Introductory article

Much ado about nothing: optical dislocation lines (phase singularities, zeros, vortices...)

Michael Berry

H H Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, United Kingdom

Often, light can be represented, approximately or exactly, by a complex scalar wave ψ, smoothly varying in space and/or time. The field ψ could be a cartesian component of the electromagnetic field or of the vector potential, or one of several scalar potentials appropriate to different circumstances[1]. This meeting has been concerned with the line singularities of the phase of ψ. Here, I wish to make some general remarks about these lines. Except where stated, all the remarks are independent of the particular wave equation that ψ satisfies.

They are lines because the only way that the phase of a smooth function can be singular is for its modulus to vanish; this in turn requires that two conditions are satisfied - $\mathrm{Re}\,\psi = 0$ and $\mathrm{Im}\,\psi = 0$ - and these define a line (think of the simplest local model $\psi = x + iy$, where the singular line of phase is the z axis). Three interpretations of the lines are: as wavefront dislocations, since the patterns of surfaces of constant phase (mod 2π) mirror those of dislocations in the arrangements of atoms in crystals[2,3]; as vortices, since the phase gradient direction (that is, the direction of the current, or Poynting vector) swirls about the singular line like fluid in an irrotational vortex[4-6]; and as zeros, that is threads of darkness.

From a broader perspective, these singularities of faint light are complementary (in the sense introduced by Bohr) to the singularities of bright light. The latter are caustics, that is, envelopes of families of rays in geometrical optics[7,8]. Apart from the obvious complementarity of bright and dark, there is the less obvious fact that measurement of each inhibits measurement of the other. To see a caustic singularity requires seeing the light on a scale large compared with the wavelength; then the dislocations, which are sub-wavelength fine structures, are too small to perceive. On the other hand, under the high magnifications required to see dislocations the caustic singularities are smoothed away by diffraction.

Continuing the comparison, I note that each singularity is a window to a deeper theory. The most striking prediction of geometrical optics is the infinite intensity at a caustic. But this is precisely where ray theory breaks down dramatically, because wave effects soften the divergence, and interference fringes are biggest - for wavelength λ (small compared with other relevant dimensions), they scale as $\lambda^{2/3}$ rather than λ itself (that is why light waves can be seen in the sky, enormously magnified, as supernumerary rainbows[9] close to the main arc which is an angular caustic). So, caustic singularities lead naturally into waves, in which there is an additional field, namely phase, and the singularities of phase are the singularities of wave optics. But these are precisely where wave optics itself breaks down, in the sense that the dark light of a dislocation reveals the photon fluctuations of quantum optics - the next (and to this day the last) level in the hierarchy of physical theories of light[10].

SPIE Vol. 3487 • 0277-786X/98/$10.00

Around a dislocation, the phase changes by a multiple of 2π. There are two natural ways to specify the strength of a dislocation in terms of this phase change. I need to discuss them carefully, for a reason that will emerge later. In the first, we let s label points on the dislocation and denote by $\mathbf{n}(s)$ one of the two (opposite) unit vectors along the dislocation. Then the strength S is $1/2\pi$ times the phase increase round the dislocation in a positive (clockwise) circuit defined by $\mathbf{n}$. Thus defined, S is a signed integer, usually ± 1. A convenient formula[11] is

$$S = sign \mathrm{Im} \nabla \psi^* \times \nabla \psi \cdot \mathbf{n} \qquad (1)$$

The choice of $\mathbf{n}$, rather than $-\mathbf{n}$, is arbitrary; often, but not always, it is convenient to make the choice at one point on the dislocation, and define $\mathbf{n}(s)$ elsewhere by continuity.

In the second way of describing the strength of a dislocation, the direction of $\mathbf{n}$ is chosen so that the phase increases in a positive circuit; the strength can then be specified simply by an arrow on the dislocation. Higher-order dislocations can be represented by multiple arrows.

