Journal of Microscopy, *Vol.* 160, *Pt* 3, *December* 1990, *pp.* 245–264.
Received 6 *July* 1989; *revised* 1 *May* 1990; *accepted* 1 *June* 1990

Generalizing the confocal microscope via heterodyne interferometry and digital filtering

by PHILIP C. D. HOBBS* *and* GORDON S. KINO, *Edward L. Ginzton Laboratory, Stanford University, Stanford, CA* 94305, *U.S.A.*

KEY WORDS. Optical microscopy, apodization, confocal, heterodyne interferometry, digital filter, interference contrast.

SUMMARY

We describe a generalized confocal optical microscope which measures both phase and amplitude, separately and simultaneously. The system is based on heterodyne inferometry, and is extremely fast and accurate. Together with the associated signal processing and data acquisition system, it can take data at the rate of 50,000 points/s, to an accuracy of 12 bits in amplitude and about the same in phase. The 1σ phase error is 0·1°, corresponding to a height uncertainty of 0·1 nm. Unlike most phase-sensitive systems of this accuracy, the phase data are not differential, and we are not restricted to small phase changes. Having phase and amplitude enables use of digital filters to modify the coherent transfer function of the microscope in almost any way we desire (within the bandwidth available), and in particular to double the spatial resolution of the system without resorting to ultraviolet light or *ad hoc* image processing procedures such as non-linear transformations or the introduction of additional assumptions about the sample. In contrast to ordinary apodizing schemes involving annular pupils and graded neutral-density filters, the digital filters used here are very easily changed to optimize the trade-off of resolution versus ringing. Computationally efficient expressions for the line-spread function and step response of a confocal microscope are developed for use in comparisons.

INTRODUCTION

The confocal microscope is finding wide use in semiconductor metrology, biology, and other fields, because of its excellent resolution, depth discrimination, and lack of edge artefacts (ringing). The classical confocal microscope measures only intensity, and like an ordinary bright-field microscope, it is rather insensitive to phase objects. Its point-spread function (PSF) is not easily modified by the user, except to increase or decrease the numerical aperture; although satisfactory for most purposes, it possesses serious drawbacks from a quantitative measurement point of view, which are discussed below.

*Present address: IBM Thomas J. Watson Research Center, PO Box 218, Yorktown Heights, NY 10598, U.S.A.

The confocal microscope was first considered by Minsky (1957), and since by many others; Sawatari (1973) was the first to build a heterodyne interference microscope, and a later version with phase sensitivity was developed by Peterson *et al.* (1984) at about the time of our early work. Both of these systems are of the single-beam, mechanically scanned type, and neither has the precision nor stability of our system. Wickramasinghe *et al.* (1982) have made a scanning a.c. differential interference contrast microscope, which measures the derivative of the phase along the scan direction; accurate reconstruction of a profile from its derivative is difficult, however.

Our system (Jungerman *et al.*, 1984; Hobbs *et al.*, 1985) uses an acousto-optic Bragg cell in a unique configuration to make an electronically scanned, two-beam, heterodyne interference microscope, capable of measuring the amplitude and phase of the received optical beam simultaneously and independently. It operates in a shearing mode; one beam is stationary and the other scans. Since both beams are reflected from the sample, any vertical vibration of the sample causes the same phase change in both beams, so the phase difference (which is what an interferometer measures) does not change, and no noise results. The superior stability afforded by this arrangement allows angstrom height measurements at d.c. in ambient air. This is in sharp contrast to most interferometers, where the $1/f$ noise is dominant, and good sensitivity is difficult to attain.

Because this is a shearing rather than a differential system, it is not limited to small phase shifts; it is equally accurate throughout 2π radians. As in any monochromatic-light interferometer, phase shifts larger than 2π introduce an ambiguity, which in this case is mitigated since depth-discrimination supplies additional height information. The electronic scanning system is very fast, accurate, and repeatable. It is coupled to a data acquisition system of our own design (Hobbs, 1987) which allows it to acquire data at a pixel rate of 50 kHz, with an accuracy of $0{\cdot}1^\circ$ of phase and 12 bits in amplitude. This performance level does not by any means represent an upper limit to what could be achieved with this technique.

Having both phase and amplitude, along with a good theoretical expression for the coherent transfer function (CTF) of the microscope allows us to use simple digital filters to eliminate the disadvantages of the type 2 (confocal) transfer function, specifically its inferior bandwidth utilization and its cusp at the origin (which causes a slow monotonic tail off in the PSF) while maintaining control over the edge artefacts. As will be seen, the trade-off between sharpness and edge ringing is very favourable; improvements in edge resolution of more than a factor of two can be obtained with very reasonable ripple and very fast settling of the step response.

It is in this regard—2π phase sensitivity, 0·1 nm height accuracy, high speed, and completely arbitrary PSF (within the bandwidth limit)—that we claim this system to be a true generalization of the confocal microscope.

THE AMPLITUDE POINT-SPREAD FUNCTION

In a coherently illuminated bright-field (type 1) microscope, the amplitude PSF is simply the amplitude distribution of the illumination system, normally an Airy pattern:

$$h_1(x, y) = 2\pi NA^2 \,\mathrm{jinc}[kNA(x^2 + y^2)^{1/2}], \tag{1}$$

where $\mathrm{jinc}\, x = \mathbf{J}_1(x)/x$. This pattern corresponds to a pupil illuminated uniformly out to a numerical aperture (NA):

$$H_1(p, q) = \mathrm{circ}[(p^2 + q^2)/NA^2], \tag{2}$$

and is typically generated by imaging an unresolved pinhole light source on to the sample. The classical confocal (type 2) microscope uses two pinholes, one on the illumination side and one on the detector side of the optical system, to produce a

microscope whose PSF is proportional to the square of the illumination pattern,

$$h_2(x, y) = 4\pi NA^2 \,\mathrm{jinc}^2[kNA(x^2 + y^2)^{1/2}]; \tag{3}$$

by the convolution theorem, the transfer function is the self-convolution of the type 1 microscope's,

$$H_2(p, q) = (2/\pi)[\cos^{-1}|\sigma| - |\sigma|(1 - \sigma^2)^{1/2}]\mathrm{circ}\,|\sigma|, \tag{4}$$

where $\sigma = 1/(2NA)\sqrt{p^2 + q^2}$, normalized for unit response at zero spatial frequency. Since H_1 is real and non-negative (at least on-focus), its self-convolution H_2 has twice the bandwidth. Unfortunately, the type 2 transfer function is not twice as good, because it is roughly conical in shape (in two dimensions), has a cusp at the origin, and thus rolls off quite steeply in the mid-frequencies. Although it is often stated that the type 2 microscope has 1·4 times the resolution of the type 1, this is somewhat misleading since this figure is the ratio of the 3-dB widths of the PSFs; the step responses have almost the same rise interval (10–90%), with the type 1 even being a bit sharper, although the type 2 settles down more rapidly and exhibits much less overshoot and ringing. It is the absence of ringing which is important in semiconductor applications, because critical dimension measurements rely on threshold algorithms, which are easily confused by such artefacts. In addition, pinhole-type microscopes are sensitive only to intensity, which tends to limit their usefulness to samples with fairly strong amplitude contrast.

The type 2 transfer function still leaves something to be desired for microscopy applications. Although it has twice the bandwidth of the type 1, it uses its bandwidth poorly; it rolls off quite sharply at mid-frequencies, compromising the edge response. In addition, while avoiding the type 1's jump discontinuity at the band edge, it exhibits a cusp at zero frequency which leads to a slow, nearly monotonic tail in the step response, making precise height and reflectivity measurements unnecessarily difficult near sample discontinuities such as the edges of integrated circuit (IC) lines.

OPTICAL SYSTEM

The optical system is shown schematically in Fig. 1. A collimated, uniform beam from a single-frequency argon-ion laser operating at 514 nm comes into a tellurium-dioxide acousto-optic (A–O) deflector, located at the pupil of a commercial microscope (Leitz Orthoplan). The two beams (zero- and first-order) from the A–O cell are focused to diffraction-limited spots on the sample, with beam axes normal to the surface. Light reflected from the two spots retraces its path, and re-emerges through the A–O cell. Since two beams enter the cell on the return path, four receive beams emerge, in two coaxial pairs. Within each pair, one beam comes from each spot. One pair, consisting of the never-diffracted beam and the twice-diffracted (once on transmit and once on receive) beam, travels back towards the laser, where minor misalignment

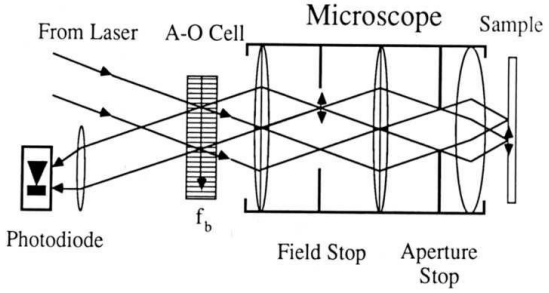


Fig. 1. Schematic diagram of the optical system.

keeps them from re-entering the laser cavity. The other two beams, namely the zero-order transmit beam which is diffracted on receive, and the first-order transmit beam which is undiffracted on receive, emerge from the cell at the Bragg angle; due to the differing diffraction geometry on transmit and receive, one of these is upshifted and one downshifted by the A–O cell drive frequency f_b. These two beams are made to interfere on a photodiode, placed at an image of the A–O cell so that the beams do not move off the diode as they scan.

