absorbing, reducing the change in M near the resist-substrate interface during exposure, thereby reducing γ. This also explains the curious observation that measured contrast shows variations with resist thickness[106] rather than being an intrinsic property of the resist chemistry.

Alternative definitions for contrast have been proposed. An expression similar to Eq. (3.16) has been derived:[107–109]

$$\gamma = \frac{R(T_0)}{T_0} \int_0^{T_0} \left(\frac{\partial \ln R}{\partial \ln E} \right) \frac{dz}{R(z)}, \tag{3.17}$$

where z is the depth in the resist film. (Again, technical rigor has been less than complete with the taking of logarithms of nondimensionless quantities.) The expression

$$\frac{\partial \ln R}{\partial \ln E} = \gamma_{th}, \tag{3.18}$$

for the "theoretical contrast" has gained acceptance as a measure of resist performance.[110,111] Although there are difficulties associated with using this metric, such as its dependence on exposure dose, it has been found to be a good indicator of how well resists will perform.[112] Equation (3.18) has several advantages over the earlier expression for contrast as introduced in Chapter 2. It is dependent solely on the dissolution properties of the photoresist and does not have the dependencies on resist film thickness found in the earlier definitions. However, in the limit of very transparent or thin resists typical of KrF resists, the two expressions for contrast are roughly equivalent.[113]

3.8 Line edge roughness

Another issue for resists that is becoming significant as feature sizes approach 100 nm, and smaller, is *line edge roughness* (LER).[114] For very small lines, structure at the molecular level contributes to roughness that can be a significant fraction of the linewidth. A measure of edge roughness is the standard deviation of the actual line edge relative to an average line edge. Edge roughness with 3σ, ~5–8 nm has been measured typically. There may be a reticle contribution to the low-spatial-frequency component of the LER,[114] while high-spatial-frequency roughness is related more to the resist process.[115,116] LER has been measured as a function of spatial frequency, and representative results from an insightful paper by Yamaguchi and coworkers are shown in Fig. 3.29.[117] The data from several resists showed similar behavior for line-edge roughness as a function of spatial frequency. In all cases, the largest contributions to LER came from lower spatial frequencies (although not necessarily at the lowest spatial frequency).

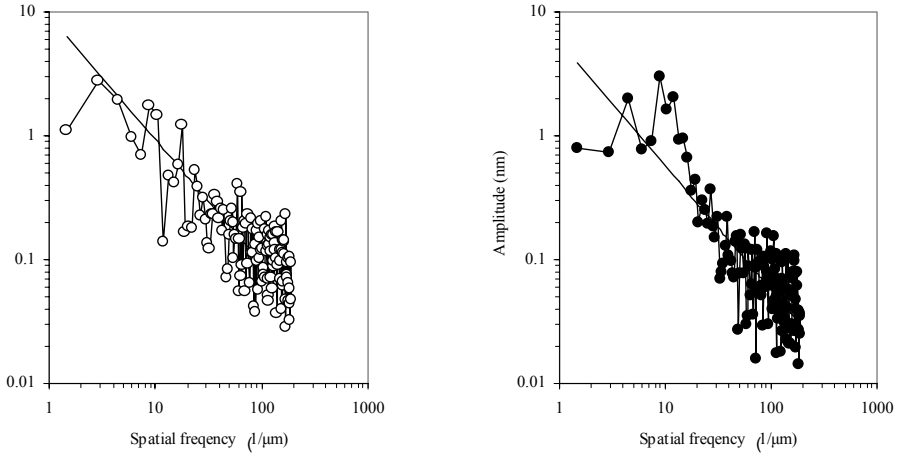

Figure 3.29 LER as a function of spatial frequency. The results on the left were from e-beam exposures, while the data on the right were from ArR exposures. The electron beam was intentionally blurred to induce greater LER.

