Multiscale characterization method for line edge roughness based on redundant second generation wavelet transform

Fei Wang, Ning Li, and Xuezeng Zhao

Citation: Journal of Applied Physics **108**, 084904 (2010); doi: 10.1063/1.3492618
View online: http://dx.doi.org/10.1063/1.3492618
View Table of Contents: http://scitation.aip.org/content/aip/journal/jap/108/8?ver=pdfcov
Published by the AIP Publishing

Multiscale characterization method for line edge roughness based on redundant second generation wavelet transform

Fei Wang,[1,a)] Ning Li,[2] and Xuezeng Zhao[1]
[1]School of Mechanical and Electronic Engineering, Harbin Institute of Technology, Harbin 150001, China
[2]Shanghai Second Polytechnic University, Shanghai 201209, China

(Received 14 June 2010; accepted 14 August 2010; published online 25 October 2010)

We introduce a multiscale characterization method for line edge roughness (LER) based on redundant second generation wavelet transform. This method involves decomposing LER characteristics into independent bands with different spatial frequency components at different scales, and analyzing the reconstructed signals to work out the roughness exponent, the spatial frequency distribution characteristics, as well as the rms value. The effect of noise can be predicted using detailed signals in the minimum space of scale. This method was applied to numerical profiles for validation. Results show that according to the line edge profiles with similar amplitudes, the roughness exponent R can effectively reflect the degree of irregularity of LER and intuitively provide information on LER spatial frequency distribution. © *2010 American Institute of Physics*. [doi:10.1063/1.3492618]

I. INTRODUCTION

According to International Technology Roadmap for Semiconductors standards for processes and measurements, photoetching technology takes into account the entire frequency spectrum containing spatial frequency components of line edge roughness (LER).[1] A particular processing step could make a difference in the distribution of LER spatial frequency.[2,3] Therefore, if the LER spatial frequency components within a particular range can be measured and controlled, more convincing measurement criteria can be obtained for the study on the effect of processing steps and conditions on LER formation. In addition, accurately and effectively separating the spatial frequency components of LER and adopting appropriate parameters to independently quantize each part yields an efficient technique for examining the effect of different LER spatial frequency components on the performance of integrated circuit (IC) devices.

The classical signal processing method based on stationary processes can produce only the statistical average results of signals from the perspective of time or frequency domain but cannot include the global and localized characteristics in time and frequency domains. Localized characteristics usually contain negligible information on roughness. Moreover, the analysis of power frequency density function based on fast Fourier transform (FFT) fails to yield a comprehensive scrutiny of the local frequency characteristics of time domain signals, although it is able to establish a connection between time and frequency domains. As a result, the FFT-based method cannot effectively work out the high-frequency components of LER, which pose a considerable effect on the performance of IC devices.

Multiscale signal processing based on wavelet transform not only recognizes important signal characteristics but also generates descriptions of signals with different scales. The second generation wavelet transform (SGWT) represents a breakthrough in wavelet studies in recent years in that it accomplishes the configuration of biorthogonal wavelet in the time domain without relying on FT, and eases the problem to be the design of predictors and updaters with different characteristics via reasonable means. The LER characteristic signals, therefore, can be regarded as the composition of random noise and signals with different spatial frequencies and amplitudes. With the multiscale analysis of SGWT, the characteristic signals of LER can be decomposed into a series of independent frequency bands with LER information on different spatial frequency components. By reprocessing information extracted through performing statistical analysis in the time domain (averages, variances, and relevance), for example, or frequency analysis, such as power spectral density function, we can effectively prove the transient characteristics inherited in the normal signals of LER and quantize specific spatial frequency distribution by combining spatial and frequency domain characteristics. As a result, SGWT represents a powerful tool for performing multiscale representation of LER characteristics with statistical, nonstationary random, and multiscale natures.

