Resolving Power and Information Capacity in Scanning Electron Microscopy

Rene Simon

Citation: Journal of Applied Physics **41**, 4632 (1970); doi: 10.1063/1.1658508
View online: http://dx.doi.org/10.1063/1.1658508
View Table of Contents: http://scitation.aip.org/content/aip/journal/jap/41/11?ver=pdfcov
Published by the AIP Publishing

JOURNAL OF APPLIED PHYSICS VOLUME 41, NUMBER 11 OCTOBER 1970

Resolving Power and Information Capacity in Scanning Electron Microscopy

RENE SIMON

Bayside Research Center of General Telephone and Electronics Laboratories, Incorporated, Bayside, New York 11360

(Received 24 February 1970; in final form 11 June 1970)

Information theory formalism has been applied to the scanning electron microscope. A theoretical model of two-dimensional secondary electron noise has been derived and the maximum capacity content per picture element of the S.E.M. has been computed as a function of its geometrical resolving power. According to the theory, it is shown that the maximum number of grey levels on the picture varies between 2 and 8 (varying with the maximum secondary electron yield) with a resolving power of 200 Å for a S.E.M. equipped with a tungsten filament gun, and 30–40 Å is equipped with a lanthanum–hexaboride gun. However, these figures are established with the assumption that the object does not present any redundant properties. The knowledge of the statistics of a general class of objects will permit the use of adapted coding to increase the information content in the object area of interest.

INTRODUCTION

In communication theory,[1,2] the transmission of the greatest possible amount of information depends mainly on two factors: (1) the statistics of the source and (2) the capacity of the channel. The statistics of the source represent in optics a new subject in itself and will not be considered. Only the capacity of the transmission channel will be treated.

In the time domain, if Q is the quantity of information transmitted during a time t, the rate of information D_0 is equal to

$$D_0 = \lim_{t\to\infty} (Q/t) \qquad (1)$$

and for a given transmission channel, i.e., for a noise function $N(t)$ characteristic of the channel, it will be possible to find a class of signal source which maximizes (1). This maximum D of D_0 is called the capacity of the channel.

In a optical channel, by analogy to (1), the rate of information D_0 will be defined as the information that the channel can transmit per single image.[3] At this stage, it is necessary to introduce the concept of degrees of freedom in an optical system: the number of picture elements necessary to specify completely an optical image. If the system has a spatial frequency bandwidth ΔW such that $-F \leq \Delta W \leq +F$, the application of the sampling theorem[3] leads to the conclusion that the picture is completely determined by giving its values at a series of discrete points (sampled points) spaced $1/2F$ apart. Under these conditions, every measurement will be independent from each other and the number of degrees of freedom will be equal to the number of measurements in one picture. The distance $1/2F$ defines the size of a picture element (or resolving power) and the rate of information D_0 can be specified in terms of rate of information C_0 per picture element. As in the time domain, it will be possible to find a class of pictures which gives a maximum value C of C_0, called the capacity of the channel. The value of C will depend on the geometrical aberrations and diffraction, as well as background and detector noise.

In 1964 Harris[4] showed that in the absence of noise and with the condition that the different aberrations coefficients are known, the *a priori* knowledge that the object is of finite extent is sufficient to ensure the exact restoration of an image degraded by aberrations and diffraction. Consequently, the main limitation imposed on the capacity of a optical channel is the noise.

In order to determine the maximum capacity of the scanning electron microscope as a communication channel, some knowledge of the statistical properties of the noise is required. For reasons which are not directly connected with noise properties, the best results in terms of resolving power have always been obtained with low-energy secondary electrons emitted by the object and collected to form the image. Under these conditions, the different sources of noise which cause a deterioration of the image are

(a) fluctuations in the primary beam (Schottky noise);
(b) fluctuations in the secondary electron emission;
(c) fluctuations in the number of electrons emitted by surface area;
(d) fluctuations in the number of secondary electrons collected;
(e) detector, amplifiers, and photographic noise.

It has been shown[5] that the main limitation in resolving power is the noise directly correlated with the object [parameters (a) and (b)] and is called the object self-noise. The noise correlated with the detector and amplifiers [(d) and (e)] can ultimately be neglected. The photographic noise will be produced by the plate grain and it is assumed that the pictures are recorded at such a magnification that, at the object level, the noise has a microscopic correlation and a very small rms value.[3]

Concerning (c), previously published studies on secondary electron noise have been concerned primarily with electrons having a Poisson distribution in time without consideration of the dispersion of secondary electron sources over a surface area S. In scanning electron microscopy, however, the secondary electron sources are distributed over the surface covered by the beam, generating a scintillation phenomena, and it will

be shown that this fact can lead to a marked departure from Poisson behavior.

In summary, as far as the noise is concerned, the only limitations involved in the maximum capacity per picture (object) element of a scanning electron microscope are introduced by the fluctuations (a)–(c) above and only they will be considered in this study.

In a scanning electron microscope, the crossover of the gun is demagnified by a series of lenses into a fine spot on the object surface. It is generally agreed that the Gaussian function $j(0)\exp(-r^2/R_0^2)$ constitutes a good approximation for the electron distribution at the crossover.[6] Then, if M is the total demagnification, the radius R_1 of the final spot is equal to $R_1=M\cdot R_0+\Delta R$, where ΔR is a term introduced by the aberrations of the objective lens. As a first approximation, Appendix A shows that the final spot has a Gaussian distribution, $j'(0)\exp(-r^2/R_1^2)$. In the rest of the paper, it has been assumed for reasons of mathematical simplicity that this spot has a uniform electron distribution inside a circle of diameter $2R$, R being defined by the equivalent surface of a Gaussian distribution. This is discussed in Appendix A.

An S.E.M. differs from a conventional optical instrument in that at any instant only one object element is being observed by the detection system. As a consequence, it is possible to visualize an idealized S.E.M. as shown in Fig. 1, in which the object, generating secondary electrons, is followed by a perfect electron-optical lens, free of geometrical aberrations. The resolving power of this lens will then be limited only by diffraction. An array of light-emitting electron detectors is located in the conjugated plane corresponding to a magnification M such that every detector area has the same size as that of the central part of the diffraction figure given by a single electron, i.e., $\lambda/2$, where λ is the de Broglie wavelength. The light is then observed through a transparent scanning aperture of diameter $2MR$. The gain of every detector is such that one collected electron will produce light of brightness B_0.

