rho T_{\text{DOE}}(\vec{\rho}) \exp(-i\vec{\beta} \cdot \vec{\rho}) \right|^2 = F(\vec{\beta})^2, \quad (27)$$

where $\vec{\beta} = (\beta_x, \beta_y)$ is spatial frequency in the xy plane and $F(\vec{\beta})^2$ is the desired source pupil fill intensity or at least a close approximation to the desired fill intensity. We assume that $\tilde{T}_{\text{DOE}}(\vec{\beta})$ can be written as $F(\vec{\beta}) \exp[i\varphi(\vec{\beta})]$ with $F(\vec{\beta})$ purely real and defined by the desired pupil fill, and $\varphi(\vec{\beta})$ being a rapidly varying function of $\vec{\beta}$ to account for the diffraction properties of the DOE. The rapid variation of $\varphi(\vec{\beta})$ implies that $\exp[i\varphi(\vec{\beta}) - i\varphi(\vec{\beta}')]]$ acts like a δ function when integrated with functions that are slowly varying compared to it, i.e., $\exp[i\varphi(\vec{\beta}) - i\varphi(\vec{\beta}')] \approx \delta(\vec{\beta} - \vec{\beta}')$. The fact that the DOE is a binary grating and that T_{DOE} is purely real means that the plus and minus diffraction orders from the DOE have the same amplitude but opposite phase, i.e.,

$$F(\vec{\beta}) \exp[i\varphi(\vec{\beta})] = F(-\vec{\beta}) \exp[-i\varphi(-\vec{\beta})]. \quad (28)$$

In general $F(\vec{\beta}) = F(-\vec{\beta})$ for most pupil fill geometries. On the other hand $\varphi(\vec{\beta}) = -\varphi(-\vec{\beta})$ means that φ at $+\vec{\beta}$ and $-\vec{\beta}$ are correlated and so we must augment the prior statistics for φ with

$$\exp[i\varphi(\beta) + i\varphi(\beta')] \approx \delta(\beta + \beta').$$

The extra contribution this makes, although present in principle, turns out to be negligible compared to regular terms, at least in the optical configuration considered here.

Finally note that convolving the DOE pattern $T(\vec{\rho})$ with itself yields the Fourier transform of the source pupil geometry, $S(\vec{R})$,

$$\begin{aligned}
 \int d^2 \rho T_{\text{DOE}}(\vec{\rho} + \vec{R}/2) T_{\text{DOE}}(\vec{\rho} - \vec{R}/2) \\
 = \int d^2 \beta d^2 \beta' F(\vec{\beta}) F(\vec{\beta}') \exp[-i\varphi(\vec{\beta}) + i\varphi(\vec{\beta}')] \\
 \times \exp[-i(\vec{\beta} + \vec{\beta}') \cdot \vec{R}/2] \delta(\vec{\beta} - \vec{\beta}') \\
 = \int d^2 \beta F(\vec{\beta})^2 \exp(-i\vec{\beta} \cdot \vec{R}) \equiv S(\vec{R}).
 \end{aligned}$$

7 Field at the Reticle (or Reticle Conjugate) Plane

It follows from the illuminator design that the field at the entrance plane of the MFE, ϕ_{MFE} , is the Fourier transform of the field at the exit plane of the DOE,

$$\phi_{\text{MFE}}(\vec{\rho}) = \int_{\text{DOE}} d^2 \rho'' \exp\left[-i \frac{k}{f_1} \vec{\rho} \cdot \vec{\rho}''\right] T_{\text{DOE}}(\vec{\rho}'') \exp[i\theta(\vec{\rho}'')]. \quad (29)$$

Here we have assumed that $A(\vec{\rho}, z) \approx 1$ across the entire area of the DOE, and we have dropped the z dependence of θ .