II. THEORY

A. Redundant SGWT

Frequency aliasing appears in SGWT during decomposing and reconstructing. That is, the decomposed wavelet not only results in aliasing of different frequency components in the frequency bands but also generates new frequency components that do not exist in the signals analyzed.[4]

There are two causes of the frequency aliasing. The nonideal cut-off frequency of the equivalent filter causes every subband to obtain partial frequency components of its neighboring subbands. Moreover, the equivalent up and down samplings caused by splitting and composing generate frequency aliasing components. For the first cause, given that there is no filter for ideal frequency cut-off, these frequency

a)Electronic mail: wangfeivc@gmail.com.

aliasing components are unavoidable. Nevertheless because the equivalent low performance filter (scale filter) and equivalent high performance filter (wavelet filter) of the SGWT have a bearing on the predictor and the updater, the aliasing between the low performance filter and high performance filter bands changes along with the predictor and the updater. Elongating the predictor remarkably improves the frequency characteristics of the wavelet filter, and reduces the aliasing between the low-frequency and high-frequency bands but has little effect on the improvement of frequency domain characteristics of the scale filter. Although a longer updater does not effectively affect the frequency characteristics of the wavelet filter, it can apparently improve the frequency characteristics of the scale filter and decrease the aliasing between different frequency bands. Given the influence of boundary effect, arithmetic complexity, and computation efficiency, however, overly lengthy predictors and updaters should be avoided.

For the second cause, the fake frequency stemming from the frequency aliasing can be avoided by omitting splitting and composing during the transform action; thus, adopting a redundant SGWT (RSGWT) inhibits frequency aliasing. Without the splitting and composing of SGWT, the RSGWT brings corresponding changes to its predictor and updater.[5,6] Assuming the predictor and the updater of the RSGWT as redundant predictor P_U and redundant updater U_U, respectively, the redundant predictor P_U^l and the redundant updater U_U^l at No. 2^l scale are obtained by inserting 2^{l-1} zeros between the coefficient of the initial predictor P and the initial updater U.

The decomposition process of RSGWT involves predicting and updating as shown in Eqs. (1) and (2).

$$d_{l+1}(z) = a_l(z) - P_U^{l+1}(z)a_l(z), \tag{1}$$

$$a_{l+1}(z) = a_l(z) + U_U^{l+1}(z)d_{l+1}(z), \tag{2}$$

where a_l and d_l are the approximation coefficient and detail coefficient, respectively; and the subscript l denotes the decomposing layer. With the abovementioned equations, the reconstructed process of RSGWT can be directly developed by restoring update and prediction as shown in Eqs. (3) and (4). However, because a reconstructed signal of $a_l(z)$ can be obtained from every step, marked $a_{la}(z)$ and $a_{ld}(z)$, the final reconstructed signal $a_l(z)$ is the average output of the two.

$$a_{la}(z) = a_{l+1}(z) - U_U^{l+1}(z)d_{l+1}(z), \tag{3}$$

$$a_{ld}(z) = d_{l+1}(z) + P_U^{l+1}(z)a_{la}(z), \tag{4}$$

$$a_l(z) = a_{la}(z)/2 + a_{ld}(z)/2. \tag{5}$$

Figure 1 shows the decomposition and reconstruction process of RSGWT in the 2^l scale. Because there is no splitting and composition, the redundant transform can effectively avoid the occurrence of frequency aliasing at each scale in SGWT, therefore freeing every decomposed component of the fake frequency.[7]

Moreover, RSGWT is translation invariant; thus, the value of wavelet coefficients remains constant and causes corresponding translation when the original signal is translated. This translation invariant characteristic can prevent amplitude vibration from occurring at the place of signal mutation or at the edge of the image when adopting wavelet transform. This is known as the fake Gibbs vibration.[8,9]

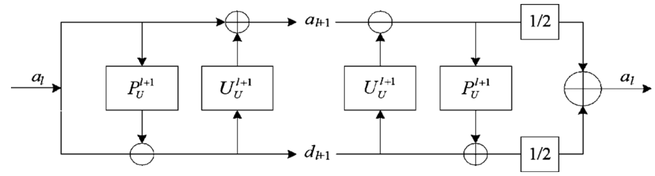

FIG. 1. Decomposition and reconstruction steps of RSGWT.

B. Calculation of roughness exponent *R*

The dramatic variation in location and direction of local points forming the line edges contain a number of detail information. By applying the multi-scale analysis to break the line edge into approximation and detail compounds, the former can indicate the trend of the line edge profile, whereas the latter components at different scales can be used to evaluate LER.

