Linewidth determination using simulated annealing

Katherine Creath
Optical Sciences Center, University of Arizona
Tucson, AZ 85721

and

Shiyu Zhang
Optical Sciences Center, University of Arizona
Tucson, AZ 85721
and
WYKO Corporation, 2650 E. Elvira Rd.
Tucson, AZ 85706

ABSTRACT

Features smaller than the coherent point spread function of an optical microscope can be detected by looking at the phase of an image rather than simply looking at the reflected intensity. To get an accurate determination of feature size, object reconstruction techniques need to be applied to remove the effects of the optical system. This paper discusses the use of simulated annealing to determine the width and average height of an object feature as well as the defocus and spherical aberration present in the imaging system. Results show that errors of less than 5% are obtained for features as small as 1/6th of the optical resolution of the system and less than 2% for features as small as 1/2 of the resolution. When noise is included in the modeling, errors of 5% are obtained for S/N of 10 and less than 2% for S/N of 100.

1. INTRODUCTION

The measurement of small object features depends upon the optical system used to measure the feature. To get an accurate determination of feature size, object reconstruction techniques need to be applied to remove the effects of the optical system. Because of the presence of noise in the measurement, object reconstruction techniques are limited in their accuracy and are used to estimate the true size of a feature as well as other pertinent information which may be desired. In the case of linewidth features created with microlithography, these techniques enable estimation of width and height of the feature as well as defocus and spherical aberration of the optical system. Object reconstruction techniques include deconvolution, least-squares estimation, maximum-likelihood estimation, and simulated annealing.3 For these techniques, the optical system has to be well known so that the point-spread function can be either determined analytically or experimentally. This paper investigates the use of the simulated annealing technique to determine object parameters and aberrations of the imaging system.

2. MICROSCOPE IMAGING SYSTEM MODEL

The microscope which will ultimately be used in this work is a 200x Linnik interference microscope with a numerical aperture (NA) of 0.95.4 The illumination is provided by a white light source with a narrowband interference filter. Because of the narrowband illumination and large angles in object space, this optical system is partially coherent. To accurately model the system, vector diffraction theory and the wavelength spread of the source need to be accounted for. However, for the purposes of determining the feasibility of using various object reconstruction techniques, we have chosen to begin our studies with a simple model of the optical system.

Since the optical system is much closer to being coherent than incoherent, this study assumes that the optical system is coherent and that linear system theory can be used for the model.5 These assumptions would be valid for NAs of less than 0.5 and narrowband filters where Δλ << λ. Another reason for using a simple model is that we wish to include the effects of aberrations in the model so that they may also be determined. In normal use, it is almost impossible to guarantee that the operator has obtained the best focus (even if it is automatically focused) and that the microscope is used at the ideal conjugates. Thus, defocus and spherical aberration play an integral part in the formation of the image being evaluated to determine critical dimensions.

To reduce computation times, this model is assumed to be 2-dimensional rather than 3-dimensional. Because of symmetry, the third dimension will not provide much additional information. We are also assuming that the impulse...
response (point-spread function PSF) of the optical system is isoplanatic so that the PSF is independent of the object point under consideration. This means that there can be no off-axis aberrations present such as coma or astigmatism. This assumption is valid for microscopes where the field of view is small. Most microscope objectives are designed to be aberration-free for the design wavelengths; however, using them at the wrong conjugates can introduce spherical and chromatic aberrations.

For a coherent optical system, the complex amplitude of the image is a convolution of the complex amplitude of the object with the impulse response or PSF of the system:

$$ g(x) = f(x) \ast h(x) , $$

where $f(x)$ and $g(x)$ are the object and image complex amplitude distributions respectively, $h(x)$ is the point spread function of the system, and $\ast$ denotes convolution. In the frequency domain, the above equation can be written as

$$ G(f_x) = F(f_x)H(f_x) , $$

where $F(f_x)$ and $G(f_x)$ are the Fourier transforms of $f(x)$ and $g(x)$, and $H(f_x)$ is the coherent transfer function of the optical system. The transfer function is related to the pupil function by another Fourier transform relationship

