

DETECTION OF PROBABILISTIC PROCESS WINDOWS

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to and the benefit of U.S. Application No. 17/472,335, filed September 10, 2021 and titled “Detection of Probabilistic Process Windows”, the entire disclosure of which is hereby incorporated by reference herein as if reproduced in full below.

[0002] This application is also related to continuation-in-part of U.S. Application No. 17/316,154, filed May 10, 2021 and titled “System and Method for Generating and Analyzing Roughness Measurements”, which is a continuation of U.S. Application No. 16/730,393, filed December 30, 2019 and titled “System and Method for Generating and Analyzing Roughness Measurements”, which is a continuation of U.S. Application No. 16/218,346, filed December 12, 2018 and titled “System and Method for Generating and Analyzing Roughness Measurements” (now U.S. Patent No. 10,522,322), which is a continuation-in-part of and claims priority to U.S. Application No. 15/892,080 filed February 8, 2018 titled “Edge Detection System” (now U.S. Patent No. 10,176,966). This application further claims priority to U.S. Provisional Patent Application No. 62/739,721 filed October 1, 2018 titled “System and Method for Generating and Analyzing Roughness Measurements” and U.S. Provisional Patent Application No. 62/678,866 filed May 31, 2018 titled “System and Method for Removing Noise From Roughness Measurements.” Further, as a continuation of U.S. Application No. 16/218,346, this patent application claims priority to U.S. Provisional Patent Application Ser. No. 62/602,152, filed April 13, 2017 and titled “Edge Detection System.” All these applications are incorporated by reference herein as if reproduced in full below.

BACKGROUND

[0003] The present disclosure relates generally to edge detection of pattern structures, and more particularly, to edge detection of pattern structures in noise-prone images, such as in images formed when using a scanning electron microscope (SEM) or other imaging apparatus that produce images including undesired noise, and even more particularly, to analyzing such roughness measurements as a function of process variations and using this analysis to optimize processes and control process tools.

SUMMARY

[0004] The present disclosure generally provides methods, systems, and computer-readable mediums for generating a probabilistic process window that accounts for measurement uncertainty.

[0005] An aspect of the present disclosure includes a computer-implemented method. The method may include selecting a first process variable represented on a first axis of a graph. The method may also include selecting a second process variable represented on a second axis of the graph. The method may also include selecting at least one response variable that is a function of the first variable and the second variable. The method may also determine a measurement uncertainty for each response variable. The method may also include determining, based on a measurement of the response variable, and the measurement uncertainty for the response variable, a plurality of probabilities representing a plurality of indications of whether a plurality of points associated with a lithography process meet a specification requirement for each response variable, wherein the plurality of probabilities represent a process window. The method further includes configuring, based on the process window, a lithography tool to manufacture a semiconductor device.

[0006] Another aspect of the present disclosure includes a system comprising, in one implementation, a lithography tool, a memory device storing instructions, and a processing device. The processing device is coupled to the memory device and the lithography tool. The processing device may execute the instructions to select a first variable that can be represented on a first axis of a graph. The processing device may also execute the instructions to select a second variable that can be represented on a second axis of the graph. The processing device may also execute the instructions to select a response variable as a function of the first variable and the second variable. The processing device may also execute the instructions to determine a measurement uncertainty for the response variable. The processing device may also execute the instructions to determine, based on a measurement of the response variable and the measurement uncertainty for the response variable, a plurality of probabilities representing a plurality of indications of whether a plurality of points associated with a lithography process meet a specification requirement. The plurality of probabilities may represent a process window. The processing device may also execute the instructions to configure, based on the process window, the lithography tool to manufacture a semiconductor device.

[0007] A further aspect of the present disclosure includes a tangible, non-transitory computer-readable medium storing instructions that, when executed, cause a processing device to select a first variable represented on a first axis of a graph. The instructions may also cause the processing device to select a second variable is represented on a second axis of the graph. The instructions may also cause the processing device to select a response variable as a function of the first variable and second variable. The instructions may also cause the processing device to determine a measurement uncertainty for the output response variable. The instructions may also cause the processing device to determine, based on a measurement of the response variable and the measurement uncertainty for the response variable, a plurality of probabilities representing a plurality of indications of whether a plurality of points associated with a lithography process meet a specification requirement. The plurality of probabilities represents a process window. The instructions further cause the processing device to configure, based on the process window, the lithography tool to manufacture a semiconductor device.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] The appended drawings illustrate only exemplary implementations of the present disclosure and therefor do not limit its scope because the inventive concepts lend themselves to other equally effective implementations.

[0009] FIG. 1A is a representation of a pattern structure that exhibits parallel line features with spaces in between the lines.

[0010] FIG. 1B is a representation of a pattern structure that includes contact hole features.

[0011] FIG. 2 shows four different rough edges, all with the same standard deviation.

[0012] FIG. 3 is a representation of power spectral density (PSD) vs. frequency on a log-log scale.

[0013] FIG. 4 is a graphic representation of power spectral density (PSD) plotted vs. frequency and depicting roughness parameters $PSD(0)$, correlation length, and roughness exponent.

[0014] FIG. 5 shows two power spectral densities (PSDs) corresponding to respective edges of a feature on a pattern structure.

[0015] FIG. 6 is a graphic representation of the tradeoff of within-feature variation and feature-to-feature variation as a function of line length.

[0016] FIG. 7 is a block diagram of a scanning electron microscope (SEM) coupled to an information handling system (IHS) that together form one implementation of the disclosed edge detection apparatus.

[0017] FIG. 8A is a representation of a feature disposed on a substrate that depicts an electron beam impinging on the center of the feature.

[0018] FIG. 8B is a representation of a feature disposed on a substrate that depicts an electron beam impinging on the feature near its edge.

[0019] FIG. 9 shows a gray scale image representation on top with a corresponding grayscale linescan along one horizontal cut being graphically plotted immediately below.

[0020] FIG. 10 shows an example of a pattern structure including a feature situated atop a substrate with varying numbers of electrons escaping from the pattern structure depending on where the electron beam impinges on the pattern structure.

[0021] FIG. 11 shows a predicted linescan of a resist step on a pattern structure such as a silicon wafer.

[0022] FIG. 12 shows another representative predicted linescan of a pattern of resist lines and spaces on a silicon wafer.

[0023] FIG. 13A is an original SEM image of a pattern structure without using the disclosed edge detection apparatus and method.

[0024] FIG. 13B is the same SEM image as FIG. 13A except using the disclosed edge detection apparatus and method.

[0025] FIG. 14 is a Raw (Biased) linewidth roughness plot vs. threshold settings showing both a prior art result (using a filter with conventional threshold edge detection), and a result using no filter and an inverse linescan model (ILM).

[0026] FIG. 15A is a power spectral density (PSD) vs. frequency plot of the right and left edges of a feature shown before noise subtraction.

[0027] FIG. 15B is a power spectral density (PSD) vs. frequency plot of the right and left edges of a feature shown after noise subtraction.

[0028] FIG. 16 shows portions of three SEM images of nominally the same lithographic features taken at different SEM electron doses.

[0029] FIG. 17A shows a typical linescan for a line feature on a wafer for a case when there is an extremely large number of electrons so that the pixel noise is negligible.

[0030] FIG. 17B shows the 1-sigma uncertainty in edge detection position for perfectly smooth features in the presence of grayscale noise, for three different X pixel sizes.

[0031] FIG. 17C shows grayscale images as an example of using a simple threshold edge detection algorithm with image filtering in the right image, and without image filtering in the left image.

[0032] FIG. 18 is a plot of linewidth roughness (LWR) PSD vs. frequency that shows the impact of two different image filters on a collection of 30 images.

[0033] FIG. 19 is a power spectral density plot vs. frequency that shows the noise subtraction process of the disclosed edge detection apparatus and method.

[0034] FIG. 20 shows PSDs of a particular resist feature type on a given wafer, measured with different frames of integration in the SEM.

[0035] FIG. 21 shows the biased and unbiased values of the 3σ linewidth roughness (LWR) measured as a function of the number of frames of integration in the SEM.

[0036] FIG. 22A shows biased linewidth roughness (LWR) power spectral densities (PSDs) as a function of different pixel sizes and magnifications employed by the SEM.

[0037] FIG. 22B shows unbiased linewidth roughness (LWR) power spectral densities (PSDs) as a function of different pixel sizes and magnifications employed by the SEM.

[0038] FIG. 23 is a flowchart that depicts a representative overall process flow that the disclosed SEM edge detection system employs to detect edges of a pattern structure.

[0039] FIG. 24A is a grayscale representation of a pattern structure of vertical lines and spaces that the disclosed metrology tool analyzes.

- [0040] FIG. 24B shows a single linescan at one Y-pixel position.
- [0041] FIG. 24C shows the averaged linescan that is generated by averaging over all Y-pixels.
- [0042] FIG. 25A shows a PSD that includes high-frequency spike artifacts.
- [0043] FIG. 25B shows the PSD with spike artifacts removed.
- [0044] FIG. 26 shows a PSD that includes mid-frequency spike artifacts and harmonics.
- [0045] FIG. 27A shows the impact of mid-frequency spike artifacts on the modeling and interpreting of the PSD.
- [0046] FIG. 27B shows the impact of removing mid-frequency spike artifacts on the modeling and interpreting of the PSD.
- [0047] FIG. 28A shows a PSD dataset that exhibits a type of bump behavior.
- [0048] FIG. 28B shows an additional PSD dataset that exhibits a type of bump behavior.
- [0049] FIG. 29A shows the modeling and analysis of a low frequency bump of type I.
- [0050] FIG. 29B shows the modeling and analysis of a low frequency bump of type II.
- [0051] FIG. 30 is a flowchart that depicts a representative process flow to detect undesired spikes in a PSD dataset, and to remove the spikes from the PSD dataset and obtain roughness parameters for a feature.
- [0052] FIG. 31 is a flowchart that depicts another representative process flow to model bumps in a PSD dataset, and to obtain unbiased roughness parameters for a feature.
- [0053] FIG. 32 is plot of an example of a Gaussian distribution of mean values.
- [0054] FIG. 33 is a diagram of an example of a heat-map illustrating probabilities that a plurality of points are in spec.
- [0055] FIG. 34 is an alternate three-dimensional view of the heat map of FIG. 33.

[0056] FIG. 35 is diagram of an example of a heat-map illustrating fractions of in spec features.

[0057] FIG. 36 is an alternate three-dimensional view of the heat map of FIG. 35.

[0058] FIG. 37 is a plot of an example of a curve for focus error and exposure error.

[0059] FIG. 38 is a flowchart that depicts a representative process flow to generate a probabilistic process window that accounts for measurement uncertainty.

[0060] FIG. 39A is a diagram of an example of a Bossung plot that combines impact of focus and exposure dose on CD.

[0061] FIG. 39B is a diagram of an example of a contour plot including a dimensional data set having contours of constant linewidth versus focus and exposure.

[0062] FIG. 39C is an example of plotting contours of CD (nominal $\pm 10\%$), 80-degree sidewall angle, and 10% resist loss all on the same graph, as a function of focus (x-axis) and exposure dose (y-axis).

[0063] FIG. 40A is an example diagram of a process window showing two maximal rectangles fit inside the process window.

[0064] FIG. 40B is an example diagram of a process window showing one maximal rectangle and one maximal ellipse fit inside the process window.

[0065] FIG. 41 is an example analysis of a process window to give exposure latitude versus depth of focus.

[0066] FIG. 42 is an example of overlapping process windows for line/space patterns of two different pitches.

[0067] FIG. 43 depicts the potential influence of measurement error on the geometric analysis of process window size.

DETAILED DESCRIPTION

[0068] Measuring the roughness of a pattern is complicated by that fact that noise in the measurement system is difficult to differentiate from the roughness being measured. It is

common to using an imaging tool, such as a microscope, to create a detailed image of an object to be measured and then analyze the information on that image to measure and characterize the roughness of one or more features of the object. In this case, noise in the acquired image can appear to be roughness of the features in the image. Described below, among other things, are techniques useful to separate the noise in the image from the actual roughness of the features in order to produce more accurate measurements of the roughness of the features.

[0069] As an example, scanning electron microscopes (SEMs) are very useful for studying the features of pattern structures, such as semiconductor devices, for example. Unfortunately, measuring feature roughness of these structures is often challenging because of the noise that is inherent in SEM images. Filtering (smoothing) of the SEM image is typically needed to achieve accurate edge detection, but such filtering undesirably changes the feature roughness that is measured. An edge detection approach is needed that reliably detects edges in very noisy SEM images without the use of image filtering (or at least without any filtering that would change the feature roughness that is measured).

[0070] Pattern roughness is a major problem in many fields. Many if not all techniques for creating patterns of various shapes produce roughness on the edges of those patterns, at least on the near molecular scale if not larger scales. For example, in advanced lithography for semiconductor manufacturing, especially for extreme ultraviolet (EUV) lithography but for other lithography methods as well, roughness of the printed and etched patterns can cause many negative effects. Reduction in roughness requires a better understanding of the sources of stochastic variation, which in turn requires better measurement and characterization of rough features. Prior art roughness measurement approaches suffer from severe bias because noise in the image adds to the roughness on the wafer. The present disclosures provides a practical approach to making unbiased roughness measurements through the use of a physics-based inverse linescan model. This enables accurate and robust measurement of roughness parameters over a wide range of SEM metrology conditions.

[0071] Before discussing implementations of the disclosed technology that address the SEM image noise problem, this disclosure first discusses lithography of pattern structures and the frequency dependence of roughness.

[0072] 1. Stochastic Effects in Lithography

[0073] Lithography and patterning advances continue to propel Moore's Law by cost-effectively shrinking the area of silicon consumed by a transistor in an integrated circuit. Besides the need for improved resolution, these lithography advances should also allow improved control of the smaller features being manufactured. Historically, lithographers focused on "global" sources of variation that affect patterning fidelity (e.g., exposure dose and focus variations, hotplate temperature non-uniformity, scanner aberrations) by attempting to minimize the sources of these variations and by developing processes with minimum sensitivity to these variations. Today's small features, however, also suffer from "local" variations caused by the fundamental stochastics of patterning near the molecular scale.

[0074] In lithography, light is used to expose a photosensitive material called a photoresist. The resulting chemical reactions (including those that occur during a post-exposure bake) change the solubility of the resist, enabling patterns to be developed and producing the desired critical dimension (CD). For a volume of resist that is "large" (that is, a volume that contains many, many resist molecules), the amount of light energy averaged over that volume produces a certain amount of chemical change (on average) which produces a certain (average) amount of dissolution to create the pattern. The relationships between light energy, chemical concentration, and dissolution rate can be described with deterministic equations that predict outputs for a given set of inputs. These models of lithography are extremely useful and are commonly used to understand and control lithography processes for semiconductor manufacturing.

[0075] This deterministic view of a lithography process (certain inputs always produce certain outputs) is only approximately true. The "mean field theory" of lithography says that, on average, the deterministic models accurately predict lithographic results. If we average over a large number of photons, a single number for light energy (the average) is sufficient to describe the light energy. For a large volume of resist, the average concentration of a chemical species sufficiently describes its chemical state. But for very small volumes, the number of atoms or molecules in the volume becomes random even for a fixed "average" concentration. This randomness within small volumes (that is, for small quantities of photons or molecules or numbers of events) is generally referred to as "shot noise", and is an example of a stochastic variation in lithography that occurs when the region of interest approaches the molecular scale.

[0076] A stochastic process is one in which the results of the process are randomly determined. At the atomic/molecular level, essentially all processes are stochastic. For

semiconductor patterning at the 20-nm node and below (with minimum feature sizes below 40 nm), the dimensions of interest are sufficiently small that stochastic effects become important and may even dominate the total variations that affect the dimensions, shapes, and placements of the patterns being fabricated. These stochastic effects can also be important for larger feature sizes under some circumstances.

[0077] The most prominent manifestation of stochastic variations in lithography (as well as etch and other parts of the patterning process) is that the patterns being produced are rough rather than smooth (FIG. 1A). In the pattern structure shown in FIG. 1A, nominally parallel vertical lines appear as bright vertical regions, while spaces appear as dark vertical regions between the lines. The roughness of the edge of a feature is called line-edge roughness (LER), and the roughness of the width of a feature is called linewidth roughness (LWR). The roughness of the centerline of the feature (the midpoint between left and right edges) is called pattern placement roughness (PPR). Another important consequence of these stochastic variations is the random variation of the size, shape, and placement of features, which are especially evident for contact hole features (FIG. 1B).

[0078] Stochastic effects in patterning can reduce the yield and performance of semiconductor devices in several ways: a) Within-feature roughness can affect the electrical properties of a device, such as metal line resistance and transistor gate leakage; b) Feature-to-feature size variation caused by stochastics (also called local CD uniformity, LCDU) adds to the total budget of CD variation, sometimes becoming the dominant source; c) Feature-to-feature pattern placement variation caused by stochastics (also called local pattern placement error, LPPE) adds to the total budget of PPE, sometimes becoming the dominant source; d) Rare events leading to greater than expected occurrence of catastrophic bridges or breaks are more probable if error distributions have fat tails; and e) Decisions based on metrology results (including process monitoring and control, as well as the calibration of optical proximity correction (OPC) models) can be poor if those metrology results do not properly take into account stochastic variations. For these reasons, proper measurement and characterization of stochastic-induced roughness is critical.