The decomposing results of RSGWT at the Jth layer are the detail coefficient $d_j (j=1,2,\ldots,J)$ and approximation coefficient a_J, in which the detail coefficient obtains the high-frequency component of the original signal X, and the approximation coefficient obtains the low-frequency components of signals. By increasing the sampling density of wavelet transform along the time axis, the redundant transform equates the lengths of the detail coefficient and the approximate coefficient at every scale with the length of the original signal X so that there are adequate coefficients at every scale to fully reflect the characteristics of signals to be analyzed.

Reconstructing the detail coefficients and approximation coefficients at every decomposing layer $j<J$ can decompose signal X into[10]

$$X = \sum_{j=1}^{J} D_j + A_J, \tag{6}$$

where D_j is the detail component of X at scale 2^{j-1} and A_J is the approximation component of X at scale 2^{J-1}.

While decomposing the LER characteristics using RSGWT, attention should be paid on the issue of decomposition layer. Scale is the inborn characteristic shared by many physical phenomena, which causes the corresponding characteristics to appear, disappear, and merge when the resolution of signal analysis changes. Thus, the sensing of signals at different resolutions brings about different results. When the scale expands to a certain roughness component wavelength, the roughness components with shorter wavelengths do not affect the roughness, and the roughness amplitude decreases; when the resolution becomes increasingly apparent (decreased scale) the roughness components of shorter wavelengths emerge. The relevance of changing the values of line edge signals in the direction of line length is likely to be evaluated through autocorrelation function of approximate components of decomposition layers. In particular, with

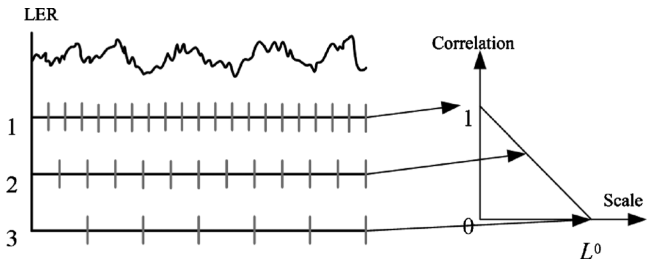

FIG. 2. Correlation of LER at different scales.

the reduced resolutions (enlarged scale), the relevance decreases and even reaches zero when the scale equates a certain physical scale (L_0) (Fig. 2).

When the decomposed layer is J, similar components contain enough information of large fluctuating quantities. That is, similar components have dominant spatial frequency components of LER characteristics, while the detail components at scale j ($j=1,2\ldots J-1$) demonstrate the distribution of high-frequency components crucial to LER characterization. Theoretically, if the dominant spatial frequency of LER is f_{LER} and the high-frequency restriction of measurement resolution is $f_{\max}$ ($=1/2$ fs) (where f_s is the spatial sampling frequency), then the decomposed layer J can be established by the formula as follows:

$$J = \text{int}\left(\log_2 \frac{f_{\max}}{f_{\text{LER}}} - 1\right). \tag{7}$$

Defining the physical scale $2^{J-1}\Delta(\text{nm})$ at the decomposed level J as the characteristic length, which determines the correlation range of LER characteristic in any position and reflects how long the scan length should be used in LER measurement. Therefore, the similar signal $A_J=X-\sum_{j=1}^{J}D_j$ at the decomposed level J is approximate enough to represent the original signal without higher frequency components, which then becomes of little concern.

Defining the repeated times of quantized changes in detail characteristics (such as sinusoidal structure) in unit length as spatial frequency (per nanometer), the LER characteristic signal can be considered a composition between the periodical components of different frequencies and amplitudes, and random noise.[7]

When the main information of signals is characterized by one or a group of characteristic quantities, it is likely to carry out information acquisition and characteristic discrimination on signals in a more intuitive, effective, and convenient manner, and provide a quantized means for evaluating and measuring the signal state. The power of detail signals at each scale can be applied to describe the high-frequency components of signals. However, for the LER characteristic signal attained from the atomic force microscopy (AFM)-measured image of different measure samples or AFM-measured images of the same source but of different scan ranges, it is impossible to describe LER objectively by depending on the detail signal power. This is because the fundamental power of the line edges attained from different measurement conditions may differ. Because the approximate signal is the average trend of the original signals after smoothness process at scale 2^{J-1}, the energy of the approximate signals are used to demonstrate the fundamental energy of line edge, whereas the irregular degree of LER characteristics is evaluated by comprehensive comparison taking the power of detail and approximate signals into consideration.