STOCHASTIC PROPERTIES OF SECONDARY EMISSION IMAGES

In the theory of information, a random process is said to be stationary if its statistical characteristics are invariant under time shifts, i.e., if they remain the same when t is replaced by $t+a$, where a is arbitrary. Then the probability densities $p_n(X_1\cdots X_n,\ t_1\cdots t_n)$ of the stochastic variables $X_1\cdots X_n$ and their moments $m_n(t_1\cdots t_n)$ and correlation functions $c(t_1\cdots t_n)$ do not depend on the absolute position of the points $t_1\cdots t_n$ on the time axis, but only on their relative configuration, i.e., only on the time differences $t_2-t_1\cdots$, t_n-t_1. By analogy to the time domain, a random process $X[\mathbf{M}(x,y)]$ in the plane $[x,y]$ will be stationary if its statistical characteristics are invariant under translation, i.e., if they remain the same when $\mathbf{M}$ is replaced by $\mathbf{M}+\mathbf{N}$, where $\mathbf{N}$ is arbitrary. Then, the probability densities $p_n(X,\mathbf{M}_1\cdots\mathbf{M}_n)$ and the moments will depend on $\mathbf{M}_2-\mathbf{M}_1\cdots\mathbf{M}_n-\mathbf{M}_1$. In the S.E.M., the brightness $B(\mathbf{M})$ of an infinitesimal element of an object within a picture or object element is generally a nonstationary process since the object will have variations in surface topography smaller than the size of the element; therefore, the conditions of invariance under translation will not be conserved. On the other hand, an analysis of the nonstationary properties of $B(\mathbf{M})$ will necessitate a knowledge of the object surface that we do not have, since this is just the purpose of the observation. Therefore, the analysis of brightness must be limited to a class of objects such that every member has a spatial frequency bandwidth equal or less than the bandwidth transmitted by the instrument. With these conditions, the invariance under translation is conserved and $B(\mathbf{M})$ can be considered as a stationary process. Then it will be possible to derive:

(1) the average value $\bar{B}$ per picture or object element.

(2) the autocorrelation function, which is given by the second-order properties of the centered variable $\beta(\mathbf{M})=B(\mathbf{M})-\bar{B}$:

$$K(\mathbf{M}_1-\mathbf{M}_2)=\langle\beta(\mathbf{M}_1)\beta(\mathbf{M}_2)\rangle. \tag{2}$$

The knowledge of $K(\mathbf{M}_1-\mathbf{M}_2)$ will give the Wiener spectrum such that;

$$K(\mathbf{M}_1-\mathbf{M}_2)=\frac{1}{2\pi}\int_\omega \exp(i\boldsymbol{\omega}\mathbf{M})d\Gamma(\boldsymbol{\omega}), \tag{3}$$

where $\boldsymbol{\omega}$ is defined in the frequency domain.

(3) the probability density $p(B,\mathbf{M})dB$.

According to the assumptions made concerning the class of objects we observe, $B(\mathbf{M})$ is an isotropic stochastic function, and the Fourier transform of Eq. (3) is

$$\Gamma(\omega)=2\pi\int_0^\infty J_0(\omega k)K(k)k\,dk, \tag{4}$$

where $\omega^2=\omega_x^2+\omega_y^2=4\pi^2\nu^2$ (in radian/L) and $k=|\mathbf{M}_1-\mathbf{M}_2|$. Let us now consider an object element with a surface πR^2 which emits N secondary electrons during an observation time t. According to the conditions of our experiment as described by Fig. (1), the location of these electrons will be known within a spatial uncertainty defined by a circle of diameter $\lambda/2$, where λ is the de Broglie wavelength associated with the electron.[7] It will then be possible to represent each electron C_j emitted by the surface by a small circle of diameter $2R_j=\lambda/2$, where the R_j's are stochastic variables independent of each other and of position on the surface: R_j will only depend on the kinetic energy of the electron considered. We will call $p(r)$ the probability distribution function of the R_j's, such that $p(r)=$

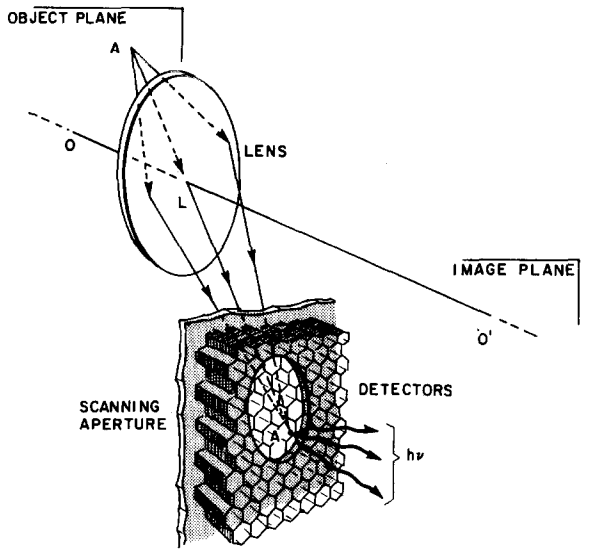

FIG. 1. Principle of image formation in a scanning electron microscope.

Prob$(R<r)$. Let us also assume that the centers 0_j of these circles are distributed on the surface according to a Poisson distribution. Then, the stochastic variable $n(\mathbf{M})$, representing the number of circles C_j covering a given surface S will have a probability density

$$p_S(n)=\bar{d}^n e^{-\bar{d}}/n!, \tag{5}$$

where $\bar{d}=N/\pi R^2$ is the average electron density.

If a brightness B_0 is associated with every electron contributing to the image, then we can define the brightness B of a point M of the object by

0 if M is not covered by any circle

B_0 if M is covered by a circle C_j

.....

$n(\mathbf{M})B_0$ if M is covered by n circles C_j.

With these hypotheses, $\bar{B}$ and $K(\mathbf{M}_1-\mathbf{M}_2)$ can be evaluated and their values will then allow the determination of the normalized variance:

$$\sigma^2(B)/\bar{B}^2=K(0)/\bar{B}^2. \tag{6}$$

All these results can be computed using the joint characteristic function of $n(\mathbf{M}_1)$ and $n(\mathbf{M}_2)$. This is easily obtained by using a method of calculation developed by M. Savelli[8] and B. Picinbono.[9] This calculation is performed in Appendix B. We then have

$$\bar{B}=\bar{d}B_0\pi r_0^2=aNB_0, \tag{7}$$

$$K(k)=2\bar{d}B_0^2F(k,r), \tag{8}$$

where r_0 is the radius of the central part of the diffraction figure in the object plane, given by an electron emitted with an average kinetic energy E_0, and k is the distance between two points 0_j.

Application of the Hankel transform shown in Eqs. (4)–(8) gives the Wiener spectrum $\Gamma(\omega)$. Figure (2) shows the variation of $\Gamma(\omega)$ as a function of ω (in radian/Å). It is clear that, for spatial frequencies above 20 Å, the noise is white.