[0079] Many other kinds of devices are also sensitive to feature roughness. For example, roughness along the edge of an optical waveguide can cause loss of light due to scattering. Feature roughness in radio frequency microelectromechanical systems (MEMS) switches can affect performance and reliability, as is true for other MEMS devices. Feature roughness can

degrade the output of light emitting diodes. Edge roughness can also affect the mechanical and wetting properties of a feature in microfluidic devices. Roughness of the features in a wire grid polarizer can affect the efficiency and transmission of the polarizer.

[0080] Unfortunately, prior art roughness measurements (such as the measurement of linewidth roughness or line-edge roughness using a critical dimension scanning electron microscope, CD-SEM) are contaminated by measurement noise caused by the measurement tool. This results in a biased measurement, where the measurement noise adds to the true roughness to produce an apparent roughness that overestimates the true roughness. Furthermore, these biases are dependent on the specific measurement tool used and on its settings. These biases are also a function of the patterns being measured. Prior art attempts at providing unbiased roughness estimates often struggle in many of today's applications due to the smaller feature sizes and higher levels of SEM noise.

[0081] Thus, there is a need for a new approach to making unbiased roughness measurements that avoids the problems of prior art attempts and provides an unbiased estimate of the feature roughness that is both accurate and precise. Further, a good pattern roughness measurement method should have minimum dependence on metrology tool settings. CD-SEM settings such as magnification, pixel size, number of frames of averaging (equivalent to total electron dose in the SEM), voltage, and current may cause fairly large changes in the biased roughness that is measured. Ideally, an unbiased roughness measurement would be independent of these settings to a large degree.

[0082] 2. The Frequency Dependence of Line-Edge Roughness (LER), Line-Width Roughness (LWR), and Pattern Placement Roughness (PPR)

[0083] Rough features are most commonly characterized by the standard deviation of the edge position (for LER), linewidth (for LWR), or feature centerline (for PPR). But describing the standard deviation is not enough to fully describe the roughness. FIG. 2 shows four different rough edges, all with the same standard deviation. The prominent differences visible in the edges make it clear that the standard deviation is not enough to fully characterize the roughness. Instead, a frequency analysis of the roughness is required. The four randomly rough edges depicted in FIG. 2 all have the same standard deviation of roughness, but differ in the frequency parameters of correlation length (ξ) and roughness exponent (H). More

specifically, with respect to FIG. 2, in case a) $\xi = 10$, $H = 0.5$; in case b) $\xi = 10$, $H = 1.0$; in case c) $\xi = 100$, $H = 0.5$; and in case d) $\xi = 0.1$, $H = 0.5$.

[0084] The standard deviation of a rough edge describes its variation relative to and perpendicular to an ideal straight line. In FIG. 2, the standard deviation describes the vertical variation of the edge. But the variation can be spread out differently along the length of the line (in the horizontal direction in FIG. 2). This line-length dependence can be described using a correlation function such as the autocorrelation function or the height-height correlation function.

[0085] Alternatively, the frequency f can be defined as one over a length along the line (FIG. 3). The dependency of the roughness on frequency can be characterized using the well-known power spectral density (PSD). The PSD is the variance of the edge per unit frequency (FIG. 3), and is calculated as the square of the coefficients of the Fourier transform of the edge deviation. The low-frequency region of the PSD curve describes edge deviations that occur over long length scales, whereas the high-frequency region describes edge deviations over short length scales. Commonly, PSDs are plotted on a log-log scale as used in FIG. 3.

[0086] The PSD of lithographically defined features generally has a shape similar to that shown in FIG. 3. The low-frequency region of the PSD is flat (so-called “white noise” behavior), and then above a certain frequency it falls off as a power of the frequency (a statistically fractal behavior). The difference in these two regions has to do with correlations along the length of the feature. Points along the edge that are far apart are uncorrelated with each other (statistically independent), and uncorrelated noise has a flat power spectral density. But at short length scales the edge deviations become correlated, reflecting a correlating mechanism in the generation of the roughness, such as acid reaction-diffusion for a chemically amplified resist. The transition between uncorrelated and correlated behavior occurs at a distance called the correlation length.

[0087] FIG. 4 shows that a typical PSD curve can be described with three parameters. $\text{PSD}(0)$ is the zero-frequency value of the PSD. While this value of the PSD can never be directly measured (zero frequency corresponds to an infinitely long line), $\text{PSD}(0)$ can be thought of as the value of the PSD in the flat low-frequency region. The PSD begins to fall near a frequency of $1/(2\pi\xi)$ where ξ is the correlation length. In the fractal region, we have what is sometimes called “1/f” noise and the PSD has a slope (on the log-log plot) corresponding to a power of

1/f. The slope is defined as $2H + 1$ where H is called the roughness exponent (or Hurst exponent). Typical values of H are between 0.5 and 1.0. For example, $H = 0.5$ when a simple diffusion process causes the correlation. Each of the parameters of the PSD curve has important physical meaning for a lithographically defined feature as discussed in more detail below. The variance of the roughness is the area under the PSD curve and can be derived from the other three PSD parameters. The exact relationship between variance and the other three PSD parameters depends on the exact shape of the PSD curve in the mid-frequency region (defined by the correlation length), but an approximate relationship can be used to show the general trend, as per EQUATION 1 below:

$$\sigma^2 \approx \frac{PSD(0)}{(2H + 1)\xi} \quad (1)$$

[0088] The differences observed in the respective four rough edges of FIG. 2 can now be easily seen as differences in the PSD behavior of the features. FIG. 5 shows two PSDs, corresponding to edge a) and edge c) from FIG. 2. While these two edges have the same variance (the same area under the PSD curve), they have different values of $PSD(0)$ and correlation length (in this case the roughness exponent was kept constant). Although the standard deviations of the roughness of edge a) and edge c) are the same, these edges exhibit different PSD behaviors. As discussed below, the different PSD curves will result in different roughness behavior for lithographic features of finite length.

[0089] 3. Impact of the Frequency Behavior of Roughness

[0090] The roughness of the lines and spaces of pattern structures is characterized by measuring very long lines and spaces, sufficiently long that the flat region of the PSD becomes apparent. For a sufficiently long feature the measured LWR (that is, the standard deviation σ of the measured linewidths along the line) can be thought of as the LWR of an infinitely long feature, $\sigma_{LWR}(\infty)$. But pattern structures such as semiconductor devices are made from features that have a variety of lengths L . For these shorter features, stochastics will cause within-feature roughness, $\sigma_{LWR}(L)$, and feature-to-feature variation described by the standard deviation of the mean linewidths of the features, $\sigma_{CDU}(L)$. This feature-to-feature variation is called the local critical dimension uniformity, LCDU, since it represents CD (critical dimension) variation that is not caused by the well-known “global” sources of error (scanner aberrations, mask illumination non-uniformity, hotplate temperature variation, etc.).

[0091] For a line of length L , the within-feature variation and the feature-to-feature variation can be related to the LWR of an infinitely long line (of the same nominal CD and pitch) by the Conservation of Roughness principle given in EQUATION 2 below:

$$\sigma_{CDU}^2(L) + \sigma_{LWR}^2(L) = \sigma_{LWR}^2(\infty) \quad (2)$$

[0092] The Conservation of Roughness principle says that the variance of a very long line is partitioned for a shorter line into within-feature variation and feature-to-feature variation. How this partition occurs is determined by the correlation length, or more specifically by L/ξ . Using a basic model for the shape of the PSD as an example, it is seen that:

$$\sigma_{CDU}^2(L) = \frac{PSD(0)}{L} \left[1 - \frac{\xi}{L} (1 - e^{-L/\xi}) \right] \quad (3)$$

[0093] Thus, EQUATIONS 1-3 show that a measurement of the PSD for a long line, and its description by the parameters $PSD(0)$, ξ , and H , enables one to predict the stochastic influence on a line of any length L . It is noted that the LCDU does not depend on the roughness exponent, making H less important than $PSD(0)$ and ξ . For this reason, it is useful to describe the frequency dependence of roughness using an alternate triplet of parameters: $\sigma_{LWR}(\infty)$, $PSD(0)$, and ξ . Note that these same relationships apply to LER and PPR as well.

[0094] It is also noted that, examining EQUATION 3, the correlation length is the length scale that determines whether a line of length L acts “long” or “short”. For a long line, $L \gg \xi$ and the local CDU behaves as per EQUATION 4 below:

$$\sigma_{CDU}(L) \approx \sqrt{\frac{PSD(0)}{L}} \text{ when } L \gg \xi \quad (4)$$

[0095] This long-line result provides a useful interpretation for $PSD(0)$: It is the square of the LCDU for a given line times the length of that line. Reducing $PSD(0)$ by a factor of 4 reduces the LCDU by a factor of 2, and the other PSD parameters have no impact (so long as $L \gg \xi$). Typically, resists have yielded correlation lengths on the order of one quarter to one half of the minimum half-pitch of their lithographic generation. Thus, when features are longer than approximately five times the minimum half-pitch of the technology node, we are generally in this long line length regime. For shorter line lengths, the correlation length begins to matter as well.

[0096] EQUATIONS 1–3 show a trade-off of within-feature variation and feature-to-feature variation as a function of line length. FIG. 6 shows an example of this relationship. For very long lines, LCDU is small and within-feature roughness approaches its maximum value. For very short lines the LCDU dominates. However, due to the quadratic nature of the Conservation of Roughness, $\sigma_{LWR}(L)$ rises very quickly as L increases, but LCDU falls very slowly as L increases. Thus, there is a wide range of line lengths where both feature roughness and LCDU are significant.

[0097] Since the Conservation of Roughness principle applies to PPR as well, short features suffer not only from local CDU problems but also from local pattern placement errors (LPPE) as well. For the case of uncorrelated left and right edges of a feature, the PSD(0) for LWR is typically twice the PSD(0) of the LER. Likewise, the PSD(0) of the LER is typically twice the PSD(0) of the PPR. Thus, in general, the LPPE is about half the LCDU. When left and right feature edges are significantly correlated, these simple relationships no longer hold.

[0098] 4. Measurements of the Roughness of Pattern Structures with a Scanning Electron Microscope (SEM)

[0099] A common way to measure feature roughness for small features is the top-down critical dimension scanning electron microscope (CD-SEM). Typical light microscopes have magnifications up to 1000X and resolutions down to a few hundred nanometers. Scanning electron microscopes use electrons to create very small spots (near 1 nm in width) that can be used to create high-resolution images, with magnifications above 20,000X. CD-SEMs are SEMs that have been optimized for measuring the dimensions of a wide range of features found on semiconductor wafers. They can measure the mean critical dimension of a rough feature with high precision, but have also proven very useful for measuring LER, LWR, PPR, and their PSDs as well. However, there are errors in the SEM images that can have large impacts on the measured roughness and the roughness PSD while having little impact on the measurement of mean CD. For this reason, the metrology approach needed for PSD measurement may be quite different than the approach commonly used for mean CD measurement.

[0100] FIG. 7 shows a block diagram of one implementation of the disclosed edge detection system 700 that determines feature roughness. The pattern structure 800 and the electron imaging optics (710, 715, 720, and 725) are situated in a vacuum chamber 703 that is evacuated by vacuum pump 702. Electrons are generated from a source such as an electron gun 705 to

form an electron beam 707. Common electron beam sources include a heated tungsten filament, a lanthanum hexaboride (LaB6) crystal formed into a thermionic emission gun, or a sharp-tipped metal wire formed to make a field emission gun. The emitted electrons are accelerated and focused using electromagnetic condenser lenses 710, 715, and 720. The energy of the electrons striking the pattern structure 800 is generally in the 200 eV to 40 keV range in SEMs, but more typically 300 eV to 800 eV for CD-SEMs. Final condenser lens 720 employs scanning coils 725 to provide an electric field that deflects the electron beam 707 toward pattern structure 800 as a focused spot. Scanning coils 725 scan the focused spot across the pattern structure 800 through final lens aperture 735 in a raster scan fashion to expose a specific field of view on the pattern structure 800. SEM 701 includes a backscatter electron detector 740 that detects backscatter electrons scattering back from pattern structure 800. SEM 701 also includes a secondary electron detector 745, as shown in FIG. 7. Prior to imaging pattern structure 800, the user places pattern structure 800 on a pattern structure receiver 732 that supports and positions pattern structure 800 within SEM 701. SEM 701 includes a controller (not shown) that controls the raster scanning of pattern structure 800 during imaging.

[0101] Referring now to FIGs. 8A and 8B, the electrons of electron beam 707 that strike pattern structure sample 800 undergo a number of processes that depend on the energy of the electron and the material properties of the sample. Electrons scatter off the atoms of the sample material, release energy, change direction, and often generate a cascade of secondary electrons by ionizing the sample atoms. Some of these secondary electrons 805 may escape from the pattern structure 800 and others may remain inside the pattern structure. Pattern structure 800 includes a substrate 810, such as a semiconductor wafer. A feature 815 is disposed atop substrate 810, as shown in FIG. 8A. Feature 815 may be a metallic line, a semiconductor line, a photoresist line or other structures on substrate 810. Feature 815 may have other shapes such as a pillar or a hole, or more complicated shapes. Feature 815 may be repeating or isolated with respect to other features on the pattern structure. The space surrounding feature 815 may be empty (vacuum or air) or may be filled with a different material. Pattern structure 800 may be a liquid crystal or other flat panel display, or other pattern semiconductor or non-semiconductor device. Feature 815 includes edges 815-1 and 815-2. The region of feature 815 where the electron beam 707 interacts with feature 815 is the interaction volume 820 that exhibits, for example, a tear-droplet-like shape as depicted in FIG. 8A.

[0102] Occasionally electrons ricochet backwards off the atom nucleus and exit out of the sample (called backscatter electrons). Some of the lower energy secondary electrons 805 can also escape out of the sample 800 (frequently through the edges of a feature, see FIG. 8B). The way in which a SEM forms an image is by detecting the number of secondary electrons and/or backscatter electrons that escape the sample for each beam position.

[0103] As the electron beam is scanned across pattern structure 800 during one linescan, it “dwells” at a specific spot for a specific time. During that dwell time, the number of electrons detected by either the backscatter electron detector 740 or secondary electron detector 745, or both, is recorded. The spot is then moved to the next “pixel” location, and the process is repeated. The result is a two-dimensional array of pixels (locations along the surface of the sample) with detected electron counts digitally recorded for each pixel. The counts are typically then normalized and expressed as an 8-bit grayscale value between 0 and 255. This allows the detected electron counts to be plotted as a grayscale “image”, such as those images shown in FIG. 1. While the image coming from a SEM reminds a viewer of an optical image as perceived through the eye, it is important to note that these grayscale images are actually just convenient plots of the collected data.

[0104] A CD-SEM measures the width of a feature using the SEM image. The first step in measuring feature width is to detect the edges of the features. For pixels near an edge of a feature, higher numbers of secondary electrons escape through the feature edge, producing bright pixels called “edge bloom” (see FIG. 8B and FIG. 9). It is this bright edge bloom that allows the feature edge to be detected. For example, in the grayscale image representation in the upper portion of FIG. 9, such edge blooms are observed at edges 905 and 910 of feature 915. A linescan is essentially a horizontal cut through a 2D SEM image that provides a grayscale value as a function of horizontal pixel position on the feature, as in the graph shown in the bottom half of FIG. 9.

[0105] The data from a single horizontal row of pixels across the sample is called a “linescan”. Note that the term linescan is used here broadly enough to include cases where an image is formed without the use of scanning. The positions of the edges of a feature can be detected from a single linescan, or from a collection of linescans representing the entire image, such as shown in the upper portion of FIG. 9. These same edges appear as peaks 905’ and 910’ in the grayscale value vs. pixel position graph in the lower portion of FIG. 9. Once the edges

of a particular feature have been determined, the width of the particular feature is the difference between the positions of these two edges.

[0106] 5. Linescan Models

[0107] Images are created through a physical process based on the microscope or other imaging tool used to acquire the image of a structure. Often these images are two-dimensional arrays of data, where the image can be thought of as a data set derived from the structure. A single one-dimensional cut through the image is called a linescan. A model of the imaging tool can predict the image for a given structure being imaged. For example, a model that describes a scanning electron microscope can predict the image that would be obtained by a SEM when imaging a given structure.

[0108] A CD-SEM converts a measured linescan or a series of measured linescans into a single dimension number, the measured CD. To better understand how the linescan relates to the actual dimensions of the feature being measured, it is important to understand how the systematic response of the SEM measurement tool to pattern structures impacts the shape of the resulting linescan. Rigorous 3D Monte Carlo simulations of SEM linescans can be extremely valuable for this purpose, but they are often too computationally expensive for day-to-day use. Thus, one approach is to develop a simplified analytical linescan model (ALM) that is more computationally appropriate to the task of quickly predicting linescans. The ALM employs the physics of electron scattering and secondary electron generation, and each term in the model has physical significance. This analytical linescan expression can be fit to rigorous Monte Carlo simulations to both validate and calibrate its use.