The field at the reticle (or reticle conjugate) plane ϕ_R is the superposition of low NA images of the field across the entrance plane of each lenslet of the MFE, appropriately magnified to cover the exposure slit area and with an angle of incidence at the reticle plane proportional to the transverse position in the MFE entrance plane, which we write as $\vec{\rho}_{\text{MFE}} = M \cdot \rho' + \vec{\rho}_\ell$. Here $\vec{\rho}'$ is the position in the reticle plane corresponding to $\vec{\rho}_{\text{MFE}}$. The term $M \cdot \rho' = M_x x' + M_y y'$ with M_x and M_y defined so that as $\vec{\rho}'$ ranges over the full reticle plane $M \cdot \rho'$ ranges only over the size of one lenslet. The lenslets are labeled by $\ell = (\ell_x, \ell_y)$, with ℓ_x and ℓ_y being integers labeling the lenslets in the MFE, and $\vec{\rho}_\ell$ is the position of the center of lenslet ℓ in the entrance plane of the MFE. To be specific, let $\ell_x = -N_x, -N_x + 1, \dots, -1, 0, 1, \dots, N_x - 1, N_x$ and $\ell_y = -N_y, -N_y + 1, \dots, -1, 0, 1, \dots, N_y - 1, N_y$, so that we have $2N_x + 1$ lenslets in the x direction and $2N_y + 1$ lenslets in the y direction. $\ell = (0, 0)$ is the center lenslet. Taking the lenslets to be rectangular with size H_x^{lenslet} and H_y^{lenslet} in the x and y directions, respectively, gives $x_\ell = \ell_x H_x$ and $y_\ell = \ell_y H_y$ with magnification factors

$$M_x = \frac{H_x^{\text{lenslet}}}{H_x^{\text{ref}}} = \frac{f_x^{\text{lenslet}}}{f_2} \ll 1, \quad M_y = \frac{H_y^{\text{lenslet}}}{H_y^{\text{ref}}} = \frac{f_y^{\text{lenslet}}}{f_2} \ll 1, \quad (30)$$

where H_x^{ret} and H_y^{ret} are the widths of the illuminated area of the reticle plane, and f_x^{lenslet} and f_y^{lenslet} are the focal lengths of the MFE lenslets in the x and y directions, respectively. The point spread function (PSF) in the reticle plane of the rectangular lenslets is

$$P_{\text{lenslet}}(x, y) = H_x^{\text{lenslet}} H_y^{\text{lenslet}} \frac{\sin\left(\frac{kH_x^{\text{lenslet}}x}{f_2}\right)}{\frac{kH_x^{\text{lenslet}}x}{f_2}} \frac{\sin\left(\frac{kH_y^{\text{lenslet}}y}{f_2}\right)}{\frac{kH_y^{\text{lenslet}}y}{f_2}}. \quad (31)$$

Using generic values for f_2 of several hundred mm, H_x^{lenslet} on the order of 1 mm with $H_x^{\text{lenslet}} \sim 2H_y^{\text{lenslet}}$ and $f_{x,y}^{\text{lenslet}}$ on the order of 2 mm with $2f_x^{\text{lenslet}} \sim f_y^{\text{lenslet}}$, we can easily get $H_x^{\text{ret}} \sim 100$ mm and $H_y^{\text{ret}} \sim 32$ mm, as required for slit illumination in the reticle (or $1 \times$ reticle conjugate plane) for a $4 \times$ reduction system. The width of lenslet PSF in the reticle plane $2\pi f_2 / kH_{x,y}^{\text{lenslet}}$ is then on the order of tens of microns in both the x and y directions.

Combining the previous factors yields

$$\phi_R(\vec{\rho}) = \sum_{\ell} \int d^2 \rho' P_{\text{lenslet}}(\vec{\rho} - \vec{\rho}') \exp \left[i \frac{k}{f_2} \vec{\rho} \cdot (M \cdot \rho' + \vec{\rho}_\ell) \right] \phi_{\text{MFE}}(M \cdot \rho' + \vec{\rho}_\ell), \quad (32)$$

where the factor

$$\exp \left[i \frac{k}{f_2} \vec{\rho} \cdot \vec{\rho}_{\text{MFE}} \right] = \exp \left[i \frac{k}{f_2} \vec{\rho} \cdot (M \cdot \rho' + \vec{\rho}_\ell) \right]$$

accounts for the angle of incidence and $\vec{\rho}$ is the position in the reticle plane.