Dislocation lines can move, either in time (e.g. for wave pulses) or as parameters vary. Then they can collide and interact in various ways[11-13]. These interactions are constrained - independently of any wave equation - by a topological conservation law[2] involving the dislocation strength. Consider an arbitrary curve C (stationary or moving) in space. It may be threaded by dislocation lines, in the sense that these may pierce an arbitrary surface Σ spanning C. Let a sense be defined on C, so that the two sides of Σ ('front' and 'back') can be distinguished. Using the second specification of dislocation strength, we define the total strength S_C as the signed number of dislocation arrows piercing Σ, with arrows pointing to the front counting +1 and arrows pointing to the back counting -1; S_C is independent of Σ. The conservation law is that S_C does not change unless a dislocation line crosses C.

Dislocation interactions are also constrained by a second, and more subtle, conservation law[14]. In three dimensions, the wavefronts, on which the phase is constant (mod 2π) are surfaces. Wavefronts intersect each spanning surface Σ in lines - phase contours. Consider the pattern of phase contours on Σ, without regard to the phase labels they carry. The patterns have singularities, of three kinds: not only dislocations but also phase saddles and phase extrema (maxima or minima). For each singularity, a Poincaré index can be defined: in a circuit of the singularity, the index is the number T of complete rotations of the direction of the phase contours in the same direction that the circuit is traversed. For dislocations, from which phase contours on Σ radiate like spokes of a wheel, they rotate in the same sense as the circuit; therefore $T = +1$ for first-order dislocations, irrespective of whether their strength is $S = +1$ or $S = -1$. For phase maxima and minima, the phase contours degenerate to points surrounded by loops, and $T = +1$ for these too. For phase saddles, the contours are hyperbolas that rotate in the opposite direction to the circuit, so $T = -1$. We define the total strength T_C as the sum of the Poincaré indices T for all the singularities on Σ. T_C is independent of Σ. The conservation law is that T_C does not change unless a dislocation line or phase saddle or phase extremum crosses C.

The subtlety arises on considering a continuous family of spanning surfaces Σ, nonintersecting except on C. On each of these surfaces, there can be phase saddles and phase extrema, and it is tempting to connect these points on different Σs, and regard the loci as line singularities of phase saddles and extrema. But this would be wrong, because, unlike dislocations, these lines have no independent existence; they are artefacts of the choice of family of Σs: for a different choice, the lines would be different. In fact, the phase saddles and extrema involved in the T_C conservation law are essentially point singularities in two dimensions (phase saddles and extrema can exist as points in three dimensions, where all three components of phase gradient vanish, but these do not appear to have any significant association with dislocations).

Notwithstanding this appearance of artificiality, the T_C conservation law is useful because it implies that phase saddles must be implicated in certain sorts of dislocation interaction. Consider for example the commonly occurring situation shown in figure 1, of a dislocation curved like a hairpin. As the height z of the surface Σ is increased, the two points where the dislocation intersects Σ collide and annihilate. Since each dislocation point has $T = +1$, the total T for the pair is +2. After the annihilation there are no dislocation points, that is the total T for dislocations is zero. By the conservation law, the collision must involve at least two saddles, whose contribution of -2 cancels that of the dislocations. This argument holds irrespective of the possible presence of phase extrema, because phase saddles are the only source of negative T.

Sometimes, the strength of a dislocation is called its topological charge[15-16]. This terminology is justified for the intersections of dislocation lines with a surface Σ, because the intersections are points analogous to electric charges in two dimensions. However, it is misleading because it can obscure essential features of the dislocation line in three dimensions. To see this, consider again the hairpin of figure 1. Along this curved dislocation, it is natural to choose **n** as illustrated, so that the strength is the same at all points (+1 in this case). But for the plane surface Σ it is natural to use a contrary convention, and define **n** at the two intersections so that both vectors point to the same side of Σ (e.g. to the 'front', with the larger value of z, as indicated by the dotted arrows). Now the two intersections appear to represent dislocations with different strengths (+1 and -1), and their collision as Σ is moved looks like the annihilation of opposite charges. Such a description obscures the fact that the collision - which simply occurs where the curved dislocation line touches one of the family of surfaces Σ - is an artefact of the choice of the family: for surfaces with a different orientation, the collision occurs at a different place.