The signal from the photodiode is proportional to the squared modulus of the beam amplitude, filtered by the frequency response of the diode circuit; since the incoming beam has components at two frequencies, the diode output current has two d.c. terms corresponding to the average power, plus an a.c. term at the beat frequency $2f_b$, whose amplitude is the product of the optical amplitudes and whose phase is the difference of the optical phases.

This signal is converted to a constant 60-MHz IF, filtered, and finally digitized.

The heterodyne interferometer mounts on a standard metallurgical microscope, and allows viewing of the beams as they scan, which is very useful for alignment, focusing, and positioning. The system could be extended to a 2-D scan without much difficulty, by including another A–O cell and a transfer lens.

In the course of a scan, the A–O cell drive frequency varies from 50 to 100 MHz, so that the rf photocurrent is at 100–200 MHz. Typical gas lasers have a mode spacing of 150–400 MHz, so care must be taken to avoid spurious signals due to the beating of the other modes (all of which would be split by the A–O cell) getting into the IF passband. Our instrument uses a single-frequency argon-ion laser operating at 514 nm. A frequency-doubled, diode-pumped Nd:YAG laser (532 nm)—with its very large mode spacing (10 GHz or so), small size, and comparatively low cost—would be a good alternative, especially since the problem of large (10%) amplitude fluctuations seems to have been solved.

The scan size is determined by the focal length of the microscope and the deflection angle of the A–O cell. The particular cell used here was a Matsushita EFLD-340.

In the past, A–O deflectors have had a bad reputation for beam distortion, which has made them somewhat unpopular in imaging applications; we have found with both devices we have tested (the other one was a one-off from Crystal Technology) that diffraction-limited resolution was easily obtained with pupil sizes of 3 or 4 mm.

ANALOG SIGNAL PROCESSING

The photodiode is essentially a current source, so for best sensitivity, its load resistance should be as large as possible. Real photodiodes have a shunt capacitance of a few picofarads at least; to achieve a 200-MHz bandwidth in an untuned circuit requires a load resistor of at most a few hundred ohms. Better results can be achieved with a matching network; a theorem of Bode (1945) states that in matching a resistance–capacitance to a pure resistance, the matching gain G_m of the network satisfies

$$-\int_0^\infty \log[1 - |G_m(\omega)|]\,\mathrm{d}\omega \leqslant \frac{2\pi}{R_L C}. \tag{5}$$

The photodiode matching network shown in Fig. 2 achieves a measured worst-case return loss of 5 dB into 50 Ω over a 100–210-MHz bandwidth, less than 0·5 dB worse than the Bode limit for a perfect rectangular frequency response.

Figure 3 is a block diagram of the analog signal processing subsystem. The signal from the photodiode at $2f_b$ is amplified, mixed with a frequency-doubled, down-shifted version of the A–O cell drive signal to yield a constant 60-MHz IF signal, and finally filtered, split into two highly isolated channels, and amplified to a nominal

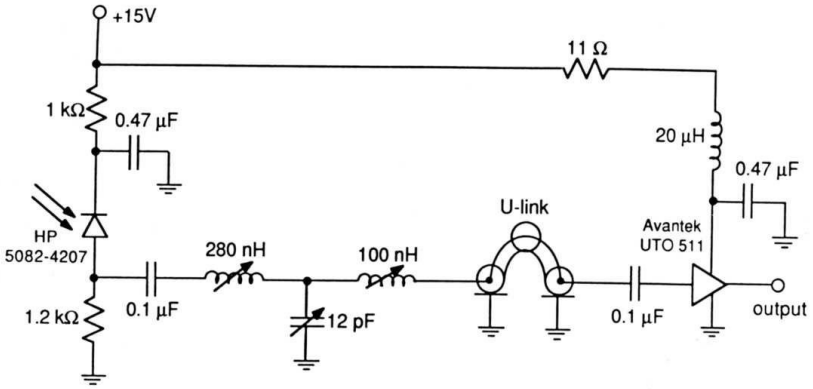


Fig. 2. Schematic diagram of the photodiode matching network.

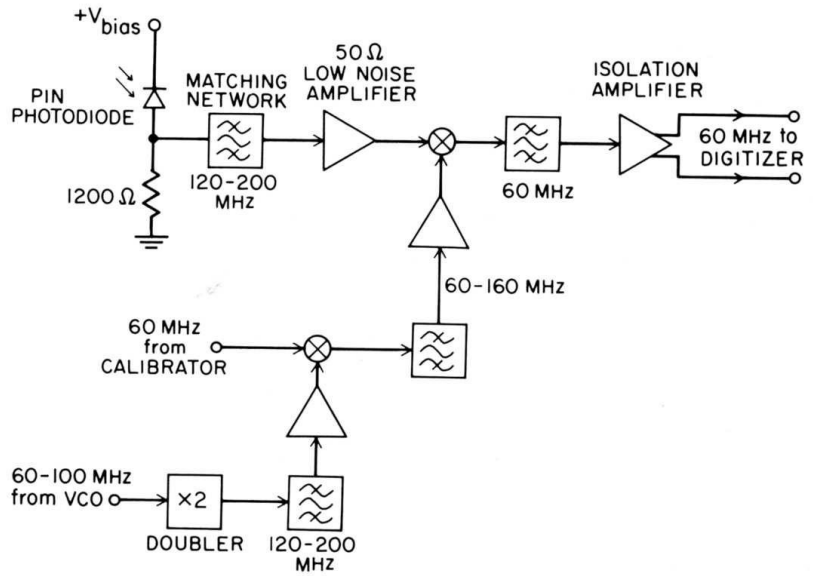


Fig. 3. Block diagram of the analog signal processing subsystem.

level of 0 dBm. The outputs go to amplitude and phase digitizers of our own design, which are described in Hobbs (1987). These digitizers operate at a 50-kHz sampling rate with a phase uncertainty of 0·1° (1σ) and an amplitude uncertainty of 0·2% of the reading—essentially the performance of a good lock-in amplifier, at 500 times the speed. The use of commercially available digitizing arrangements, such as a down-converter followed by a lock-in amplifier, or in-phase and quadrature detection, was rejected because such schemes lack the accuracy and/or speed to match the capabilities of the microscope.

The A–O cell is driven by a voltage-controlled oscillator (VCO) which in turn is controlled by computer via a digital-to-analog converter (DAC). The scan is linearized by performing a secant-method search through the DAC code space (using a GPIB-controlled frequency counter) to find N codes which generate N equally spaced frequencies.

There is one somewhat subtle problem with the system as described so far. The A–O cell's time–bandwidth product determines the number of resolvable spots, so with a bandwidth of 50 MHz, one must contend with a delay of about 6 μs in the A–O cell. As the frequency changes, this delay causes a large linear phase shift (amounting to

hundreds of cycles over the scan range) between the photocurrent and the doubled VCO signal. This large delay must be cancelled somehow if the phase sensitivity is to be useful. At present, the phase slope is removed numerically, by forcing the average phases of the first five and last five points in the scan to be equal; this method is convenient, but places very stringent demands on the stability and repeatability of the VCO. Other ways to accomplish this would be to use a bulk-acoustic wave delay line after the frequency doubler, and then trim the null by translating the A–O cell along the acoustic propagation direction, or to replace the VCO with a direct digital synthesizer, which has the required accuracy and tuning speed.

IN-FOCUS TRANSFER FUNCTION AND PSF

Although this instrument has no pinholes, it is nevertheless of type 2. In this section, we develop the theory for this instrument, in order to elucidate the origin of the lateral resolution and to justify the approximations made.