As far as the reconstructed signals at different scales are concerned, the power of detail signals and approximate signals can be calculated with Eqs. (8) and (9):

$$E_d = \sum_j \sum_n |D_j[n]|^2, \tag{8}$$

$$E_a = \sum_n |A_J[n]|^2, \tag{9}$$

where $j=1,2\ldots J-1$ and $n=0,1,\ldots,N$, N is the number of sampling points. The ratio of the power of detail signals and approximate signals is defined as the roughness exponent R, expressed as

$$R = E_d/E_a. \tag{10}$$

Therefore, R value represents the power distribution of detail signals versus approximate signals in the LER characteristics. The bigger the R value, the more important the high-frequency of LER is.

Considering this from another point of view, according to the number of deconstructed layers and the spatial sampling frequency, the original signal X is decomposed into detail components at different scales and approximate components at scale J. With the multiscale analysis based on RSGWT, the different spatial frequency components of signals can be accurately decomposed into corresponding scales. That is, the reconstructed detail signals and approximate signals at different scales carry different LER spatial frequency components. By performing frequency analysis on the reconstructed signals at each scale, the advantages of wavelet transformation and FT are integrated, making the LER spatial frequency analysis more comprehensive without hiding the LER high-frequency components that are necessary for IC manufacture.[7]

The spatial frequency information of signal characteristics is embodied in signals A_J and $D_{J-1},\ldots,D_2,D_1$ (from low to high) that are reconstructed with wavelet coefficients. The PSD function of every reconstructed signal is calculated at the scale of $2^{j-1}(j=1,2,\ldots,J-1)$ to obtain the spatial frequency distribution of reconstructed signals at all scales. Compared with the direct PSD analysis of original signals, the high-frequency information of signals is not covered and the detail signal at all scales can clearly describe the distribution characteristics of high-frequency components of signals at different scales.

As a significant parameter of LER amplitude characterization, rms roughness reflects the amplitude vibration of LER characteristics against zero. The rms e of reconstructed signals at different scales is defined as

$$e_j = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(D_j(i)-\bar{D}_j)^2}, \quad e_{j+1} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(A_J(i)-\bar{A}_J)^2}. \tag{11}$$