[0109] The general application for the ALM has been the typical forward modeling problem: Given material properties (for the feature and the substrate) and a geometric description of the feature (width, pitch, sidewall angle, top corner rounding, footing, etc.), the ALM predicts the linescan that would result. The mathematical details of the ALM are found in the publications: Chris A. Mack and Benjamin D. Bunday, "Analytical Linescan Model for SEM Metrology", *Metrology, Inspection, and Process Control for Microlithography XXIX, Proc.*, SPIE Vol. 9424, 94240F (2015), and Chris A. Mack and Benjamin D. Bunday, "Improvements to the Analytical Linescan Model for SEM Metrology", *Metrology, Inspection, and Process Control for Microlithography XXX, Proc.*, SPIE Vol. 9778, 97780A (2016), the disclosures of both

publications being incorporated herein by reference in their entireties. Other models with similar inputs and outputs can also be used.

[0110] The analytical linescan model (ALM) is briefly reviewed below. The mathematical modeling begins by assuming the interaction of the electron beam with a flat sample of a given substance produces an energy deposition profile that takes the form of a double Gaussian, with a forward scattering width and a fraction of the energy forward scattered, and a backscatter width and a fraction of the energy deposited by those backscattered electrons. The model also assumes that the number of secondary electrons that is generated within the material is in direct proportion to the energy deposited per unit volume, and the number of secondary electrons that escape the wafer (and so are detected by the SEM) are in direct proportion to the number of secondary electrons near the very top of the wafer.

[0111] The secondary electrons that reach the detector will emerge some distance r away from the position of the incident beam. From the assumptions above, the number of secondary electrons detected will be a function as given in EQUATION 5 below:

$$f(e) = ae^{-r^2/2\sigma_f^2} + be^{-r^2/2\sigma_b^2} \quad (5)$$

[0112] where σ_f and σ_b are the forward and backscatter ranges, respectively, and a and b are the amounts of forward scattering and backscattering, respectively.

[0113] SEMs detect topography because of the different number of secondary electrons that escape when the beam is in the space between features compared to when the beam is on top of the feature. FIG. 10 shows that secondary electrons have trouble escaping from a space (especially if it is small), making spaces appear relatively dark. When an electron beam is focused to a spot in a space between lines, scattered electrons interact with feature 815 which absorbs some of the escaping secondary electrons. The detected secondary electron signal is reduced as the beam approaches the feature edge within the space.

[0114] The absorption by the step (i.e. feature 815) can be modeled to produce a prediction of the shape of the linescan in the space region. If a large feature has a left edge 815-1 at $x = 0$, with the feature 815 to the right (positive x), the detected secondary electron signal as a function of position ($SE(x)$) will be given by EQUATION 6 below:

$$\text{For } x < 0, \quad \frac{SE(x)}{SE(-\infty)} = 1 - \alpha_f e^{x/\sigma_f} - \alpha_b e^{x/\sigma_b} \quad (6)$$

[0115] where α_f is the fraction of forward scatter secondary electrons absorbed by the step and α_b is the fraction of backscatter secondary electrons absorbed by the step.

[0116] However, when the beam is on top of feature 815, the interaction of the scattered electrons with the feature is very different, as accounted for in EQUATION 7 below. As illustrated in FIG. 8, two phenomena occur as when the beam is closer to the edge compared to further away. First, secondary electrons from both forward and backscattered electrons can more easily escape out of the edge 815-1. This causes the edge bloom already discussed above. To account for this effect, a positive term $\alpha_e e^{-x/\sigma_e}$ is added to account for the enhanced escape of forward-scattered secondary electrons where σ_e is very similar to the forward scatter range of the step material. Additionally, the interaction volume itself decreases when the beam is near the edge 815-1, so that there are fewer secondary electrons being generated. Thus, the term $\alpha_v e^{-x/\sigma_v}$ where $\sigma_v < \sigma_e$ is subtracted to give EQUATION 7 below which is the linescan expression for the top of the large feature 815:

$$\text{For } x > 0, \quad \frac{SE(x)}{SE(-\infty)} = 1 + \alpha_e e^{-x/\sigma_e} - \alpha_v e^{-x/\sigma_v} \quad (7)$$

[0117] FIG. 11 shows an example of the result for this model. More specifically, FIG. 11 shows a predicted linescan of a left-facing resist step 815 (large feature with left edge 815-1 at $x = 0$) on a substrate such as a silicon wafer. The calibrated model 1105 is superimposed on the rigorous Monte Carlo simulation results 1110. The calibrated model 1105 agrees so closely with the Monte Carlo simulation results 1110 that the two curves appear together almost as one line.

[0118] The above discussion involves modelling an isolated left-facing edge 815-1. Adapting the model to include a right-facing edge involves translating and reversing the edge and adding the resulting secondaries (i.e., secondary electrons). Some complications arise if the two edges are close enough to interact, resulting in additional terms. Additionally, the impact of non-vertical sidewalls and rounded corners at the top and bottom of the feature edge may be included in the model (FIG. 12).

[0119] FIG. 12 shows a representative predicted linescan of a pattern of resist lines and spaces on a silicon wafer. The calibrated model 1205 is superimposed on the rigorous Monte Carlo simulation results 1210. Again, the calibrated model 1205 agrees so closely with the Monte Carlo simulation results 1210 that the two curves appear together almost as one line. A final model (ALM) includes 15 parameters that depend on the properties of the materials of the wafer and feature, and the beam voltage. To validate the model and to calibrate these parameters, rigorous first principle Monte Carlo simulations can be used to generate linescans for different materials and feature geometries. The ALM can then be fit to the Monte Carlo results, producing best-fit values of the 15 unknown parameters.

[0120] 6. Inverse Linescan Model

[0121] Linescan or image models, such as the analytical linescan model (ALM) discussed above, predict an image or the shape of an image linescan for a particular pattern structure (such as a feature on a wafer). The ALM solves a forward modelling problem wherein the model receives geometry information for the particular feature as input, and provides the predicted shape of a respective SEM linescan of the particular feature as output.

[0122] In contrast to ALM, the disclosed edge detection system 700 includes a reverse model that receives as input “measured linescan information” from SEM 701 that describes a particular feature on the wafer. In response to the measured linescan information describing the particular feature, edge detection system 700 employs its reverse model to generate as output “feature geometry information” that describes the feature geometry that would produce the measured linescan. Advantageously, edge detection system 700 has been found to be effective even when the measured linescan information from SEM 701 includes a significant amount of image noise. In one implementation, the outputted feature geometry information includes at least feature width. In another implementation, the outputted feature information includes feature width and/or other geometry descriptors relative to the geometry of the particular feature, such as sidewall angle, feature thickness, top corner rounding, or bottom footing. It is noted that a feature disposed on a semiconductor wafer is an example of one particular type of pattern structure to which the disclosed technology applies.

[0123] Like many models of imaging systems, the ALM is inherently nonlinear. To address the nonlinear nature of the ALM, edge detection system 700 numerically inverts the ALM or a similar forward model and fits the resulting inverse linescan model to a measured linescan to

detect feature edges (e.g. to estimate the feature geometry on the wafer). The disclosed edge detection system apparatus and edge detection process include the ability to detect and measure feature roughness. The disclosed apparatus and methodology may apply as well to other applications in general CD metrology of 1D or 2D features, such as the precise measurement of feature width (CD) and edge position or placement.

[0124] It is first noted that the ALM (and similar models as well) has two types of input parameters, namely material-dependent parameters and geometry parameters. Material-dependent parameters include parameters such as forward and backscatter distances, while geometry parameters include parameters such as feature width and pitch. In one implementation, for a repeated edge detection application, the material parameters will be fixed and only the geometry parameters will vary. In the simplest case (that is, for simple edge detection), it is assumed that only the edge positions for the feature are changing, such that sidewall angle, corner rounding, etc., are assumed to be constant. Thus, the use of a linescan model for edge detection in edge detection system 700 involves two steps: 1) calibrating the parameters that are assumed to be constant across the entire image, and then 2) finding the feature edge positions that provide a best fit of the measured linescan to the linescan model for each measurement.

[0125] In one implementation, in the first step, calibration is accomplished by comparing the linescan model to rigorous Monte Carlo simulations. The goal in this step is to find material parameters over the needed range of applications, and to ensure the fitting is adequate for the needed range of feature geometries. When finished, this calibrated linescan model can serve as the starting point for the generation of an inverse linescan model. The Inverse Linescan Model (ILM) should be calibrated to the specific SEM images that are to be measured. Since image grayscale values are only proportional to secondary electron signals, at the very least a mapping to grayscale values is required. In real-world applications, material properties in the experimental measurement will not be identical to those assumed in the Monte Carlo simulations such that some calibration of those parameters will also be required.

[0126] 7. Calibration of the Inverse Linescan Model

[0127] Before using the ILM for edge detection, the ILM is first calibrated. Some parameters of the model (such as material-dependent parameters) are assumed to be constant for the entire image. However, geometry parameters, such as the positions of the edges, feature width and

pitch, are assumed to vary for every linescan. The goal of ILM calibration is to determine the parameters that are constant for the whole image, regardless of the exact positions of the feature edges. It is a further goal of ILM calibration to accurately determine these parameters in the presence of image noise. These goals are accomplished by averaging along an axis of symmetry for the feature being measured, thus averaging out both the image noise and the actual feature roughness.

[0128] By averaging the linescan along an axis of symmetry (such as the direction parallel to a long line or space feature), information about the actual edge positions is lost, but information about the material parameters of the linescan model remain. Further, noise in the image is mostly averaged out in this way. Calibrating the ILM to the average linescan produces a set of material parameters (or any parameters assumed constant throughout the image) specific to this image.

[0129] Many features to be measured exhibit an axis of symmetry appropriate for ILM calibration. For example, a vertical edge has a vertical axis of symmetry. Averaging all pixels in a vertical column of pixels from the image will average away all vertical variation, leaving only horizontal information, in a direction perpendicular to the edge of the feature. The result of this averaging is a one-dimensional linescan called the average linescan. Likewise, a nominally circular contact hole or pillar is ideally radially symmetric. Averaging through polar angle about the center of the feature will produce an average linescan that removes noise and roughness from the image. An elliptical hole shape can also be so averaged by compressing or expanding the pixel size in one direction in proportion to the ratio of major to minor axes of the ellipse. Other axes of symmetry exist for other features as well.

[0130] One measured image (for example, one SEM image) may contain one or more features in the image. For example, FIG. 1A shows multiple vertical line features and multiple vertical space features. FIG. 1B shows multiple contact holes. For such a case, each feature can be separately averaged along an axis of symmetry to form an average linescan for that feature. For the example of FIG. 1A, the SEM image can be partitioned into vertical stripes, each stripe containing only one line feature, where the stripe extends horizontally from approximately the center of one space to approximately the center of the next space. For the example of FIG. 1B, the image can be partitioned into separate rectangular regions, each containing exactly one contact hole with the center of the contact hole approximately coinciding with the center of the rectangular region. The averaged linescan for that contact hole is then determined from that

rectangular region of the image. Alternately, each of the averaged linescans from each feature in an image can themselves be averaged together to form a single averaged linescan applicable to the entire image.

[0131] For a repeated edge detection application (such as the detection of all the edges on a single SEM image), the material parameters will be fixed and only the geometry parameters will vary. In the simplest case (that is, for simple edge detection), one can assume that only the edge positions for the feature are changing, so that feature thickness, sidewall angle, corner rounding, etc., are assumed constant. Thus, the use of the ILM for edge detection will involve two steps: calibrating one time for the parameters that are assumed to be constant (i.e., material and fixed geometry properties) using the average linescan, and then finding the feature edge positions that provide a best fit of the measured linescan to the linescan model for each linescan. Optionally, calibration is first accomplished by comparison of the linescan model to rigorous Monte Carlo simulations, as has been previously described. The goal of this initial step is to find material parameters over the needed range of applications, and to ensure the model is adequate for the needed range of feature geometries. When finished, this partially calibrated linescan model must still be fully calibrated to the specific SEM images that are to be measured using the average linescan.

[0132] Once the ILM has been calibrated to the given SEM image or sets of images, it is then used to detect edges. Due to the non-linear nature of linescan models such as the ALM model, numerical inversion is needed, for example using non-linear least-square regression to find the values of the left and right edge positions that best fit the model to the data. For simpler linescan models, a linear least-squares fit may be possible. Other means of “best fit” are also known in the art. The ILM as an edge detector allows the detection of edges in a high noise environment without the use of filters. FIGs. 13A and 13B demonstrate the reliable detection of edges for a very noisy image without the use of any filtering or image smoothing. More particularly, FIG. 13A is an original SEM image of a pattern structure that exhibits 18 nm lines and spaces before edge detection with an ILM. FIG. 13B is the same image after edge detection using an ILM.

[0133] Gaussian filters are common image smoothing filters designed to reduce noise in an image. Other filters such as box filters and median filters are also commonly used for this purpose. To illustrate the impact of image filtering on roughness measurement, TABLE 1 below shows the measured 3σ linewidth roughness (LWR) as a function of Gaussian filter x-

and y-width (in pixels). For each case, the ILM edge detection method was used, so that the difference in the resulting LWR is only a function of the image filter parameters. The range is almost a factor of two, showing that many different roughness measurements can be obtained based on the arbitrary choice of filter parameters. In all cases, the ILM edge detection was used. If a conventional threshold edge detection method is used, the range of resulting 3σ roughness values is much greater (TABLE 2). Similar results are obtained if other filter types (box or median, for example) are used.

[0134] TABLE 1

[0135] The raw (biased) 3σ LWR (nm) as a function of Gaussian filter x- and y-width (in pixels), using ILM edge detection.

	y-width = 1	y-width = 2	y-width = 3	y-width = 4
x-width = 1	4.99	4.67	4.03	3.82
x-width = 3	4.92	4.02	3.48	3.28
x-width = 5	4.85	3.82	3.28	3.00
x-width = 7	4.79	3.69	3.13	2.84
x-width = 9	4.73	3.59	3.08	2.80
x-width = 11	4.68	3.54	3.07	2.80

[0136] TABLE 2

[0137] The raw (biased) 3σ LWR (nm) as a function of Gaussian filter x- and y-width (in pixels), using conventional threshold edge detection.

	y-width = 1	y-width = 2	y-width = 3	y-width = 4
x-width = 1		11.17	8.52	7.28
x-width = 3	9.58	5.22	4.02	3.72
x-width = 5	8.12	4.62	3.83	3.49
x-width = 7	7.44	4.50	3.78	3.42
x-width = 9	7.03	4.45	3.77	3.41
x-width = 11	6.77	4.44	3.77	3.41

[0138] While the arbitrary choice of image filter parameters has a large impact on the measurement of roughness of the pattern structure, the impact of threshold value depends on the specific edge detection method used. For the case of a simple threshold edge detection

after image filtering, there is one threshold value that minimizes the 3σ roughness measured, with other values changing the roughness quite dramatically (see FIG. 14). For the case of the ILM, the choice of threshold has almost no impact on the measured LWR (in FIG. 14, the LWR varies from 5.00 nm to 4.95 nm as the threshold is changed from 0.25 to 0.75). Thus, for the conventional prior art method of detecting edges the arbitrary choice of threshold value can cause a large variation in the measured roughness. For the ILM, there are essentially no arbitrary choices that affect the measurement of roughness.

[0139] While the disclosed ILM system achieves accurate detection of edges in the presence of high levels of noise, the noise still adds to the measured roughness. For a linescan of a given edge slope, uncertainty in the grayscale values near the line edge translates directly into uncertainty in the edge position. A major difference, though, is that the impact of noise can be measured for the case without filtering. The noise floor of an unfiltered image can be subtracted out from the PSD (power spectral density), producing an unbiased estimate of the PSD (and thus the roughness). For the case of a filtered image, the noise floor is mostly smeared away, so that it cannot be detected, measured, or removed.

[0140] FIGs. 15A and 15B show LER power spectral densities from many rough features with right and left edges combined separately. More specifically, FIG. 15A shows raw PSDs after edge detection using the disclosed ILM technology, while FIG. 15B shows PSDs after noise subtraction.

[0141] Consider the results shown in FIG. 15A, where the line-edge roughness (LER) for the left and right edges of a feature on a pattern structure are compared. The raw PSDs indicate that the two edges behave differently. However, these differences are an artifact of the SEM, caused by a scan-direction asymmetry (such as charging) that makes the right linescan slope lower than the left linescan slope. In fact, there is no difference between right and left edge on the wafer for this sample. By measuring the noise floor for each edge separately, subtracting the noise produces a common left/right LER (FIG. 15B) that is an unbiased estimate of the true PSD.

[0142] Once the noise has been subtracted, reliable analysis of the PSD can lead to reliable estimates of the important roughness parameters, such as the zero-frequency $PSD(0)$, the correlation length λ , and the roughness exponent H . The unbiased 3σ roughness can also be

obtained. Without removing the noise, extraction of these parameters from the empirical PSD is problematic and prone to systematic errors.