Consider the instantaneous power incident on the detector, which is considered to be infinite in extent (in practice it need only be bigger than the beam); if the input scalar field in reciprocal space is $\psi(\mathbf{k})$, the incident power is:

$$P = \iint \frac{\mathrm{d}^2 k'}{(2\pi)^2} |\psi(\mathbf{k}')|^2 \frac{\mathbf{n}\cdot\mathbf{k}'}{|\mathbf{k}'|}. \tag{6}$$

Since the numerical aperture at the detector is small, $\mathbf{n}\cdot\mathbf{k}'$ is approximately k'. Using Rayleigh's theorem to transform into real space, we obtain for the photocurrent

$$I = R \iint_{\text{pupil}} \mathrm{d}^2 x'' |\psi(\mathbf{x}'')|^2. \tag{7}$$

In order to get Eq. (7) into an integral over the sample surface, we use the far-field form of the Rayleigh–Sommerfeld integral formula,

$$\psi(\mathbf{x}) = (1/i\lambda) \iint_{\text{sample}} \frac{\exp(ik|\mathbf{x}-\mathbf{x}'|)\ \mathbf{n}\cdot(\mathbf{x}-\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|^2} \psi(\mathbf{x}')\, \mathrm{d}^2 x'. \tag{8}$$

Consider the fields on the surface of a sphere of radius a. If we define angular variables $p = x/a$, $q = y/a$, then

$$|\mathbf{x}-\mathbf{x}'| = (a^2 + \mathbf{x}'\cdot\mathbf{x}' - 2\mathbf{x}\cdot\mathbf{x}')^{1/2} \simeq a(1 - \mathbf{x}\cdot\mathbf{x}'/a^2) + O(a^{-1}). \tag{9}$$

For large a, $\mathbf{x}\cdot\mathbf{x}'/a \ll a$ so that this term can be dropped in the denominator; however, this term is never small compared to unity, so it must be retained in the exponent. If we discard the large constant phase term $\exp(ika)/a$, then the angular spectrum is the Fourier transform of the source distribution, multiplied by the hemispherical window function $\mathrm{circ}[(p^2+q^2)/NA^2]\sqrt{(1-p^2-q^2)}$. The window function arises from the obliquity factor $\mathbf{n}\cdot(\mathbf{x}-\mathbf{x}')$ in Eq. (8), and from the finite range of angles available, not from any effect of lenses. As a corollary, if we wish to calculate the scattered field when a beam is incident on the surface, we must put in the inverse of this obliquity factor; in most cases of interest this means that the obliquity factors cancel out, at least approximately.

There is no general theory for what a real lens does. The unphysical ray model is used in lens design, the thin lens (for small NA) and Fourier transformer (for large NA) models in most imaging theory. Sometimes an extra $\cos^{1/2}\theta$ apodization factor is added for energy conservation; this seems a needless elaboration in an approximate theory, especially since energy conservation in propagating from the source to the detector (for a perfect reflector) is ensured anyway by the cancellation of the obliquity factors.

Since we are interested in the large-NA limit, we consider an aplanatic, non-vignetting lens of unit focal length. Apart from its finite NA, it performs an ideal unit scaling operation from the angular coordinates p and q to pupil plane spatial coordinates x'' and y'', provided the sample is at the front focal plane. If the NA were unity, the ideal unit focal length objective could be expressed as $O[\psi(p, q)] = \psi(x'', y'')$. Hereafter, the distinction between (p, q) and (x'', y'') will be dropped. A lens with $NA < 1$ merely cuts off those spatial frequency components which fall outside the circle $p^2 + q^2 > NA^2$. Re-writing Eq. (7) for the photocurrent gives

$$I = R\iint_{-\infty}^{\infty} \mathrm{d}p\,\mathrm{d}q\,|\psi_r(p, q)|^2\,\mathrm{circ}[(p^2 + q^2)/NA^2], \tag{10}$$

where ψ_r is the reflected field. In general, real samples will have topography; they will not be perfectly flat. Nevertheless, for analytical tractability, the sample is modelled as an infinite plane, with a complex amplitude reflection coefficient $\rho(x, y)$, which includes both reflectivity and height information. In this case $\psi_r(x, y) = \psi_{\mathrm{illum}}(x, y)\rho(x, y)$, so that $\psi_r(p, q) = \psi_{\mathrm{illum}}(p, q)\star\rho(p, q)$, where $\star$ denotes convolution.

If looked at too closely, ρ is a somewhat poorly defined quantity, since we expect it to depend on both $\mathbf{k}$ and $\mathbf{x}$. Dependence on $\mathbf{x}$ arises from spatial inhomogeneities in the sample (features), while $\mathbf{k}$ dependence comes from both intrinsic variations of ρ with $\mathbf{k}$ (e.g. the Fresnel formulae for dielectrics) and the different phase shifts $\exp(ik_z z)$ suffered by different plane waves due to height variations in the sample. It is thus analogous to a classical variable in quantum mechanics, whose validity depends on its not varying too rapidly with $\mathbf{x}$. In this case, we can integrate over the $\mathbf{k}$ dependence, arriving at an average amplitude and phase factor to apply to all plane wave components of the incident beam; this factor is allowed to vary slowly with $\mathbf{x}$ as well. The assumption of constant phase shifts is equivalent to infinite depth of field, and hence restricts us to small height changes, $\Delta z \ll \lambda/(1 - \cos\theta)$ where $\sin\theta = NA$; this is physically reasonable since if the sample is going in and out of focus during scanning, the infinite plane model is clearly not applicable.

First, consider a single-beam microscope, in which we illuminate the sample with

$$\psi_{\mathrm{illum}}(p, q) = \exp[-ik(px' + qy')]\,\mathrm{circ}[(p^2 + q^2)/NA^2], \tag{11}$$

and detect only $(p, q) = (0, 0)$ plane wave component of the returned beam.

Provided ρ is sufficiently narrowband, an incoming plane wave component will be scattered only into 'nearby' plane waves, so that the obliquity factors coming and going cancel, and we can write

$$I = R\iint_{-\infty}^{\infty} \mathrm{d}p\,\mathrm{d}q\,|[\psi_{\mathrm{illum}}(p, q)\star\rho(p, q)]|^2\,\delta(p, q). \tag{12}$$

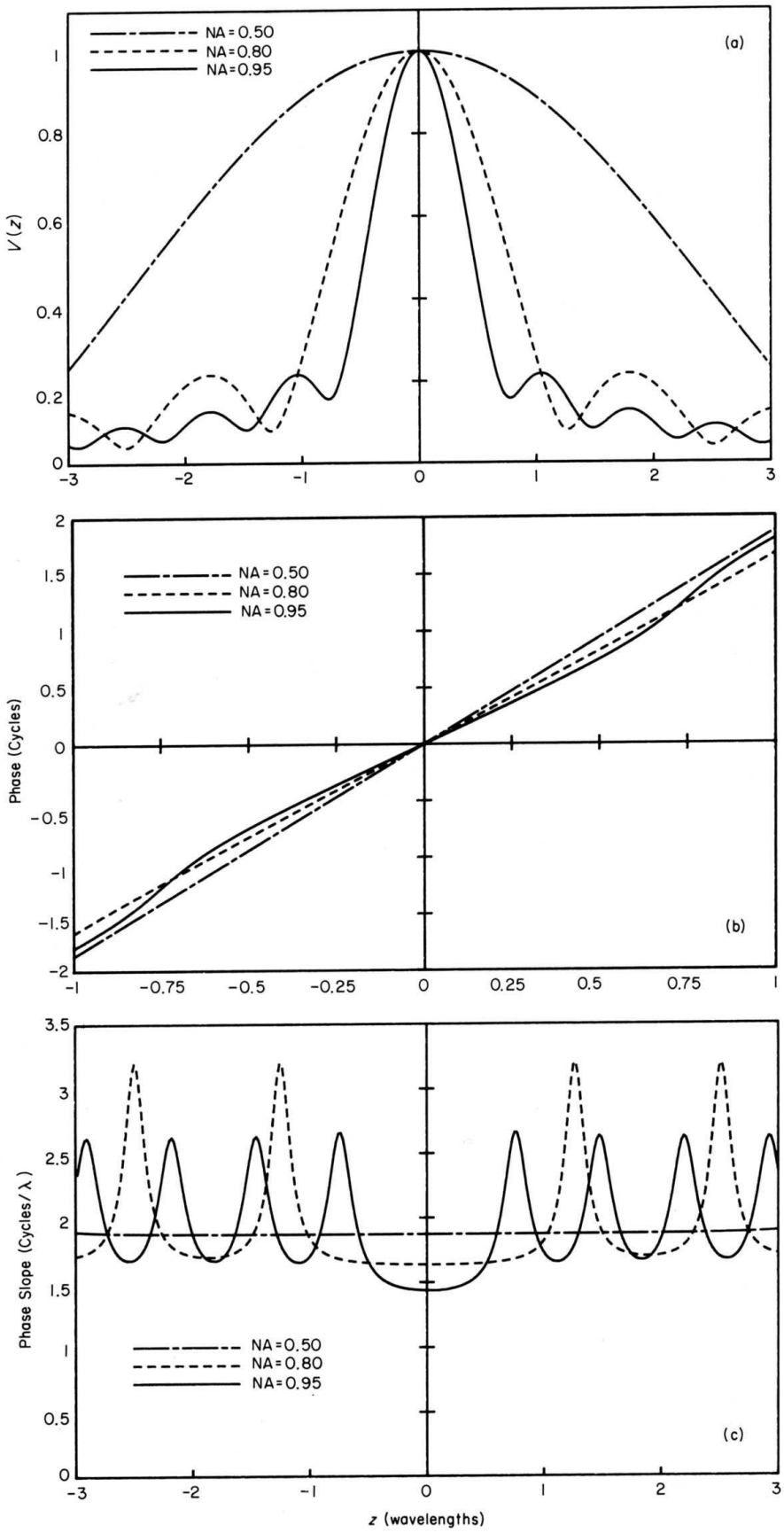
The central value of the convolution equals the infinite integral of the product of the transforms, and thus on the sample,

$$I = \kappa^2 R\left|\iint_{-\infty}^{\infty} \mathrm{d}^2x\, 2\pi NA^2\,\mathrm{jinc}(kNA|\mathbf{x} - \mathbf{x}'|)\rho(\mathbf{x})\right|^2, \tag{13}$$

where $\kappa = 2\pi NA^2$ and $\mathrm{jinc}(x) = \mathbf{J}_1(x)/x$. This shows that the microscope (up to the detector) is a linear system in complex amplitude, with a PSF of $h_1(x, y) = \kappa\,\mathrm{jinc}[k(x^2 + y^2)^{1/2}NA]$.