The values of $\bar{B}$ and $\sigma^2(B)$ are the first moment and the variance of the probability density $p(B\mid N,\mathbf{M})dB$ over a very small surface element centered on point M. This function is easily computed with the same argument used by Rice.[10] It is found to be a gamma distribution

$$p(B\mid N,\mathbf{M})=\frac{1}{B_0}\frac{1}{\Gamma(aN)}\left(\frac{B}{B_0}\right)^{aN-1}e^{-B/B_0}. \tag{9}$$

As explained above, N is a stochastic variable representing the number of secondary electrons emitted by the object element during the time t that the primary beam remains on the same location. N depends on the statistics of secondary-electron emission and on the Schottky noise of the primary beam. If an average $\bar{N}_1$ primary electrons per second reach the object, then the probability $p(x)$ of having x primary electrons after a time t is

$$p(x)=[(\bar{N}_1t)^x\exp(-\bar{N}_1t)]/x!. \tag{10}$$

If we assume that the probability $p(N/x)$ of generating N secondary electrons from a surface with a secondary-electron yield $\bar{\gamma}$ follows a binomial distribution

$$p(N\mid x)=[x!/N!(x-N)!]\bar{\gamma}^N(1-\bar{\gamma})^{x-N}, \tag{11}$$

then the probability of producing N secondary electrons from an average of $\bar{N}_1$ primary electrons is

$$p(N)=\sum_{x=N}^{\infty}p(N\mid x)p(x)=\frac{(\bar{\gamma}\bar{N}_1t)^N}{N!}\exp(-\bar{N}_1\bar{\gamma}t) \tag{12}$$

$$p(B,\mathbf{M})=\sum_{N=1}^{\infty}p(B\mid N)p(N)$$

$$=\exp(-B/B_0-\bar{N}_1\bar{\gamma}t)\frac{1}{B_0}\sum_{N=1}^{\infty}\frac{1}{\Gamma(aN)}\left(\frac{B}{B_0}\right)^{aN-1}\frac{(\bar{N}_1\bar{\gamma}t)^N}{N!}. \tag{13}$$

Since this series is uniformally convergent for all values of B/B_0, we have

$$\bar{B}=\int_0^{\infty}Bp(B,\mathbf{M})dB=\int_0^{\infty}B\exp(-B/B_0-\bar{N}_1\bar{\gamma}t)\frac{1}{B_0}\times\left[\sum_{N=1}^{\infty}\frac{1}{\Gamma(aN)}\left(\frac{B}{B_0}\right)^{aN-1}\frac{(\bar{N}_1\bar{\gamma}t)^N}{N!}\right]dB$$

$$=B_0\sum_{N=1}^{\infty}\frac{(N_1\gamma t)^N}{N!}\frac{\exp(-\bar{N}_1\bar{\gamma}t)}{\Gamma(aN)}\times\int_0^{\infty}\left(\frac{B}{B_0}\right)^{aN}\exp(-B/B_0)d(B/B_0). \tag{14}$$

Finally

$$\bar{B}=\bar{d}\pi r_0^2 B_0 \tag{15}$$

which is equal to the same value as in (7) as expected. By the same argument,

$$\langle B_x^2\rangle = B_0^2 \sum_{N=1}^{\infty} \frac{(\bar{N}_1\bar{\gamma}t)^N}{N!}\frac{\exp(-\bar{N}_1\bar{\gamma}t)}{\Gamma(aN)} \times \int_0^{\infty}\left(\frac{B}{B_0}\right)^{aN+1}\exp(-B/B_0)d\left(\frac{B}{B_0}\right)$$

$$=\bar{d}\pi r_0^2 B_0^2(1+r_0^2/R^2)+\bar{d}^2\pi^2 r_0^4 B_0^2 \tag{16}$$

and the new value σ^2, (B) is

$$\sigma_1^2(B)=\bar{d}\pi r_0^2 B_0^2(1+r_0^2/R^2). \tag{17}$$

Obviously, the spatial-correlation function does not depend on time and is equal to:

$$K(k)=2\bar{d}(1+r_0^2/R^2)B_0^2 F(k,r). \tag{18}$$

SIGNAL-TO-NOISE RATIO IN VIDEO SIGNAL

The scanning aperture will have low-pass filtering properties for noise. If $\sigma_{im}^2(B)/\bar{B}^2$ is the normalized variance of the signal at the output, we have, in the Fourier domain:

$$\frac{\sigma_{im}^2(B)}{\bar{B}^2}=\int_{\omega}\Gamma(\omega)\,|\,T(\omega)\,|^2 d\omega \Big/ \int_{\omega}|\,T(\omega)\,|^2 d\omega, \tag{19}$$

where $T(\omega)$ is the Fourier transform of the scanning aperture transmittance. On the spatial domain, Eq. (19) corresponds to the following relationship:

$$\frac{\sigma_{im}^2(B)}{\bar{B}^2}=\frac{1}{\bar{B}^2\pi^2R^4}\int_{\mathbf{k}}K(k)\left(\int_{\mathbf{M}}t(\mathbf{M})t(\mathbf{M}+\mathbf{k})d\mathbf{M}\right)d\mathbf{k}, \tag{20}$$

where $t(\mathbf{M})$ is the transmittance of the aperture as a function of $\mathbf{M}(x,y)$. According to our hypothesis, $t(\mathbf{M})=1$ for $|M|\leq R$ and $=0$ for $|M|>R$. The convolution product of Eq. (20) is equal to:

$$\int_0^R t(\mathbf{M})t(\mathbf{M}+\mathbf{k})d\mathbf{k}=R^2T(k/2R), \tag{21}$$

where

$$T(k\,|\,2R)=\pi-2\left[\arcsin\left(\frac{k}{2R}\right)+\frac{k}{2R}\left(1-\frac{k^2}{4R^2}\right)^{1/2}\right] \qquad k\leq 2R$$

$$=0 \qquad k>2R$$

and by Eqs. (8) and (21)

$$\frac{\sigma_{im}^2(B)}{\bar{B}^2}=\frac{1}{\pi r_0^2}\frac{4(1+r_0^2/R^2)}{\bar{d}\pi r_0^2\pi R^2}\int_0^R F(k,r)T\,(k\,|\,2R)\,k dk. \tag{22}$$

The relation (22) gives the signal-to-noise ratio in one image element.

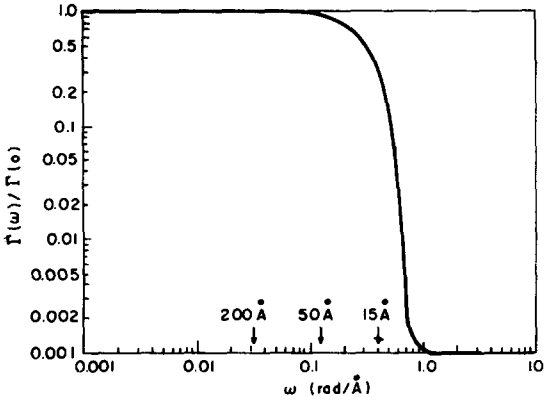

FIG. 2. Power spectrum of spatial noise. The arrows indicate the frequencies corresponding to, respectively, 200, 50, and 15 Å.

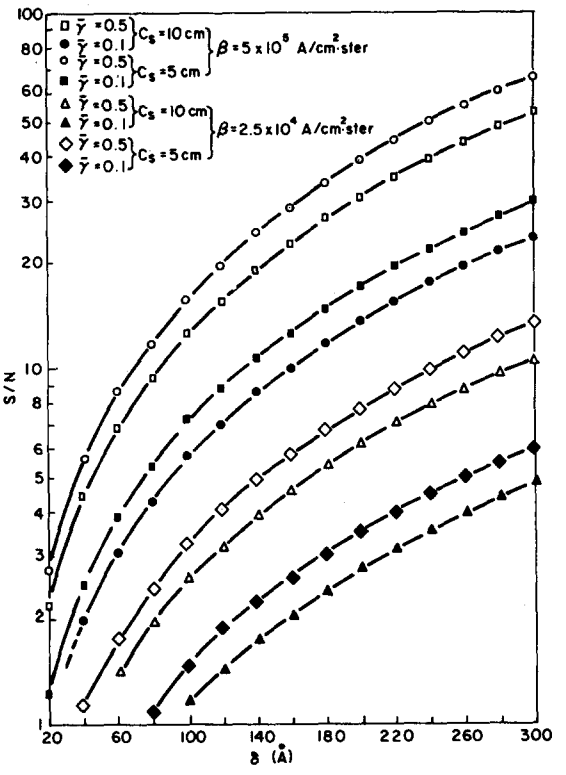

FIG. 3. Signal-to-noise ratio in video signals as a function of resolving power.