As one might expect, diffusion that occurs during post-exposure bake serves to reduce line-edge roughness. This is related to the original motivation for chemical amplification—addressing the productivity of exposure tools with low light intensity. When exposures become low, statistical variations in the number of photons involved in exposing the resist can contribute to line-edge roughness. This results from basic photon statistics (shot noise), where the root-mean-square variation in the number of photons Δn is related to the average number of photons $\hat{n}$ by the following expression:[118]

$$\frac{\Delta n}{\hat{n}} = \sqrt{\frac{1}{\hat{n}} + 1}\,. \tag{3.19}$$

When the number of photons becomes small, the statistical fluctuation can be appreciable. Consider, for example, ArF light passing through the top of a resist film in an area that is 10 nm × 10 nm. For an intensity of 1 mJ/cm^2, approximately 1000 photons on average enters this area. From Eq. (3.19), this light intensity fluctuates approximately ±3% (1σ). At the edges of features, the light intensity is typically ~1/3 of that found in the middle of large features, so the effect of photon statistics is greater at line edges. (See Problem 3.2.)

In chemically amplified resists, the photo-acid diffuses with a diffusion length that is typically 40–60 nm. For example, the amount of line blurring in KrF ESCAP resists was measured at ~50 nm (FWHM).[119,120] Diffusion over such a length may serve to smooth out the roughness induced by photon statistics.

While diffusion during the post-exposure baking of chemically amplified resists reduce line-edge roughness, a problem arises when the amount of blurring becomes comparable to the half-pitch. At this point, the optical image originally

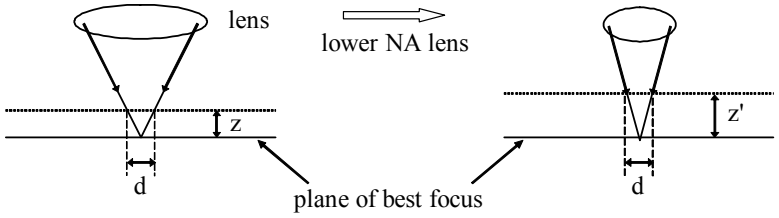

Figure 8.3 Illustration of the depth-of-focus in the context of geometrical optics.

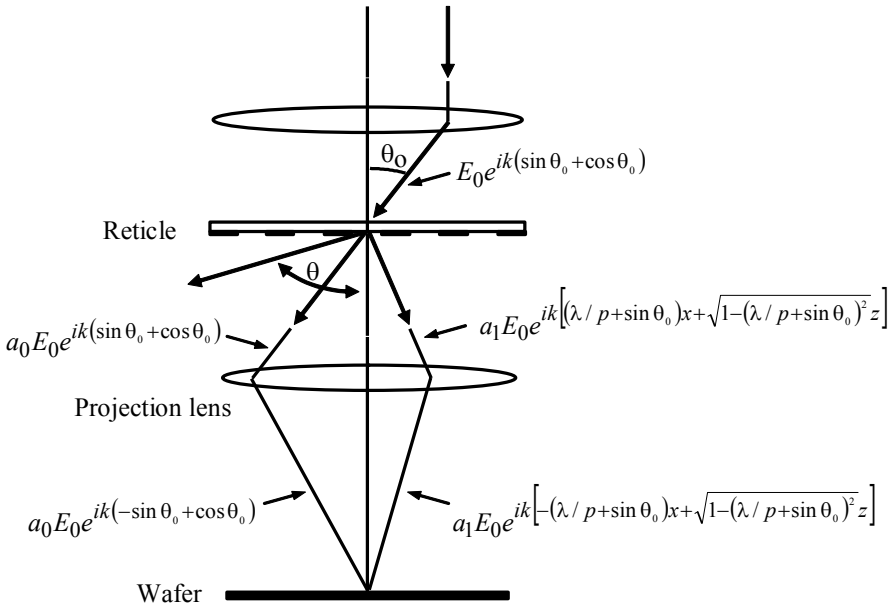

Figure 8.4 Light rays diffracted by a diffraction grating of pitch *p*. $k = 2\pi/\lambda$. This example is appropriate for 1:1 optics. Angles and dimensions will be scaled by the lens reduction for reduction optics.

which occurs when

$$\sin(\theta_0) = \frac{\lambda}{2p}. \tag{8.5}$$

Equation (8.5) shows an important characteristic of image enhancement when using off-axis illumination: the optimum parameters, such as the angle of incidence for the illumination, are pitch dependent.[2] This indicates that process window enhancement can be obtained for a particular pitch when using a given

angle for the off-axis illumination, but that the enhancement may be less or nonexistent for other pitches.