For standard pupil fills $F(\vec{\beta})$, many lenslets, both on and off axis, are illuminated. The spatial scale of the speckle is set by the maximum value of $k(M \cdot \rho' + \vec{\rho}_\ell) / f_1$, and because many lenslets are illuminated, this is much shorter than the width of the lenslet PSF. Also, $M \cdot \rho' \ll \vec{\rho}_{\ell \text{max}}$, and so we can drop the $M \cdot \rho'$ terms relative to the $\vec{\rho}_\ell$ terms. Also, we take the pulse amplitude to be uniform, so that $A(\vec{\rho}) = 1$ inside the pulse and zero outside. Combining all this and letting $T_{\text{DOE}}(\vec{\rho})A(\vec{\rho}) = T(\vec{\rho})$, the field at the reticle (or reticle conjugate) is given by

$$\begin{aligned} \phi_R(\vec{\rho}) &= \sum_{\ell} \int d^2 \rho' P_{\text{lenslet}}(\vec{\rho} - \vec{\rho}') \exp \left[i \frac{k}{f_2} \vec{\rho} \cdot \vec{\rho}_\ell \right] \phi_{\text{MFE}}(\vec{\rho}_\ell) \\ &\approx \sum_{\ell} \exp \left[i \frac{k}{f_2} \vec{\rho} \cdot \vec{\rho}_\ell \right] \phi_{\text{MFE}}(\vec{\rho}_\ell) \int d^2 \rho' P_{\text{lenslet}}(\vec{\rho} - \vec{\rho}') \\ &\approx \sum_{\ell} \exp(ik\vec{\rho} \cdot \vec{\rho}_\ell / f_2) \int d^2 \rho' T(\vec{\rho}') \exp[i\theta(\vec{\rho}') - ik\vec{\rho}' \cdot \vec{\rho}_\ell / f_1]. \end{aligned} \quad (33)$$

The factor $\int d^2 \rho' P_{\text{lenslet}}(\vec{\rho} - \vec{\rho}')$ is a constant overall factor and so has been dropped.

8 Statistics of the Illumination Intensity: Residual Speckle

We now evaluate the mean and rms variation of the illumination intensity in the reticle plane. For the mean intensity, we evaluate the effect of the coherence floor, which, as expected, leads to a fixed nonchanging speckle pattern. For the rms variation, only the direct speckle averaging term is evaluated.

The average intensity is given by

$$\begin{aligned} \langle I_R(\vec{\rho}) \rangle &= \langle \phi_R(\vec{\rho})^* \phi_R(\vec{\rho}) \rangle \\ &= (1 - \Gamma_0) \sum_{\ell, m} \exp[-ik\vec{\rho} \cdot (\vec{\rho}_\ell - \vec{\rho}_m) / f_2] \\ &\quad \times \int d^2 \eta_1 d^2 \eta_2 T(\vec{\eta}_1)^* T(\vec{\eta}_2) \exp(ik\vec{\eta}_1 \cdot \vec{\rho}_\ell / f_1 - ik\vec{\eta}_2 \cdot \vec{\rho}_m / f_1) G(\vec{\eta}_1 - \vec{\eta}_2, \vec{\xi}) \\ &\quad + \Gamma_0 \sum_{\ell, m} \exp[-ik\vec{\rho} \cdot (\vec{\rho}_\ell - \vec{\rho}_m) / f_2] \\ &\quad \times \int d^2 \eta_1 d^2 \eta_2 T(\vec{\eta}_1)^* T(\vec{\eta}_2) \\ &\quad \times \exp(ik\vec{\eta}_1 \cdot \vec{\rho}_\ell / f_1 - ik\vec{\eta}_2 \cdot \vec{\rho}_m / f_1), \end{aligned} \quad (34)$$

after substituting

$$\begin{aligned} \langle \exp[-i\theta(\vec{\eta}_1) + i\theta(\vec{\eta}_2)] \rangle &= \Gamma(\vec{\eta}_1 - \vec{\eta}_2, \vec{\xi}) \\ &= (1 - \Gamma_0) G(\vec{R}, \vec{\xi}) + \Gamma_0. \end{aligned}$$

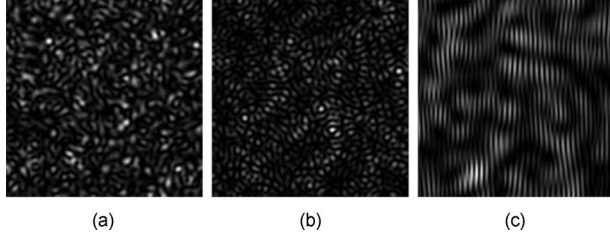


Fig. 6 Speckle patterns generated by (a) top-hat, (b) annular, (c) and x -dipole fill geometries. The top-hat case shows speckles that vary in size as well as orientation, corresponding to the range of angles and spatial frequencies in the top hat. The annular case shows random orientation but essentially fixed size, since the annulus corresponds to a fixed radial spatial frequency with a complete range of angles. The high frequency ripples in the dipole case come from the interference between the dipoles and the large scale structure from the small size of each dipole.

