Notes on "Image Methods in Electrostatics" (A Computer-Animated Film)

RONALD BLUM*
The College and Department of Physics, University of Chicago, Chicago, Illinois
(Received 30 October 1967; revision received 15 January 1968)

The computation and production of a computer-animated film is described. The film depicts a point charge slowly approached by (1) a grounded conducting sphere, and (2) an ungrounded (net charge equals zero) conducting sphere.

INTRODUCTION

The human organism is not a digital machine. It continuously monitors its environment by means of its senses, notably sight, a sense which is remarkably well adapted and acute. In conceptual areas as well, we instinctively rely upon pictures to develop and coordinate ideas; extremely complicated relations can be comprehended literally *in einem Augenblick* through the use of a picture. In addition to these obvious formal advantages, there is a fundamental fascination with images which seems to underlie the current rapid development of computer animation. Many a scholar opens his Sunday newspaper to the colored comics section first. One may speculate that the time is not far distant when the "average" research scientist will be able to program complicated three-dimensional stereo representations in motion with the same ease with which he would attempt a sketch. As a possible antidote to the paper explosion, filmed computer output is already in use, while the advent of small, portable cartridge-loading film projectors may be expected to lead to vastly increased use of short film loops as education and information-storing devices.

One topic in physics which lends itself to a pictorial treatment is the concept of fields and field lines. This concept is widely used in electromagnetism and plasma physics, yet direct physical observation of field lines is generally limited to sprinkling iron filings on a bar magnet, although more sophisticated techniques exist.[1] Hence this film, "Image Methods in Electrostatics," designed to demonstrate, on the level of junior-year physics, something of the nature of field lines, their elasticity, and the way in which they can form and re-form. The film, produced at Argonne National Laboratory on a Data Display 80A cathode-ray tube (CRT) on line with a CDC 3600 computer, depicts the familiar problem of a point charge in the vicinity of a grounded conducting sphere, and the less-familiar problem of a point charge and a conducting sphere which is *not* grounded, but maintains a net charge of zero.[2] During the animated sequences the spheres are shown slowly moving toward the point charge—so slowly that time-dependent effects may be ignored.

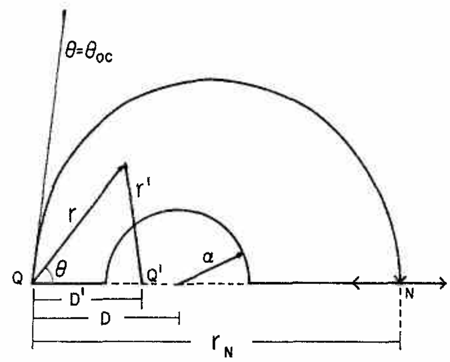

FIG. 1. Grounded conducting sphere.

I. POINT CHARGE AND GROUNDED CONDUCTING SPHERE

In the case of a grounded conducting sphere of radius a separated from a point charge Q by a distance D (see Fig. 1) one must solve Laplace's equation for electrostatic potential V, in the region outside the sphere, with a singularity at Q and vanishing potential on the boundary of the sphere. This is accomplished by replacing the sphere with an image charge $Q'=-aQ/D$ located a distance $D'=D-a^2/D$ from Q. This is the physical equivalent of constructing Green's function for the

* Work performed under the auspices of the U.S. Atomic Energy Commission, while a visiting scientist in the Applied Mathematics Division, Argonne National Laboratory, Argonne, Ill. 60440.

[1] P. Cavanagh, Am. J. Phys. **34**, 1034 (1966).

[2] Copies of the film will be loaned or sold at nominal cost to those interested, through the Argonne Film Center. Four-minute 8-mm loops are available as well as 16-mm film.

This article is copyrighted as indicated in the article. Reuse of AAPT content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 138.87.160.11 On: Tue, 09 Dec 2014 19:30:54

Laplacian in this geometry. The obvious conceptual simplicity of the problem generally precludes any serious attempt to construct the field lines, so the analysis will be presented in some detail.