[0143] 8. Unbiased Measurement of PSD

[0144] The biggest impediment to accurate roughness measurement is noise in the CD-SEM image. Among other noise sources, SEM images suffer from shot noise, where the number of electrons detected for a given pixel varies randomly. For the expected Poisson distribution, the variance in the number of electrons detected for a given pixel of the image is equal to the expected number of electrons detected for that pixel. Since the number of detected electrons is proportional to the number of electrons that impinge on the sample location represented by that pixel, relative amount of noise can be reduced by increasing the electron dose that the sample is subjected to. For some types of samples, electron dose can be increased with few consequences. But for other types of samples (such as photoresist), high electron dose leads to sample damage (resist line slimming, for example). Other types of samples, such as biological specimens, can also suffer from electron damage. Thus, to prevent sample damage electron dose is kept as low as possible, where the lowest dose possible is limited by the noise in the resulting image.

[0145] FIG. 16 shows portions of three SEM images of nominally the same lithographic features taken at different electron doses. More specifically, FIG. 16 shows portions of SEM images of nominally identical resist features with 2, 8, and 32 frames of integration (respectively, from left to right). Doubling the frames of integration doubles the electron dose per pixel. Since the dose is increased by a factor of 4 in each case, the noise goes down by a factor of 2.

[0146] SEM image noise adds to the actual roughness of the patterns on the wafer to produce a measured roughness that is biased higher. Typically, we obtain a biased roughness as given by EQUATION 8A.

$$\sigma_{biased}^2 = \sigma_{unbiased}^2 + \sigma_{noise}^2 \quad (8A)$$

[0147] where σ_{biased} is the roughness measured directly from the SEM image, $\sigma_{unbiased}$ is the unbiased roughness (that is, the true roughness of the wafer features), and σ_{noise} is the random error in detected edge position (or linewidth) due to noise in the SEM imaging and edge detection. EQUATION 8A assumes that the noise is statistically independent of the roughness

on the feature being measured. If this is not the case, more complicated noise models can be used, as further described below. Since an unbiased estimate of the feature roughness is desired, the measured roughness can be corrected by subtracting an estimate of the noise term.

[0148] Pixel noise in the SEM creates edge detection noise depending on the shape of the expected linescan for the feature. For example, FIG. 17A shows a typical linescan (grayscale value versus horizontal position, $g(x)$) for a line feature on a wafer when there is an extremely large number of electrons so that the pixel noise is negligible. The result is the “expected” linescan, that is, the expectation value of the linescan signal from a statistical perspective. By defining a threshold grayscale level, the edge position can be determined. But noise in the grayscale values results in noise in the detected edge position. For a given grayscale noise σ_{gray} , the edge position uncertainty σ_{noise} will depend on the slope of the linescan at the edge dg/dx . For small levels of noise,

$$\sigma_{noise} \sim \frac{\sigma_{gray}}{dg/dx} \quad (8B)$$

Thus, the level of edge detection noise is a function of the pixel grayscale noise and the slope of the linescan at the feature edge.

[0149] This EQUATION 8B is strictly only valid for small levels of noise and an infinitely small pixel size. To explore the impact of greater amounts of noise and a non-zero pixel size, simulation of SEM images was employed. Perfectly smooth lines and spaces (25 nm width, 50 nm pitch) were used as inputs to the Analytical Linescan Model in order to create synthetic SEM images. Then the resulting grayscale values (which range from 0 to 255) of each pixel were treated as the mean of a normal distribution with a given standard deviation (σ_{gray}) and a random grayscale number was assigned to each pixel drawn from this normal distribution. These SEM images were then treated as experimental SEM images and measured using an inverse linescan model to detect the edge positions of each feature. The 1-sigma LER measured from these images is the detected edge position uncertainty due to the grayscale pixel noise. FIG. 17B shows the 1-sigma uncertainty in edge detection position for these perfectly smooth features in the presence of grayscale noise. In this graph, the edge detection noise, for three different X pixel sizes, is plotted as a function of grayscale noise for simulated synthetic SEM images (average of 100 images, each with 20 dense lines/space features of width 25nm and pitch 50nm). The edge detection used an inverse linescan model and the resulting line-edge roughness of the features was considered to be the edge detection noise. The result is somewhat nonlinear, with higher levels of pixel noise producing ever greater edge detection noise.

Further, smaller X pixel sizes produce lower levels of edge detection noise. In fact, the edge detection variance σ_{noise}^2 is directly proportional to the X pixel size for low levels of grayscale noise.

[0150] Pixel noise is not the only source of edge detection noise. During operation the electron beam is scanned from left to right using beam steering electronics. Errors in the beam steering can place the beam at an incorrect position, which produces an edge error. Charging of the sample during electron exposure will deflect the beam to an incorrect position. While some of the charging effects will be systematic, there will also be random or pseudo-random components that will appear as random variation in the detected edge position.

[0151] While several approaches for estimating the SEM edge position noise and subtracting it out have been proposed in the prior art, these approaches have not proven successful for today's small feature sizes and high levels of SEM image noise. The problem is the lack of edge detection robustness in the presence of high image noise. More particularly, when noise levels are high, edge detection algorithms often fail to find the edge. The solution to this problem is typically to filter the image, smoothing out the high frequency noise. For example, if a Gaussian 7x3 filter is applied to the image, then for each rectangular region of the image 7 pixels wide and 3 pixels tall, the grayscale values for each pixel are multiplied by a Gaussian weight and then averaged together. The result is assigned to the center pixel of the rectangle. Box (mean) filters and median filters can also be used and produce similar results. This smoothing makes edge detection significantly more robust when image noise is high. FIG. 17C shows an example of using a simple threshold edge detection algorithm with image filtering in the right image and without image filtering in the left image. Without image filtering, the edge detection algorithm is mostly detecting the noise in the image and does not reliably find the edge.

[0152] The use of image filtering can have a large effect on the resulting PSD and measured roughness. FIG. 18 shows the impact of two different image filters on the PSD obtained from a collection of 30 images, each containing 12 features. All images were measured using an inverse linescan model for edge detection. The power spectral densities were averaged from these 360 rough features with images preprocessed using a 7x2 or 7x3 Gaussian filter, or not filtered at all, as labelled in the drawing. As can be appreciated, the high-frequency region is greatly affected by filtering. But even the low frequency region of the PSD shows a noticeable change when using a smoothing filter. Filtering in the y-direction smoothes out high-frequency

roughness. Filtering in the x-direction lowers the slope of the linescan, which can affect measured low-frequency roughness. As will be described next, the use of image filtering makes measurement and subtraction of image noise impossible.

[0153] If edge detection without image filtering can be accomplished, noise measurement and subtraction can be achieved by contrasting the PSD behavior of the noise with the PSD behavior of the actual wafer features. We expect resist features (as well as after-etch features) to have a PSD behavior as shown in FIG. 19 as the “True PSD” (and also shown earlier in FIG. 4). Correlations along the length of the feature edge reduce high-frequency roughness so that the roughness becomes very small over very short length scales. SEM image noise, on the other hand, can often be assumed to be white noise, so that the noise PSD is flat over all frequencies. Other models of the SEM image noise are also possible, for example using linescan-to-linescan correlation to describe the noise, as further described below. Thus, at a high enough frequency the measured PSD will be dominated by image noise and not actual feature roughness (the so-called “noise floor”). Given the grid size along the length of the line (Δy), SEM edge detection white noise affects the PSD according to EQUATION 9 below:

$$PSD_{biased}(f) = PSD_{unbiased}(f) + \sigma_{noise}^2 \Delta y \quad (9)$$

[0154] Thus, measurement of the high-frequency PSD (in the absence of any image filtering) provides a measurement of the SEM edge detection noise. FIG. 19 illustrates this approach for the case of a white SEM noise model. Clearly, this approach to noise subtraction cannot be used on PSDs coming from images that have been filtered, because such filtering removes the high-frequency noise floor (see FIG. 18).

[0155] EQUATION 9 assumes a white noise model, where the noise found in any pixel of the image is independent of the noise found in any other pixel. This may not always be the case. For example, the noise in each pixel may be correlated somewhat with its nearest neighbors, affecting σ_{gray} in EQUATION 8B. Alternately, the grayscale slope in EQUATION 8B may be correlated from one row of pixels to the next, possibly caused by the interaction volume of the electrons as shown in FIG. 8. If a correlation model is assumed or measured, a suitable noise expression for the PSD can be used to replace EQUATION 9, as further described below.

[0156] FIG. 19 shows one implementation of the noise subtraction process of the disclosed edge detection apparatus and method. In the disclosed edge detection method, the method first

detects the positions of the edges using the ILM without the use of any image filtering (for example, using an inverse linescan method). From these detected edges a biased PSD is obtained, which is the sum of the actual wafer roughness PSD and the SEM noise PSD. Using a model for the SEM image noise (such as a constant white noise PSD), the amount of noise is determined by measuring the noise floor in the high-frequency portion of the measured PSD. The true (unbiased) PSD is obtained by subtracting the noise level from the as-measured (biased) PSD. The key to using the above approach of noise subtraction for obtaining an unbiased PSD (and thus unbiased estimates of the parameters $\sigma_{LWR}(\infty)$, $\text{PSD}(0)$, and ξ) is to robustly detect edges without the use of image filtering. This can be accomplished using an inverse linescan model. An inverse linescan model was used to generate the no-filter PSD data shown in FIG. 18.

[0157] An example method for subtracting white noise will now be described. First, edges are detected from a SEM image without using any image filtering (for example, using an inverse linescan model). The power spectral densities of one or more edges are calculated in the usual way. Since the PSD of a single edge is quite noisy, it is extremely valuable to measure many edges and average the PSDs. Often hundreds or thousands of edges are measured and their PSDs averaged. This averaged PSD is called the biased PSD. From the average biased PSD, the highest frequencies are inspected to determine if a flat noise floor is observed. Such a noise floor is observed whenever the y pixel size is sufficiently smaller than the correlation length of the true roughness. Typically, a y-pixel size that is 20% of the correlation length or smaller is adequate. If a noise floor is observed, the average PSD value in the flat region is calculated. This is the noise floor. This number is then subtracted from the biased PSD at every frequency to produce the unbiased PSD. The biased PSD is our best estimate of the true PSD of the roughness on the wafer.

[0158] Other SEM errors can influence the measurement of roughness PSD as well. For example, SEM field distortion can artificially increase the low-frequency PSD for LER and PPR, though it has little impact on LWR. Background intensity variation in the SEM can also cause an increase in the measured low-frequency PSD, including LWR as well as LER and PPR. If these variations can be measured, they can potentially be subtracted out, producing the best possible unbiased estimate of the PSD and its parameters. By averaging the results of many SEM images where the only common aspect of the measurements is the SEM used, determination of SEM image distortion and background intensity variation can be made.

[0159] 9. Sensitivity to Metrology Tool Settings

[0160] Not all noise in measured PSDs is white noise. White noise occurs when the measurement noise of the edge position from each linescan is completely independent of all other linescans (and in particular, its nearest neighbors). White noise occurs in the absence of correlations that connect the errors in one linescan to the errors in the neighboring linescans. Any small correlations in edge errors along the length of the line would cause “pink noise”, a noise signature that is not perfectly flat over the entire frequency region.

[0161] The settings of the SEM metrology tool can impact the measured roughness of a feature in a pattern structure. These settings include the magnification and pixel size of SEM 701. These two parameters can be changed independently by changing the number of pixels in the image (from 512×512 to 2048×2048, for example). Additionally, the number of frames of integration (the electron dose) when capturing an SEM image can be adjusted. To study the impact of this setting, the number of frames of integration can be varied from 2 to 32, representing a 16X variation in electron dose, for example.

[0162] Total electron dose is directly proportional to the number of frames of integration. Thus, shot noise and its impact on edge detection noise is expected to be proportional to the square root of the number of frames of integration. FIG. 20 shows PSDs of a particular resist feature type on a given wafer, measured with different numbers of frames of integration. In this case, the PSDs correspond to 18 nm resist lines and spaces where only the number of frames of integration was varied. SEM conditions used were 500 eV, 49 images per condition, 21 features per image, pixel size = 0.8 nm square, and image size = 1024×1024 pixels. The cases of 8 or more frames of integration produce PSDs that exhibit a fairly flat high-frequency noise region. For 2 and 4 frames of integration the noise region is noticeably sloped. Thus, the assumption of white SEM noise is only approximately true, and becomes a more accurate assumption as the number of frames of integration increases and noise level decreases. This observation has been borne out in other circumstances: High noise cases are more likely to exhibit non-flat noise floors.

[0163] FIG. 21 shows the biased and unbiased values of the 3σ linewidth roughness measured as a function of the number of frames of integration. All conditions were the same as described in FIG. 20, and error bars represent 95% confidence interval estimates. The biased roughness varies from 8.83 nm at two frames of integration to 5.68 nm at 8 frames and 3.98 nm at 32

frames. The unbiased roughness, on the other hand, is fairly stable after 6 frames of integration, varying from 5.25 nm at two frames of integration to 3.25 nm at 8 frames and 3.11 nm at 32 frames. While the biased roughness is 43% higher at 8 frames compared to 32, the unbiased roughness is only 4% higher at 8 frames compared to 32. Since the assumption of white SEM noise is not very accurate at 2 and 4 frames of integration, the noise subtraction of the unbiased measurement using a white noise model is not completely successful at these very low frames of integration. A correlated noise model can produce better noise subtraction especially for the low frames of integration, as is more fully described below. While the results shown are for LWR, similar results are obtained for the measurement of line edge roughness (LER) and pattern placement roughness (PPR).

[0164] One possible cause of correlations in edge noise would be correlations in the pixel noise. To test this possibility, isolated edges were measured in the CD-SEM. The edge allows the SEM to perform its imaging functions in a typical way, but at a distance left or right from the edge the field is flat and featureless. In this region the only variation in pixel grayscale values comes from image noise. The correlation coefficient between neighboring pixels can then be calculated. Performing these calculations, the average correlation between neighboring pixels in the x-direction was 0.12, but the average correlation in the y-direction was only 0.01, essentially zero. These correlations coefficients were determined for edges measured at 2 to 32 frames of integration. There was little variation in the pixel-to-pixel correlation as a function of the number of frames of integration. Thus, correlated pixel noise is not responsible for the pink noise observed at low frames of integration. However, it is possible that the linescan slope in equation 8B is responsible for the noise correlations.

[0165] A possible cause of noise correlations in the linescan slope comes from the interaction of the beam with the sample. Electrons striking the sample undergo a number of processes that depend on the energy of the electron and the material properties of the sample. Electrons scatter off the atoms of the sample material, release energy, change direction, and often generate a cascade of secondary electrons by ionizing the sample atoms. Occasionally electrons ricochet backwards off the atom nucleus and exit out of the sample (called backscatter electrons). Some of the lower energy secondary electrons can also escape out of the sample (frequently through the edges of a feature, see FIGs. 8A and 8B). The way in which a SEM forms an image is by detecting the number of secondary electrons and/or backscatter electrons that escape the sample for each beam position.

[0166] When forming an image using an SEM, a small spot of electrons dwells at a specific point on the sample (i.e., a pixel) while the number of escaping secondary electrons is counted by the secondary electron detector. When the spot is a long way from a feature edge, as in FIG. 8A, the number of detected secondary electrons 805 is small (and the pixel is dark). When the spot is near a feature edge, as in FIG. 8B, secondary electrons 805 from the interaction volume readily escape from the feature edge producing a bright pixel.

[0167] The interaction volume of the electrons can be one to a few tens of nanometers in diameter, depending on the beam voltage and the sample material properties. This interaction volume means that electrons impinging on one spot on the sample are influenced by the sample shape over a range determined by the interaction volume. Thus, the slope of the linescan at one row of pixels will not be independent of the slope of the linescan at neighboring pixels whenever the interaction volume radius is greater than the y pixel size. This dependency can be the cause of correlations in the noise, with a noise correlation length affected by the electron beam interaction volume.

[0168] 10. Detecting and Removing Spikes from a Power Spectral Density

[0169] In addition to noise interfering with the signal in typical images of rough features, other errors can be present in the images that have a very different frequency behavior as compared to white noise or pink noise, and as compared to the roughness being measured. Some such errors produce large but narrow spikes in a PSD. FIG. 25A shows one example of high frequency “spikes” that intermittently are found in datasets. One cause for such spikes can be electrical interference in the scanning electronics of the imaging tool. If the interference is at a frequency in a range that allows one or more interfering events within a full scan of the image, this interference can result in a slight but regular “jitter” of the scanning beam position. For highly precise scanning, even a sub-nanometer jitter can result in one or more large spikes in the measured PSD. Depending on the mechanism, such interference spikes may be present in the line-edge roughness (LER) and pattern placement roughness (PPR) but not the linewidth roughness (LWR) PSD. Alternately, the interference may cause spikes at the same frequencies in all three PSDs.

[0170] For example, electrical interference at a frequency of 50Hz or 60Hz can cause noticeable spikes in a measured PSD when the measurement tool captures images at a standard “TV” scan rate or small multiples of this rate. Additionally, electrical interference at normal

audio frequencies can cause spikes that are visible at higher PSD frequencies in typical measurement tool images.

[0171] The presence of spikes in the PSD can be undesirable for a number of reasons depending on their quantity, their amplitude, and their frequency. For the case of high-frequency spikes as seen in FIG. 25A, the spikes can affect the noise removal process described above, resulting in an overestimation of the amount of white or pink noise in the image.