In practice, infinite depth-of-focus is not obtained, because the illumination consists of more than two rays of light coming from angles $\pm\theta_0$. Instead, the illumination is comprised of light incident over ranges of angles around the optimum. This occurs for several reasons. First, maintaining a reasonable level of light intensity requires that illumination be collected over a finite range of angles. Additionally, a finite range of angles for the illumination results in light that is spread out as it propagates through the lens. Spreading the light out within the projection optics has a number of benefits. One advantage is reduced sensitivity to aberrations in the projection optics. This occurs because phase and amplitude errors of one particular ray of light are averaged with the errors from other rays. The resulting image quality will not be determined by worst-case errors.[3] In addition, with finite collection angles, the light is less concentrated as it propagates through the projection optics, reducing the potential for glass damage. (See Chapter 5, Section 4.) Finally, light from a point source is very coherent, which leads to interference issues, as discussed in Chapter 5. For all of these reasons, the illumination is incident over a range of angles. Since this results in less-than-optimal performance, there is pressure on the exposure tool manufacturers to reduce lens aberrations and enhance illuminators to provide high light intensity even when the illumination is directed over a narrow range of angles.

Illumination that contains light incident on the mask in a cone from all angles from zero to a maximum defined by the partial coherence, σ, is called *conventional illumination*, while illumination from which the normally and near-normally incident light rays are excluded is called *off-axis illumination*. The type of illumination that has been described thus far, where the light is incident on the mask from just two opposing directions, is a particular type of off-axis illumination called *dipole illumination*.

Illumination is typically described by partial coherence, σ, rather than an angle of illumination. From the defining equation for σ [Eq. (A11) in Appendix A] and Eq. (8.5), one can see that the optimum illumination angle for dipole illumination is given by

$$\sigma = \frac{1}{\mathrm{NA}}\frac{\lambda}{2p}. \tag{8.6}$$

Because light is collected over a finite range of angles, off-axis illumination is usually described by a pair σ values, one representing the outer angles ("σ_{outer}") and the other value signifying the inner angles ("σ_{inner}"). The optimum value given by Eq. (8.6) is intermediate between σ_{outer} and σ_{inner}.

The benefits of off-axis illumination for enhancing image contrast have been long known among optical microscopists. Motivated by experience in microscopy, the concept of oblique or off-axis illumination was introduced to

use of shorter wavelengths, but the large fraction of the depth-of-focus budget from device topography still limited the use of high-numerical-aperture optics.

The advent of chemical-mechanical polishing reduced the problems associated with device topography significantly, by eliminating most device topography. The requisite depth-of-focus of the optics was decreased, because it was no longer necessary to image well throughout the height of device features. This enabled the use of higher numerical aperture optics and allowed the extension of optical lithography further than earlier thought possible, because the depth-of-focus requirement was reduced by the use of chemical-mechanical polishing.

However, wafers can be flattened only once. The benefit of chemical-mechanical polishing has nearly been fully realized, and this process cannot provide additional help for extending optical lithography. As with resist and wavelength, only small gains are possible for extending optical lithography by the means of reducing wafer topography, as used in the past. Optical lithography has been extended beyond earlier expectations, but the means by which this extension occurred provide little further benefit. However, there are new techniques such as phase-shifting masks that have not yet been fully exploited, enabling a move to lower k_1. The question of how low in k_1 it is possible to go is now considered.

10.6 How low can k_1 go?

An inspection of Fig. 2.13 shows that image contrast becomes zero when

$$\frac{\lambda}{2d\mathrm{NA}} = 2,$$

where the pitch is $2d$. This means that there is no image contrast when the half pitch is $0.25\lambda/\mathrm{NA}$ or smaller. This is true even when using powerful resolution enhancement techniques such as dipole illumination or alternating phase-shifting masks. Thus, there is a limit on k_1 of 0.25 imposed by the laws of physics. It only need be assumed that the optical imaging is linear and that a single exposure step is involved for this to be a true limit imposed by the laws of nature. How close to $k_1 = 0.25$ a manufacturing process can be is a matter of technology.