In mks units the equation for the potential is

$$4\pi\epsilon_0 V=(Q/r)+(Q'/r'), \tag{1}$$

and for the electric fields $\mathbf{E}=-\nabla V$.

$$4\pi\epsilon_0 E_r=(Q/r^2)+[Q'(r-\mu D')/r'^3],$$

$$4\pi\epsilon_0 E_\theta=Q'D'\sigma/r'^3, \tag{2}$$

where we use the notation $\mu=\cos\theta$, $\sigma=\sin\theta$ for convenience. The field-line equation

$$r(d\theta/dr)=E_\theta/E_r \tag{3}$$

is easily integrated by noting that, since $\nabla\cdot\mathbf{E}=0$ implies

$$\partial(r^2\sigma E_r)/\partial r=-\partial(r\sigma E_\theta)/\partial\theta, \tag{4}$$

i.e., the integrating factor of the field-line equation is $r\sigma$, its solution, for $a=1$, is

$$F(r,\theta)=\mu+[(D'-\mu r)/Dr']=B, \tag{5}$$

where the constant B determines the particular field line. If the field line leaves Q at an angle θ_0, then

$$B=\mu_0+1/D. \tag{6}$$

The field-line equation gives rise to a quadratic in r which has the solution, for $0\leq\theta\leq\pi$,

$$r=D'(A^2-\mu^2)^{-1}[(A^2-1)\mu+A\sigma(1-A^2)^{1/2}], \tag{7}$$

where $A=D(B-\mu)$. If $A=+\mu$, $r=D'/2\mu$; if $A=-\mu$, $r=\infty$. For real solutions, $-1\leq A\leq 1$, or $\mu_0\leq\mu\leq\mu_0+2/D$ for the field line characterized by μ_0; for $-\pi\leq\theta\leq 0$ replace the $(+)$ sign before the radical in Eq. (7) with a $(-)$. The result is symmetric about the axis; hence this region is not shown in the film.

As r, $r'\to\infty$, Eq. (5) shows that the field lines asymptotically approach

$$\mu_\infty=\frac{\mu_0+(1/D)}{1-(1/D)}\leq 1, \tag{8}$$

implying that for some $\mu_0\geq\mu_{0c}=1-2/D$ the field lines never arrive at ∞, but end on the image charge instead. This is another way of saying that field lines originating in a solid angle of $4\pi(Q'/Q)$ steradians about the polar axis must end on Q'—a check on the correctness of the calculations.

The fact that there are two different types of field lines emanating from Q—those which end at infinity and those which end on the sphere—causes no embarrassment, because the explicit dependence $r(\theta)$ is known. When the dependence is only implicit the distinction is important, as is shown next. It is important to note the behavior of the limiting field line, μ_{0c} (see Fig. 1) which "splits" at N. A good tactic when dealing with problems of this sort is to look immediately for neutral points in the field. N must be such a point, and by solving the neutral-point equations $E_r=0=E_\theta$ one finds:

$$\mu_N=1,$$

$$r_N=(1+D)[1+1/(D)^{1/2}]. \tag{9}$$

When these values are substituted back into Eq. (5), one finds that $\mu_0=1-2/D=\mu_{0c}$, as expected.

II. POINT CHARGE AND UNGROUNDED CONDUCTING SPHERE

This case is mathematically very similar to the previous case, but with a fundamental difference: the sphere is ungrounded and so must have a net charge of zero although the potential on the sphere is positive due to charge separation induced by the point charge Q. To do this we require two image charges, $Q'=+aQ/D$ and $Q''=+aQ/D$. The first charge is placed as before and serves to fix the sphere at $V=0$; the second charge is placed at the center of the sphere, thus achieving charge neutrality while maintaining the sphere as an equipotential at $V=aQ/4\pi\epsilon_0 D$. Furthermore, the net amount of electric flux entering the sphere must be zero; i.e., every time a field line is pulled into the conductor from one side it must emerge again on the other side.