[0172] PSD spikes can be caused by phenomenon other than electrical interference within the imaging tool. The object being measured may include periodic or semi-periodic structures other than the rough features that are to be measured. For example, a set of vertically oriented rough features of the object may be on top of a periodic set of horizontal features resulting in topography below the rough features that are slightly visible in the image. Such underlying topography can result in a mid-frequency spike to the PSD (with higher-frequency harmonics possible as well). FIG. 26 shows an example of this phenomenon.

[0173] Another phenomenon that can give rise to spikes in the PSD would be the presence of grains of a small size range within the material of the features on the object to be measured. Grains of similar size packed tightly together can produce a nearly periodic appearance that results in a spike in the measured PSD.

[0174] Roughness measurements can also be performed on images taken of photomask features, where said photomasks are used in a lithography process. Photomasks are typically fabricated using a direct-write lithography tool with limitations such as a non-zero address grid and rectangular shots to make up the image. For some features, such as a line oriented at 45 degrees to the direction of the writing grid of the tool used to print the photomask, the result will be small, regularly spaced jogs along the edge of the photomask feature. These jogs will produce a spike (or a main spike plus harmonic spikes) in the PSD of the measured photomask roughness.

[0175] Spikes such as those found in FIG. 26 can be very disadvantageous to the measurement of roughness parameters from the biased or unbiased PSD. FIG. 27A shows how a PSD with spikes can alter the model that is fit to the unbiased PSD, including modeling parameters such as $PSD(0)$, correlation length, and roughness exponent. In contrast, FIG. 27B shows how a PSD with spikes removed can affect the model that is fit to the unbiased PSD, including modeling parameters such as $PSD(0)$, correlation length, and roughness exponent.

[0176] For these and other reasons, it is desirable to remove spikes in the PSD when the cause of those spikes is thought to be from a mechanism different from the mechanisms that give rise to the roughness of the features being measured. In other words, it is desirable to separate the PSD artifacts caused by one mechanism (such as spikes caused by electrical interference) from the PSD artifacts caused by other mechanisms (such as the stochastic effects that give rise to roughness). This can be done much like the noise removal described above, by recognizing the different frequency signatures of the different mechanisms.

[0177] As mentioned above, white noise (or pink noise) can be separated from the true (unbiased) roughness PSD since the noise frequency signature (flat or near flat at high frequencies) is very different from the frequency signature of the true roughness (a power-law decreasing at high frequencies). Likewise these so-called spikes in the PSD have frequency signatures that are very different from the frequency signature of the feature roughness itself. In particular, a so-called spike has a high amplitude over a very narrow frequency range.

[0178] A procedure for detecting and removing spikes will now be described. First, the definition of a “spike” can be established as being a frequency response that rises and falls over a frequency range smaller than a threshold (the “threshold range”) and has a height great than a threshold (the “threshold height”).

[0179] Next, a baseline can be established as being the best estimate of the PSD without the spike. For example, the threshold range for spike detection can be set to three frequency increments in the PSD data (which typically is sampled at a constant frequency increment). Other threshold ranges are also possible. A baseline can be determined by smoothly connecting PSD values separated by the threshold range plus one increment (using a straight line on a linear or logarithmic scale, for example, or by using a model for the expected PSD behavior). This baseline is then subtracted from the actual PSD data within this threshold range to arrive at an estimate of the non-baseline PSD behavior within this frequency range. If the non-baseline PSD behavior rises to a value greater than the threshold height (expressed either in absolute terms or as a multiple of the baseline PSD value), then a spike has been identified. To remove the spike, the calculated baseline behavior can be used to replace the actual PSD values within the threshold range. A search for spikes can cover the entire PSD frequency range if desired.

[0180] The threshold range can be chosen in such a way as to only detect (and possibly remove) spikes that occur due to specific types of mechanisms. For example, interference at exactly a single frequency will most likely cause a spike in the PSD that is up to two frequency increments wide (since the spike is unlikely to be at a frequency that exactly coincides with the sampled frequencies of the PSD). A threshold range of two to three frequency increments wide will be effective in detecting such “single frequency” interference events. A wider threshold range will detect other, broader-band interference events.

[0181] The threshold height can also be adjusted based on the mechanisms that are desired to be detected. But the minimum threshold height is also a function of the overall noise in the PSD. Since a PSD measures, by definition, the randomness in a random rough sample, PSD measurement is inherently noisy. It is well known that the PSD of a single measured feature has a statistical uncertainty of 100% (1-sigma). That is, the statistical uncertainty in any given PSD value at any given frequency is 100% for the measurement of a single feature. For that reason, many features are typically measured and averaged together so that the uncertainty in the PSD can be reduced by one over the square root of the number of features being measured.

[0182] But for any given number of features measured and averaged, the PSD will have a statistical uncertainty that is inherent in the sample size. The threshold height for spike detection should be chosen to be significantly higher than the inherent noise level of the PSD. Otherwise, the detection of spikes would be frequently triggered not by physical spikes but rather by noise in the PSD data. Alternately, the threshold height can be chosen to be a multiple of the measured or calculated PSD noise (for example, 5X).

[0183] FIG. 25A shows several PSDs (linewidth roughness (LWR PSD 2502), line-edge roughness (LER PSD 2504), and pattern placement roughness PPR PSD 2506) which exhibit several high-frequency spikes (spike artifacts 2507). FIG. 25B shows the same PSDs (e.g., LWR PSD 2502 as LWR PSD 2508, LER PSD 2504 as LER PSD 2510, and PPR PSD 2506 as PPR PSD 2512) with the spikes removed using the procedure outlined in the previous paragraphs. For this removal, the threshold range was set to three frequency increments, and the threshold height was set to be three times the baseline PSD value. Effective removal of the spikes was accomplished using these settings.

[0184] FIGs. 27A and 27B show another case of spike removal, this time for mid-frequency spikes. The left-hand graph FIG. 27A shows the PSDs (biased and unbiased) before spike

removal. The presence of the spikes has a deleterious effect on the modeling the PSD and the extraction of PSD measured values. The right-hand graph, FIG. 27B shows the same PSDs with the spikes removed using the procedure outlined in the previous paragraph. For this removal, the threshold range was set to three frequency increments, and the threshold height was set to be three times the baseline PSD value. Effective removal of the spikes was accomplished using these settings. The resulting PSD modeling and PSD measurement more accurately reflects the feature roughness PSD behavior excluding the mechanism that gave rise to the spikes.

[0185] An alternate procedure of removing spikes will now be described. Spikes can be removed from a PSD by passing the PSD through a low-pass filter. Using well-known techniques, the PSD can be Fourier transformed, multiplied by a low-pass frequency filter, then inverse Fourier transformed. The cut-off frequency of the low-pass filter can be set to only filter away spikes narrower than a set limit. Other approaches to low-pass filtering known in the field can also be applied.

[0186] Other methods for detecting and removing spikes based on the different frequency characteristics of a spike compared to the more slowly varying true roughness PSD will be known to those skilled in the art.

[0187] Referring to FIG. 7, the Information Handling System (IHS 750) can be modified to include the detection and/or removal of spikes using one of the exemplary methods described here. Information about each detected spike, such as its center frequency, amplitude, area, and/or width, can be recorded and output to output device 770. This information can be useful for identifying the root cause of the spike formation and thus can assist in the process of reducing or eliminating such root cause mechanism.

[0188] 11. Detection and Measurement of PSD Bumps

[0189] Other phenomenon can give rise to PSD behavior that appears as a “bump” in the PSD that otherwise has the typical shape shown in FIG. 3. Such bumps generally occur at relative low frequencies. These bumps are distinguished from spikes by covering a relative wide range of frequencies, as opposed to the narrow frequency confines of a spike. FIGs. 28A and 28B show two examples of this so-called bump behavior in PSD, labeled as Bump Type I and Bump Type II.

[0190] Bump Type I (FIG. 28A) is a large rise in the low-frequency PSD behavior above what would normally be considered the flat low-frequency regime characterized by $PSD(0)$. Several mechanisms can give rise to this bump, such as the presence of photomask roughness that is then transferred to the wafer during a photolithography step. Uncompensated field distortions in the imaging tool used to capture the images being measured can also give rise to this kind of bump. Other mechanisms are possible as well.

[0191] Bump Type II (FIG. 28B) occurs at low-to-mid frequencies such that the PSD behavior at frequencies higher and lower than the bump follows the expected behavior (as seen, for example, in FIG. 3). When this type of PSD bump is found in the line-edge roughness PSD but not in the linewidth roughness PSD, the effect is sometimes called “wiggle” since it can be noticeable as a wiggle in the feature of the image. Such wiggle can be caused, for example, from stress or tension in the films used to make the features. Photolithography and subtractive etching of the film to form the features can relieve stress and allow the relaxed remaining film to wiggle. Other mechanisms for causing wiggle are also possible.

[0192] Like white noise and spikes, bumps in the PSD are thought to arise through mechanisms separate from the stochastic mechanism that gave rise to the rest of the PSD. Thus, it is desirable to separate out the effects of the bump from the rest of the PSD. It is possible to use a procedure similar to spike detection and removal for bump detection and removal. However, this approach becomes problematic when the width of the bump is large due to the difficulty in defining a baseline PSD behavior over a large frequency range. While the larger frequency range of the bump means it is possible to distinguish bumps from spikes, it also means that different procedures for detecting and measuring bumps are likely required.

[0193] A separate technique of bump detection, measurement, and removal involves the use of a model for the bump. Like white noise and pink noise, the bump model adds directly to the typical PSD of the feature roughness. Thus, the bump model can be fit to the PSD simultaneously with the typical PSD model that does not include bump behavior.

[0194] A useful form for a bump model is given in Equation 10 below:

$$PSD_{bump}(f) = Ae^{-(f-f_c)^2/2\sigma_w^2} \quad (10)$$

[0195] where A is the amplitude of the bump, f_c is the center frequency of the bump, and σ_w is the width of the bump. For a Type I bump (FIG. 28A), the center frequency can be zero.

Other models may also be used. Alternate parameterizations of the model can also be used, such as the area and center frequency of the bump.

[0196] The area of the bump above the baseline PSD, as determined for example from the best fit model, is a useful measure of the magnitude of the phenomenon that gave rise to the bump. For example, for the case of wiggle (a Bump Type II example, FIG. 28B), the area represents the variance of the wiggle that adds to the variance caused by stochastic roughness. In other words, this approach for bump detection and measurement allows the total variance of the feature to be separated into a wiggle variance plus a stochastic roughness variance.

[0197] Referring to FIG. 7, the IHS 750 can be modified to include the detection and/or removal of bumps using one of the exemplary methods described here. Information about each detected bump, such as its center frequency, amplitude, area, and/or width, can be recorded and output to output device 770. This information can be useful for identifying the root cause of the bump formation and thus can assist in the process of reducing or eliminating such root cause mechanism. By subtracting the bump behavior from the total PSD, the remaining PSD can be characterized (using, for example, parameters such as PSD(0), correlation length, and roughness exponent) so that this remaining PSD reflects more accurately the mechanisms that gave rise to the PSD exclusive of the bump mechanism.

[0198] Referring now to FIG. 30, an example method 3000 to detect undesired spikes in a PSD dataset, and for removing spikes in a PSD dataset is illustrated. The method 3000 starts (block 3002) and generates, using an imaging device, a set of one or more images, each image of the set including an instance of a feature within a respective pattern structure, each image including measured linescan information corresponding to the pattern structure that includes noise (block 3004). Next the method proceeds to detect edges of the features within the pattern structure of each image of the set without filtering the images (block 3006) and generates a power spectral density (PSD) dataset representing feature geometry information corresponding to the edge detection measurements of the set of images (block 3008). If desired, an unbiased PSD data set can be generated from the biased PSD data set by subtracting SEM noise. Next, the method defines a threshold range and a threshold height (block 3010) and generates a baseline for a portion of the PSD dataset, by smoothly connecting a first PSD value of the portion of the PSD dataset to a second PSD value, wherein the first PSD value and the second PSD value are separated by the threshold range (block 3012), determines that a difference between a third PSD value of the portion of the PSD dataset and the baseline is greater than a

threshold height (block 3014), and replaces the portion of the PSD dataset with the baseline for the portion of the PSD dataset (block 3016). Thereafter, the method ends (block 3018).

[0199] Referring now to FIG. 31, an example method 3100 to model bumps in a PSD dataset is illustrated. The method 3100 starts (block 3102) and generates, using an imaging device, a set of one or more images, each image of the set including an instance of a feature within a respective pattern structure, each image including measured linescan information corresponding to the pattern structure that includes noise (block 3104). Next the method proceeds to detect edges of the features within the pattern structure of each image of the set without filtering the images (block 3106) and generate a biased power spectral density (PSD) dataset representing feature geometry information corresponding to the edge detection measurements of the set of images (block 3108). If desired, an unbiased PSD data set can be generated from the biased PSD data set by subtracting SEM noise. A first bump is evaluated in the PSD dataset to create a bump model (block 3110); and fits a typical PSD model and the bump model to the PSD dataset to create a best fit model (block 3112). Thereafter, the method ends (block 3114).

[0200] The flowcharts of FIG. 30 and FIG. 31 include the steps that can be performed using the edge detection system 700 depicted in FIG. 7, including certain steps that can be carried out by the SEM 701 and certain other steps that can be carried out by the IHS 750 and its included processor 755 and storage 760, both as described in detail herein. Instructions can be stored in storage 760 that, when executed by the processor, cause the processor to perform the methods disclosed herein and described by the flowcharts of FIG. 30 and FIG. 31, in analogous fashion as other instructions stored in storage 760 that implement the inverse linescan model metrology tool 765 described herein.

[0201] 12. Influence of Pixel Size and Magnification

[0202] With respect to the pixel size and magnification employed by SEM 701, FIGs. 22A and 22B show the biased and unbiased power spectral densities (PSDs), respectively, for a pattern of 16 nm lines and spaces for different magnifications and pixel sizes, assuming a white noise model. For a given number of frames of integration, changing the pixel size changes the electron dose per unit wafer area and the noise in the SEM image. Under this range of conditions, the biased LWR varied by 0.63 nm (14%), while the unbiased LWR varied by only

0.07 nm (2%). The unbiased LWR is essentially unaffected by these metrology tool settings. Similar results are obtained for the measurement of LER and PPR.

[0203] FIGs. 22A and 22B show power spectral densities as a function of pixel size and magnification. More particularly, FIG. 22A shows the biased LWR PSD and FIG. 22B shows the unbiased LWR PSD after noise has been measured and subtracted off. The SEM conditions for these results used a landing energy of 500 eV, 3 images per condition, and 16 nm resist lines and spaces.

[0204] TABLE 3 below shows the measured PSD parameters for the PSDs shown in FIGs. 22A and 22B.

[0205] TABLE 3

[0206] Biased and unbiased 3σ LWR (nm) measurements as a function of pixel size and magnification.

	Pixel 0.8nm 82kX	Pixel 0.8nm 164kX	Pixel 0.5nm 130kX	Pixel 0.5nm 264kX	Pixel 0.37nm 180kX
Biased LWR (3-sigma, nm)	5.10	4.99	4.67	4.61	4.47
Unbiased LWR (3-sigma, nm)	3.66	3.65	3.70	3.67	3.63
Unbiased LWR PSD(0) (nm ³)	15.95	16.18	17.2	16.25	16.35
LWR Correlation Length (nm)	5.08	5.05	5.31	5.11	5.38

[0207] It has been found that the difference between biased and unbiased LWR is not constant, but varies with metrology tool settings, feature size, and process. Likewise, the ratio between biased and unbiased LWR varies with metrology tool settings, feature size, and process. TABLE 4 below shows the difference and ratio of biased to unbiased LWR for a variety of conditions. For these conditions, the ratio of biased to unbiased LWR varies from

1.09 to 1.66. The difference between biased and unbiased LWR varies from 0.32 nm to 2.19 nm in this particular example.

[0208]

TABLE 4

[0209] The relationship between biased and unbiased LWR for a variety of processes.

Process	3σ LWR: Biased/Unbiased	3σ LWR (nm): Biased – Unbiased
193i litho, 84nm pitch, 500V, 512 rect pixels	1.20	0.76
193i etch, 84nm pitch, 800V, 512 rect pixels	1.14	0.43
EUV litho, 32nm pitch, 500V, 2048 0.8nm pixels	1.39	1.44
EUV litho, 32nm pitch, 500V, 1024 0.8nm pixels	1.37	1.34
EUV litho, 32nm pitch, 500V, 2048 0.5nm pixels	1.26	0.97
EUV litho, 32nm pitch, 500V, 1024 0.5nm pixels	1.26	0.94
EUV litho, 32nm pitch, 500V, 1024 0.37nm pixels	1.23	0.84
EUV litho, 36nm pitch, 500V, 1024 0.8nm pixels	1.52	1.86
EUV litho, 32nm pitch, 500V, 1024 rect pixels	1.66	2.19
EUV etch, 32nm pitch, 800V, 1024 rect pixels	1.09	0.32

[0210] 13. Edge Detection implementations

[0211] FIG. 23 is a flowchart that depicts a representative overall process flow that the disclosed SEM edge detection system employs to detect edges of a pattern structure. For discussion purposes, the process described in the flowchart of FIG. 23 is applied to sample 2400 of FIG. 24A. Sample 2400 is a pattern structure that may also be referred to as pattern structure 2400. The flowchart of FIG. 23 includes the steps carried out by inverse linescan model metrology tool 765 to determine the edges of the pattern structure.