The value of k_1 used in manufacturing has been decreasing for years. Since k_1 must always be larger than 0.25, it is interesting to see how $k'_1 = k_1 - 0.25$ trends over time. A graph of k'_1 is shown in Fig. 10.3. As can be seen, considerable progress has been made in reducing k_1 over the years, and k_1 has descended from a value of ~ 0.8 in the mid-1980s to nearly 0.35 today. This decrease in k_1 alone accounts for a reduction in feature sizes of over 2×, which represents two nodes

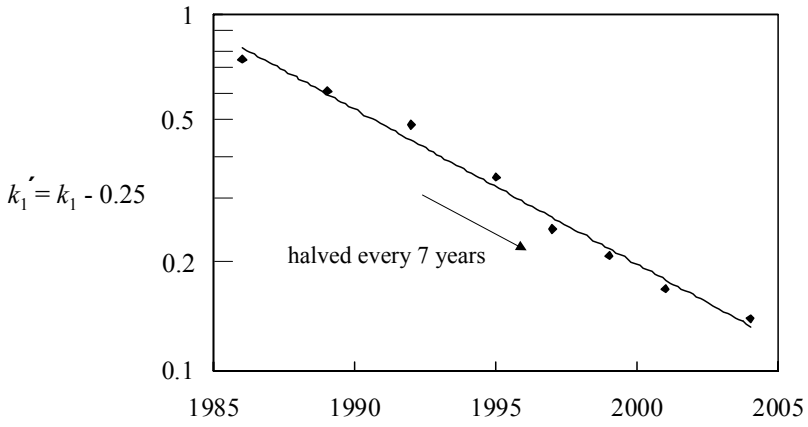

Figure 10.3 The evolution of k_1 over time, using DRAM half pitches to ascertain k_1.

of progress on the International Technology Roadmap for Semiconductors*. However, a similar reduction in the future is not possible, since k_1 cannot become smaller than 0.25. The most reduction that is physically possible is

$$\frac{0.25}{0.35} = 0.71, \tag{10.3}$$

and this requires taking resolution to the absolute limit allowed by the laws of physics. A more likely minimum value is somewhat larger, perhaps k_1 = 0.28 or 0.29. Thus, only modest decreases in the half pitch of single exposures will be possible in the future through reduction in k_1.

As has been described thus far in this chapter, nearly all of the improvements that have been introduced in the past to advance lithography—better resists, lenses with lower aberrations, flatter wafers, and shorter wavelengths—have been almost completely exploited. The only remaining opportunity for substantial decreases in feature size is the numerical aperture. With numerical apertures of 0.92 and greater, the geometrical limit of 1.0 for lenses imaging in air (Fig. 2.11) has nearly been reached. The potential for NA > 1 can only be realized by immersion imaging. It is to the subject of immersion lithography that attention is now turned.

10.7 Immersion lithography and maximum numerical aperture

In Chapter 2, where the parameter *numerical aperture* was introduced, it was defined as $n\sin\theta$, where n is the index of refraction of the medium between the

* http://public.itrs.net

In addition to commercially produced electron beam systems, there has been a long-term electron-beam direct-write program at IBM that has produced a succession of tools, EL-1 through EL-4,81, 82 based upon shaped beams.83 This technology evolved into the EL-5, which has been used to make 1× x-ray masks that have the same resolution requirements as direct-write tools.[84]

Another application of direct write e-beam lithography is in technology development, where wafer throughput does not need to be very high. For example, a Leica SB 320-50 SW shaped beam tool was used to fabricate SRAM cells with at 65 nm design rules at a time when 90-nm technology had not yet reached manufacturing.[85] E-beam patterning was used on the critical active, gate, contact, and metals layers. Throughput was very low, 18–24 hours for a single 200-mm wafer, but this was nevertheless useful for early development of advanced technologies, where few wafers were needed. A negative resist, Sumitomo NEB-33, was used for clear-field active and gate layers, while a positive resist, Fujifilm FEP-171, was used for contact and metal layers. The use of a negative resist for clear-field layers minimized writing time.