The electric fields and the equation of the field lines are given by (see Fig. 2)

$$4\pi\epsilon_0 E_r=\frac{Q}{r^2}+\frac{Q'(r-\mu D')}{r'^3}+\frac{Q''(r-\mu D)}{r''^3} \tag{10}$$

$$4\pi\epsilon_0 E_\theta=(Q'D'\sigma/r'^3)+(QD\sigma/r''^3) \tag{11}$$

$$F(r,\theta)=\mu+[(D'-\mu r)/Dr']-[(D-\mu r)/Dr'']$$

$$=\mu_0,\ \text{const}, \tag{12}$$

This article is copyrighted as indicated in the article. Reuse of AAPT content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 138.87.160.11 On: Tue, 09 Dec 2014 19:30:54

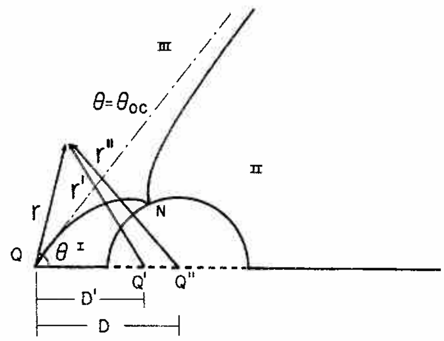

FIG. 2. Ungrounded conducting sphere.

where μ_0 determines the particular field line leaving Q at angle θ_0.

From the equations $E_r=0=E_\theta$, it is not difficult to show that

$$r_N=\nu D, \qquad r_N'=\nu, \qquad r_N''=1, \tag{13}$$

and

$$\mu_N=(D^2-1+\nu^2D^2)/2\nu D^2, \tag{14}$$

where $\nu^3=D'/D<1$. Thus, the neutral point is on the surface of the sphere. Substituting into Eq. (12) shows that the critical field line is given by

$$\mu_{0c}=\nu\mu_N-(1-\nu^2)<\mu_N \tag{15}$$

and divides up the region outside the sphere into three separate portions (see Fig. 2):

(I) $\mu>\mu_{0c}$, $r<r_N$, containing field lines from Q which terminate on the sphere;

(II) $\mu>\mu_{0c}$, $r>r_N$, containing field lines leaving the sphere and going to infinity;

(III) $\mu<\mu_{0c}$, containing field lines which do not touch the sphere but go on to infinity, although with more or less visible distortion.

Since Eq. (12) could not be solved for $r(\theta)$ directly, iterative solutions were obtained by the Newton–Raphson method, using a simple predictor to hasten convergence. Consequently, it proved necessary to test for and treat the three regions separately in the computations.

III. COMPUTATION

The animated sequences were to show the distortion of field lines for varying positions of grounded and ungrounded conducting spheres relative to a point charge fixed at the origin of coordinates. It was necessary to depict the sphere slowly (quasistatically) approaching the point charge Q, lines of force emerging from Q at intervals $\theta=0°$, $10°$, $20°$, $\cdots$, $180°$, and the image charges as the sphere moved in toward Q from off screen. Instead of using some fixed value of $\delta\theta$ or δr as the increment for describing a field line, it was decided to take a fixed value of arc length

$$\delta s=(\delta r^2+r^2\delta\theta^2)^{1/2} \tag{16}$$

as increment to insure uniformity in the drawing and a more consistent relationship between distances on the screen and the number of iterations in the program. The increment δs was chosen small enough to give a smoothly continuous representation of field lines, but not too small. The CRT screen is divided into a mesh of 1024×1024 raster points; as a general rule, an arc of length less than 10–20 rasters can give rise to a jagged or "lumpy" effect.