[0212] Process flow commences at start block 2300 of FIG. 23. As seen in FIG. 7, the IHS 750 is coupled to SEM 701 to receive SEM linescan image information from SEM 701. IHS 750 includes a processor 755 and storage 760 coupled thereto. Storage 760 may include volatile system memory and non-volatile permanent memory such as hard drives, solid state storage devices (SSDs) and the like that permanently store applications and other information. Storage 760 stores the inverse linescan model (ILM) metrology tool 765 disclosed herein and described by the flowchart of FIG. 23. SEM 701 includes a controller (not shown) that the IHS 750 instructs to perform image acquisition on pattern structure 800 and that provides linescan information from SEM 701 to IHS 750.

[0213] As per block 2305, SEM 701 sends an SEM image of pattern structure 800 to IHS 750, and in response, IHS 750 loads this SEM image into system memory within storage 760. IHS 750 preprocesses the pattern structure image from the SEM 701, as per block 2310. For example, this preprocessing of the loaded SEM image may include adjusting grayscale values and subtracting out background tilts of intensity levels. Optionally, as per block 2315, IHS 750 may perform filtering of the loaded image, although this is generally not preferred.

[0214] In the case of a pattern structure such as the vertical lines and spaces seen in the pattern structure 2400 of FIG. 24A, the inverse linescan model metrology tool 765 averages vertically over the axis of symmetry to generate an average linescan, as per block 2320. An average linescan may be a grayscale value as a function of horizontal position wherein all of the vertical pixels have been averaged together. This averages out much of the SEM noise contained in the SEM image and produces a linescan that is more representative of the physical processes that generate a linescan without noise. FIG. 24B shows a single linescan at one Y-pixel position. FIG. 24C shows the averaged linescan that is generated by averaging over all Y-pixels.

[0215] While the example shown here is for vertical lines and spaces, any pattern with an axis of symmetry can be so processed to produce an average linescan. For example, long lines, long spaces, or long isolated edges can be so processed whenever the length of the line is sufficient to allow adequate averaging. Contact holes or pillars, with circular or elliptical symmetry, can also be averaged in a radial direction to produce an average linescan.

[0216] As per block 2325, tool 765 calibrates the inverse linescan model to the averaged linescan that was obtained in the manner described above. It is noted that the linescan model includes two kinds of parameters, namely 1) parameters that depend upon the materials and the properties of the SEM, and 2) parameters that depend on the geometry of the feature on the sample. Tool 765 can calibrate all of these parameters. Tool 765 finds the best fit of the model to the average linescan of FIG. 24C, as per block 2325. The values of the best fit parameters of the model are then the calibrated values.

[0217] That calibrated model is applied to a single linescan as shown in FIG. 24B. The best fit of the model to the single linescan of FIG. 24B is found, however, in this case the inverse linescan model metrology tool 765 fixes all of the parameters that relate to the materials and SEM imaging tool. In this scenario, tool 765 varies only the parameters related to the geometry of the feature of the pattern structure in order to find the best fit of the calibrated model to a single linescan.

[0218] In a simplified scenario, the only parameters varied in block 2330 would be the positions of the edges of the feature. In one implementation, it is assumed that the vertical dimension of the feature exhibits a predetermined thickness and that only the edge positions of the feature are varying. Next, the calibrated inverse linescan model is fit to every single horizontal cut through the 2D image of the feature, as per block 2330. We take the top horizontal row of pixels, and then the next row of pixels that are one pixel down, and then the next horizontal row of pixels down, and so forth. An example of one such single linescan is shown in FIG. 24B. The resulting best fit edge positions are the detected edges.

[0219] After the edges of the feature are detected in the manner described above, tool 765 may detect that the sample was rotated slightly during image acquisition, resulting in parallel tilted lines (that is, lines that are not perfectly vertical). Such tilting or rotation may contribute to inaccuracy of the detected edges by changing the average linescan and thus the calibrated ILM. Image rotation can be detected by fitting all the edges in the image to a set of parallel

lines and determining their slope compared to vertical. If the slope is sufficiently different from the vertical case, the rotation should be removed. One possible criterion would be to compare the pixel position of the best fit line at the top of the image to the pixel position of the best fit line at the bottom of the image. If these pixel positions differ by some threshold, such as two pixels, then the image rotation is considered to be sufficiently large that its removal is required.

[0220] If such tilting/rotation is detected, as per block 2335, then the prior calibration is considered to be a first pass calibration and calibration is repeated. More particularly, if such tilting/rotation is detected, the rotation is subtracted out by shifting some rows of pixels to bring the edges into vertical alignment, as per block 2345, and calculating a new average linescan. Calibration of the model is then repeated as per block 2350 and 2325. Another fitting is performed as well, as per block 2330. Ultimately, tool 765 outputs geometry feature information (such as edge positions) describing the geometry of the feature that corresponds to the linescan image information provided to tool 765.

[0221] Like image rotation, the roughness of the features themselves contributes inaccuracies to the calibration of the ILM. Optionally, after a first pass edge detection, each row of pixels can be shifted to not only subtract out image rotation, but to subtract out the feature roughness as well. The final result after the shifting of each row of pixels is a vertical edge where the edge position varies by less than one pixel from a perfect vertical line. These shifted rows of pixels can then be averaged vertically to produce a more accurate average linescan for use in ILM calibration.

[0222] In actual practice, the IHS 750 may include an interface 757 coupled between processor 755 and an output device 770 such as a display, printer, or other user interface so that the user may observe the feature edges determined by inverse linescan model metrology tool 765. Interface 757 may be a graphics interface, a printer interface, network interface, or other hardware interface appropriate for the particular type of output device 770.

[0223] While the implementations described above make reference to the measurement of structures found on semiconductor wafers, as used in the manufacture of semiconductor devices, the invention is not limited to these applications. The present disclosure can be usefully employed to measure the roughness of feature edges found on flat panel displays, microelectromechanical systems, microfluidic systems, optical waveguides, photonic devices,

and other electronic, optical, or mechanical devices. Further, the invention can be used to measure the feature edge characteristics of naturally occurring structures such as crystals or minerals, or manmade structures such as nanoparticles or other nanostructures. Further, the invention can be used to measure the feature edge characteristics of biological samples as well.

[0224] While the implementations described above make reference to measurements using a scanning electron microscope, the invention is not limited to that imaging tool. Other imaging tools, such as optical microscopes, stimulated emission and depletion (STED) microscopes, x-ray microscopes, transmission electron microscopes (TEM), focused ion beam microscopes, and helium ion microscopes, can also be used. Other forms of microscopes, such as scanning probe microscopes (atomic force microscopes) (AFMs) and scanning near-field optical microscopes (SNOMs), for example can be used as well.

[0225] 14. Prior Art Methods for Process Window Determination

[0226] In lithography, two major variables are frequently adjusted to maintain control of the final printed features: exposure dose (also called exposure energy, the product of light intensity and exposure time) and focus (often described as the position of the plane of best focus of the projected image relative to the top surface of the photoresist as coated on the wafer). Both of these process variables can be changed using the lithography projection imaging tool employed to print the features (such as a stepper, scanner, or step-and-scan tool). Other lithography techniques such as direct-write patterning, proximity printing, and electron-beam lithography can also adjust exposure dose and focus or their equivalent.

[0227] In order to obtain the best lithographic results, it is desirable to set the exposure dose and focus of the lithography tool to optimal values, sometimes called best dose and best focus. The meaning of “best” can vary depending on the goals of the lithography process. One common goal would be to have the mean feature size of a collection of specified printed patterns on the wafer or wafers match the target value for that feature size, to within some tolerance. Another goal would be to minimize the variation of the feature size across some collection of features. Another goal would be to improve the pattern fidelity of the printed patterns. Another goal would be to minimize the sensitivity of the printed feature size or fidelity to process variations across some collection of features.

[0228] The collection of features examined could include features printed at different points of the imaging field (or different points within the scanner slit area for the case of scanning

lithography), different points across the chip being manufactured, different points across the wafer, and across different wafers or different lots of wafers. The collection of features examined could also include features printed at different times. The collection of features examined could also include different feature types, features that have different proximity to other features, and features with different target feature sizes.

[0229] In many lithographic processes, feature size varies monotonically with exposure dose. For example, over a small range of dose the feature size can vary approximately linearly with dose. For some processes, feature size may be approximately proportional to one over the exposure dose. For focus, feature size often varies approximately quadratically with focus. Further, focus and dose generally interact to cause a variation in the printed patterns. The impact of dose on a feature depends on the focus setting of the exposure tool, and the impact of focus on a feature depends on the exposure dose setting of the exposure tool.

[0230] Characterizing the impact of both focus and exposure dose on printed patterns is generally accomplished through the use of a focus-exposure matrix (FEM). Descriptions of the FEM can be found in the prior art, for example in the textbook *Chris A. Mack, Fundamental Principles of Optical Lithography: The Science of Microfabrication, John Wiley & Sons, (London: 2007)*.

[0231] In general, depth of focus (DOF) can be thought of as the range of focus errors that a process can tolerate and still give acceptable lithographic results. A change in focus results in two major changes to the final lithographic result: the photoresist profile changes (including its dimensions), and the sensitivity of the process to other processing errors is changed. Typically, photoresist profiles are described using parameters such as the linewidth (or critical dimension, CD), edge placement error, the sidewall angle of the profile, the final height of the feature, the line-edge or linewidth roughness, and other metrics known in the art. The variation of these parameters with focus can be determined for any given set of conditions using various measurement methods. For example, CD can be measured using a scanning electron microscope.

[0232] Additionally, as an image goes out of focus, the process becomes more sensitive to changes in other processing variables such as exposure dose, post-exposure bake time and temperature, develop time, developer concentration and temperature, underlying filmstack properties, etc. Exposure dose is commonly chosen to represent these other process responses.

The impact of focus on the sensitivity of the patterning to exposure dose can be characterized by varying both exposure dose and focus, each over some range of settings, and measuring characteristics of the printed patterns, such as CD, linewidth roughness, etc.

[0233] One way to display the combined impact of focus and exposure dose on CD is using what is called a Bossung plot, as seen in Figure 39A. What is not shown in Fig. 39A is the experimental error (or experimental uncertainty) that is always present in experimental data. Experimental error/uncertainty (often expressed as a multiple of the standard error of the measured value, such as two times the standard error) can be one to several percent of the CD at best focus and exposure, but the error can be much higher when out of focus. To better analyze focus-exposure CD data, one common approach in the art is to fit the data to a reasonable empirical equation in order to reduce data noise and eliminate flyers (also called outliers). A drawback of this approach is that the choice of the empirical expression adds a level of arbitrariness to the resulting analysis.

[0234] One output as a function of two inputs can be plotted in several different ways. For example, the Bossung curves could also be plotted as exposure latitude curves (linewidth versus exposure dose) for different focus settings. Another very useful way to plot this two-dimensional data set is a contour plot – contours of constant linewidth versus focus and exposure (Figure 39B). For noisy experimental data, it is also common to plot a smoothed version of the data in these contour plots. For example, curve fitting functions can be used to smooth the data, as mentioned above, and then contours based on the curve fits plotted instead of contours from the raw data.

[0235] The contour plot form of data visualization is especially useful for establishing the limits of exposure and focus that allow the final image to meet certain specifications. Rather than plotting all of the contours of constant CD, one could plot only the two CDs corresponding to the outer limits of acceptability – the minimum CD and the maximum CD allowed based on the CD specifications. Because of the nature of a contour plot, other variables can also be plotted on the same graph. Figure 39C shows an example of plotting contours of CD (nominal $\pm 10\%$), 80-degree sidewall angle, and 10% resist loss all on the same graph. The result is a process window – the region of focus and exposure that keeps the final printed pattern within all applied specifications.

[0236] Many different measurement results can be combined into a process window. For feature sizes and lithography processes where stochastic variation is significant, stochastic-related results can be plotted on the process window contours. For example, linewidth roughness could be plotted using a single specification – the maximum allowed LWR. Defectivity could be used as a stochastic metric as well.

[0237] The focus–exposure process window is useful since it shows how exposure and focus work together to affect linewidth and/or other metrics used to judge the printed results. The process window can be thought of as a process capability – how the process responds to changes in focus and exposure. An analysis of the error sources for focus and exposure in a given process will give a process requirement. If the process capability exceeds the process requirements, yield of the devices being fabricated will be high. If, however, the process requirement is too large to fit inside the process capability, yield or device performance may suffer.

[0238] It is often useful to assess the maximum range of focus and exposure (that is, the maximum process requirement) that can fit inside the process window. A simple way to investigate this question is to graphically represent errors in focus and exposure as a rectangle on the same plot as the process window. The width of the rectangle represents the built-in focus errors of the processes, and the height represents the built-in dose errors. The problem then becomes one of finding the maximum rectangle that fits inside the process window.

[0239] There may be many possible rectangles of different widths and heights that are ‘maximal’, i.e., they cannot be made larger in either direction without extending beyond the process window (Figure 40A). Each maximum rectangle represents one possible trade-off between tolerance to focus errors and tolerance to exposure errors. Larger DOF can be obtained if exposure errors are minimized. Likewise, exposure latitude can be improved if focus errors are small. The result is a trade-off between exposure latitude and DOF.

[0240] If all focus and exposure errors were systematic, then a graphical (or geometric) representation of those errors would be a rectangle. The width and height would represent the total ranges of the respective errors. If, however, the errors were randomly distributed, then a probability distribution function would be needed to describe them. It is common to assume that random errors in exposure and focus are caused by the summation of many small sources

of error, so that by the central limit theorem the overall probability distributions for focus and dose will be approximately Gaussian (a normal distribution).

[0241] In order to graphically represent the errors of focus and exposure, one should describe a surface of constant probability of occurrence. All errors in focus and exposure inside the surface would have a probability of occurring which is greater than the established cutoff. For a Bi-Gaussian distribution (a Gaussian distribution for two independent variables), the surface of constant probability is an ellipse (Figure 40B).

[0242] Using either a rectangle for systematic errors or an ellipse for random errors, the size of the errors that can be tolerated for a given process window can be estimated using this geometric approach. Taking the rectangle as an example, one can find the maximum rectangle that will fit inside the process window. Figure 41 shows an analysis of the process window where every maximum rectangle is determined and its height (the exposure latitude) plotted versus its width (depth of focus). Likewise, assuming Gaussian errors in focus and exposure, every maximum ellipse that fits inside the process window can be determined. The horizontal width of the ellipse could represent a six-sigma error in focus (plus and minus three-sigma variation about best focus), while the vertical height of the ellipse would give a six-sigma error in exposure. Plotting the height versus the width of all the maximum ellipses gives the second curve of exposure latitude versus DOF in Figure 41.

[0243] The definition of depth of focus also leads naturally to the determination of best focus and best exposure. The DOF value corresponding to one point on the exposure latitude-versus-DOF curve corresponds to one maximum rectangle or ellipse that fits inside the process window. The center of this rectangle or ellipse would then correspond to best focus and exposure for this desired operating point. Knowing the optimum focus and dose values is essential to being able to use the full process window. If the process focus and dose settings deviate from this optimum, the range of focus and dose errors that can be tolerated will be reduced accordingly.

[0244] Although all of the above results describe the focus and exposure response of one critical feature, in reality a number of mask features must be printed simultaneously. For example, features of different nominal size and pitch or different proximities to other features could be present within the same design. For a such a case, the overall process window would be the overlapping of process windows for each feature size or type under consideration. Focus

and dose settings inside the overlapping process window would allow each feature type and/or size to meet specifications. A more important measure of performance than the DOF of each individual feature is the overlapping DOF of multiple critical features.

[0245] Just as multiple profile metrics were overlapped to form one overlapping process window in Figure 39C, process windows from different features can be overlapped to determine the DOF for simultaneously printing those multiple features. Figure 42 shows such an example for line/space patterns of two different pitches. Ideally, process windows for all critical feature sizes and pitches would be overlapped.

[0246] Systematic variations in patterning as a function of field position can also be accounted for with overlapping process windows. If the same feature were printed at different points in the lithography tool image field (typically the center and the four corners of the field are sufficient), process windows can be overlapped to create the usable process window for that feature. The resulting depth of focus is sometimes called the useable depth of focus (UDOF).

[0247] FEM analysis can be used in other ways as well. One possible output of process window analysis is the isofocal bias. Generally, there is one exposure dose that makes a printed feature have minimum sensitivity to focus changes. This dose is called the isofocal dose. At this dose, the printed CD is generally different from the target or desired CD. This difference is called the isofocal bias. Process windows can be used to assess the isofocal bias.

[0248] The prior art geometric methods for process window determination and analysis as described above have several problems. First, metrology errors or metrology uncertainty are not taken into account. The contours plotted in Fig. 39C or Fig. 40A depict sharp boundaries delineating the region inside the process window as separate from the region outside the process window. Such sharp boundaries produce a hard cut-off between regions that are deemed “good” (inside the process window) and “bad” (regions outside the process window). In reality, the boundary between good and bad is fuzzy due to metrology uncertainty.