In a frequently investigated special case, ψ represents a monochromatic wave with a well defined propagation direction (+z, say) - for example, that generated by a paraxial superposition of laser beams[17,18]. Then it is natural to consider the surfaces Σ as planes of constant z (the circuits C being regarded as at infinity), and think of increasing z as analogous to increasing 'time'. The stationary dislocation lines appear as 'moving' points on Σ, and this is one of the situations, referred to above, where the designation topological charge is legitimate - although the essential nature of dislocations as lines in three dimensions should never be forgotten.

In addition to the topological indices S and T, and the direction **n**, there are geometrical measures of dislocations. Three of these are the local wavevector (whose orientation with respect to **n** measures the edge-screw character of the dislocation), the local frequency (useful for non-monochromatic waves) and the phase of the dislocation[3].

As a final remark about scalar waves, it is worth noting that although there are no phase extrema in strictly two-dimensional monochromatic waves satisfying a wave equation[14], in paraxial three-dimensional waves these extrema can occur. For example, in a gaussian beam, the concave/convex wavefronts form phase maxima/minima in planes Σ before/after the focus.

In vector electromagnetic fields, several other sorts of line singularities exist, analogous to disclinations in liquid crystals. Their properties have been studied in detail[19-25]. From the general perspective I outlined earlier, these are singularities of the new physical property introduced in the generalization from scalar to vector waves, namely polarization.

ACKNOWLEDGEMENTS

I thank Professor M.S. Soskin for generous hospitality at the conference on Singular Optics, and Professors N.R. Heckenberg, J.F. Nye, M.S. Soskin, and M.V. Vasnetsov for illuminating discussions.

REFERENCES

1. Born, M. and Wolf, E., 1959, *Principles of Optics* (Pergamon, London).
2. Nye, J. F. and Berry, M. V., 1974, Dislocations in wave trains, *Proc. Roy. Soc. Lond.* **A336,** 165-90.
3. Nye, J. F., 1981, The motion and structure of dislocations in wavefronts, *Proc. Roy. Soc. Lond.* **A378,** 219-239.
4. Hirschfelder, J. O., Christoph, A. C. and Palke, W. E., 1974, Quantum mechanical streamlines. 1. Square potential barrier, *J. Chem. Phys.* **61,** 5435-5455.
5. Hirschfelder, J. O. and Tang, K. T., 1976, Quantum mechanical streamlines. III Idealized reactive atom-diatomic molecule collision, *J. Chem. Phys.* **64,** 760-785.
6. Hirschfelder, J. O. and Tang, K. T., 1976, Quantum mechanical streamlines. IV. Collision of two spheres with square potential wells or barriers, *J. Chem. Phys.* **65,** 470-486.
7. Berry, M. V. and Upstill, C., 1980, Catastrophe optics: morphologies of caustics and their diffraction patterns, *Progress in Optics* **18,** 257-346.
8. Berry, M. V., 1981, Singularities in Waves and rays, in *Les Houches Lecture Series Session 35*, eds. Balian, R., Kléman, M. & Poirier, J.-P. (North-Holland: Amsterdam), pp. 453-543.
9. Nussenzveig, H. M., 1992, *Diffraction effects in semiclassical scattering* (University Press, Cambridge).
10. Park, D., 1997, *The fire within the eye: a historical essay on the nature and meaning of light* (University Press, Princeton).
11. Berry, M. V., 1997, Dislocation reactions in nonparaxial gaussian beams, *J. Mod. Opt.*, (in the press).
12. Karman, G. P., Beijersbergen, M. W., van Duijl, A. and Woerdman, J. P., 1997, Creation and annihilation of phase singularities in a focal field, *Optics Letters* **19,** 1503-1505.
13. Nye, J. F., 1997, Unfolding of higher-order wave dislocations, *J. Opt. Soc. Amer. A.*, submitted.
14. Nye, J. F., Hajnal, J. V. and Hannay, J. H., 1988, Phase saddles and dislocations in two-dimensional waves such as the tides, *Proc. Roy. Soc. Lond.* **A417,** 7-20.
15. Soskin, M. S., Gorshkov, V. N., Vasnetsov, M. V., Malos, J. T. and Heckenberg, N. R., 1997, Topological charge and angular momentum of light beams carrying optical vortices, *Phys. Rev. A* **56,** 4064-4075.
16. Beijersbergen, M., 1996, Phase singularities in optical beams, *in Huygens Laboratory* (Leiden), p. 98.
17. Basistiy, I. V., Bazhenov, V. Y., Soskin, M. S. and Vasnetsov, M. V., 1993, Optics of light beams with screw dislocations, *Opt. Comm.* **103,** 422-428.
18. Basistiy, I. V., Soskin, M. S., and Vasnetsov, M. V., 1995, Optical wavefront dislocations and their properties, *Opt. Comm.* **119,** 604-612.
19. Nye, J. F., 1983, Polarization effects in the diffraction of electromagnetic waves: the role of disclinations, *Proc. Roy. Soc. Lond.* **A387,** 105-132.
20. Nye, J. F., 1983, Lines of circular polarization in electromagnetic wave fields, *Proc. Roy. Soc. Lond.* **A389,** 279-290.
21. Nye, J. F. and Hajnal, J. V., 1987, The wave structure of monochromatic electromagnetic radiation, *Proc. Roy. Soc. Lond.* **A409,** 21-36.
22. Nye, J. F., 1997, Singularities in wave fields, *Phil. Trans. Roy. Soc. Lond.* **355,** 2065-2069.
23. Hajnal, J. V., 1987, Singularities of the transverse fields of electromagnetic waves. I. Theory, *Proc. Roy. Soc. Lond.* **A414,** 433-446.
24. Hajnal, J. V., 1987, Singularities in the transverse fields of electromagnetic waves. II. Observations on the electric field., *Proc. Roy. Soc. Lond.* **A414,** 447-468.
25. Hajnal, J. V., 1990, Observation of singularities in the electric and magnetic fields of freely propagating microwaves., *Proc. Roy. Soc. Lond.* **A430,** 413-421.