First, evaluate the second term, which comes from the coherence floor, by substituting the Fourier representation of $T(\vec{\rho}) = T_{\text{DOE}}(\vec{\rho})A(\vec{\rho})$, which is the convolution of $F(\vec{\beta})\exp[i\varphi(\vec{\beta})]$ with $\tilde{A}(\vec{\beta})$. But assuming $A(\vec{\rho})$ covers a large area $\tilde{A}(\vec{\beta}) \sim \delta(\vec{\beta})$, and so the Fourier transform of $T(\vec{\rho})$ is essentially just the Fourier transform of $T_{\text{DOE}}(\vec{\rho})$. Then using the standard Fourier definition of the Dirac delta function, and as usual ignoring overall factors of 2π , etc., we get

$$\begin{aligned}
 & \Gamma_0 \sum_{\ell,m} \exp[-ik\vec{\rho} \cdot (\vec{\rho}_\ell - \vec{\rho}_m)/f_2] \\
 & \quad - \vec{\rho}_m/f_2] \int d^2\eta_1 d^2\eta_2 T(\vec{\eta}_1)^* T(\vec{\eta}_2) \exp(ik\vec{\eta}_1 \cdot \vec{\rho}_\ell/f_1 \\
 & \quad - ik\vec{\eta}_2 \cdot \vec{\rho}_m/f_1) = \Gamma_0 \sum_{\ell,m} \exp[-ik\vec{\rho} \cdot (\vec{\rho}_\ell - \vec{\rho}_m)/f_2] \\
 & \quad \times \int d^2\beta_1 d^2\beta_2 F(\vec{\beta}_1) F(\vec{\beta}_2) \exp[-i\varphi(\vec{\beta}_1) + i\varphi(\vec{\beta}_2)] \\
 & \quad \times \int d^2\eta_1 d^2\eta_2 \exp(ik\vec{\eta}_1 \cdot \vec{\rho}_\ell/f_1 - ik\vec{\eta}_2 \cdot \vec{\rho}_m/f_1 - i\vec{\beta}_1 \cdot \vec{\eta}_1 \\
 & \quad + i\vec{\beta}_2 \cdot \vec{\eta}_2) = \Gamma_0 \left| \sum_{\ell} F(k\vec{\rho}_\ell/f_1) \right. \\
 & \quad \times \exp[-ik\vec{\rho} \cdot \vec{\rho}_\ell/f_2 - i\varphi(k\vec{\rho}_\ell/f_1)] \left. \right|^2 \equiv \Gamma_0 \left| \sum_{\ell} F_{\ell} \right. \\
 & \quad \times \exp(-ik\vec{\rho} \cdot \vec{\rho}_\ell/f_2 - i\varphi_{\ell}) \left. \right|^2. \tag{35}
 \end{aligned}$$

In the last step we have made an obvious simplification in the notation. With φ_{ℓ} effectively a random phase term coming from the DOE, this corresponds to a single static, i.e., time independent nonaveraging speckle pattern. Noting that F_{ℓ} corresponds to the fill geometry and that F_{ℓ} is 1 for a lenslet that is illuminated and 0 for a lenslet that is not, we see that the speckle pattern depends directly on the fill geometry as expected. Example patterns are shown in Refs. 3 and 4 and experimental results are given in Ref. 6. Figure 6 shows examples of the speckle patterns created by top-hat, annular, and x -dipole pupil fill.