TYPE 1 AND TYPE 2 MICROSCOPES

We have so far described one implementation of a type 1 microscope, since a δ-function detector at the pupil can most easily be implemented with a large-area coherent detector at the image plane. In our heterodyne system, a uniform reference


NA = 0.50
NA = 0.80
NA = 0.95
(a)
V(z)
1
0.8
0.6
0.4
0.2
0
-3
-2
-1
0
1
2
3
NA = 0.50
NA = 0.80
NA = 0.95
(b)
Phase (Cycles)
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-1
-0.75
-0.5
-0.25
0
0.25
0.5
0.75
1
NA = 0.50
NA = 0.80
NA = 0.95
(c)
Phase Slope (Cycles/λ)
3.5
3
2.5
2
1.5
1
0.5
0
-3
-2
-1
0
1
2
3
z (wavelengths)

beam $\psi_{\text{bias}} \equiv \psi_{\text{illum}} \exp(-i2\omega_b t)$ is introduced which is coherent with the received beam, but with a frequency offset of $2f_b$; as described in the previous section, this is derived from a second beam reflected from another point on the sample. In this case, under the same assumptions, the photocurrent is given by

$$I = R \iint_{\text{pupil}} |\psi_{\text{illum}}(p, q) \exp[-i(\omega + \omega_b)t] + [\psi_{\text{illum}}(p, q) \star \rho(p, q)] \exp[-i(\omega + \omega_b)t]|^2 \, dp \, dq. \quad (14)$$

Under the same assumptions as before, and using Rayleigh's theorem, the a.c. photocurrent is given by

$$I_{\text{a.c.}}(\mathbf{x}) = \frac{\kappa^2 R}{\pi} \mathbf{Re}\left[\iint \text{jinc}^2(kNA|\mathbf{x} - \mathbf{x}'|)\rho(\mathbf{x}')\mathrm{d}^2x'\right]. \quad (15)$$

This is a convolution of $\rho(x, y)$ with a jinc^2 function, which shows that this microscope is of type 2 (i.e. confocal).

DEFOCUS RESPONSE

Besides the jinc^2 PSF, the most salient feature of the confocal microscope is depth discrimination, or optical sectioning: the images of out-of-focus features are much reduced in amplitude. In a conventional confocal microscope (with pinholes), this property is easily understood: a scanned spot reflected from an out-of-focus feature will produce a large spot in the image, so most of the light will miss the pinhole and hence not contribute to the detected image.

Our heterodyne system exhibits depth discrimination as well, but only for *differential* defocus. Since both the reference and receive beams are reflected from the sample, if the sample moves out of focus, both beams are affected identically and the interference term $\psi_{\text{bias}}\psi^{\star}_{\text{illum}}$ is still perfect, so the rf signal does not decrease. For differential defocus (provided the reference beam remains in focus), the depth-discrimination properties of this microscope are identical to those of the conventional confocal system.

CALCULATION OF $V(z)$

If only the probe beam is displaced by a distance δz, each plane wave component (p, q) will be phase shifted by $\Delta\phi = 2k_z\delta z = 2k\delta z\sqrt{(1 - p^2 - q^2)}$. If it is assumed that the sample is a perfect reflector, $\rho(x, y) \equiv 1$, and that the pupil function is uniform, the rf photocurrent is given by

$$I_{\text{a.c.}} = \frac{\kappa^2 R}{\pi} \text{Re}\left[\exp(-i2\omega_b t) \iint_{\text{pupil}} \exp[i2kz(1 - p^2 - q^2)^{1/2}] \, dp \, dq\right]. \quad (16)$$

Writing $\beta = 2kz$, $\sin\theta = NA$, $\alpha = \beta \sin^2(\theta/2)$, and $\gamma = \beta\cos^2(\theta/2)$, this becomes

$$I_{\text{a.c.}} = 2\pi\kappa^2 R \, \text{Re}\left[\exp(-i\omega_b t) \int_{\cos\theta}^{1} r \, dr \exp(i\beta r)\right], \quad (17)$$

and therefore if $V(z) \equiv I_{\text{a.c.}}(z)/I_{\text{a.c.}}(0)$,

$$V(z) = \frac{4}{\beta^2 NA^2}\left\{\begin{array}{l} \alpha\sin\beta + \sin\alpha(\beta\cos\theta\cos\gamma - \sin\gamma) \\ + i[-\alpha\cos\beta + \sin\alpha(\beta\cos\theta\sin\gamma + \cos\gamma)] \end{array}\right\}. \quad (18)$$

Fig. 4. Vertical response $V(z)$ of the confocal microscope, for NA = 0·5, 0·8, and 0·95: (a) magnitude, (b) phase, (c) phase derivative $\mathrm{d}\phi/\mathrm{d}z$.

Figure 4(a), (b) and (c) shows the magnitude, phase, and phase derivative $d\phi/dz$ for NA = 0·5, 0·8 and 0·95. Note the peculiar structure of the phase slope curve; this can introduce errors into height measurements if care is not exercised.

For small NA, r is close to 1, so we neglect $(r - 1)$ in order to simplify the integral: then we find that

$$V(z) \simeq \mathrm{sinc}[2\pi z(1 - \cos\theta)/\lambda] \exp[ikz(1 + \cos\theta)]. \tag{19}$$

This is a good approximation for the phase shift in the vicinity of focus for small NA, with an error of about 0·2% up to NA = 0·5; it swiftly becomes a poor approximation for larger NA, which is the case of greatest interest; its error is −5% at 0·9 NA and −8% at 0·95.

LINE-SPREAD FUNCTION AND STEP RESPONSE

When imaging 1-D structures such as IC lines, it is important to know the line-spread function (LSF) and step response of the microscope. One way to obtain the LSF is to treat the integral along the line as a convolution with a line delta function $\delta(x)$; in the frequency domain that is a multiplication by $\delta(q)$, so that the line spread is the inverse Fourier transform of a central cross-section of the 2-D transfer function. For a type 1 microscope with a uniform pupil function $H_1(p, q) = \mathrm{circ}[(p^2 + q^2)/NA^2]$, this is a rectangle, so that the LSF is

$$l_1(x) = 2\sin(kxNA)/kx \tag{20}$$

and the step response is

$$s_1(x) = \frac{1}{2} + \frac{1}{\pi}\mathrm{Si}(kxNA). \tag{21}$$

For the type 2 microscope, the transfer function is

$$H_2(p, q) = (2/\pi)[\cos^{-1}(\sigma) - \sigma(1 - \sigma^2)^{1/2}]\,\mathrm{rect}(\sigma/2), \tag{22}$$

where $\sigma = (p^2 + q^2)^{1/2}/(2NA)$ and the normalization has been chosen to make the central value unity; the LSF $l_2(x)$ and step response $s_2(x)$ are not elementary, and their calculation is treated in Appendix 1. Figure 5 is a plot of $l_2(x)$ and $l_1(x)$, which shows the much improved side lobes and somewhat narrower main lobe (3 dB width about 30% smaller) of the type 2 system. Figure 6 is a plot of $s_2(x)$ and $s_1(x)$, showing the greatly reduced edge ringing of the type 2 system. This improvement is due to the continuity of the transfer function; instead of the abrupt cutoff of the type 1 transfer function $H_1(p)$, $H_2(p)$ is tangential to the axis. The ultimate rolloff of the type 2 LSF l_2 goes as ξ^{-2} because of the cusp in H_2 at the origin (see for example Bracewell, 1978).

It should be noted that the edge rise interval of the type 2 step response is not very different from that of the type 1—indeed the type 1 is slightly faster from 10–90%. The fundamental advantage of the type 2, especially in technological applications, is its freedom from edge artefacts due to ringing. These artefacts can be very confusing, especially when the sample has many sharp edges and small features. In IC metrology applications, this problem is made even more difficult by the widespread use of threshold algorithms for measuring critical dimensions; such algorithms are easily fooled by artefacts of this kind.