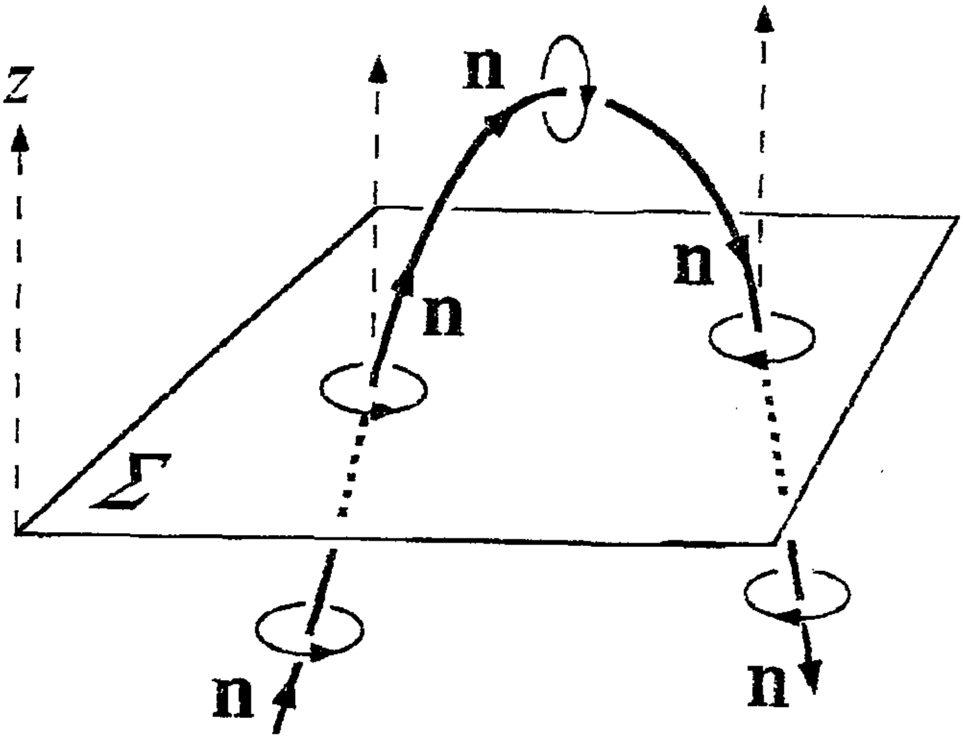

Figure 1. Curved dislocation line piercing a surface Σ. Loops show the sense in which phase increases, and **n** labels the choice of unit vector, continuous along the dislocation.