Now evaluate the first, term by substituting the Fourier representations of $T(\vec{\rho})$ and $G(\vec{R}, \vec{\xi})$ and again using the Fourier definition of the Dirac delta function, we get

$$\begin{aligned}
 & (1 - \Gamma_0) \sum_{\ell,m} \exp[-ik\vec{\rho} \cdot (\vec{\rho}_\ell - \vec{\rho}_m)/f_2] \\
 & \quad \times \int d^2\eta_1 d^2\eta_2 T(\vec{\eta}_1)^* T(\vec{\eta}_2) \exp(ik\vec{\eta}_1 \cdot \vec{\rho}_\ell/f_1 \\
 & \quad - ik\vec{\eta}_2 \cdot \vec{\rho}_m/f_1) G(\vec{\eta}_1 - \vec{\eta}_2, \vec{\xi}) = (1 - \Gamma_0) \\
 & \quad \times \sum_{\ell,m} \exp[-ik\vec{\rho} \cdot (\vec{\rho}_\ell - \vec{\rho}_m)/f_2] \\
 & \quad \times \int d^2\beta_1 d^2\beta_2 d^2\beta F(\vec{\beta}_1) F(\vec{\beta}_2) \\
 & \quad \times \exp[-i\varphi(\vec{\beta}_1) + i\varphi(\vec{\beta}_2)] \tilde{G}(\vec{\beta}, \vec{\xi}) \\
 & \quad \times \int d^2\eta_1 d^2\eta_2 \exp[ik\vec{\eta}_1 \cdot \vec{\rho}_\ell/f_1 - ik\vec{\eta}_2 \cdot \vec{\rho}_m/f_1 - i\vec{\beta}_1 \cdot \vec{\eta}_1 \\
 & \quad + i\vec{\beta}_2 \cdot \vec{\eta}_2 + i\vec{\beta} \cdot (\vec{\eta}_1 - \vec{\eta}_2)] = (1 - \Gamma_0) \\
 & \quad \times \int d^2\beta \left| \sum_{\ell} \exp(-ik\vec{\rho} \cdot \vec{\rho}_\ell/f_2) F(k\vec{\rho}_\ell/f_1 + \vec{\beta}) \right. \\
 & \quad \times \exp[-i\varphi(k\vec{\rho}_\ell/f_1 + \vec{\beta})] \left. \right|^2 \tilde{G}(\vec{\beta}, \vec{\xi}). \tag{36}
 \end{aligned}$$

The last line shows that this term produces a superposition of speckle patterns with the consequent averaging that reduces the contrast to below the 100% value, which holds for a single speckle pattern. We can estimate the amount of averaging by noting first that the transform of $T(\vec{\rho}) = T_{\text{DOE}}(\vec{\rho})A(\vec{\rho})$ is the convolution of $F(\vec{\beta})\exp[i\varphi(\vec{\beta})]$ with the Fourier transform of $A(\vec{\rho})$, which smooths the nominally rapidly varying DOE phase $\varphi(\vec{\beta})$, so that it varies on a scale of $1/A_{\text{beam}}$, which is the scale of the width of the Fourier transform of $A(\vec{\rho})$. The Fourier transform of $G(\vec{R}, \vec{\xi})$ given by $\tilde{G}(\vec{\beta}, \vec{\xi})$ varies on a scale of $1/(2\xi_x 2\xi_y) \gg 1/A_{\text{beam}}$, where factors of 2 are to account for the fact that the correlation lengths correspond to roughly the half-width of $G(\vec{R}, \vec{\xi})$, and we want the full width. The number of independent speckle patterns N can be estimated as the ratio of these two scales with the result $N \approx (1/(4\xi_x \xi_y))/(1/A_{\text{beam}}) = A_{\text{beam}}/(4\xi_x \xi_y)$. To verify this, we also use a more direct approach to estimate the residual speckle.

Finally, note that the previous result shows that the speckle in the reticle plane in this system is periodic in the x and y with period $P_x = f_2 \lambda / H_x^{\text{lenslet}}$ and $P_y = f_2 \lambda / H_y^{\text{lenslet}}$, respectively. This follows directly from the fact that the speckle pattern depends spatially on $\exp[-ik\vec{\rho} \cdot \vec{\rho}_\ell/f_2] = \exp[-i2\pi(x\ell_x H_x^{\text{lenslet}}/\lambda f_2 + y\ell_y H_y^{\text{lenslet}}/\lambda f_2)]$, which is periodic for $x \rightarrow x + n\lambda f_2 / H_x^{\text{lenslet}}$ and $y \rightarrow y + m\lambda f_2 / H_y^{\text{lenslet}}$ with n and m integers.