[0249] A second problem with the prior art geometric methods is the need for curve-fitting or smoothing of the measurement data before constructing the process window. Just one aberrant data point near the process window edge could produce a large deviation in the contour position (Fig. 43). The result could be a very marked difference in DOF determined from this process window, and/or a large difference in best focus and exposure.

[0250] The use of smoothing or curve fitting of the original data adds a level of arbitrariness to the analysis of the data. There is no *a priori* correct amount of smoothing or correct fitting function to use, and different choices produce different results.

[0251] A third problem is the choice between using rectangles, ellipses, or some other geometric shape to represent the probabilities of focus and exposure dose errors about the best focus and dose values. As Fig. 41 shows, different choices produce different results. Further, using geometric shapes to represent these probabilities and then finding the maximum shape that fits inside the process window means that only a handful of points in the process window shape affect the outcome. As shown in Fig. 40B, the inscribed ellipse and rectangle touch the process window at only a few points. This makes the size of those inscribed shapes especially sensitive to metrology errors that might impact just those few points of the contour-based process window.

[0252] All of the above problems are heightened when overlapping more metrics of pattern size or fidelity or roughness, or overlapping process windows from different features.

[0253] A consequence of the problems of geometric process window analysis is inaccurate and/or imprecise measurement of exposure latitude and depth of focus. A consequence of inaccurate or imprecise measurement of exposure latitude or depth of focus is poor optimization of the patterning process. Often material choices (different photoresists, different underlayers below the photoresist, etc.) are made based on (or take into account) which material provides the greatest depth of focus, for example. Additionally, process tool settings, such as the size and shape of the illumination used by the optical projection tool, take into account the resulting depth of focus.

[0254] Optical Proximity Correction (OPC) is used to modify the chip design as it is represented on the photomask to improve the fidelity of the final printed patterns. Sometimes this OPC takes into account the influence of these photomask modifications on the process window. Thus, inaccurate or imprecise determination or analysis of the process window can have a deleterious impact on the results of OPC.

[0255] Another consequence of the problems of geometric process window analysis is inaccurate and/or imprecise measurement of best focus and best exposure. The output of best focus and best exposure from process window analysis can be used to control the lithographic exposure tool during manufacturing. The measured values of best focus and dose are sent to

the lithography tool and used to adjust the dose and focus settings of that tool for subsequent use in manufacturing product wafers. Inaccurate and/or imprecise measurement of best focus and best exposure could then have a negative impact on device yield or device performance as a consequence.

[0256] Further, the exposure setting of a lithography tool can have an impact on the throughput of that tool. Generally, a lower exposure dose produces higher throughput of the exposure tool. For this reason, it can be desirable to lower the exposure dose as much as possible while still maintaining adequate exposure latitude and depth of focus as determined by the process window analysis. Inaccurate and/or imprecise measurement of the process window makes such improvement in throughput difficult to achieve without putting device yield or device performance at risk.

[0257] A further use of process window analysis is for process monitoring. Best focus, best exposure, exposure latitude, and depth of focus can all be monitored over time use standard statistical techniques such as trend charting or statistical process control (SPC). Deviations of the current behavior of the charted variable from its historical behavior can be flagged for further investigation or other actions. Inaccurate and/or imprecise measurement of the process window and its related measures reduces the efficacy of trend charting or other forms of process control.

[0258] Further, the prior art geometric process window approach does not produce uncertainty estimates for the various analysis outputs such as best focus, best exposure, depth of focus, and exposure latitude. Without uncertainty estimates, the usefulness of these outputs is reduced. For example, using values of best dose and focus as determined from process window analysis to control the dose and focus settings of a scanner in manufacturing should only result in a change to the scanner settings if that change were larger than the uncertainty estimates for the process window best dose and/or focus.

[0259] For these and other reasons, there is a desire to find a different method for process window measurement and analysis that addresses one or more of the problems with the prior art process window approaches.

[0260] 15. Probabilistic Process Windows

[0261] A new method embodiment for process window measurement and analysis, and its use for process optimization and control, will now be described. This new method embodiment, called the Probabilistic Process Window (PPW), can solve many of the problems of the prior art methods (such as the geometric process window) as described above. This PPW approach can be used when two or more interacting process variables are varied in order to determine their impact on one or more output results. In the main example used below, two process variables will be discussed. Further, the main example below will use the two process variables of exposure dose and focus during the process step of lithography patterning. However, the PPW approach is general enough to handle more than two process variables dealing with lithography, etch, deposition, or other processing steps.

[0262] In some implementations, probabilistic process windows are used to account for, among other things, uncertainty in measurements to determine where probabilities of indications of associations within a lithography process meet a specification requirement. For example, instead of (or in addition to) the ILM metrology tool 765, the system 700 may include a probabilistic process window tool (not shown). A two-dimensional (2D) design of experiments (DoE) may include a list of X Values (possibly irregularly spaced) representing a first DoE factor, ordered increasing. This first DoE factor would be a process variable intentionally changed in order to determine its impact on some output or outputs. The 2D DoE may also include a list of Y Values (possibly irregularly spaced) representing a second DoE factor, ordered increasing. This second DoE factor would also be a process variable intentionally changed in order to determine its impact on that output or outputs.

[0263] The 2D DoE may also include a set of specifications s (with a minimum and/or maximum acceptable value) applied to some process response or output. For example, in a lithography process one output could be the critical dimension (CD) of a printed feature. A specification on that output would describe the minimum acceptable value for that output, and/or the maximum acceptable value for that output. As an example, specifications on the CD could be the target (desired) value $\pm 10\%$. For a target value of the CD of 20 nm, the specifications (minimum and maximum values, respectively) could be set to 18 nm and 22 nm.

[0264] For some points $(x \in X, y \in Y)$, the 2D DoE may also include a z -value (a measurement result) and a standard error corresponding to that z -value (an expression of measurement uncertainty) for each spec $s \in S$. In other words, the measurement data can be represented by two three-dimensional arrays, one for the z -values and one for the standard

errors, indexed by (x, y, s) . To determine the probability that some point (x, y) is in spec, a process can be used to generate a probabilistic process window. In some implementations, the process may include two main steps. In a first step, the probability that each given z -value is actually in spec s is determined using the standard error (or other measure of uncertainty) for that point, as further described below. In a second step, a bivariate probability distribution centered at some test points (x, y) values) across the field generated in the first step is applied to determine the probability that test point is within spec, given some errors in x and y , as further described below.

[0265] The experimental portion of PPW determination varies the two or more process variables in a matrix. Consider two process variables (called x -input and y -input here) such as exposure dose and focus. Each of exposure dose and focus are varied over some range, for example using a constant step size in dose and a constant step size in focus. Variable step sizes in dose and focus can also be used. Patterns are printed using this two-dimensional matrix of input values on one or more wafers. Those wafers are then measured to determine one or more measurement outputs (called z -outputs here). For example, CD, unbiased roughness, defectivity, local CDU, metrics of pattern shape or fidelity, or other metrics of the output of the process can be measured.

[0266] In most cases, the measurement of an output such as CD includes an estimate of the measurement uncertainty for that output. For example, a single SEM image of a wafer may include, for example, 20 lines and spaces. In such a case it is common to produce, as a measurement output, the mean CD as the desired output, that is, the mean value of the CDs of each individual line on the image. The standard error of the mean can then be used as an estimate of the measurement uncertainty of that mean CD. Other metrics of the uncertainty of the mean, as known in the field of statistics, could also be used. Sometimes a multiple of the standard error (2 times, for example) is used as an estimate of the uncertainty. Other estimates of the total measurement uncertainty can also be used.

[0267] Sometimes, for various reasons, one or more data points in the array of outputs as a function of input values can be missing. Thus, prior to the first step in PPW determination, missing data points may be filled in (their values approximated), for example, by interpolation. For example, TABLE 5 and TABLE 6 included below are measured values with missing data points. Other methods for handling missing data points may also be employed.

[0268]

TABLE 5

[0269] Mean Line CD (spec 1).

	x-value = 1	x-value = 2	x-value = 3	x-value = 4	x-value = 5
y-value = 10	20.8	22.2	23.6	25	26.9
y-value = 9	20.6	21.9	23.4	24.8	26.7
y-value = 8	20.4	21.6		24.6	26.5
y-value = 7	20.2	21.3	23.2	24.4	26.3
y-value = 6	20	21	23	24.2	26

[0270]

TABLE 6

[0271] Unbiased LER average (spec 2).

	x-value = 1	x-value = 2	x-value = 3	x-value = 4	x-value = 5
y-value = 10	9	11	13	15	17
y-value = 9	8	10	10.5	11	15
y-value = 8	7	7.5	9		14
y-value = 7	6	7	8	9	10
y-value = 6	5	5.5	6	6.5	7

[0272] With a grid for each spec with as many filled in values as can be determined, for example, by interpolation, in the first step of the process, it is determined whether an output value meets its specification (that is, the value is said to be “within specifications” or is “in spec”). In a first part of the first step of the process, for each measured output (mean z -value) and uncertainty for that output (such as σ_z , the standard error of the mean z -value), and for each spec s , the probability that output is actually in spec is determined. For example, one might assume that the distribution of the output follows a normal (i.e., Gaussian) distribution. Other probability distributions are also possible. The calculation of the probability that the output is actually in spec will now be described.

[0273] Given a mean output value z , a standard error of that mean σ_z , and a spec s (minimum and maximum spec values, for example, s_{min} and s_{max}), we want to determine the probability that the measured mean value z is in spec. Assuming, for example, a Gaussian distribution on the mean value z , the area under the normal curve can be calculated with $\mu = z$ and $\sigma = \sigma_z$ from s_{min} to s_{max} in order to determine the probability that the true value of z is in

$[s_{min}, s_{max}]$. An example normal distribution cumulative distribution function (CDF) is given in EQUATION 11 below:

$$\text{Normal Distribution CDF} = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x-\mu}{\sigma\sqrt{2}} \right) \right] \quad (11)$$

[0274] where erf is the well-known Error Function. The definition of the Error Function is given in EQUATION 12 below:

$$\operatorname{erf}(a) = \frac{2}{\sqrt{\pi}} \int_0^a e^{-t^2} dt \quad (12)$$

[0275] The probability that a z value is in spec may be determined using EQUATION 13 below:

$$P(z \text{ in spec}) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{s_{max} - z}{\sigma_z \sqrt{2}} \right) \right] - \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{s_{min} - z}{\sigma_z \sqrt{2}} \right) \right] \quad (13)$$

[0276] If there is no maximum spec value, the first term equals 1. Further, if there is no minimum spec value, the second term equals 0. FIG. 32 illustrates an example of interpreting the probability that the measured value is in spec as the appropriate area under the probability density curve.

[0277] Without taking metrology uncertainty into account, in a traditional process window approach, either an output value is in spec or not. In other words, a probability of 1 is effectively attributed to an in-spec output, and a probability of 0 is effectively attributed to an out-of-spec output.

[0278] In a second part of the first step, the probability that the output is within all of the specs is determined. The output of the first step may be called a “Probabilistic Process Window.” Given that $P(z \text{ in spec } s_i | x, y), \forall s_i \in S$, for each point (pair of input values) (x, y) , the probability that each point is within all the specs is determined. For example, based on an assumption that each P is independent (that is, the measurement uncertainty for one output response is independent of the measurement uncertainty of a second output response), all of the probabilities for all the specs are multiplied together as given in EQUATION 14 below:

$$P(z | x, y) = \prod_{s_i=0}^S P(z \text{ in spec } s_i | x, y) \quad (14)$$

[0279] After starting with $|s|$ grids of probability in spec, the end result of the first step is one grid of probability in spec. In some implementations, the processor 755 of system 700 is

configured to present the plurality of probabilities as one or more graphical elements of a graph displayed on a user interface (e.g., on the output device 770). For example, FIG. 33 is an example of a heat-map representing an end result of the first step. The contour lines included in the heat-map illustrated in FIG. 33 represent the traditional (geometric) process window and are drawn for comparison purposes. An alternate view of the three-dimensional end result heat-map is illustrated in FIG. 34.

[0280] In the second step of the process, using the grid of probabilities from the first step, a bivariate probability distribution of input values is applied to the grid, using some mean μ_x, μ_y and some error σ_x, σ_y . The mean μ_x, μ_y represents the process set point, that is, the nominal values of the input variables x and y used by the process. The errors σ_x, σ_y represent variations in the input variables that exist in the process, such as variations across the scanner field, across the chip, across the wafer, or from wafer to wafer. Often, the two input variables x and y (such as focus and exposure dose) can be considered as independent variables for the purposes of generating the bivariate probability distribution, but this is not a requirement. The bivariate distribution with using some mean μ_x, μ_y and some error σ_x, σ_y represents the probability density that the two input variables x and y will have specific values.

[0281] In a first part of the second step, for each grid cell between points, the volume under the bivariate distribution center at some mean μ_x, μ_y with some error σ_x, σ_y is determined. Next, the determined volume is weighted by the output of a function fit to the probabilities of the corner points of the current grid cell generated in the first step of the process. For example, using a Gaussian distribution, the volume may be determined using EQUATION 14 below:

$$P((x, y)) = \left(\frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x_{stop} - x}{\sigma_x \sqrt{2}} \right) \right] - \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x_{start} - x}{\sigma_x \sqrt{2}} \right) \right] \right) \\ \times \left(\frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{y_{stop} - y}{\sigma_y \sqrt{2}} \right) \right] - \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{y_{start} - y}{\sigma_y \sqrt{2}} \right) \right] \right) \quad (14) \\ \times f_{P \text{ in } spex}(x, y)$$

[0282] Next, all of the values in the grid are summed to determine the overall probability for this mean μ_x, μ_y with σ_x, σ_y . For example, all of the values in the grid may be summed using EQUATION 15 below:

$$P = \sum_{x=0}^X \sum_{y=0}^Y P((x, y)) \quad (15)$$

[0283] In a second part of the second step, the first part of the second step is repeated for different values of μ_x, μ_y to determine the probability in-spec with means centered at that point. For example, using the original grid from the first step illustrated in FIGs. 33 and 34, each value of μ_x, μ_y is determined, though there is no requirement that μ_x, μ_y be set to the values of the original input data grid. FIG. 35 is a chart of an example of the first part of the second step repeated $|X| \times |Y|$ times. Each square illustrated in FIG. 35 represents a bivariate Gaussian centered at that point. An alternate chart view of the chart illustrated in FIG. 35 is illustrated in FIG. 36.

[0284] In a third step of the process, a standard non-linear optimization algorithm (e.g., Nelder-Mead or Gauss-Newton) is used to determine the μ_x, μ_y with the highest probability for some error σ_x, σ_y . This represents the “best” values for the process settings for the input variables x and y . Here, “best” would mean the process settings (the values μ_x, μ_y) that produce the highest fraction of in-spec features given the errors σ_x, σ_y built into the process. For the case of an FEM process window, this would be best focus and best exposure. Setting the lithography tool to use this focus and exposure dose may produce the best printing results.

[0285] In a fourth step of the process, a σ_y vs. σ_x curve is generated. A step size for σ_x is selected. As σ_x is incremented, σ_y is adjusted and the search method in the third step is repeatedly used to determine the best value for σ_y (where ‘best value’ denotes, e.g., the point with the lowest probability above some cutoff, e.g., 99.73 %). The incrementing of σ_x is stopped when no σ_y meets the cutoff criteria. The determined points are plotted to generate the σ_y vs. σ_x curve. In some implementations, σ_x is the depth of focus and σ_y is the exposure latitude. FIG. 37 is an example of a σ_y vs. σ_x curve. A traditional process window analysis may be geometrically based, finding the largest rectangle or ellipse that fits inside the process window. Similar, but more rigorous curves, can be generated using the Probabilistic Process Window using the steps outlined above.

[0286] Once a probabilistic process window is determined for a certain dataset with certain specs, the probabilistic process window may be combined with any number of other probabilistic process windows, as long as the two or more input process variables are the same

and with similar bounds. Other probabilistic process windows may have different underlying output values, feature types, and specifications and still be combined, due to the unit-less nature of probabilities. To combine probabilistic process windows, a test point on each one may be evaluated, and the resulting probabilities are multiplied together to determine a combined probability across probabilistic process windows.

[0287] Referring now to FIG. 38, an example method 3800 is provided to generate a probabilistic process window that accounts for measurement uncertainty. The method 3800 starts (block 3802) and selects a first variable represented on a first axis of a graph (3804). Next, the method 3800 proceeds to select a second variable represented on a second axis of a graph (block 3806). In some implementations, the first variable and the second variable are associated with a stochastic process window. In some implementations, the first variable comprises exposure dose and the second variable comprises focus of a scanning lithography process. In alternate implementations, the first variable comprises etch time and the second variable comprises wafer temperature. Many other combinations of process variables are possible.

[0288] Next, the method 3800 proceeds (block 3808) to select at least one response variable (or output response variable) that is a function of the first variable and second variable. These output response variables are measured as a function of the first and second variables. Each output response variable can be graphed as a function of the first and second variables. For each output response variable, specifications are set for values of the output response considered acceptable.