A cardinal rule for economy in computer animation is to substitute intelligence for computation whenever possible. A useful tactic when changing configurations by small amounts is to avoid calling library functions by using recursion relations, identities, calculus of finite differences, and storage of frequently used function values. In the present case, trigonometric functions of θ were avoided by treating $\cos\theta=\mu$ as the angle variable. As the computations proceeded, the field-line equation was solved for $r(\mu)$ or $\mu(r)$, either exactly or by the Newton–Raphson method to within 0.5% accuracy, as the occasion warranted, and increments chosen from

$$\delta\mu=\pm\delta s[(\delta r/\delta\mu)^2+(r/\sigma)^2]^{-1/2} \tag{17}$$

$$\delta r=\pm\delta s[1+(r/\sigma)^2(\delta\mu/\delta r)^2]^{-1/2}. \tag{18}$$

In order to begin the calculation for a given field line μ_0 starting from Q, we note that the field lines must be approximately radial, so that the first point on the field line has coordinates $(\delta s, \mu_1)$. The variation in the cosine is second order in δs, and expanding r' and r'' around the origin, retaining terms in δs^2,

$$\mu_1=\mu_0+(\sigma_0^2\delta s^2/2DD'^2),$$

grounded sphere,

$$\mu_1=\mu_0+(\sigma_0^2\delta s^2/2DD'^2)(1-\nu^6),$$

ungrounded sphere. (19)

This article is copyrighted as indicated in the article. Reuse of AAPT content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 138.87.160.11 On: Tue, 09 Dec 2014 19:30:54

The film was made in three sequences: (a) the grounded sphere moving from $D=20$ to $D=1.4$ in steps of $\delta D=-0.05$, slowing down to $\delta D=-0.025$ at $D=8$, $\delta s=0.3$. The CRT screen scale was chosen to be $-10.1\leq x\leq 10.1$, $-3.1\leq y\leq 17.1$ and plotting was restricted to the region $-10.0\leq x\leq 10.0$, $-3.0\leq y\leq 17.0$ by appropriate tests; (b) the ungrounded sphere moving from $D=12$ to $D=3$ in steps of $\delta D=-0.05$. $\delta s=0.3$, with the same scale as in (a); (c) the ungrounded sphere moving from $D=7$ to $D=1.4$ in steps of $\delta D=-0.02$, $\delta s=0.15$. At $D=3$ the step was changed to $\delta D=-0.01$, and the scale was shifted and enlarged to $-0.4\leq x\leq 8.1$, $-1.0\leq y\leq 7.4$. The time rate of change of D is irrelevant, except that it is much less than the speed of light.

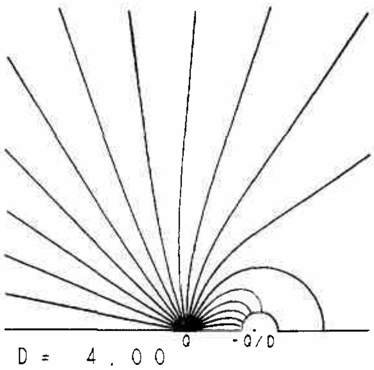

FIG. 3. Sample negative, sequence (a).

In the case of the ungrounded sphere, a field line $\mu_0>\mu_{0c}$ which enters the sphere from region I must emerge again in region II, and approach a radial asymptote at angle μ_0. This condition plus the condition $r''=1$ yields the starting point for the continuation of the emerging field line in region II:

$$\mu_2=D(\mu_0-1)+1;$$

$$r_2^3-(1+D^2+2\mu_0D^2)r_2+2DD'=0, \qquad (20)$$

where the second equation is solved for r_2 by iteration, starting from

$$r_2^2=1+2D+D^2(2\mu_0-1) \qquad (21)$$

as a first approximation.

Figures 3, 4, and 5 show negatives from sequences (a), (b), and (c), respectively. The effect of animation is to fill up blanks and smooth out

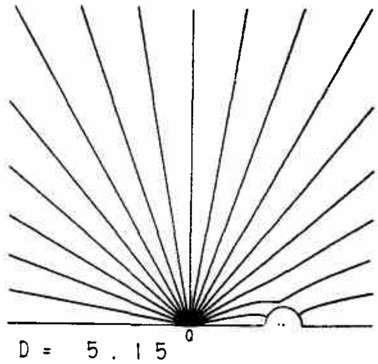

FIG. 4. Sample negative, sequence (b).

nonuniformities, giving a more polished appearance than indicated by these figures.