To estimate the number speckle patterns of N directly, we use the following approach. Instead of treating $\theta(\vec{\rho})$ as a smoothly varying function, treat it as constant over trans-

verse areas corresponding to the transverse coherence, and over longitudinal positions corresponding to the longitudinal coherence. This is a standard approach for turning a continuum statistics problem into a discrete problem.^{1,8–10} The transverse coherence area is on the order of $a \approx (2\xi_x) \times (2\xi_y) = 4\xi_x\xi_y$. We label each coherence area or patch with $c = 1, 2, \dots, N$, where N is the number of coherence areas in the beam $A_{\text{beam}}/a = N$, and we take θ_c to be uncorrelated from area to area, $\langle \exp(i\theta_c - i\theta_{c'}) \rangle = \delta_{c,c'}$. Using this approach, the field at the reticle takes the form

$$\begin{aligned} \phi_R(\vec{\rho}) &\approx \sum_c \exp(ik\vec{\rho} \cdot \vec{\rho}_c/f_2) \sum_c \exp(i\theta_c) \\ &\quad \times \int_c d^2\rho' T(\vec{\rho}') \exp(-ik\vec{\rho}' \cdot \vec{\rho}_c/f_1) \\ &\approx \sum_c \exp(ik\vec{\rho} \cdot \vec{\rho}_c/f_2) \sum_c \exp(i\theta_c) F(k\vec{\rho}_c/f_1) \\ &\quad \times \exp[i\varphi_c(k\vec{\rho}_c/f_1)], \end{aligned} \quad (37)$$

where the notation \int_c indicates integration over the area of coherence patch c . In the second line we have made the crucial assumption that DOE has been designed so that each coherence patch on its own produces the full fill geometry $F(\vec{\beta})$ but with of course a different phase distribution $\varphi_c(\vec{\beta})$. The average intensity is given by

$$\begin{aligned} \langle I_R(\vec{\rho}) \rangle &\approx \sum_c \left| \sum_c F(k\vec{\rho}_c/f_1) \exp[-ik\vec{\rho} \cdot \vec{\rho}_c/f_2] \right. \\ &\quad \left. - i\varphi_c(k\vec{\rho}_c/f_1) \right|^2. \end{aligned} \quad (38)$$

Under the assumption that the phase distributions φ_c are uncorrelated from each coherence patch c , this is a sum of N speckle patterns, and so the residual speckle is on the order of $100\%/\sqrt{N}$, where N is the number of spatially incoherent patches or transverse areas in the beam. Note that if the DOE is designed so that the φ_c are correlated, then the speckle averaging would be less. For example, if all the phases were exactly the same, then there would be no averaging at all and the speckle contrast would remain 100%.

9 Discussion and Conclusions

The results show that in the limit where each pulse is many longitudinal coherence lengths long (and σ is large enough so that the coherence floor $\Gamma_0 \approx 0$), then the transverse spatial coherence controls the amount of averaging, with the result that for a single pulse the residual speckle or contrast is on the order of

$$1/\sqrt{N} = \sqrt{\frac{(2\xi_x)(2\xi_y)}{A_{\text{beam}}}}. \quad (39)$$

For example, if ξ_x and ξ_y are both on the order of $200 \mu\text{m}$ and the beam area is on the order of $20 \times 20 \text{ mm} = 400 \text{ mm}^2$, then the expected contrast or rms value for a single pulse is on the order of 2%. Of course for a smaller beam area and/or larger coherence lengths, the value will be larger. Since the light is propagating through the same