$$ H(f_x) = \mathfrak{F}\{P(x)\} , $$

where $\mathfrak{F}\{\cdot\}$ denotes a forward Fourier transform. The pupil of the system is assumed to transmit 100% of the illumination inside an aperture of diameter $D$, and none outside. Aberration terms are introduced as phase variations across the pupil. For this microscope, the pupil function can be written as

$$ P(x) = \begin{cases} i\left(\frac{2\pi}{\lambda}\right)[W_{020}(2x/D)^2 + W_{040}(2x/D)^4] & |x| \leq D/2 \\ 0 & |x| > D/2 \end{cases} $$

where $\lambda$ is the illumination wavelength, $W_{020}$ is the defocus coefficient, and $W_{040}$ is the spherical aberration coefficient. The transfer function then becomes

$$ H(f_x) = \begin{cases} i\left(\frac{2\pi}{\lambda}\right)[W_{020}(f_x/f_c)^2 + W_{040}(f_x/f_c)^4] & |f_x| \leq f_c \\ 0 & |f_x| > f_c \end{cases} $$

where the cutoff frequency of the system is defined as

$$ f_c = \frac{NA}{\lambda} . $$

$NA$ is the numerical aperture of the optical system. The cutoff frequency is an indication of the optical resolution of the system.

Since we are interested in the specific application of line width and height determination, the object used for this study is composed of a single feature of reflectivity $r_a$ on a substrate with reflectivity $r_B$. These reflectivities are complex quantities. The object function can be written as
where \( L \) is the width of the entire object, \( \ell \) is the width of the feature, and the phase \( \phi_o(x) \) of the object is related to the surface height profile \( Z_o(x) \) by

\[
\phi_o(x) = \frac{4\pi Z_o(x)}{\lambda}.
\]

To find the complex amplitude of the image \( g(x) \), the spectrum of the object \( G(f_x) \) is multiplied by the transfer function \( H(f_x) \) and an inverse Fourier transform is performed. Because the microscope we are using determines surface profiles from phase data, we need to determine the surface height data from the complex amplitude distribution of the image. The surface profile \( Z_i(x) \) is related to the phase by

\[
z_i(x) = \frac{\lambda \phi_i(x)}{4\pi},
\]

where

\[
\phi_i(x) = \tan^{-1}\left( \frac{\Im[g(x)]}{\Re[g(x)]} \right).
\]

Note that the phase distribution is only determined modulo \( 2\pi \) so the phase must be unwrapped to remove discontinuities to provide an accurate height profile.

### 3. EFFECT OF ABERRATIONS ON IMAGES

Figures 1 through 5 show height profiles of images simulated for an object feature 0.5 \( \mu \)m wide and 0.2\( \lambda \) high. The feature and substrate both have 100% reflectivities for these calculations. The profiles for a number of different values for defocus and spherical aberration are shown. As the amount of either aberration changes, the shape of the image profile changes a lot. Because the feature has height and not just amplitude variations, the signs of the aberrations make a difference. In general, the best image with spherical aberration present is not the same as the best focus if there were no spherical aberration. The minimum blur size is obtained when defocus coefficient is \(-2/3\) the value of the spherical aberration coefficient. Since it is not always possible to take data at the optimum focus, the variations in the image profiles will lead to variations in the measured width and height of a feature. The width of the feature as determined as the full-width half-maximum (FWHM) value is plotted versus spherical aberration in Figure 6 for a number of different defocus values. There clearly are preferred values for the aberrations. The error with a small amount of spherical aberration (1.5\( \lambda \)) can be as much as 90%. Clearly aberrations need to be accounted for when determining feature dimensions.