INFORMATION CONTENT OF THE IMAGE

Let x and y represent, respectively, the received and transmitted signals in a communication channel. The total mutual information $I(X,Y)$ or the information capacity of the communication channel per degree of freedom can be represented by the expression:

$$C=I(X,Y)=H(Y)-H(Y\,|\,X), \tag{23}$$

where $H(Y)$ stands for the entropy of y, $H(Y\,|\,X)$ for

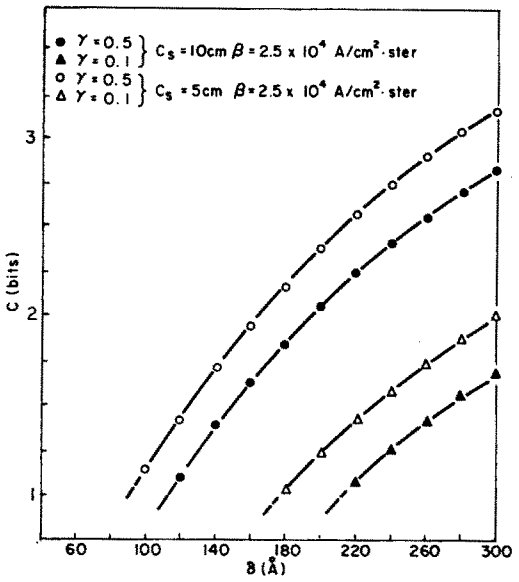

FIG. 4. Information capacity of an S.E.M. with a tungsten filament gun as a function of resolving power.

the conditional entropy when the signal x is known and transmitted through a noisy channel.[2] If $p(y)$, $p(x)$ are defined as the probability densities of y and x, respectively, and $p(y\,|\,x)$ as the conditional probability density of y when x is known, then Eq. (23) becomes[1,2]

$$C=-\int p(y)\ln(p(y))dy + \iint p(x)p(y\,|\,x)\ln(p(y\,|\,x))dxdy. \quad (24)$$

In the application of this formalism to the scanning electron microscope, x will be the average brightness $\bar{B}$ of an object element and y will be the brightness B which will be perceived by the observer's eye. Because of the stochastic properties of B, y can be different from the average value, and C will be the information capacity of the S.E.M. per picture element. The conditions for maximization of Eq. (24) depends upon the statistics of the noise and of the object, and is treated below.

Concerning the noise, if the psychophysical uncertainty in the brightness perception is neglected, then the probability density $p(y\,|\,x)=p(B\,|\,\bar{B})$. For signal-to-noise ratios larger than 3 to 4, which, according to Rose,[11] is the differential threshold for contrast, the number of electrons contributing to a picture element is large enough for the application of the central theorem of probability theory. Hence, $p(B\,|\,\bar{B})$ can be considered a good approximation to a Gaussian function:

$$p(B\,|\,\bar{B})=\frac{1}{\sigma_{im}(2\pi)^{1/2}}\exp[-(B-\bar{B})^2/2\sigma_{im}{}^2] \quad (25)$$

with $\bar{B}$ and $\sigma_{im}{}^2(B)$, respectively, given by Eqs. (15) and (22).

Concerning the object, $p(x)=p(\bar{B})$, the brightness of every element lies between 0 and $\bar{B}_{\max}=\bar{d}_m\pi r_0{}^2B_0$. The second part of Eq. (24), which is the conditional entropy $H(Y\,|\,X)$, is maximized when all values of B inside this range are equally probable, i.e., $p(\bar{B})=1/\bar{B}_{\max}$. The signal received, however, is strictly speaking, represented by B, and the first part of Eq. (24) is maximized only for a uniform distribution $p(B)=1/B_{\max}$. However, if B has a uniform probability, then the probability $p(\bar{B})$ is not uniform. It does not matter at all, however, because whatever the resultant distribution for B may be, the capacity of the optical channel must be evaluated with that probability distribution which maximizes $H(Y)$. This would be precisely a uniform distribution, provided B is limited to a well-defined range. The last condition is not strictly true, since the upper limit is infinite. However, it is clear that only a small error will result if it is assumed that the upper limit is fixed and equal to $\bar{B}_{\max}$, and for each picture element:

$$H(Y)=\ln(\bar{d}_m\pi r_0{}^2)+\ln B_0 \quad (26)$$

and

$$H(Y\,|\,X)=\frac{1}{\bar{B}_{\max}}\Bigg[\int_0^{\bar{B}_{\max}}\int_0^{\infty}\frac{1}{\sigma_{im}(2\pi)^{1/2}}\ln\left(\frac{1}{\sigma_{im}(2\pi)^{1/2}}\right)\times\exp\left(\frac{-(B-\bar{B})^2}{2\sigma_{im}{}^2}\right)dBd\bar{B} - \int_0^{\bar{B}_{\max}}\int_0^{\infty}\frac{1}{\sigma_{im}(2\pi)^{1/2}}\frac{1}{2\sigma_{im}{}^2}(B-\bar{B})^2\times\exp\left(\frac{-(B-\bar{B})^2}{2\sigma^2}\right)dBd\bar{B}\Bigg], \quad (27)$$

which gives after transformation,

$$H(Y\,|\,X)=\frac{1}{\bar{B}_{\max}}\int_0^{\bar{B}_{\max}}\Bigg[\tfrac{1}{2}\ln\left(\frac{1}{\sigma_{im}(2\pi)^{1/2}}\right)+\tfrac{1}{4}+\ln\left(\frac{1}{\sigma_{im}(2\pi)^{1/2}}\right)\frac{1}{\sigma_{im}(2\pi)^{1/2}}\int_0^{\bar{B}}\exp\left(\frac{-X^2}{2\sigma_{im}{}^2}\right)dX - \frac{\bar{B}}{2\sigma_{im}(2\pi)^{1/2}}\exp\left(\frac{-\bar{B}^2}{2\sigma_{\rm im}{}^2}\right)+\frac{1}{2\sigma_{im}(2\pi)^{1/2}}\times\int_0^{\bar{B}}\exp\left(\frac{-X^2}{2\sigma_{im}{}^2}\right)dX\Bigg]d\bar{B}. \quad (28)$$

In the cases of interest $(S/N\geq 3)$, $\operatorname{erf}(\bar{B}/\sigma_{im}\sqrt{2})\approx 0.5$ and, by Eqs. (24), (26), and (28), the maximum information capacity C is equal to:

$$C=\ln\left[\frac{e}{A}\left(\frac{\bar{d}_m\pi r_0{}^2}{2\pi}\right)^{1/2}\right]-\frac{A}{(\bar{d}_m\pi r_0{}^2)^{1/2}(2\pi)^{1/2}}\times\exp\left(\frac{-\bar{d}_m\pi r_0{}^2}{2A^2}\right)+\frac{A^2}{\bar{d}_m\pi r_0{}^2}\operatorname{erf}[(\bar{d}_m\pi r_0{}^2)^{1/2}/A\sqrt{2}], \quad (29)$$

where the parameter A is such that $\sigma_{im}{}^2=A^2\bar{d}\pi r_0{}^2B_0{}^2$.