To circumvent the need to convey $\geq 10 \times 10^{12}$ bits through a single pathway, approaches have been proposed that involve multiple electron optical columns[86] or multiple beams.[87] This will be a challenging technology to implement, requiring extremely good reliability and innovations to produce adequate calibrations among all of the beams and columns. While this approach is in the early development stage, it is worth noting that several commercial companies, Etec, Canon, and Ion Diagnostics, have described activities in this area.

12.4 Electron-projection lithography—EPL

To address the throughput issue of electron-beam direct-write lithography, without multiple columns and the attendant complexity, electron-projection lithography (EPL) involving masks has been developed. These EPL systems fall into two classes. The first, for which there are already commercially available systems, involves the use of stencil masks and very small fields ($\leq 5\ \mu m \times 5\ \mu m$). A stencil mask consists of a thin membrane through which holes are etched to provide regions of the mask "transparent" to electrons. The second method allows for much larger masks (≥ 20 mm $\times$ 20 mm) and involves electron scattering to create the "opaque" regions on the mask.

12.4.1 Small-field EPL systems

A small-field electron system, the HL-800D, is available commercially from Hitachi.[88,89] Such tools have limited applicability, because of their small fields (5 $\mu m \times 5\ \mu m$), but they have nevertheless found use in particular niche applications. If their resolution is 100 nm, then systems with a 5-$\mu m \times 5$-μm field size, on the wafer, represent an increase in the "parallelism" of electron-beam

lithography by a factor of 2500 over direct-write systems. With electron optics, reduction imaging is possible, again avoiding the 1× problem of x-ray lithography. The masks for these systems are fabricated using thin film techniques, similar to those used for making x-ray masks, described earlier in this chapter. Unfortunately, the field size is far too small for such systems to have general applicability to the patterning of integrated circuits. For certain applications, such as patterning the repetitive core cells of memories, these projection electron lithography systems are useful, and they have been used to make prototype high-bit-count DRAMs[90] before adequate optical lithography capability became available. While quite useful for technology development, these small-field systems have limited applicability.

12.4.2 Large-field EPL systems

One of the potential limitations of projection electron lithography is mask heating. To achieve adequate throughput, high currents are needed. For masks that create opaque regions by simply blocking and absorbing the electrons, the masks become hot.[91] To overcome this problem, a very clever scheme has been proposed,[92] where nearly all electrons are allowed to pass through the mask. Instead of absorbing electrons in those portions of the mask that are supposed to correspond to unexposed areas on the wafer, the electrons are scattered by high-atomic-number materials, such as tungsten, tantalum, or other materials used for x-ray masks. At a focal plane within the electron optics is a physical aperture, through which unscattered electrons pass. However, only a very small fraction of the scattered electrons pass through the aperture (Fig. 12.18). Those portions of the mask corresponding to regions of the design that are supposed to be exposed on the wafer must allow electrons to penetrate with little scattering. This can be accomplished by using stencil masks.[93] Alternatively, the scattering materials are placed on a thin membrane of low-atomic-number material, such as silicon nitride or diamond-like carbon.[94] This combination of a scattering mask, in conjunction with a focal plane aperture in the electron optics, has been given the name SCALPEL, the acronym for SCattering with Angular Limitation-projection Electron-beam Lithography. The SCALPEL approach has demonstrated some very impressive patterning. Shown in Fig. 12.19 are 80-nm contacts created using SCALPEL.[95]

This stitching is not a trivial challenge for the SCALPEL technology, since the left half of features, written in one field, need to line up with the right half, with a tolerance that is a small fraction of the linewidth. What complicates this stitching is the heating of the mask and wafer that occurs during exposure. While very little energy is deposited into the SCALPEL mask, the membranes, by being so thin, have very little thermal mass, and they heat appreciably during exposures even though little total energy is deposited. Simulations show that the temperature increase, which occurs nonuniformly across the mask, exceeds 7°C,