### 4. SIMULATED ANNEALING ALGORITHM

Many methods have been developed to extract information about an object from an unresolved image.\(^1-3\) A lot of these algorithms assume a linear dependence between the object and image. Most of them require a lot of computation. Since we know the basic shape of the object features, it is much easier to extract certain parameters describing these features than reconstructing the entire object. One drawback to object reconstruction techniques is that the optical system needs to be accurately modeled so that the transfer function or PSF is known. The simulations for this paper provide an illustration of how well the simulated annealing technique will work. For use with real data, the microscope system needs to be modeled more accurately than the simple model we presented above.
Figure 1. Image profiles for varying amounts of defocus with $1.5\lambda$ spherical aberration.

Figure 2. Image profiles for varying amounts of defocus with no spherical aberration.

Figure 3. Image profiles for varying amounts of defocus with $-1.5\lambda$ spherical aberration.

Figure 4. Image profiles for varying amounts of spherical aberration with no defocus.

Figure 5. Image profiles for varying amounts of spherical aberration with $1\lambda$ defocus.

Figure 6. Calculated linewidth versus spherical aberrations for varying amounts of defocus.
If only one or two parameters about the object need to be determined, the easiest way is the brute force approach where the entire solution space is sampled and searched. This is the fastest way to determine quantities such as the linewidth and the feature height. However, because of the variation in the image with spherical aberration and defocus present and our inability to be able to determine them accurately for each measurement, there really are four parameters rather than two which we must determine. These four are linewidth, height, defocus, and spherical aberration. The advantage of the object reconstruction techniques are that the number of parameters can be easily varied and that system aberrations don't need to be known because they can be determined along with the object parameters.

The simulated annealing algorithm is a heuristic technique for global optimization of large scale problems that uses methods from statistical mechanics. The optimization problem is visualized as a physical system undergoing an annealing schedule. In the beginning the system is at a high temperature. The temperature is lowered gradually until the system settles down to a globally minimum state.

If the entire object is to be reconstructed, the energy of the system (also known as the cost function) is defined as

$$E = \|g'(x) - f(x) \ast h(x)\|,$$

(11)

where $\|\cdot\|$ denotes the norm of the function. It is the difference between the measured image $g'(x)$ and a calculated image using the parameters which will be varied randomly and the simple system model explained in the previous section. Because the measured data for this study will be surface heights determined from phase data, the cost function to reconstruct the object would be defined as

$$E = \|z_{meas}(x) - z_1(x)\|,$$

(12)

where $z_{meas}(x)$ is the measured phase data and $z_1(x)$ is defined by Eq. (9). When parameters are of interest rather the reconstruct the entire image, the cost function is constructed as a sum of the squares of the differences between the measured $q_{meas}$ and estimated parameters $q_i$.

$$E = \sum_{j=1}^{n} (q_{meas} - q_i)^2$$

(13)

The starting values for the parameters linewidth $\ell$, average height $Z_0$, defocus $W_{020}$, and spherical aberration $W_{040}$ are chosen randomly within a preset allowable range for each parameter which is known a priori. Subsequent iterations involve randomly choosing new values for each of the parameters and recalculating the cost function. If the new values cause the energy to be reduced, they are accepted. To ensure that the solution does not get trapped in a local minimum, values which increase the energy of the system are accepted with a probability given by the Boltzmann law

$$P(accept) = e^{-E/T},$$

(14)

where $T$ is the effective temperature of the system, and the Boltzmann constant $k$ is assumed to be unity. Accepting some changes which increase the energy of the system will move the solution out of a local minimum so that the global minimum may be found. When $T$ is high, the probability to accept a higher energy is large and as $T$ decreases, the probability decreases. Only small changes which increase the energy will be accepted at low temperatures. This process continues until the energy level is small enough to be acceptable. If there is no noise present, the energy will be reduced to zero for an exact estimation.

For this simulation, initial values for the parameters are chosen randomly to calculate the original energy value. Then the measured data are assumed to be a random starting point. The initial temperature of the system is set so that $90\%$ of the energy increases will be accepted,

$$P(accept) = e^{-E/T_{init}} = 0.9$$

(15)
This corresponds to an initial temperature $T_{\text{init}} \equiv 10E$. At each temperature an arbitrary number of iterations $N_T$ are done. (50 iterations at each temperature were found to be sufficient for convergence.) Then the temperature is lowered and the process repeated. The temperature is lowered by a factor $\alpha$ so that

$$T_{n+1} = \alpha T_n .$$

(16)

To reduce the temperature gradually, $\alpha = 0.9$ is used to reduce the temperature 10% between cycles. At the same time the temperature is lowered, the ranges for the parameter values are also decreased slowly. The values of the parameters which produced the lowest energy value for the $N_T$ iterations at the current temperature are used as the starting point for the next temperature.