PRACTICAL APPLICATION

In this section calculations of both signal-to-noise ratio and information capacity of an S.E.M. viewed as a communication channel as a function of resolving power is performed.

As is generally the case (Appendix A), the stigmatic and chromatic aberrations of the objective lens of a scanning electron microscope can be neglected and the maximum current for a beam of diameter $2R$ is given by[12]:

$$i_p = \frac{3\pi^2\beta}{16}\left(\frac{(2R)^{8/3}}{C_s^{2/3}} - \tfrac{4}{3}(1.22\lambda)^2\right), \qquad (30)$$

where C_s is the lens spherical aberration, β the gun brightness, and λ is the de Broglie wavelength. However, because of the lateral diffusion of the electrons in the object, the emission area for the secondary electrons is larger than the primary-beam cross section. It is very difficult to have an exact estimation of the lateral diffusion, since, among other factors, this phenomenon depends on the electron beam kinetic energy and on the various energy losses which are not very well known for kinetic energies used in an S.E.M. However, a reasonable order-of-magnitude[13] is obtained by setting $D_{\text{opt}} = 2.20R$; Eq. (30) therefore becomes:

$$i_p = \frac{3\pi^2\beta}{16}\left(\frac{(1.80R)^{8/3}}{C_s^{2/3}} - \tfrac{4}{3}(1.22\lambda)^2\right). \qquad (31)$$

It is assumed that every picture element is observed during an interval of time of 10^{-4} sec (this value corresponds to a 100-sec recording time for a S.E.M. with 1000 lines/frame). If $\bar{\gamma}$ is the maximum secondary electron yield of the object observed, the average of the maximum brightness is:

$$\bar{B}_m = \bar{d}_m \pi r_0^2 B_0 \approx \frac{3\pi^2(1.80)^{8/3}R^{2/3}}{26(C_s)^{2/3}} r_0^2\beta\bar{\gamma}B_0 \times 10^{15}. \qquad (32)$$

The values of $\bar{\gamma}$ chosen to compute the expressions (22) and (29) are 0.1 and 0.5, corresponding to the large majority of conductive samples observed with primary energies on the 20–30 KeV range.[5]

Concerning the gun brightness β, two values have been chosen: one corresponding to a normal gun with a tungsten hairpin filament at 20 KeV (2.5×10^4 A/cm^2·sr), the other corresponding to a lanthanum-hexaboride gun (5×10^5 A/cm^2·sr) at 12 KeV as published in the literature.[14]

Two cases will be considered: one with the objective lens spherical-aberration coefficient C_s, 5 cm, and the other 10 cm which is typical of the electromagnetic lenses used in conventional scanning electron microscopy. The results of these calculations are shown in the figures.

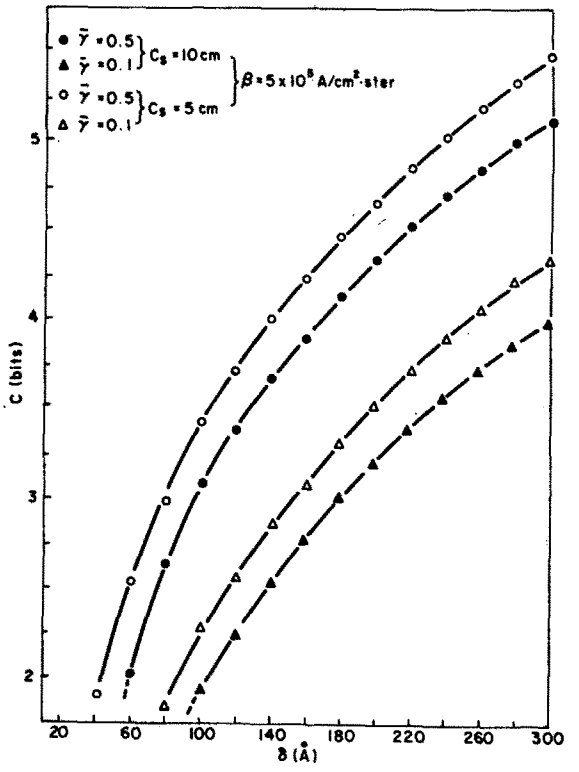

FIG. 5. Information capacity of an S.E.M. with a lanthanum-hexaboride gun as a function of resolving power.

DISCUSSION

Figures (3)–(5) represent the video signal-to-noise ratio S/N and the maximum information capacity per picture element of a scanning electron microscope as a function of geometrical resolving power. The only source of noise taken into consideration is the object self-noise. Concerning the validity of this approximation, it has been shown[5] that at least 90% of the limitations in resolving power by physical phenomena are due to the noise properties of the primary beam and the secondary electrons. The introduction of detector and amplifier noise in expressions (22) and (29) will probably change the results by only a few percent; however, if the observer's perception is taken into account, as it must be ultimately, the Weber–Fechner phenomena will introduce a more severe limitation on S/N and C, particularly for values of S/N near the psychophysical threshold. In any event, the values shown by these figures indicate the limitations of information capacity of an S.E.M. due to the physical phenomena that are ultimately beyond human control. These results show that for resolving powers of 200 Å obtained with a tungsten filament, or 40–50 Å obtained with a lanthanum-hexaboride gun, the average number of grey levels expected will be approximately 3–4, even though the S/N ratio is of the order of 6–10, i.e., greater than the threshold defined by Rose. Thus the classical

concept of resolving power in an optical instrument is modified by the introduction of signal statistics and of a noise factor correlated with the spatial resolution. In other terms, if we apply the Neyman and Pearson decision theory to the cases where C is smaller than one, even if the signal-to-noise ratio is equal to 3, it is possible to show that an ideal threshold detection device placed in an optimum position to choose between two different grey levels will have a 50% chance of making a wrong decision. Under these conditions, the concept of geometrical resolving power considered alone is meaningless and any attempts to process the picture to get such resolution, for example by color coding, will fail. It is possible to overcome this difficulty by increasing the total recording time of a picture; but with the present technology in the design of stabilized high voltage and current power supplies, a 100-sec total recording time seems to represent an upper limit. Beyond this time, the lack of stabilization of these power supplies becomes significant.