IV. PRACTICAL MATTERS

The major questions which arose in making the present film were: what to do about titles; how many frames (individual pictures) to compute; how much discontinuity was tolerable between frames; how will the end result look on 8-mm film loops? It is preferable to have titles and textual material done in one strip by professionals. One is thus spared the tedium of reading in, storing, and displaying titles on the CRT, a fruitful source of programming errors. The author of a title is also apt to make only half as many frames as his audience requires for comfortable reading. All characters should be filmed at high intensity, in either medium or large size; in this film all line segments were redrawn three times, and characters from three to six times in order that they be sharp and clear after reduction to 8 mm (and they still left something to be desired). No title or text is short enough to be read in less than 5 sec.

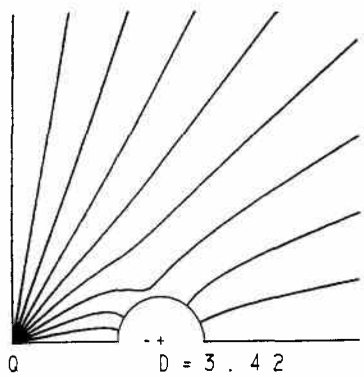

FIG. 5. Sample negative, sequence (c).

This article is copyrighted as indicated in the article. Reuse of AAPT content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 138.87.160.11 On: Tue, 09 Dec 2014 19:30:54

At the beginning of each animated sequence, an "establishing shot" is desirable; i.e., repeating the first frame for about 5 sec or so to give the audience time to orient themselves. It is best to do this on the CRT, because doing it manually from a single frame is expensive, while the coordinates for a single frame can be stored in core or on tape and recalled each time the frame is plotted. If one wishes to have one sequence fade into another, a minimum of 20 extra frames should be left at the end of the first sequence and the beginning of the second for a "lap dissolve." In this optical process the light is dimmed while exposing the end of the first sequence, turned on while exposing the beginning of the second, and the two films are overlapped and printed together.

The question of how much discontinuity is to be allowed between frames depends upon the nature of the individual motion, how long it is desired to keep the sequence before the audience, and the cost of computing each frame. The most widely used projectors are 16 mm. With sound they run at 24 frames/sec (40 frames = 1 ft of 16-mm film), silent at 16 frames/sec, as do 8-mm projectors. The faster the apparent motion, the more discontinuity one can allow. It is common practice in animation to duplicate each frame in printing ("double framing"), which can decrease computation time by a factor of two. However, if the motion is very slow, this may give rise to a noticeable jerkiness. For a very rapid motion one may be able to double frame twice, decreasing computation time by a factor of 4, without noticeable degradation of the continuity. Processing equipment also may allow for "skip framing," printing every other frame in order to shorten a film, and "every-other-frame-twice" printing, which produces a sound version from a silent 16-mm film (not a recommended procedure). In the present film approximately 1000 frames were computed to total, after double framing, two minutes of viewing time (16-mm silent). If single frames are very expensive to compute, one may give the illusion of motion by duplicating, for example, a dozen separate frames enough times so that separate configurations can be connected by lap dissolves. The effect is roughly analogous to that of neon signs which simulate motion. Another tactic is to interpolate between computed configurations. If a film is to be distributed, the original should *not* be viewed on a projector because it will acquire scratches, sprocket wear, and other damage.

In general, a film will need processing, particularly if the original is a black line on transparent background, since black lines on the original tend to "wash out" in processing. One must make a reversal print or negative to obtain white lines on a black background. A colored filter can be used, as in the present case, to make the picture more attractive. The result is edited and spliced together with titles and texts to make a work print which is duplicated to make a splice-free master print and a viewing copy. If there are many splices, better quality can be obtained by making two work prints, an "*A*" and a "*B*" print. These prints show complementary alternating sequences with blank spaces between them. They are then matched up by edge numbers on the film and made into a single splice-free master print. One beneficial result of all this processing is that it increases the contrast and makes the white lines more intense.

In some installations the CRT may be unstable or the camera may not have "positive" or "pin" registration; in that case the picture will not appear to remain in the same place when viewed. One may avoid the issue by providing some small fixed marks on the frame, or a pair of fixed coordinate axes, for example, so that the film can be manually registered frame by frame when it is duplicated. This is expensive and not to be recommended, but it is possible.