DOE with the same phase $\varphi(\vec{\rho})$, the expectation is that in the nonscanned case the residual speckle pattern will be very similar from pulse to pulse, and hence for static, i.e., nonscanned exposures, increasing the number of pulses will not improve the contrast. Of course this needs to be verified by more detailed calculations, but it does seem to follow from the experimental results.^{5,6} In the scanned case the speckle pattern is shifting from pulse to pulse nominally by distances much greater than the transverse coherence length, and so the patterns should average with the result that when scanned, the contrast should be $1/\sqrt{N_{\text{pulses}}}$ times the nonscanned result, where N_{pulses} is the number of pulses used to expose. On the other hand, as shown before, at least for the illuminator considered here, the speckle pattern is periodic with a period $\lambda f_2/H^{\text{lenslet}}$. For f_2 on the order of 200 mm and H^{lenslet} on the order of 1 mm, it follows that the periodicity of the speckle pattern is on the order of $200\lambda \sim 40\text{-}\mu\text{m}$ reticle scale or $10\text{-}\mu\text{m}$ wafer scale for $\lambda = 193 \text{ nm}$. For a wafer stage scanning at 100 mm/sec and a laser replate of 5000 Hz, the stage moves on the order of $20 \mu\text{m}$ per pulse, and so there is a possibility, although highly unlikely, that the stage would move an integer number of periods and so decrease the effect of the pulse averaging due to scanning. Overall we expect scanning to reduce the nonscanned speckle by a factor of $1/\sqrt{N_{\text{pulses}}}$, which for 30 pulses and 2% residual speckle gives about 0.4% scanned residual speckle.

The remaining question is how does the residual speckle affect the features printed into photoresist. After all, several percent residual speckle 3σ obviously far exceeds the specification on dose uniformity, and one would expect that this would cause unacceptable feature size variation that would show up as LER or at least an enhancement to LER. We now briefly explain that in spite of the expectation that several percent 3σ dose variation is egregious, it is in fact a small contributor to LER when compared to the main cause, which is, to use the term loosely, acid shot noise. (Details of this argument will be presented in a later publication.) The work done over the past several years to elucidate and explain the origin of LER^{11–15} shows that the dominant cause comes from the fact that chemically amplified photoresist captures the continuous smoothly varying image pattern as a discrete distribution of acids, with the probability of the number of acids released in a given small volume of resist being proportional to the dose in that volume. This discrete distribution of acids is effectively a random pattern of dots and so has essentially 100% contrast, i.e., either an acid is released at a given position or it is not, even if the image is perfect and speckle-free. Hence the only effect of the speckle is to slightly increase or decrease the probability of acid release. This is the reason speckle does not add much to LER as seen experimentally.⁶ This also explains why several percent residual speckle has not become a major issue for chip manufacturers. The same argument, i.e., that speckle is a small effect compared to acid statistics, holds as well for contact holes.

Finally, it is worth noting that due to its wide bandwidth, EUV does not have this type of speckle problem. But it does have an issue with mask roughness generated speckle.¹⁶

Appendix

Consider the case where $\langle \theta(\vec{r})\theta(\vec{r}') \rangle$ is spherically symmetric with correlation length ξ and constant rms value $\langle \theta(\vec{r})^2 \rangle = \sigma^2$. Sample $\theta(\vec{r})$ at positions \vec{r}_n with $n = 1, 2, \dots, N$ and $|\vec{r}_n - \vec{r}_m| \gg \xi$ for all n and m . Then the Gaussian statistics reduce to

$$P[\theta] \Rightarrow P(\theta_1, \theta_2, \dots, \theta_N) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{p=1}^N \theta_p^2\right],$$

where each θ_n ranges from $-\infty$ to $+\infty$. This probability distribution gives

$$\begin{aligned} \langle \exp[i\theta_n] \rangle &= \frac{1}{(2\pi\sigma^2)^{N/2}} \int d^N \theta \exp\left[i\theta_n - \frac{1}{2\sigma^2} \sum_{p=1}^N \theta_p^2\right] \\ &= \exp[-\sigma^2/2] \end{aligned}$$

$$\begin{aligned} \langle \exp[i\theta_n - i\theta_m] \rangle &= \frac{1}{(2\pi\sigma^2)^{N/2}} \int d^N \theta \\ &\quad \times \exp\left[i\theta_n - i\theta_m - \frac{1}{2\sigma^2} \sum_{p=1}^N \theta_p^2\right] \\ &= \delta_{nm} + \exp[-2\sigma^2/2](1 - \delta_{nm}) \end{aligned}$$