However, even if the images represent a large amount of information (5×10^{12} bits/cm^2 in the most favorable cases), they have a tremendous amount of redundancy. The capacity C computed above is an upper limit due to the fact that we considered every grey level as equiprobable. This hypothesis was corroborated by experiment as far as the amplitude statistics are concerned,[15] and the number of bits per picture element necessary for coding such information will be at least equal to the values of C in Fig. (4) and (5). However, it has been shown in the same reference that the differences between the brightness level of every picture element of a general class of images have stationary properties: the small variations from element to element are the most probable and the spectrum of "jumps" of different amplitudes between elements fit very well with a Gaussian curve. If we assume that this property holds for micrographs (and there is no apparent reason why it should not hold), the images must be transmitted through the channel by using a differential coding.[16] It is however necessary to be careful about an indiscriminate use of the bandwidth compression techniques discussed or used in T.V. transmission, where the object is to transmit an esthetic image at the expense of a nonnegligeable amount of truncation errors. In micrograph pictures, the useful scientific information must be conserved, even if the coding leads to a picture with less esthetic value. A coding system which conserves the scientific information would be a system implementation in which the scanning speed of the beam will be accelerated on the "flat" area of the object, and slowed down on the areas which show amplitude variations. For example, if it is assumed that only 10% of the object conveys useful information and that the time for recording the total picture is constrained to approximately 100 sec with a resolving power of 40 Å, one can in principle arrange conditions so that the analysis time is 10^{-3} sec/element instead of 10^{-4} sec/element on the areas of interest while the analysis time is 10^{-5} sec/element for the areas which convey no information. The average capacity of picture elements of interest will thus increase from 2 to 3.5 bits (from 4 grey levels to 10–12). This case clearly shows the value of a differential coding system; it is possible to pass from 4 grey levels, where the geometrical resolving power has little significance, to a condition where the observer can better comprehend the full information content of the picture.

ACKNOWLEDGMENTS

The author is grateful to J. Wagner for helpful discussions on noise and probabilities. Useful comments from T. Emma, W. Jensen, M. Kornblau, and V. Meyers are acknowledged. The computations of numerical applications were made by L. Hadley and C. Higginbotham.

APPENDIX A

The minimum diameter of the beam in an S.E.M. is estimated by assuming that the aberration disks and Gaussian image of the objective lens add in quadrature.[18] The beam diameter d is then:

$$d^2=d_g^2+d_c^2+d_s^2+d_d^2+\text{etc}\cdots, \tag{A1}$$

where d_g is the diameter of the Gaussian image, d_s is the minimum disk of confusion due to the spherical aberration C_s, d_c is the disk of confusion due to the chromatic aberration C_c, and d_d is the disk of confusion due to diffraction. Hence

$$d^2=\frac{4i_p}{\pi^2\beta\alpha^2}+\frac{C_s^2\alpha^6}{4}+\epsilon^2C_c^2\alpha^2+\frac{(1.22\lambda)^2}{\alpha^2}+\cdots \tag{A2}$$

and

$$i_p=\frac{\pi^2\beta}{4}\left(d^2-\frac{C_s^2\alpha^6}{4}-\epsilon^2C_c^2\alpha^2-\frac{(1.22\lambda)^2}{\alpha^2}-\cdots\right)\alpha^2, \tag{A3}$$

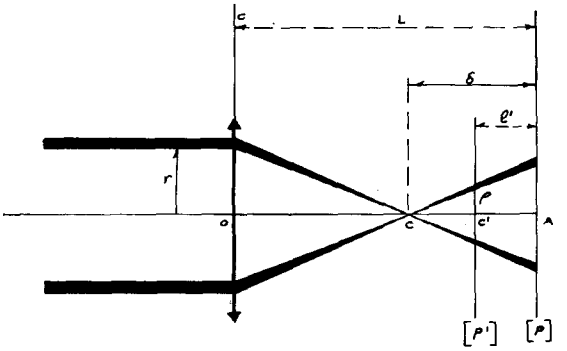

FIG. 6. Computation of impulse response of a lens with spherical aberration.

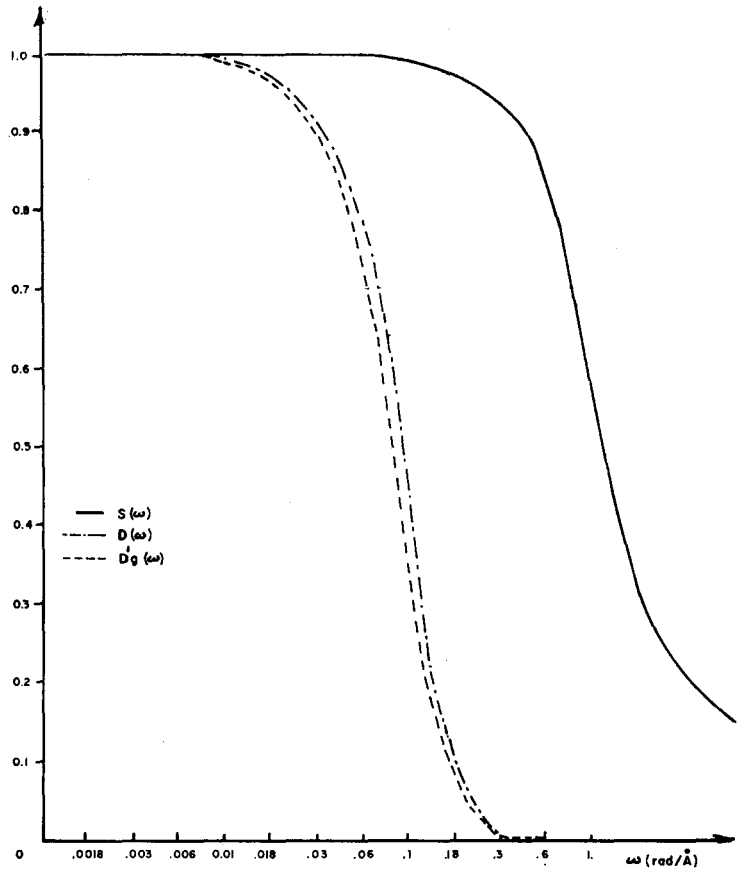

FIG. 7. Impulse response of the objective lens of a S.E.M.: $C_s=5$ cm; $\alpha=3.10^{-3}$ rad; beam size $2R_1=40$ Å. The conditions shown correspond to a frequency bandwidth $\Delta\omega=0.08$ rad/Å $(1/2F=39$ Å).

where

- i_p primary beam current
- α beam semiaperture
- β gun brightness
- λ de Broglie wavelength
- ϵ all fluctuations responsible for chromatic aberration.

The beam current has a maximum value for $\alpha=\alpha_{opt}$, which is the solution of the cubic equation

$$X^3+2\epsilon^2C_c^2/C_s^2X-d^2/C_s^2=0 \qquad X=\alpha_{opt}^2. \quad \text{(A4)}$$

This has one real solution since

$$(d^4/4C_s^4)+(8/27)(\epsilon^6C_c^6/C_s^6)>0. \quad \text{(A5)}$$

Practically, the working distance in an S.E.M. is not smaller than 0.5–1 cm, which assures the optimum conditions for the collection of secondary electrons, and the demagnification m of the objective lens is such that $1/50<m<1/20$. Under these conditions, the computed value of C_s and C_c are[19,20]: 5 cm$<C_s<$10 cm and $C_c\simeq 1$ cm. Concerning ϵ, the present state of the art in electron emission systems and voltage–current stabilization shows that $\epsilon\simeq 10^{-5}$ for accelerating voltages commonly used in S.E.M. The use of these numerical values in Eq. (A5) shows that the chromatic aberration has some influence for beam sizes of 15–30 Å, but can be neglected for 40 Å and up, and $\alpha_{opt}=(d/C_s)^{1/3}$.