LIMITS OF FOURIER OPTICS

As we have shown, the paraxial assumption made in the usual derivations of Fourier optics is unnecessary; the formalism can be extended to large-NA imaging provided

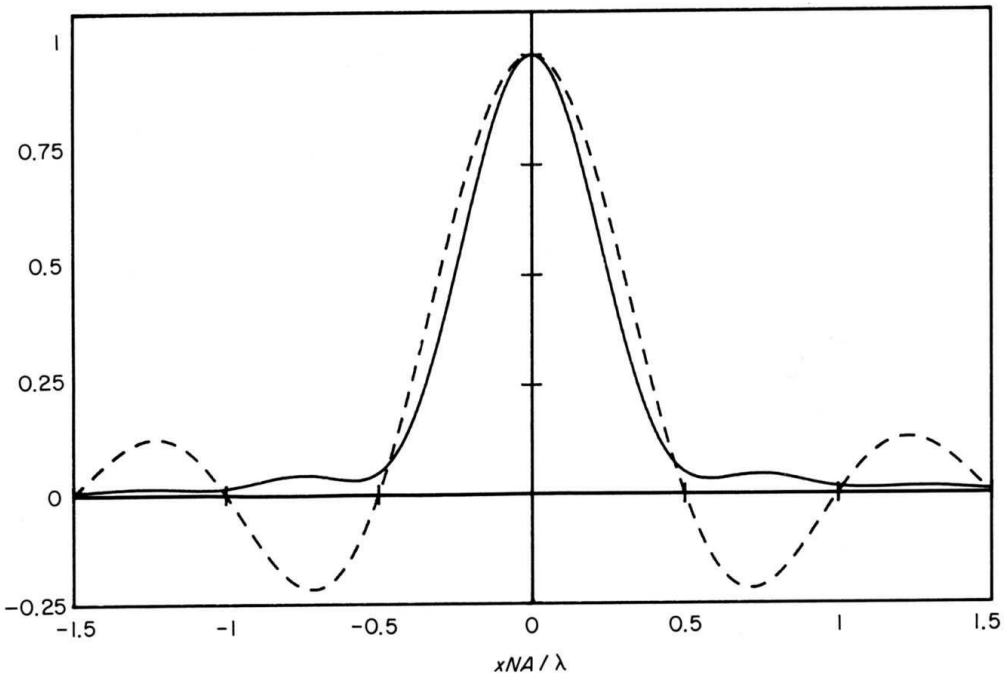


Fig. 5. Type 1 (– – –) and type 2 (——) line spread functions $l_1(x)$ and $l_2x)$, showing the improved performance of the type 2 system.

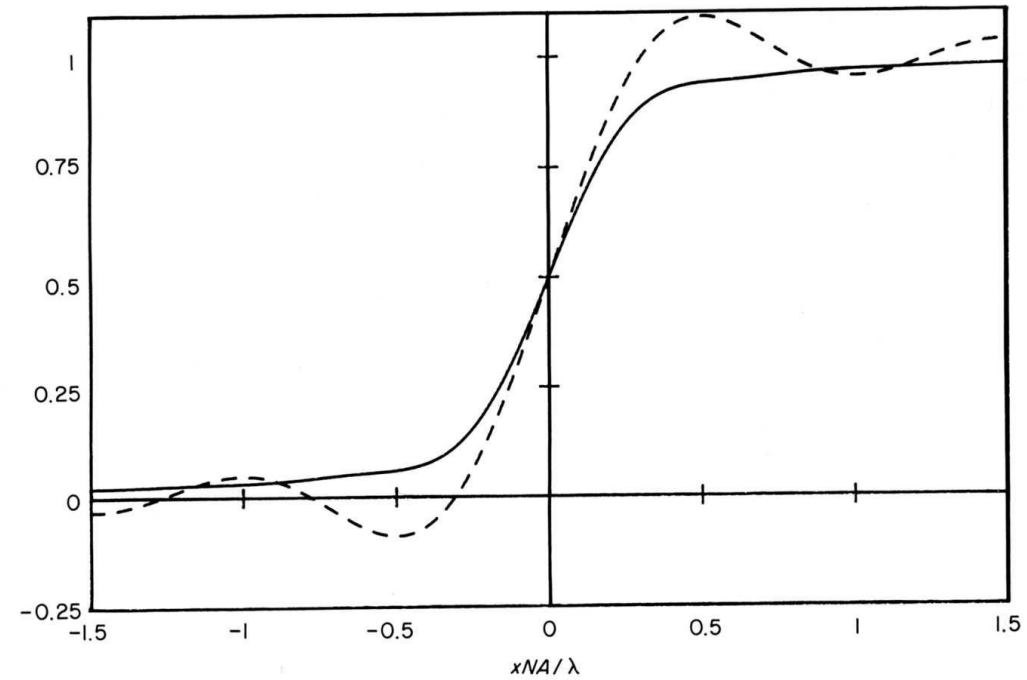


Fig. 6. Type 1 (– – –) and type 2 (——) step responses $s_1(x)$ and $s_2(x)$. Note that the type 1 is actually sharper, although it rings badly.

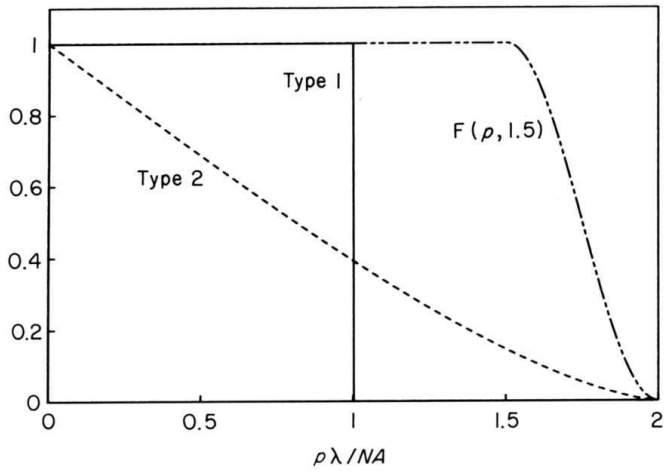


Fig. 7. Comparison of transfer functions: type 1 $H_1(p)$, type 2 $H_2(p)$, and $F(p; 1{\cdot}5)$. Note the smoothness and good bandwidth utilization of $F(p; 1{\cdot}5)$ compared with the others.

the sample is not too rapidly varying and the lens is a Fourier transformer between pupil and image planes (many lenses, such as camera lenses and infinity-corrected microscope objectives, are optimized for such service). The Fourier approximation breaks down when the convolution theorem fails to hold: in (12) we assumed this theorem even when evanescent waves are involved. We recall that in (12) we used the convolution of $\rho(p, q)$ with $\psi_{\text{illum}}(p, q)$ as the reflected field; however, in propagating from the sample to the detector, all evanescent components are lost, and the convolution theorem does not hold. If $\rho(x, y)$ varies sufficiently rapidly, this effect will be important enough to invalidate our approximation. If the sample exhibits large tilts, then the true vectorial boundary conditions will not be uniformly similar to the Dirichlet conditions used to derive the Rayleigh–Sommerfeld integral formula (8) and the various components of **E** and **H** will become mixed, in principle, invalidating the scalar approach, although in practice the predicted response still usually agrees well with experiment. Related to this is the defocus problem. If the sample goes in and out of focus, the impulse response is no longer translationally invariant; this may also indicate that the side walls of the peaks and valleys are no longer small enough to be ignored. These problems will be addressed in a subsequent paper.

DIGITAL DECONVOLUTION

Since our heterodyne interference microscope produces simultaneous measurements of optical amplitude and phase in digital form, it is perfectly legitimate (in the sense of preserving information while introducing no new assumptions) to process the data with a digital filter to obtain almost any transfer function desired, within the bandwidth limit; hence it is natural to ask whether the type 2 transfer function $H_2(p)$ makes optimal use of its extra bandwidth. Figure 7 shows $H_1(p)$ and $H_2(p)$, as well as a function $F(p; 1{\cdot}5)$, a representative of a family of transfer functions chosen for good bandwidth utilization; each curve of the family is constant to a frequency p_0, where it changes smoothly over to a raised cosine curve chosen to be tangential to the axis at $p = 2\text{NA}$. These curves are denoted $F(p; p_0/\text{NA})$, and their inverse transforms (LSFs) by $f(x; p_0/\text{NA})$ for clarity. All, except of course the rectangle $F(p; 2)$, have continuous first derivatives; this guarantees that the sidelobes of the LSFs fall off at least as fast as x^{-3} for large x. The adjustable parameter p_0 allows a trade-off of edge sharpness against ripple and noise bandwidth. $F(p; 0)$ is the familiar Von Hann ('Hanning') window, which has excellent ripple properties but whose LSF has a main