[0289] Next, the method 3800 proceeds to determine a measurement uncertainty for each of the output response variables (block 3810). Next, the method 3800 proceeds (block 3812) to determine, based on a measurement of the one or more output response variables, and the measurement uncertainty for the at least one output response variable, a plurality of probabilities representing a plurality of indications of whether a plurality of points associated with a lithography process meet a specification requirement for each output response variable (block 3810). The plurality of probabilities represents a process window. Next, the method 3800 proceeds to configure, based on the process window, a lithography tool to manufacture a semiconductor device.

[0290] In some implementations, a third input variable is selected. In some implementations, the third variable is temperature. Further, based on a measurement of the output response as a function of the third variable in combination with variations in the first and second variables, a second plurality of probabilities representing a second plurality of indications of whether a second plurality of points associated with the lithography process meet the specification requirement is determined. The second plurality of probabilities represent a second process window.

[0291] The plurality of probabilities may be modified to achieve a desired cost optimization of manufacturing a semiconductor device. For example, in some implementations, one or more graphical elements are presented that are configured to enable a user to change the first and/or second variable to modify the plurality of probabilities in real-time or near real-time. Real-time may refer to less than or equal to 2 seconds. Near real-time may refer to any interaction of a sufficiently short time to enable two individuals to engage in a dialogue via such user interface, and will generally be less than 10 seconds (or any suitable proximate difference between two different times) but greater than 2 seconds.

[0292] In some embodiments, process range variations are selected for the first and second variables. A fraction of features is determined that meets a specification requirement given a setting and process range for each of the first and second variables. Based on a fraction of features that meet a specification requirement, a determination may be made of the settings of the first and second variables that produce a maximum fraction of features that meet a specification.

[0293] 16. Using the Probabilistic Process Window for Process Control

[0294] Controlling focus and dose is an important part of keeping the critical dimensions of the printed patterns in control during semiconductor manufacturing. The first step in understanding how to control focus and dose is to characterize the response of critical features to variations using the focus-exposure process window, as discussed above. Proper analysis of CD (and other output responses) versus focus and dose data allows for the calculation of the process window, measurement of the process window size to produce the exposure latitude-DOF plot, and determination of a single value for the depth of focus. As discussed above, the best analysis method uses the probabilistic process window.

[0295] Best focus and dose are also determined with this analysis as the process settings that maximize the tolerance to focus and dose errors based on the PPW. Once best focus and dose are determined, the next goal is to keep the process centered at this best focus and dose condition as production wafers pass through the lithography cell. Since almost all errors that occur in lithography act like either an effective dose error or an effective focus error, properly adjusting dose and focus after an appropriate interval can provide for much tighter patterning control. In turn, improved patterning control can lead to higher manufacturing yields and better performance of the devices being fabricated.

[0296] Often, the variables dose and focus, as well as other process variables that could be used in a probabilistic process window analysis, are monitored and controlled using well-known Advanced Process Control (APC) methodologies. These APC methods include feed-forward and feedback loops for controlling process tools. These process control methods would be enhanced by replacing the traditional process window analysis with the more rigorous and accurate PPW analysis described above.

[0297] Process window analysis can also be used for wafer disposition purposes. Wafer disposition takes the output of a process window analysis and decides whether the wafers represented by the process window should be advanced to subsequent processes, or rejected due to expected poorly printed patterns. If rejected, these wafers might be reworked or scrapped. Incorrect decisions during wafer disposition can be very costly, passing on bad wafers for subsequent processing, or rejecting good wafers. Improved accuracy in process window analysis, for example using the PPW process described above, can improve the wafer disposition accuracy.

[0298] One approach to process control uses machine learning. A machine learning algorithm would seek to make predictions, for example related to yield or device performance, based on a number of process variables used in the manufacture of the devices. A predictive machine learning model is first trained using measured outputs (the things that are to be predicted) as a function of measured inputs (the process variables and intermediate results, such as the CD of a specific feature at a specific step in the fabrication process). Machine learning algorithms can use process window information for both training and predictions. As such, the improved accuracy and precision of the PPW approach could result in improved machine learning predictions.

[0299] While the implementations described above make reference to top-down images of nominally planar pattern structures to measure edge roughness, the present disclosure is not limited to such pattern structure geometries. Three-dimensional structures, non-flat structures, curved surfaces, or tilted structures can be measured using this invention. Besides edge roughness, surface roughness can be measured and analyzed using similar techniques as described in the present disclosure.

[0300] While the implementations described above make reference to the measurement of roughness, the present disclosure can be used to make other measurements as well. For example, highly accurate determination of pattern structure edges can be used in the measurement of feature width, feature placement, edge placement, and other similar measures. Contours of measured features can be used for many purposes, such as modeling or controlling the performance of the measured device. By collecting and statistically averaging the measurement of many samples, even greater accuracy (lower uncertainty) can be obtained.

[0301] Consistent with the above disclosure, the examples of systems and methods enumerated in the following clauses are specifically contemplated and are intended as a non-limiting set of examples.

[0302] Clause 1. A computer-implemented method comprising:
selecting a first variable represented on a first axis of a graph;
selecting a second variable represented on a second axis of the graph;
selecting at least one response variable that is a function of the first variable and second variable;
determining a measurement uncertainty for each response variable;
determining, based on a measurement of the response variables, and the measurement uncertainty for the response variables, a plurality of probabilities representing a plurality of indications of whether a plurality of points associated with a lithography process meet a specification requirement for each output, wherein the plurality of probabilities represent a process window; and
configuring, based on the process window, a lithography tool to manufacture a semiconductor device.

[0303] Clause 2. The computer-implemented method of claim 1, wherein configuring comprises transmitting control signals to the lithography tool to set a first operating parameter based on the first variable and to set a second operating parameter based on the second variable.

[0304] Clause 3. The computer-implemented method of any clause herein, further comprising presenting the plurality of probabilities as graphical elements of the graph on a user interface of a computing device.

[0305] Clause 4. The computer-implemented method of any clause herein, wherein the presentation of the plurality of probabilities is a heat-map or 3D plot or contour plot on the user interface.

[0306] Clause 5. The computer-implemented method of any clause herein, wherein the measurement uncertainty of the one or more response variables is represented by a Gaussian normal probability distribution.

[0307] Clause 6. The computer-implemented method of any clause herein, wherein the first and second variables are associated with a stochastic process window.

[0308] Clause 7. The computer-implemented method of any clause herein, wherein the first variable comprises exposure dose and the second variable comprises focus of a scanning lithography process.

[0309] Clause 8. The computer-implemented method of any clause herein, further comprising presenting one or more graphical elements that are configured to enable a user to change properties of the first variable, the second variable, or both to modify the plurality of probabilities in real-time or near real-time.

[0310] Clause 9. The computer-implemented method of any clause herein, further comprising presenting one or more graphical elements that are configured to enable a user to change properties of the first variable, the second variable, or both to modify the plurality of probabilities in real-time or near real-time, wherein modifying the plurality of probabilities is associated with a desired cost optimization of manufacturing the semiconductor device.

[0311] Clause 10. A system comprising:
a lithography tool;
a memory device storing instructions; and

a processing device, coupled to the memory device and the lithography tool, the processing device executes the instructions to:

select a first variable represented on a first axis of a graph,

select a second variable represented on a second axis of the graph,

select a response variable that depends on the first and second variable

determine a measurement uncertainty for the response variable,

determine, based on a measurement of the response variable, and the measurement uncertainty for the response variable, a plurality of probabilities representing a plurality of indications of whether a plurality of points associated with a lithography process meet a specification requirement, wherein the plurality of probabilities represent a process window, and
configure, based on the process window, the lithography tool to manufacture a semiconductor device.

[0312] Clause 11. The system of any clause herein, wherein configuring comprises transmitting control signals to the lithography tool to set a first operating parameter based on the first variable and to set a second operating parameter based on the second variable.

[0313] Clause 12. The system of any clause herein, wherein the processing device is further configured to present the plurality of probabilities as graphical elements of the graph on a user interface of a computing device.

[0314] Clause 13. The system of any clause herein, wherein the presentation of the plurality of probabilities is a heat-map on the user interface.

[0315] Clause 14. The system of any clause herein, wherein the measurement uncertainty of the response variable is a Gaussian normal probability distribution.

[0316] Clause 15. The system of any clause herein, wherein the first and second variables are associated with a stochastic process window.

[0317] Clause 16. The system of any clause herein, wherein the first variable comprises exposure dose and the second variable comprises focus of a scanning lithography process.

[0318] Clause 17. The system of any clause herein, wherein the processing device is further configured to:

select process range variations for the first and second variables,

determine a fraction of features that meet a specification requirement given a setting and process range for each of the first and second variables,
determine, based on a the fraction of features that meet a specification requirement, the settings of the first and second variables that produce a maximum fraction of features that meet a specification.

[0319] Clause 18. A tangible, non-transitory computer-readable medium storing instructions that, when executed, cause a processing device to:

select a first variable represented on a first axis of a graph;

select a second variable represented on a second axis of the graph;

select at least one response variable that is a function of the first variable and second variable;

determine a measurement uncertainty for each response variable;

determine, based on a measurement of the response variable, and the measurement uncertainty for the response variable, a plurality of probabilities representing a plurality of indications of whether a plurality of points associated with a lithography process meet a specification requirement, wherein the plurality of probabilities represent a process window; and

configure, based on the process window, the lithography tool to manufacture a semiconductor device.

[0320] Clause 19. The computer-readable medium of any clause herein, wherein configuring comprises transmitting control signals to the lithography tool to set a first operating parameter based on the first variable and to set a second operating parameter based on the second variable.

[0321] Clause 20. The computer-readable medium of any clause herein, wherein the processing device is further configured to present the plurality of probabilities as graphical elements of the graph on a user interface of a computing device.

[0322] The terminology used herein is for the purpose of describing particular implementations only and is not intended to be limiting of the present disclosure. As used herein, the singular forms “a,” “an,” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms “comprises” and/or “comprising,” when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof.

[0323] Note that not all of the activities described above in the general description or the examples are required, that a portion of a specific activity may not be required, and that one or more further activities can be performed in addition to those described. Still further, the order in which activities are listed are not necessarily the order in which they are performed.

[0324] It can be advantageous to set forth definitions of certain words and phrases used throughout this patent document. The term “communicate,” as well as derivatives thereof, encompasses both direct and indirect communication. The terms “include” and “comprise,” as well as derivatives thereof, mean inclusion without limitation. The term “or” is inclusive, meaning and/or. The phrase “associated with,” as well as derivatives thereof, can mean to include, be included within, interconnect with, contain, be contained within, connect to or with, couple to or with, be communicable with, cooperate with, interleave, juxtapose, be proximate to, be bound to or with, have, have a property of, have a relationship to or with, or the like. The phrase “at least one of,” when used with a list of items, means that different combinations of one or more of the listed items can be used, and only one item in the list can be needed. For example, “at least one of: A, B, and C” includes any of the following combinations: A, B, C, A and B, A and C, B and C, and A and B and C.

[0325] The description in the present disclosure should not be read as implying that any particular element, step, or function is an essential or critical element that must be included in the claim scope. The scope of patented subject matter is defined only by the allowed claims. Moreover, none of the claims invokes 35 U.S.C. § 112(f) with respect to any of the appended claims or claim elements unless the exact words “means for” or “step for” are explicitly used in the particular claim, followed by a participle phrase identifying a function. Use of terms such as (but not limited to) “mechanism,” “module,” “device,” “unit,” “component,” “element,” “member,” “apparatus,” “machine,” “system,” “processor,” or “controller” within a claim is understood and intended to refer to structures known to those skilled in the relevant art, as further modified or enhanced by the features of the claims themselves, and is not intended to invoke 35 U.S.C. § 112(f).

[0326] Benefits, other advantages, and solutions to problems have been described above with regard to specific implementations. However, the benefits, advantages, solutions to problems, and any feature(s) that can cause any benefit, advantage, or solution to occur or become more pronounced are not to be construed as a critical, required, or essential feature of any or all the claims.

[0327] After reading the specification, skilled artisans will appreciate that certain features are, for clarity, described herein in the context of separate implementations, can also be provided in combination in a single embodiment. Conversely, various features that are, for brevity, described in the context of a single embodiment, can also be provided separately or in any sub-combination. Further, references to values stated in ranges include each and every value within that range.

[0328] The corresponding structures, materials, acts, and equivalents of all means or step plus function elements in the claims below are intended to include any structure, material, or act for performing the function in combination with other claimed elements as specifically claimed. The description of the present disclosure has been presented for purposes of illustration and description, but is not intended to be exhaustive or limited to the invention in the form disclosed. Many modifications and variations will be apparent to those of ordinary skill in the art without departing from the scope and spirit of the present disclosure. The implementations was chosen and described in order to best explain the principles of the present disclosure and the practical application, and to enable others of ordinary skill in the art to understand the present disclosure for various implementations with various modifications as are suited to the particular use contemplated.

CLAIMS

What is claimed is:

1. A computer-implemented method comprising:
selecting a first variable;
selecting a second variable;
selecting at least one response variable that is a function of the first variable and second variable;
determining a measurement uncertainty for each response variable;
determining, based on a measurement of the response variable, and the measurement uncertainty for the response variable, a plurality of probabilities representing a plurality of indications of whether a plurality of points associated with a lithography process meet a specification requirement for each response variable, wherein the plurality of probabilities represent a process window; and
configuring, based on the process window, a lithography tool to manufacture a semiconductor device.
2. The computer-implemented method of claim 1, wherein configuring comprises transmitting control signals to the lithography tool to set a first operating parameter based on the first variable and to set a second operating parameter based on the second variable.
3. The computer-implemented method of claim 1, further comprising presenting the plurality of probabilities as graphical elements of the graph on a user interface of a computing device.
4. The computer-implemented method of claim 3, wherein the presentation of the plurality of probabilities is a heat-map, three-dimensional plot, or contour plot on the user interface.
5. The computer-implemented method of claim 1, wherein the measurement uncertainty of the response variable is represented by a Gaussian normal probability distribution.

6. The computer-implemented method of claim 1, wherein the first variable is represented on a first axis of a graph, and the second variable is represented on a second axis of the graph.

7. The computer-implemented method of claim 1, wherein the first variable comprises exposure dose and the second variable comprises focus of a scanning lithography process.

8. The computer-implemented method of claim 1, further comprising presenting one or more graphical elements that are configured to enable a user to change properties of the first variable, the second variable, or both to modify the plurality of probabilities in real-time or near real-time.

9. The computer-implemented method of claim 1, further comprising presenting one or more graphical elements that are configured to enable a user to change properties of the first variable, the second variable, or both to modify the plurality of probabilities in real-time or near real-time, wherein modifying the plurality of probabilities is associated with a desired cost optimization of manufacturing the semiconductor device.

10. A system comprising:
a lithography tool;
a memory device storing instructions; and
a processing device, coupled to the memory device and the lithography tool, the processing device executes the instructions to:
select a first variable;
select a second variable;
select a response variable that depends on the first and second variable;
determine a measurement uncertainty for the response variable;
determine, based on a measurement of the response variable, and the measurement uncertainty for the response variable, a plurality of probabilities representing a plurality of indications of whether a plurality of points associated with a lithography process meet a specification requirement for each response variable, wherein the plurality of probabilities represent a process window; and
configure, based on the process window, the lithography tool to manufacture a semiconductor device.

11. The system of claim 10, wherein configuring comprises transmitting control signals to the lithography tool to set a first operating parameter based on the first variable and to set a second operating parameter based on the second variable.
12. The system of claim 10, wherein the processing device is further configured to present the plurality of probabilities as graphical elements of the graph on a user interface of a computing device.
13. The system of claim 12, wherein the presentation of the plurality of probabilities is a heat-map on the user interface.
14. The system of claim 10, wherein the measurement uncertainty of the response variable is a Gaussian normal probability distribution.
15. The system of claim 10, wherein the first variable is represented on a first axis of a graph, and the second variable is represented on a second axis of the graph.
16. The system of claim 10, wherein the first variable comprises exposure dose and the second variable comprises focus of a scanning lithography process.
17. The system of claim 10, wherein the processing device is further configured to:
select process range variations for the first and second variables,
determine a fraction of features that meet a specification requirement given a setting and process range for each of the first and second variables,
determine, based on the fraction of features that meet a specification requirement, the settings of the first and second variables that produce a maximum fraction of features that meet a specification.
18. A tangible, non-transitory computer-readable medium storing instructions that, when executed, cause a processing device to:
select a first variable;
select a second variable;
select at least one response variable that is a function of the first variable and second variable;
determine a measurement uncertainty for each response variable;

determine, based on a measurement of the response variable, and the measurement uncertainty for the response variable, a plurality of probabilities representing a plurality of indications of whether a plurality of points associated with a lithography process meet a specification requirement for each response variable, wherein the plurality of probabilities represent a process window; and
configure, based on the process window, the lithography tool to manufacture a semiconductor device.

19. The computer-readable medium of claim 18, wherein configuring comprises transmitting control signals to the lithography tool to set a first operating parameter based on the first variable and to set a second operating parameter based on the second variable.

20. The computer-readable medium of claim 18, wherein the processing device is further configured to present the plurality of probabilities as graphical elements of the graph on a user interface of a computing device.