Then, since the only aberration of importance is C_s, the electron distribution within the beam at the object plane is given by the convolution between $d_g(\mathbf{r})$, the Gaussian image, which has a Gaussian distribution, and $s(\mathbf{r})$, the impulse response of the objective lens due to the spherical aberration:

$$d(\mathbf{r})=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} d_g(\mathbf{r}')s(\mathbf{r}-\mathbf{r}')d\mathbf{r}'. \quad \text{(A6)}$$

Convolutions are difficult to perform and it is easier to use the linear system formalism where, on the frequency domain, Eq. (A6) is transformed:

$$D(\boldsymbol{\omega})=D_g(\boldsymbol{\omega})S(\boldsymbol{\omega}), \quad \text{(A7)}$$

where $D(\boldsymbol{\omega})$, $D_g(\boldsymbol{\omega})$, and $S(\boldsymbol{\omega})$ are, respectively, the Fourier transforms of $d(\mathbf{r})$, $d_g(\mathbf{r})$, and $s(\mathbf{r})$. The behavior of $D(\boldsymbol{\omega})$ will give useful indications on the behavior of $d(\mathbf{r})$.

$D_g(\boldsymbol{\omega})$ is the Fourier transform of a Gaussian distribution and it is equal to:

$$D_g(\boldsymbol{\omega})=Cte\cdot\exp(-R_1^2\omega^2/4). \quad \text{(A8)}$$

The computation of $S(\boldsymbol{\omega})$ is easily performed. Let us consider (Fig. 6) a lens with a spherical aberration coefficient C_s illuminated by a parallel electron beam. The plane P is the Gaussian plane situated at a distance L from the lens. An infinitely thin ring of radius r and

width dr is focussed into a point C whose coordinate δ depends on C_s and r. Then if N is the electron density at the entrance pupil of the lens and $N(\rho)$ the electron density in the plane P' situated at a distance l' from P, the following relations hold:

$$2\pi N r dr = N(\rho) 2\pi\rho d\rho \tag{A9}$$

$$N(\rho) = [(L-\delta)^2/(l'-\delta)^2]N \tag{A10}$$

$$r = [(l'-\delta)/(L-\delta)]\rho \tag{A11}$$

and $S(\omega)$ is then given by the Hankel transform of (A10):

$$S(\omega) = 2\pi K \int_0^\infty \rho N(\rho) J_0(\omega\rho) d\rho, \tag{A12}$$

where K is the normalization factor. Introducing Eqs. (A10) and (A11) in (A9) gives

$$S(\omega) = 2\pi N K \int_0^{r_{\max}} r J_0\left(\omega \frac{l'-\delta}{L-\delta} r\right) dr. \tag{A13}$$

If the plan P' is in the plane of least confusion, and assuming an aperture $\alpha = r_{\max}/L$, Eq. (A13) is

$$S(\omega) = 2\pi N K \int_0^{r_{\max}} r J_0\left(\omega \frac{C_s r^3}{4L^3}\right) dr \tag{A14}$$

and for the normalized function:

$$S(\omega) = \frac{2}{{r_{\max}}^2} \int_0^{r_{\max}} r J_0\left(\omega \frac{C_s r^3}{4L^3}\right) dr. \tag{A15}$$

It is difficult in general to find an analytical expression for (A15). However, as an example, Fig. (7) shows $S(\boldsymbol{\omega})$ and $D(\boldsymbol{\omega})$ for a lens with a focal distance $f=15$ mm, a spherical aberration $C_s=50$ mm and an aperture diameter $2r_{\max}=100\,\mu$. With these conditions, $d_s=7$ Å and, if $d_g=33$ Å, $d\approx 40$ Å. The Fourier transform $D_g'(\boldsymbol{\omega})$ of a Gaussian beam with a "diameter" $2R_1=40$ Å is also shown in Fig. (7). At a first approximation, it is then possible to consider the electron distribution on the sample as Gaussian, and the "equivalent surface area" for such a beam is

$$\pi R^2 = 2\pi \int_0^\infty \exp\left(\frac{-r^2}{{R_1}^2}\right) r dr \tag{A16}$$

which gives the value of $R=R_1$.

APPENDIX B

Let us consider in Fig. (8) an element ΔS of the surface of a plane (P) where the center A_j is such that $A_jM_1=r_1$ and $A_jM_2=r_2$ and let us also consider two stochastic variables X_1 and X_2 taking values of 1 or 0 depending on the fact that one of the points M_1 or M_2 are, respectively, covered or not by a circle C_j with a center 0_j in ΔS. n_1 and n_2 will then be the sums of X_1 and X_2, respectively, extended to the totality of the plan P. It is always possible to choose ΔS small enough so that the probability of having more than one point 0_j in ΔS can be considered negligible by comparison with the probabilities of having 0 or 1 point 0_j in ΔS. According to the Poisson distribution, it is then possible to write:

$$Pr(0) = Pr(0 \text{ point in } \Delta S) = \exp(-\bar{d}\Delta S) \sim 1 - \bar{d}\Delta S$$

$$Pr(1) = Pr(1 \text{ point in } \Delta S) = \bar{d}\Delta S \exp(-\bar{d}\Delta S) \sim \bar{d}\Delta S$$

$$Pr(2) = Pr(2 \text{ points in } \Delta S) = [(\bar{d}\Delta S)^2 \exp(-\bar{d}\Delta S)]/2 \sim 0.$$

Because of the independency in the Poisson distribution of the disjoint, ΔS, X_1 and X_2 are independent stochastic variables and the characteristic function $\phi_n(u_1, u_2)$ associated with (n_1, n_2) is the product of the characteristic functions $\phi(u_1, u_2)$ associated to (X_1, X_2) and relative to each ΔS. If ΔS is on the right side of Fig. 7, Table I can be written. It is now possible to find the characteristic function associated with the right side of (P):

$$\phi_R(u_1, u_2) = 1 + Pr(1)\{p(r_1) - 1 + e^{iu_1}[p(r_2) - p(r_1)] + e^{i(u_1+u_2)}[1 - p(r_2)]\}$$

$$= 1 + f(u_1, u_2; r_1, r_2)\Delta S_R \tag{B1}$$

and if ΔS is in the left side

$$\phi_L(u_2, u_1) = 1 + f(u_2, u_1; r_2, r_1)\Delta S_L. \tag{B2}$$

Since $\phi_n(u_1, u_2) = \boldsymbol{\Pi}\phi(u_1, u_2)$, it is easier to compute the second characteristic function $\psi(u_1, u_2) = \ln\phi(u_1, u_2)$ and $\Psi_n(u_1, u_2) = \boldsymbol{\Sigma}\psi(u_1, u_2)$.