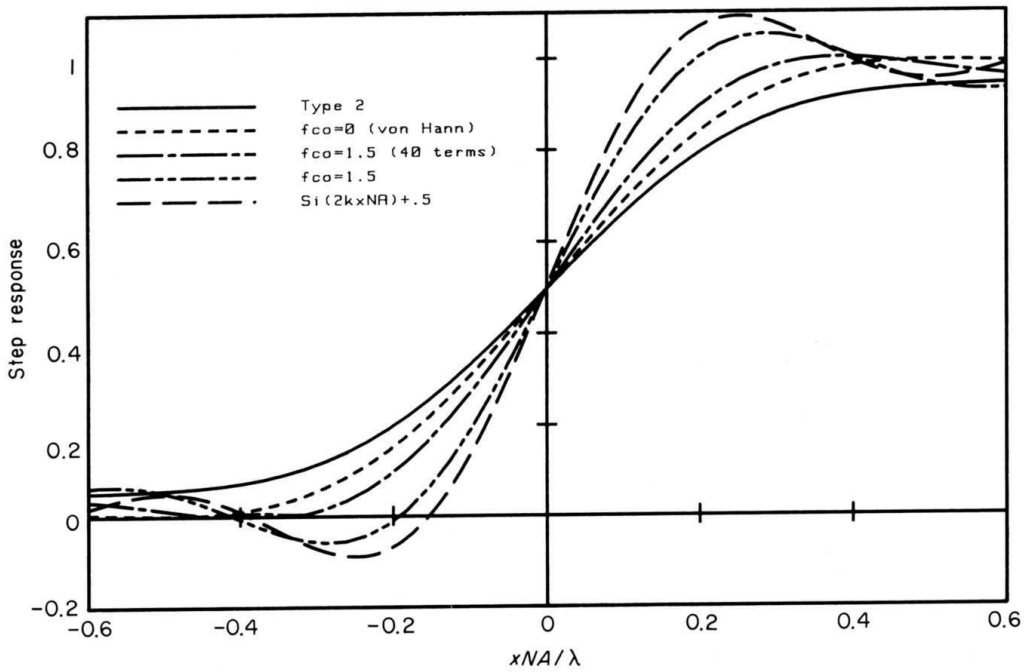


Fig. 8. Type 2 step responses $s_2(x)$, compared with those of $f(x; 0), f(x; 1{\cdot}5), f(x; 2)$ (the response of a type 1 microscope operating at the second harmonic), and s_2 processed with a filter which was designed using $F(p; 1{\cdot}5)$ and a window of length $2{\cdot}25\lambda/\mathrm{NA}$.

lobe twice as wide as that of $f(p; 2)$. As p_0 is increased, the main lobe narrows and the ripple increases; the trade-off between edge sharpness and sensitivity to fine detail near the edge (which may be masked by ringing) and the numerical properties of the filter dictates how far it is possible to go. For data with a good signal-to-noise ratio, it is profitable to go as far as $F(p; 1{\cdot}5)$; beyond this, ripple increases quite dramatically, and the edge sharpness hardly changes.

The disadvantages of the type 2 response stem primarily from the cusp at $p = 0$; it produces a positive, slowly decreasing (as x^{-2}) tail in the LSF, and hence an x^{-1} tail in the step response, which makes images of small structures somewhat more rounded than they need be. This is a significant limitation in a phase-sensitive system, since in order to perform a model-independent measurement of the height of a feature, the feature must be wide enough for the edge artefacts to disappear in the middle. In addition, although the ripple performance of the type 2 is quite good, its nearly triangular transfer function is not the best way to achieve this; other functions with better mid-frequency response achieve equivalent ripple with better edge definition and no long tail.

Figure 8 compares the step responses $s_2(x)$ of the type 2 to those of $f(x; 0), f(x; 1{\cdot}5)$, $f(x; 2)$ (the response of a type 1 microscope operating at the second harmonic), and s_2 processed with a filter designed with $F(p; 1{\cdot}5)$ and then truncated to a width of nine samples (projected down on the sample that is $2{\cdot}25\ \lambda/\mathrm{NA}$ wide), with a Kaiser window, $f_a = 4{\cdot}5$.

The improvement in edge sharpness with any of the f family over s_2 is immediately apparent, and the trade-off of edge response against ripple is also easily seen. Note that, compared with the others, the $f(x; 0)$ and truncated $f(x; 1{\cdot}5)$ step responses settle quickly to their final values; they are good choices for measuring the heights of narrow structures. The improvement in the tails is a direct result of eliminating the cusp in $H_2(p)$. Adjusting p_0 controls noise gain, while adjusting the number of samples

Table 1. Abscissas and heights of the peaks, and total noise gain, of the deconvolution filters used for resolution enhancement.

p_0	p_{pk}	Peak value	Noise gain (dB)
0·0	0·818	1·29	0·8
0·4	1·075	1·77	3·2
0·8	1·321	2·68	5·9
1·2	1·555	4·83	9·5
1·5	1·726	9·64	13·6
1·6	1·781	13·43	15·5
1·7	1·837	20·60	17·9
1·8	1·892	37·70	21·4
1·9	1·946	106·10	27·4

the edge speed versus ripple trade-off with a given noise gain. As p_0 increases, the edge resolution approaches that of a type 1 microscope operating at the second harmonic, allowing resolution equivalent to that of an ultraviolet microscope, while still using visible light.

No one member of the F family is clearly best; the choice depends on the object of interest. The design and use of these filters is so easy, requiring a total of 1 min to design, compute, and use a new filter on a line scan, that it is entirely reasonable to use different filters to accentuate different features in the same data set.

FILTER IMPLEMENTATION

The filtering is done non-recursively. The filter itself is generated by dividing the chosen $F(p)$ by $H_2(p)$, sampling, zero-padding, and taking the inverse discrete Fourier transform (DFT). Aliasing error introduced by this procedure is minimized by using about five times as many zeros as data points, and in any case $F(p)$ is somewhat arbitrary so minor departures are not objectionable. The resulting filter is truncated to the desired length using a Kaiser window (a Kaiser parameter f_a of 4·5 works well). The particular family $F(p)$ was chosen for simplicity and good bandwidth utilization, as mentioned above, but also to ensure that the digital filters are well conditioned; the filters selectively amplify high spatial frequencies, exhibiting quite a large peak near $p = \mathrm{NA} + p_0/2$, especially when p_0 is near 2NA. Table 1 lists the positions and heights of the peaks and the total noise gains of filters for various values of p_0.

For p near 2NA, $H_2(2\mathrm{NA}\text{-}\varepsilon)$ goes as $\varepsilon^{3/2}$, while F goes as ε^2, which means that the filter tends to zero like $\varepsilon^{1/2}$. Although there is no singularity, the noise gain of the filter becomes quite large as p_0 gets close to 2NA, because of the large peak; the choice of p_0 will be influenced by this consideration as well as by the ripple. The noise gain *per se* is not necessarily the most important thing; as the singular case $p_0 = 2\mathrm{NA}$ is approached, the peak becomes tall and narrow, causing the amplified noise increasingly to resemble ringing, a particularly undesirable characteristic. Since our microscope has such a good signal-to-noise ratio (better than 60 dB when laser noise is discounted), we are free to use such strongly peaked filters; this is one of the key advantages of the heterodyne interference microscope.

EXPERIMENTAL PERFORMANCE

The heterodyne interference microscope lives up to its potential. Figure 9 shows some scan data taken with an early version of the microscope (four-beam system with a lock-in time constant of 1 ms), with a numerical aperture of 0·9 and $\lambda = 514{\cdot}5$ nm; the sample was an 80 ± 5-nm-thick aluminium film on aluminium. The diamond symbols represent the raw experimental data, the solid line the theoretical response to a 78-nm-high phase step. The phase data agree extremely well, although the amplitude

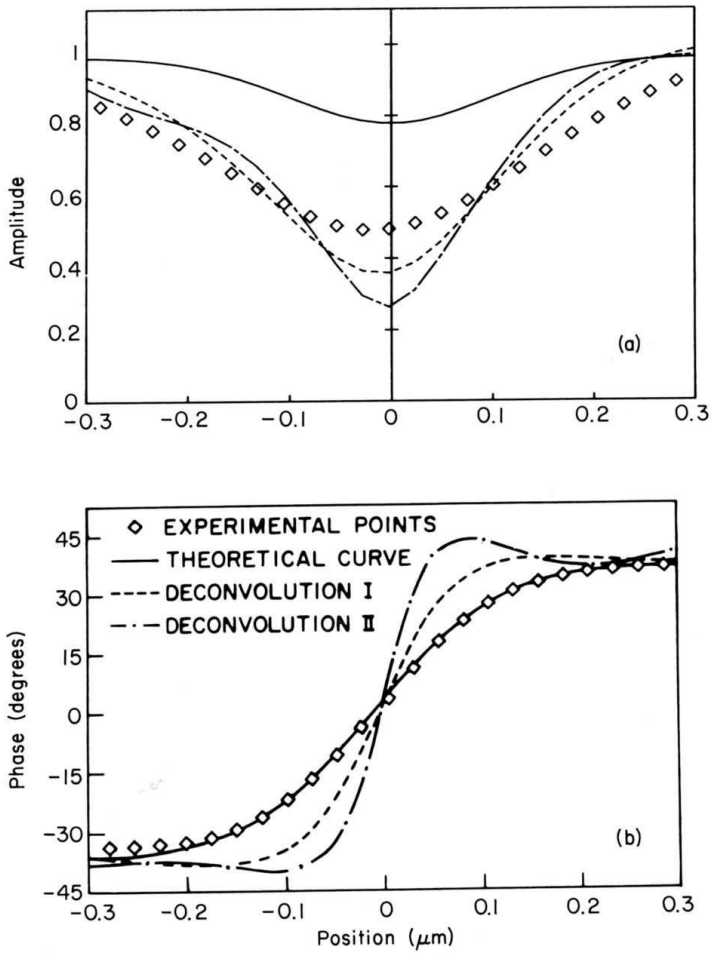


Fig. 9. Single scan of 80 ± 5-nm-tall Al step, taken with λ = 514·5 nm and NA = 0·9, compared with theory for a 78-nm step: (a) amplitude, (b) phase.

data do not (for reasons as yet unknown). The other curves are the experimental data filtered to $F(p; 1{\cdot}5)$ with a long filter (best sharpness at the expense of ripple) and with the same filter truncated to a width of nine samples (2·25 λ/NA on the sample). These filters are the same as those used in Fig. 7. As before, they improve the edge response and increase the ripple, with the trade-off favouring the somewhat milder filter. The edge sharpness (10–90%) is 230 nm (0·45 λ) for the raw data, 130 nm (0·25 λ) for the milder filter, and 95 nm (0·18 λ) for the long filter. The filtering has improved the edge sharpness by a factor of two, even though the imaging conditions were apparently not ideal. The discrepancy between the amplitude and phase data is not uncommon experimentally for metal edges; this is an advantage of phase sensitivity.