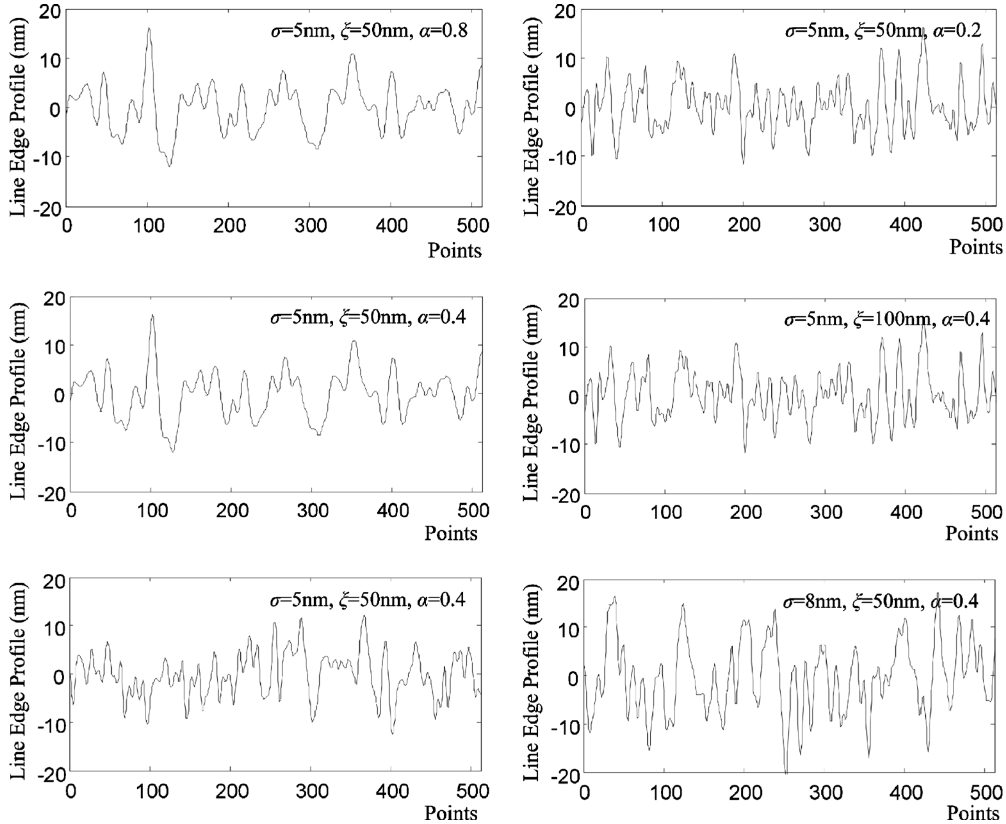

FIG. 3. Random line profiles with different roughness generated by numerical simulation.

From the rms value e_j of reconstructed signals at each scale, we can obtain the quantized amplitude values of individual spatial frequency components at different scales. From the deconstructed scales with larger rms value, the physical scale of spatial frequency compounds can be obtained with relatively high power. Considering the greater effect of noise on detail signal power, the detail signals of more noise at minimum scales have a relatively larger rms value.

III. RESULTS AND DISCUSSION

Two groups of line edge profiles (16/group, 32 target lines of edge profiles in total) that meet the requirements of the Gaussian height distribution of zero means and self similarity has been generated through a numerical simulation method,[7] in which the rms roughness σ, relevant length ξ, and roughness coefficient α $(\alpha=2-D)$ were chosen as the fractal characterization parameters of line edge profile.[11–14] The profile value σ of group 1 is invariably 5 nm, ξ is 20, 50, 80, and 100 nm, and α is 0.2, 0.4, 0.6, and 0.8; in group 2, profile α is 0.4, ξ is 20, 50, 80, and 100 nm, and α is 2, 5, 8, and 10 nm. To demonstrate the significance of α, ξ, and σ, we drew eight of the 32 profile curves (Fig. 3). As the figure illustrates, the high-frequency distribution of LER increases as α decreases; the waveform structural width of LER increases with ξ; and the spectrum value increases in proportion to σ.

Suppose the line edge profile length $L=3000$ nm; because the sampling point $N=512$, the scan interval Δ is 5.87 nm and the spatial sampling frequency f_s is 0.2 nm^{-1}. Calculating the roughness exponent R_s of the above-mentioned two groups of profiles, the results are shown in Fig. 4.

From the figures above, parameter α shows the measurement of high-frequency components of roughness; hence, the lower the value of α, the more important the high-frequency components of roughness are. With low-frequency components being relatively consistent as α increases, the power of detail signals in the multiscale analysis gradually decreases along with the decrease in approximation signals and the roughness exponent R; the low-frequency change in the line edge width along with the relatively larger relevant length ξ becomes more apparent, that is, as the power of approximate signals becomes bigger, the roughness exponent R becomes relatively smaller. As parameter σ (the change degree of line edge amplitude) increases, the power of detail signals and approximate signal both increase; hence, roughness exponent R does not demonstrate an obvious change trend. These two parameters are therefore independent of each other, i.e., σ is the quantized description of general LER amplitude change, whereas R denotes the quantized characterization of the high frequency of LER characteristics against the power of low-frequency components. Therefore, as far as the line edges with similar amplitudes are concerned, roughness exponent R effectively reflects the irregular degree of line edges, and provides the distribution information of LER spatial frequencies.

IV. SUMMARY

We introduced a multiscale characterization method for LER based on RSGWT. This method involves decomposing