$$\psi(u_1, u_2) = \begin{pmatrix} f(u_1, u_2; r_1, r_2) \text{ if } \Delta S \text{ is in } [R] \\ f(u_2, u_1; r_2, r_1) \text{ if } \Delta S \text{ is in } [L] \end{pmatrix} \tag{B3}$$

The sum of all the elementary contributions $\psi(u_1, u_2)$ for all $[P]$ is equal to:

$$\Psi_n(u_1, u_2) = \bar{d}\left((e^{iu_1} + e^{iu_2} - 2)\int_R [1 - p(r_1)]\Delta S + (2e^{i(u_1+u_2)} - e^{iu_2})\int_R [1 - p(r_2)]\Delta S\right). \tag{B4}$$

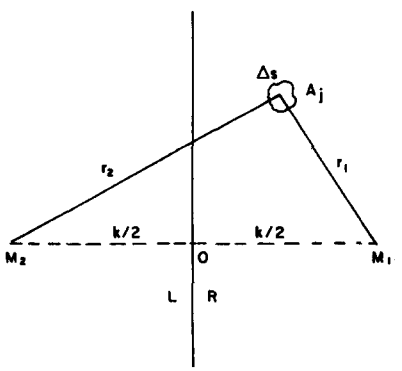

FIG. 8. Computation of average brightness and autocorrelation function.

Since $B=n(M)B_0$,

$$\Psi_B(v_1, v_2)=\Psi_n(v_1B_0, v_2B_0).$$

Finally

$$\phi_B(v_1, v_2)=\exp[\Psi_B(v_1, v_2)]. \quad \text{(B5)}$$

The computation of the moments is then obtained. By application of the theorem of characteristic functions

$$\bar{B}_1=\bar{B}_2=\frac{1}{i}\frac{\partial\phi_B}{\partial v_i}\bigg|_{v_1=v_2=0} \quad \langle E(B_1B_2)\rangle=\frac{1}{i^2}\frac{\partial^2\phi_B}{\partial v_1\partial v_2}\bigg|_{v_1=v_1=0}$$

and

$$\bar{B}=\bar{d}B_0\int_{\text{all plane}}[1-p(r)]\Delta S \quad \text{(B6)}$$

$$K(k)=2\bar{d}B_0^2\int_R[1-p(r_2)]\Delta S. \quad \text{(B7)}$$

Concerning the determination of $p(r)$, it has been assumed above that the image of one electron leaving a surface with a kinetic energy E is a circle of diameter $2R=\lambda/2$, where λ is the electron wavelength. Consequently:

$$R=A/(V)^{1/2}(A=12.25/4,\ V=\text{Volts},\ R=\text{Å}). \quad \text{(B8)}$$

According to the experimental results[21] the energy distribution of the secondary electrons can be approximated by a Rayley distribution

$$N(E)dE=\frac{E}{\sigma^2}\exp(-E^2/2\sigma^2)dE\begin{cases}E_{\max}=\sigma\\ \bar{E}=1.25\sigma\end{cases}. \quad \text{(B9)}$$

An easy transformation with (B8) and (B9) gives the probability distribution of R:

$$p(R)dR=(2A^4/\sigma^2)(1/R^5)\exp(-A^4/2\sigma^2R^4)dR \quad \text{(B10)}$$

TABLE I. Computation of characteristic function.

Number of 0_j in ΔS	X_1	X_2	$\exp[i(u_1X_1+u_2X_2)]$	Prob.
0	0	0	1	$Pr(0)\sim 1-Pr(1)$
1	0	0	1	$p(r_1)Pr(1)$
1	1	0	$\exp(iu_1)$	$[p(r_2)-P(r_1)]Pr(1)$
1	1	1	$\exp[i(u_1+u_2)]$	$[1-p(r_2)]P(r_1)$
2	0	0	1	$Pr(2)\sim 0$

and

$$p(r)=\int_0^r p(R)dR. \quad \text{(B11)}$$

The integral in expression (B6) is easily obtained by noticing that it represents the sums of ΔS weighed by the probability that ΔS is inside a circle of radius r centered on the origin. The integral in expression (B7) represents the sums of ΔS on the right side of plane P weighed by the probability that ΔS is inside a circle of radius r centered on the point M_2. After a few transformations we have:

$$\bar{B}=\bar{d}\pi r_0^2B_0 \quad \text{(B12)}$$

$$C(k)=2dB_0^2F(k, r) \quad \text{(B13)}$$

$$F(k, r)=\int_{k/2}^{\infty}\left[\pi-2\arcsin\left(\frac{k}{2r}\right)\right]\times\left[1-\frac{2A^4}{\sigma^2}\int_0^r\frac{1}{R^5}\exp\left(-\frac{A^4}{2\sigma^2R^4}\right)dR\right]rdr, \quad \text{(B14)}$$

where r_0 is the average value of the probability distribution (B10). For the majority of metals $E_{\max}=\sigma^2=2$ eV and $r_0=2.27$ Å.

[1] C. E. Shannon, Bell Telephone System, Tech. Publ., Monographs B-1598 (1948).
[2] N. I. Klyuev, U.S. Department of Commerce Nat. Bur. Std. Ad 686 Rept. No. 380, 1969.
[3] A. Blanc-Lapierre, Ann. Inst. Henri Poincare **13**, 245 (1953).
[4] J. L. Harris, J. Opt. Soc. Amer. **54**, 931, 606 (1964).
[5] R. Simon, J. Appl. Phys. **40**, 2851 (1969).
[6] P. Grivet, *Electron Optics* (Pergamon New York, 1965), p. 197.
[7] L. Brillouin, *Science and Information Theory* (Academic, New York, 1956), p. 156.
[8] M. Savelli, Compt. Rend. Acad. Sci. **244**, 1710 (1957).
[9] B. Picinbono, Compt. Rend. Acad. Sci. **240**, 2296 (1955).
[10] O. Rice, *Selected Papers on Noise and Scholastic Processes*, edited by N. Wax (Dover, New York, 1954), p. 225.
[11] A. Rose, Advan. Electron. Electron Phys. **1**, 131 (1948).
[12] T. Mulvey, *Focussing of Charged Particles*, edited by A. Septier (Academic, New York, 1965), p. 476.
[13] P. Duncumb, Brit. J. Appl. Phys. **10**, 420 (1959).
[14] N. Broers, J. Sci. Instrum. **2**, 273 (1969).
[15] D. Estournet, Onde Electrique **49**, 848 (1969).
[16] It will preferable to use a digital system instead of an analog, the analog differentiation systems being noise sensitive.[17]
[17] R. E. Graham, Acta Electron. **2**, 333 (1957).
[18] M. Von Ardenne, Z. Tech. Phys. **19**, 407 (1938); **20**, 238 (1940).
[19] P. Durandeau and C. Fert, Rev. Opt. **36**, 205 (1957).
[20] G. V. Der-Shvats and I. S. Makarova, Radio Eng. Electron. **1**, 161 (1967).
[21] O. Hackenberg and W. Braver, Advan. Electron. Electron Phys. **11**, 413 (1959).