Figure 10 shows a scan taken with the system as described in this paper. The sample was a 2-GHz acoustic resonator, consisting of a periodic array of 0·4-μm lines and spaces on ST-cut quartz; the lines are 25-nm-thick aluminium. This sample, with its strong amplitude contrast and relatively weak phase contrast, is a very stringent test of the microscope's performance; such samples confuse traditional phase contrast schemes completely. In addition, any amplitude- to phase-modulation (AM–PM) conversion in the electronics would create errors in the phase data. The filter used was similar to the milder filter of Fig. 7. The agreement is excellent, both before and after filtering, apart from the 5-nm-high 'ears' on the line; we think these are real, since the

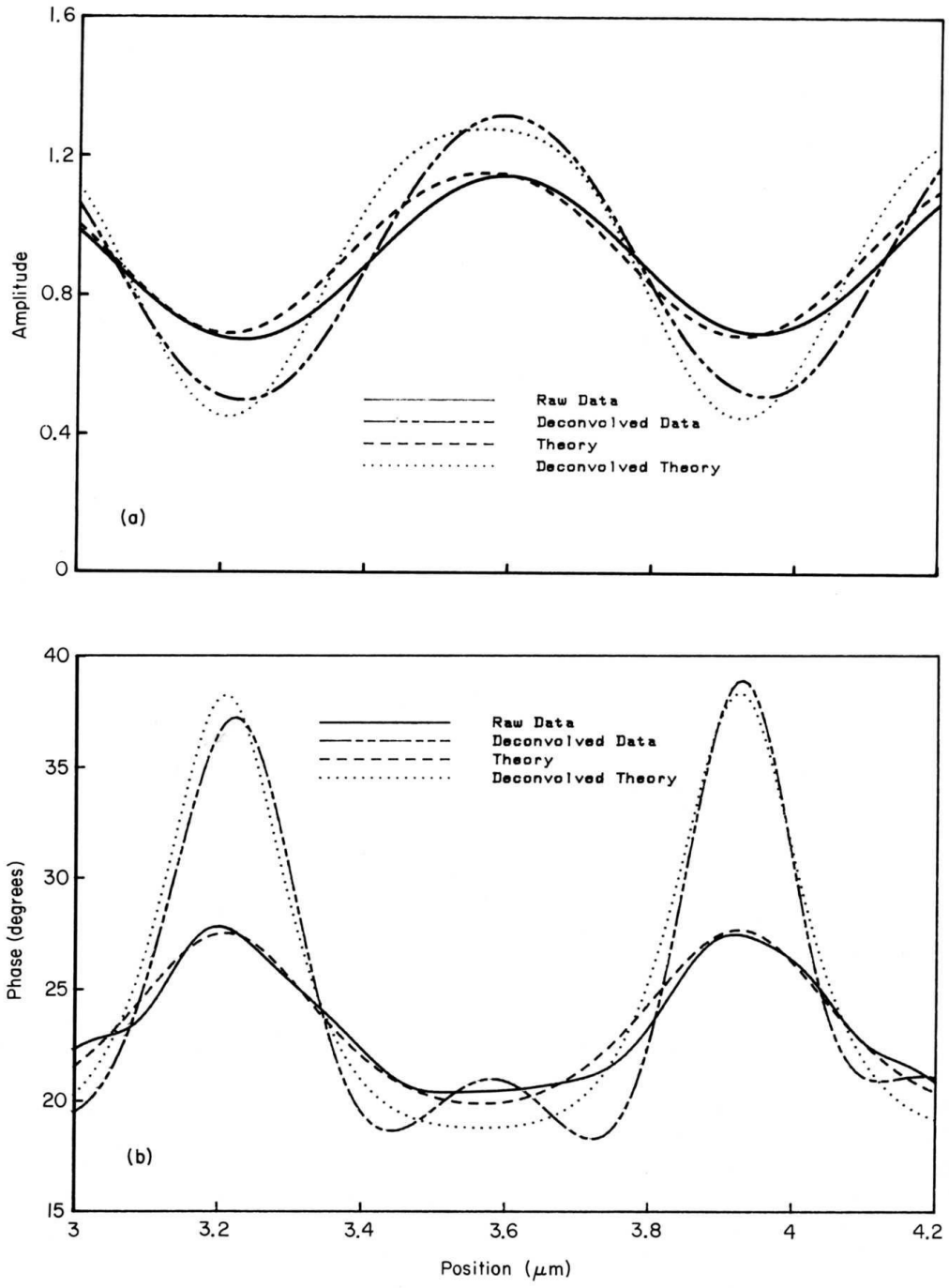


Fig. 10. Scan of a 2-GHz acoustic resonator, 0·4 μm wide × 25 mm thick. Al lines on ST-cut quartz, 0·8-μm pitch, before and after deconvolution, with theoretical plots for comparison. λ = 514·5 nm, NA = 0·95: (a) amplitude, (b) phase.

lines were made by a lift-off process, and they are apparent from the square shape of the raw data, compared to the theory. There is a slight lateral shift between the experimental and theoretical curves, arising from fitting the two over the entire array rather than this one line. The apparent asymmetry of the lines and spaces in the phase plot is due to the much higher reflectivity of the line. In order for the phase to slew halfway

to the substrate value, half the power in the returning beam must come from the substrate; because the substrate signal is comparatively weak, this means that most of the spot must be off the line, thus explaining the apparent asymmetry. In the filtered data, the apparent spot size is smaller, so that the phase is able to slew further before the edge of the next line is encountered.

CONCLUSIONS

The already good performance of the visible-light confocal microscope can be improved by a factor of two in edge sharpness, and heights can be measured with very high accuracy at high speed, by using a combination of heterodyne interferometry and digital filtering. The filters are based on the scalar theory of microscope imaging, and exhibit good performance even when used for submicrometre features. The agreement between experiment and theory is extremely good, which allows confident interpretation of scan data. Filtering the data is much easier and more flexible than attempting to apodize the pupil function with physical filters or annular pupils, and the filters can be tailored to the application easily and quickly. The flexibility and ease of use of this system recommend its use in any situation where precise dimensional measurements and the highest possible resolution are required.

REFERENCES

Bode, H.W. (1945) *Network Analysis and Feedback Amplifier Design*. Van Nostrand, New York (section 16.3).

Bracewell, R.N. (1978) *The Fourier Transform and Its Applications*, 2nd edn, pp. 143–146. McGraw-Hill, New York.

Hobbs, P.C.D. (1987) High performance amplitude and phase digitizers at 60 MHz. *Rev. Sci. Inst.* **58,** 1518–1522.

Hobbs, P.C.D., Jungerman, R.L. & Kino, G.S. (1985) A phase sensitive scanning optical microscope. *Micron and Submicron IC Metrology* (ed. by K. M. Monahan), *SPIE Proc.* **565,** 71–80.

Jungerman, R.L., Hobbs, P.C.D. & Kino, G.S. (1984) A phase sensitive scanning optical microscope. *Appl. Phys. Lett.* **45,** 846–848.

Minsky, M. (1957) *Microscopy Apparatus*. U.S. Patent 3013467.

Peterson, R.W., Robinson, G.M., Carlsen, R.A., Englund, C.D., Moran, P.J. & Wirth, W.M. (1984) Interferometric measurements of the surface profile of moving objects. *Appl. Opt.* **23,** 1464–1466.

Sawatari, T. (1973) Optical heterodyne scanning microscope. *Appl. Opt.* **12,** 2768–2772.

Wickramasinghe, H.K., Ameri, S. & See, C.W. (1982) Differential phase contrast optical microscope with a 1 Å depth resolution. *Elec. Lett.* **18,** 973–975.

APPENDIX

Calculation of the line spread and step response

This appendix gives details of the calculation of the series representations for the LSF and step response of a type 2 microscope, in order to develop the computationally efficient and well-conditioned expressions for these functions which were used to generate the theoretical curves in the preceding sections.