When the energy becomes less than 10% of the temperature, the temperature will be reset to ten times the energy and a new cycle will be started. This will speed up the process when the system is converging quickly. On the other hand, if the energy is larger than the temperature, it will not be reduced for the next cycle. This is likely to be the case with noisy data. If there is still no improvement in the energy after another few cycles, the variable range is reduced slightly for the next cycle. This will aid in speeding up convergence. The iterating process is stopped if the energy is very low, if the variable range is very small, if the temperature range is small, or if a certain maximum number of iterations have been performed.

5. RESULTS

The simulations for this study were performed on an HP workstation using the program PVWave. No effort has been made to optimize and speed up the calculations. The number of iterations required for the simulated annealing technique to converge depends upon the number of parameters. Figure 7 shows a plot of the percent error in each of the 4 parameters versus the number of iterations. This example assumes $\ell = 0.5$, $Z_0 = 0.2\lambda$, $W_{020} = 0.2\lambda$, and $W_{040} = 0.5\lambda$. After 20,000 iterations, the errors in each of the parameters are less than 5%, and after 25,000 iterations, the errors are within 2%.

Figure 8 shows a plot of percent error in each of the 4 parameters versus the actual linewidth. The values of height, defocus, and spherical aberration are the same as for Fig. 7. Each point represents the results of a single run. Subsequent runs for the same parameter values may show more or less error. The optical resolution of the microscope system is on the order of a third of a micrometer. Linewidths as small as 1/6 of the optical resolution can easily be determined using simulated annealing even in the presence of aberrations.

Since no imaging system produces noiseless data, a number of runs were also made by adding Gaussian noise to the simulated measured image. The variance of the Gaussian noise is assumed to be $\sigma^2 = \frac{Z_{\text{max}}}{S/N}$, where $Z_{\text{max}}$ is the maximum of $Z(x)$ and $S/N$ is the signal-to-noise ratio. For noisy data, assuming that the object is located in the center of the image, the algorithm can be improved by weighting the center of the image more than the edge.

A sample image with a signal-to-noise ratio of 10 is shown in Fig. 9. This assumes the same parameter values as Fig. 7. As the signal-to-noise ratio decrease, the error in the estimated values of the parameters increases as shown in Fig. 10. A signal-to-noise ratio of 10 is necessary to provide reasonable results and a signal-to-noise ratio of 100 provides very good results with the simulated annealing technique.

6. CONCLUSIONS

The object linewidth and average height can easily be determined using the simulated annealing technique even in the presence of noise. Because the image does not change as much when defocus and spherical aberration are changed, the aberration coefficients can not be determined as accurately as the width and height. This is fine since we really are only interested in the object parameters. The aberration parameters are used to aid us in obtaining the best estimate of the object parameters. It is possible to determine a linewidth as small as 0.05 µm with a system resolution of 0.34 µm to an accuracy of 5%. The ultimate limit on how small of a feature we can detect depends upon the number of pixels across the feature. At this point, object heights are limited to less than 0.25µm because of the limitations imposed on calculating surface profiles.
using an arctangent function. Measurements at more than one wavelength would be needed to get around this. The biggest drawback to this technique is that it takes 25,000 iterations to get a good result; however, this is small compared to the number of calculations necessary if the entire solution space were searched.

ACKNOWLEDGMENTS

This work has been supported by the WYKO Corporation and a Phase II SBIR Grant from the National Science Foundation. The authors wish to thank Dr. Yiping Xu of WYKO Corporation and Prof. Harry Barrett of the Optical Sciences Center for their helpful discussions.
REFERENCES