One final word about esthetics: The artistic scientist or educator must keep an open mind about his mathematics. Often one may save significant amounts of time and trouble by taking advantage of the latitude which the visual process allows; accuracy must not be a sacred cow. In the present film it was found that terminating field lines at $r''=0.97$, rather than 1.00 (the surface of the sphere) gave a better image near the sphere, and that drawing straight lines left of the point charge in the second sequence gave results not visibly different from the correct formula. The intense glow around the point charge Q, which appeared as the result of too many field lines converging on one point, was purposely left in for "dramatic effect." With a bit more courage on the part of the author, 2% accuracy would have been good enough for the Newton–Raphson computations, rather than 0.5%. Pedagogically, the im-

This article is copyrighted as indicated in the article. Reuse of AAPT content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 138.87.160.11 On: Tue, 09 Dec 2014 19:30:54

portant point is not to draw a mathematically accurate image, but to convey a physically accurate idea.

ACKNOWLEDGMENTS

The author wishes to acknowledge the generous support extended him by the Applied Mathematics Division, Argonne National Laboratory, and particularly the assistance of Ronald Krupp of the Applied Mathematics Division and Stanley Andrews, of the Motion Picture Section. The author is also grateful to the Commission on College Physics for their considerable help and encouragement in pursuing this project.

AMERICAN JOURNAL OF PHYSICS VOLUME 36, NUMBER 5 MAY 1968

Density of States in a Sphere and Cylinder

R. H. LAMBERT

Department of Physics, University of New Hampshire, Durham, New Hampshire

(Received 24 August 1967; revision received 8 January 1968)

The density of states for the solutions to the wave equation are derived for a spherical and a cylindrical volume. The number of states per unit volume, in the limit of short wavelength, is known to be independent of the shape of the surface enclosing the volume. It is hoped the general result, first proved by Weyl, will appear more plausible if explicitly proved for several different shapes.

The density of eigenvalues, or states as it is often called, for solutions of the wave equation is of use in many contexts. Probably the most famous occurrence is in the derivation of the Rayleigh–Jeans and Planck radiation laws. Conventionally the derivation is based on counting the number of solutions found to the wave equation in a parallelepiped either with periodic boundary conditions or with the requirements that the functions vanish on the boundary. If the wavelengths of interest are much smaller than the smallest dimension characterizing the volume containing the waves, the resulting number density should be of the form[1]

$$dN = Vk^2[C_0 + C_1(1/kL)^p + \cdots]dk,$$

with $p>0$, dN the number of eigenvalues lying in the range dk about k $(k=2\pi/\lambda=\omega/c)$, V the volume containing the waves, and L is a characteristic linear dimension of the cavity. The leading coefficient C_0 has been shown, in a rather sophisticated proof[2] to have the value $1/2\pi^2$ independent of the shape of the volume. As an alternate to the rather inaccessible general proof, it seems worthwhile to give explicit proofs for the sphere and the cylinder since these shapes, in conjunction with the parallelepiped, will tend to make the general result more plausible. It should perhaps be mentioned that this work was the result of assigning the problem to a class with the thought that the derivation must be reasonably accessible. For the sphere a solution utilizing the WKB method has been given by Pauli[3] but it is felt that an alternate solution, not requiring the WKB method, would also be useful.

I. SPHERE

The solutions of the wave equation in spherical coordinates which are regular at the origin are

[1] E. A. Power, *Introductory Quantum Electrodynamics* (Longmans Green and Co., Ltd., London, 1964).

[2] H. Weyl, Math. Ann. **71**, 441 (1912). See also R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience Publishers, Inc., New York, 1953), Theorem 18, p. 442.

[3] W. Pauli, *Handbuch der Physik*, S. Flügge, Ed. (Springer-Verlag, Berlin, 1958), Vol. 5, p. 92.

This article is copyrighted as indicated in the article. Reuse of AAPT content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 138.87.160.11 On: Tue, 09 Dec 2014 19:30:54