$$\begin{aligned} \langle \exp[i\theta_n + i\theta_m] \rangle &= \frac{1}{(2\pi\sigma^2)^{N/2}} \int d^N \theta \exp\left[i\theta_n + i\theta_m - \frac{1}{2\sigma^2} \sum_{p=1}^N \theta_p^2\right] \\ &= \exp[-2\sigma^2/2], \end{aligned}$$

etc. In the limit as $\sigma \rightarrow \infty$, this reduces to

$$\langle \exp[i\theta_n] \rangle \rightarrow 0,$$

$$\langle \exp[i\theta_n - i\theta_n] \rangle \rightarrow \delta_{nm},$$

$$\langle \exp[i\theta_n + i\theta_m] \rangle \rightarrow 0,$$

etc., which matches the results for the expectation values obtained using the uniform distribution $P_{\text{uniform}}(\theta_1, \theta_2, \dots, \theta_N) = 1/(2\pi)^N$ with each θ_n , in this case ranging only from 0 to 2π .

References

1. R. Loudon, *The Quantum Theory of Light*, 2nd ed., Oxford University Press, Oxford, UK (1983).
2. J. Goodman, *Speckle Phenomena in Optics*, 1st ed., Roberts and Co. Publishers, Greenwood Village, CO (2006).
3. C. Rydberg, J. Bengtsson, and T. Sandström, "Dynamic laser speckle as a detrimental phenomenon in optical projection lithography," *J. Microlithogr., Microfabr., Microsyst.* **5**, 033004 (2006).
4. C. Rydberg and J. Bengtsson, "Performance of diffractive optical elements for homogenizing partially coherent light," *J. Opt. Soc. Am. A* **24**, 3069–3079 (2007).
5. O. Kritisun, I. Lalovic, S. Rokitski, B. Partlo, B. La Fontaine, N. Farrar, and H. Levinson, "Improving lithography pattern fidelity and line-edge roughness by reducing laser speckle," *J. Vac. Sci. Technol. B* **26**, 2145–2149 (2008).
6. O. Noordman, A. Tychkov, J. Boselmans, J. Tsacoyeanes, G. Politi, M. Patra, V. Blahnik, and M. Menl, "Speckle in optical lithography and the influence on line width roughness," *Proc. SPIE* **7274**, 72741R (2009).
7. A. Zee, *Quantum Field Theory in a Nutshell*, Princeton University Press, Boston (2003).
8. Y. Lin and J. Buck, "Numerical modeling of the excimer beam," *Proc. SPIE* **3677**, 700–711 (1999).
9. Y. Lin, G. N. Lawrence, and J. Buck, "Characterization of excimer lasers for application to lenslet array homogenizers," *Appl. Opt.* **40**, 1931–1941 (2003).
10. J. Goodman, "Speckle with a finite number of steps," *Appl. Opt.* **47**, A111–118 (2008).
11. R. Brainard, P. Treforas, J. H. Lammers, C. A. Catler, J. F. Mackevich, A. Trefonas, and S. A. Robertson, "Shot noise, LER and quantum efficiency of EUV photoresists," *Proc. SPIE* **5374**, 74–85 (2004).
12. G. Gallatin, "Resist blur and line edge roughness," *Proc. SPIE* **5754**, 38–52 (2005).
13. R. Bristol, "The tri-lateral challenge of resolution, photospeed and LER: scaling below 50 nm?" *Proc. SPIE* **6519**, 65190W (2007).
14. J. Biafore, M. D. Smith, C. A. Mack, J. W. Thackeray, R. Gronheid, S. A. Robertson, T. Graves, and D. Blankenship, "Statistical simulation of photoresists at EUV and ARF wavelengths," *Proc. SPIE* **7273**, 727343 (2009).
15. C. Mack, "Stochastic approach to modeling line edge roughness in photoresist," *Proc. SPIE* **7273**, 727326 (2009).
16. P. Naulleau, C. N. Anderson, L.-M. Baclea-an, P. Denham, S. George, K. A. Goldberg, M. Goldstein, B. Hoef, R. Nualyma, G. Jones, C. Koh, B. LaFontaine, B. McClinton, R. Miyakawa, W. Montgomery, J. Roller, T. Wallow, and S. Wurm, "The SEMATECH Berkeley microfield exposure tool: learning at the 22 nm node and beyond," *Proc. SPIE* **7271**, 72710W (2009).

Biographies and photographs of the authors not available.