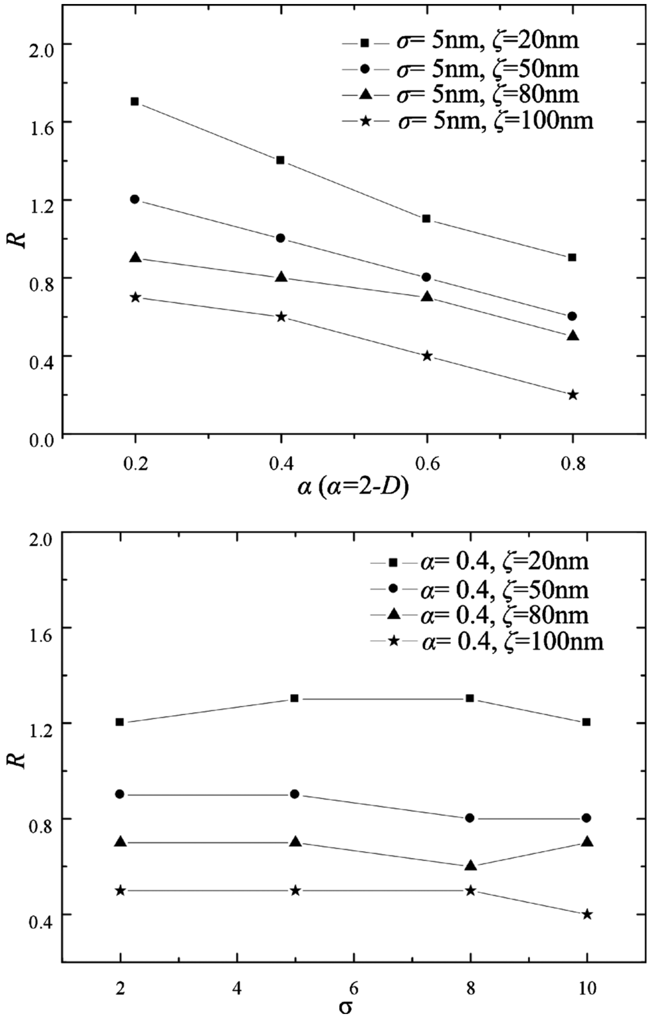

FIG. 4. Roughness exponent R of random line edge with different roughness generated by numerical simulation.

LER characteristics into independent bands with different spatial frequency components at different scales, and analyzing the reconstructed signals to work out the roughness exponent, the spatial frequency distribution characteristic of reconstructed signals, and rms value at each scale. The effect of noise can be predicted using detail signals in the minimum space of scale and the LER multiscale characterization method is applied to numerical data of random profiles. Results show that this method can directly provide the distribution information of LER spatial frequencies, effectively analyzing the LER high-frequency components that obviously influence the performance of IC instruments, and can provide quantized characterization of LER amplitude within the range of specific spatial frequency. Ultimately, this method furnishes a more comprehensive measurement basis for process supervision and optimized design of semiconductor manufacturing.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (Project No. 10902031) and China Postdoctoral Science Foundation (Project No. 20090451001).

[1] International Technology Roadmap for Semiconductors, 2008 edition, Metrology, The Semiconductor Industry Association, http://public.itrs.net

[2] A. R. Pawloski, A. Acheta, S. Bell, B. La Fontaine, T. Wallow, and H. J. Levinson, Proc. SPIE **6153**, 615318 (2006).

[3] E. Pargon, M. Martin, J. Thiault, O. Joubert, J. Foucher, and T. Lill, J. Vac. Sci. Technol. B **26**, 1011 (2008).

[4] W. Bao, R. Zhou, J. G. Yang, D. R. Yu, and N. Li, Mech. Syst. Signal Process. **23**, 1458 (2009).

[5] A. Stoffel, Signal Process. **69**, 177 (1999).

[6] C. S. Lee, C. K. Lee, and K. Y. Yoo, Electron. Lett. **36**, 1894 (2000).

[7] See supplementary material at http://dx.doi.org/10.1063/1.3492618 for the anti-aliasing property of RSGWT, reconstructed signal frequency analysis and generating method of random profile.

[8] D. L. Donoho, IEEE Trans. Inf. Theory **41**, 613 (1995).

[9] R. R. Coifman and D. L. Donoho, Lecture Notes in Statistics **103**, 125 (1995).

[10] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis* (Cambridge University Press, London, UK, 2000).

[11] G. P. Patsis, V. Constantoudis, A. Tserepi, E. Gogolides, G. Grozev, and T. Hoffmann, Microelectron. Eng. **67–68**, 319 (2003).

[12] V. Constantoudis, G. P. Patsis, and E. Gogolides, J. Microlithogr., Microfabr., Microsyst. **3**, 429 (2004).

[13] V. Constantoudis, G. P. Patsis, A. Tserepi, and E. Gogolides, J. Vac. Sci. Technol. B **21**, 1019 (2003).

[14] V. Constantoudis and G. P. Patsis, J. Vac. Sci. Technol. B **22**, 1974 (2004).