Beginning with the type 2 transfer function, given in Eq. (22), we set up the 1-D inverse Fourier transform:

$$l_2(x) = \left(\frac{4NA}{\pi}\right)\int_{-\infty}^{\infty} \exp(i\sigma\xi)\,\mathrm{rect}(\sigma/2)[\cos^{-1}|\sigma| - (1-\sigma^2)^{1/2}|\sigma|]\mathrm{d}\sigma. \qquad (23)$$

Integrating by parts, all the boundary terms vanish, leaving

$$l_2(x) = \left(\frac{16\,NA}{\pi}\xi\right)\int_0^1 \sin(\sigma\xi)(1-\sigma^2)^{1/2}\mathrm{d}\sigma. \qquad (24)$$

This integral is tabulated, both as a series and in closed form:

$$l_2(x) = \frac{16\,NA}{\pi}\sum_{m=0}^{\infty}(-1)^m\frac{(\xi^{2m})}{(2m+1)!!(2m+3)!!} = \frac{8\pi\,NA}{\xi^2}\mathbf{H}_1(\xi), \qquad (25)$$

where $\mathbf{H}_1(x)$ is the first-order Struve function and $\xi = 2kxNA$. The series was first obtained by Rayleigh.

The series is moderately ill-conditioned, as is typical of power series for oscillatory functions (especially ones which die away for large x); using Stirling's formula we estimate that the largest term (that for which $m \simeq \xi/2 - 1$) is about $a_{\mathrm{pk}} \simeq (8NA/\pi)\exp(\xi)/\xi^3$, whereas (since the transform is continuous) l_2 dies off at least as rapidly as ξ^{-2}. Its formal convergence is relatively good; we can estimate that m terms of the Maclaurin series will be accurate to better than ε for x satisfying

$$|kxNA| < m\frac{\varepsilon^{1/2m}}{e}. \tag{26}$$

Since this series will be inadequate for large ξ, we must seek an asymptotic series, valid for $\xi \gg 1$.

The integral in Eq. (24) is ideal for the method of steepest descents. By transforming the contour of integration in the σ plane from [0, 1] to the union of [0, $0 + iR$], [$0 + iR$, $1 + iR$] and [$1 + iR$], and letting R go to infinity, the integral along the 'crossbar', [$0 + iR$, $1 + iR$] is exponentially small and so may be neglected. Writing $\zeta = i\sigma$, expanding the square roots in Taylor series about the finite endpoints of integration and applying Watson's lemma, we obtain the asymptotic series.

$$\begin{aligned} l_2(x) \sim \frac{16}{\pi}\Bigg[\xi^{-2} + \sum_{m=0}^{\infty}(-1)^m\frac{(2m+1)!!\,(2m-1)!!}{\xi^{2m+4}}\Bigg] \\ + 8\frac{(\cos\xi + \sin\xi)}{\sqrt{\pi\xi^5}}\Bigg[-1 + \sum_{m=0}^{\infty}(-1)^{m+1}\frac{(4m+1)!!(4m+5)!!}{8^{2m+2}(2m+2)!\xi^{2m+2}}\Bigg] \\ + 8\frac{(\cos\xi - \sin\xi)}{\sqrt{\pi\xi^5}}\Bigg[\sum_{m=0}^{\infty}(-1)^m\frac{(4m-1)!!(4m+3)!!}{8^{2m+1}(2m+1)!\xi^{2m+1}}\Bigg], \end{aligned} \tag{27}$$

where we set $(-1)!! \equiv 1$.

This is quite a well-behaved series, yielding accurate values for quite small ξ.

A simple rule for the calculation of $l_2(x)$, with an absolute error less than 3×10^{-8} is to use the asymptotic series for $|\xi| > 10$, and a rational function approximation for $|\xi| < 10$. In the asymptotic series, oscillatory terms up to $\xi^{-5.5}$ and monotonic terms up to ξ^{-10} should be kept. The rational function is:

$$l_2(x) \simeq \frac{\sum_{i=0}^{6} C_i\xi^{2i}}{\sum_{i=0}^{3} D_i\xi^{2i}}, \tag{28}$$

where the C_i and D_i are given in Table 2.

The rational function in Table 2 was obtained by using a fast Fourier transform technique to generate a scaled Chebyshev polynomial fit to $l_2(x)$ in terms of ξ, then

Table 2. Coefficients of the rational function approximation to $l_2(x)$; $-10 < \xi < 10$.

i	C_i	D_i
0	1·00000000000D + 00	5·890486138792D − 01
1	−5·65454175509D − 02	5·961925426661D − 03
2	1·27597452411D − 03	2·706903550000D − 05
3	−1·39150408948D − 05	6·226934890341D − 08
4	7·99755530925D − 08	
5	−2·37047416995D − 10	
6	2·91074747525D − 13	

Table 3. Coefficients of the rational function approximation to $s_2(x)$, $-17 < \xi < 17$.

i	C_i	D_i
0	1	7·40220263756681
1	−1·4792667233478D − 2	5·49957770867583D − 2
2	2·4092520834109D − 4	1·85489554940277D − 4
3	−2·0005193306711D − 6	3·43811730882858D − 7
4	1·0405064986847D − 8	
5	−3·3692625377515D − 11	
6	6·6748052987163D − 14	
7	7·4519270720155D − 17	
8	3·6149769084954D − 20	

transforming this into a ratio of two Chebyshev sums, and finally scaling and rewriting numerator and denominator in standard form. The function values used in the fit were obtained from the Maclaurin series, evaluated in double precision to eliminate the effects of round-off error in computation of this single-precision approximation.

The resulting approximation is very nearly equiripple, and is much easier to generate than a true equiripple approximation. Use of this rational function approximation is much more efficient than summing the Maclaurin series. For $\xi = 10$, twenty-four terms of the Maclaurin series would be required to get this accuracy; here there are only ten free parameters, and in addition, the rational representation is well-conditioned.

Step response

The step response can be calculated by directly integrating the series representations; these operations can be justified rigorously since the power series alternates (and hence is uniformly convergent) and the large-ξ series is obviously asymptotic and so integratable term-by-term. Integrating the Maclaurin series with respect to s/λ, gives

$$s_2(x) = \frac{1}{2} + \frac{4}{\pi^2}\sum_{m=0}^{\infty}(-1)^m \frac{(4\pi x NA/\lambda)^{2m+1}}{(2m+1)(2m+1)!!(2m+3)!!}. \tag{29}$$

A rational function approximation to $s_2(x)$ similar to that for $l_2(x)$ is given in Table 3. Its maximum absolute error for ξ in $(-17, 17)$ is 5×10^{-8}. The approximation is

$$s_2(x) \simeq \frac{1}{2} + \frac{\sum_{i=0}^{8} C_i x^{2i+1}}{\sum_{i=0}^{3} D_i x^{2i}}. \tag{30}$$

We can integrate the monotonic terms of the asymptotic series by inspection; the oscillatory terms are most conveniently obtained by recursion. The special form of these terms, negative powers of ξ multiplied by sines and cosines, is preserved under integration by parts; integrating each term produces a boundary term and a term which is identical (except for a numerical factor) to the next term in the series. Writing the LSF and step response as

$$l_2(x) \sim \frac{16}{\sqrt{\pi}}\sum_{m=0}^{\infty}\frac{A_m \sin\xi + B_m \cos\xi}{\xi^{m+\alpha}} + \sum_{m=0}^{\infty} C_m/\xi^{2m+2} \tag{31}$$

$$s_2(x) \sim \frac{1}{2} + \frac{4}{\pi^{3/2}}\sum_{m=0}^{\infty}\frac{A'_m \sin\xi + B'_m \cos\xi}{\xi^{m+\alpha}} - \sum_{m=0}^{\infty}\frac{C_m}{(2m+1)\xi^{2m+1}} \tag{32}$$

Table 4. Coefficients of the asymptotic series (32) for the step response of a type 2 microscope.

m	A'_m
0	-1
1	$17/8$
2	$937/2^7$
3	$-33{,}627/2^{10}$
4	$-5{,}913{,}627/2^{15}$
5	$307{,}435{,}839/2^{18}$
6	$36{,}889{,}462{,}845/2^{22}$
7	$-2{,}508{,}416{,}581{,}635/2^{25}$
8	$-38{,}161{,}423{,}775{,}989/2^{29}$

(in this case, $\alpha = 2{\cdot}5$), we derive the recursion relations for A'_m and B'_m:

$$\begin{gathered} A'_0 = B_0 \qquad B'_0 = -A_0 \\ B'_{m+1} = -A_{m+1} - (m+\alpha)A'_m \qquad A'_{m+1} = B_{m+1} + (m+\alpha)B'_m. \end{gathered} \tag{33}$$

Because of the form of A_m and B_m, $B'_m = (-1)^{m+1}A'_m$; Table 4 gives A'_m for $m < 9$.

This series is somewhat poorer than the one for $l_2(x)$. To achieve an accuracy of 5×10^{-8}, the patch point had to be moved out to $\xi = 17$; for this accuracy, keep oscillatory terms up to A'_4 and monotonic terms up to C_6.