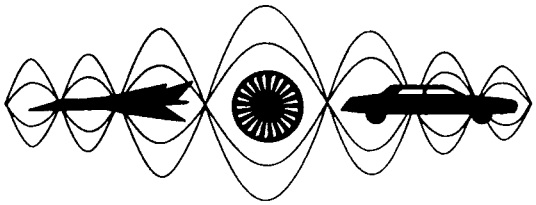

FIFTH INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
DECEMBER 15-18, 1997
ADELAIDE, SOUTH AUSTRALIA

Distinguished Keynote Paper

A CENTURY OF SHOCK WAVE DYNAMICS

Sir James Lighthill
(Department of Mathematics, University College London,
Gower St., London WC1E 6BT, UK.)

Abstract

The recent three-volume history 'Twentieth Century Physics' [1] includes my 118 page 'Chapter 10. Fluid Dynamics' about a field where the 20th century's first decade saw big breakthroughs in the analysis of nonlinear problems for which the physicist's standard perturbation methods break down – and which would later be named singular perturbation problems. Besides the 1904 elucidation of boundary-layer structure by Prandtl, these included the 1910 elucidation of shock-wave structure in complementary investigations by Rayleigh and by Taylor. Subsequent advances in shock wave dynamics brought crucial new discoveries on the structure and propagation of weak shocks, and also on shock-wave/boundary-layer interaction, both with major aeronautical implications; along with still harder investigations into the structure and propagation of strong shock waves such as appear in explosions and implosions and also around spacecraft re-entering the earth's atmosphere. It may, perhaps, be worth noting that questions which remain relatively simple for weak waves, like the nature of reflexion and diffraction by a solid body, raise formidable and intriguing difficulties for strong shock waves.

In this introductory paper looking back over a century of shock wave dynamics, I highlight (i) key analytical approaches for both weak and strong shock waves, (ii) beautiful and effective optical methods for use in wind tunnels and shock tubes and (iii) powerful techniques for accurate shock capturing in computational fluid dynamics.

1. INTRODUCTION

A century ago, students of Acoustics and Vibration depended above all on Rayleigh's superb treatise 'The Theory of Sound' which had just appeared in an expanded second edition [2]. Yet in section 253 of this great work Rayleigh brilliantly showed how consideration of just the simplest problem in the nonlinear theory of sound posed an enigma which it was quite impossible to resolve with the knowledge then available. Existing suggestions for a resolution, incorporating those discontinuous waves or shock waves which Mach [3] had photographed around bullets in flight, appeared on Rayleigh's analysis to contradict fundamental physical principles; in short, the 'century of shock wave dynamics' had not yet begun.

In surveying the dynamics (rather than the whole physics) of shock waves, space may be saved by concentrating on waves propagated through a perfect gas with constant specific heats in a ratio γ; for which scientists since Laplace [4] had appreciated how, in any sound wave, those changes of pressure p and density ρ whose ratio is the square of the sound speed c must satisfy the adiabatic relationship

$$\frac{p}{p_0} = \left(\frac{\rho}{\rho_0}\right)^{\gamma} \text{ giving } c^2 = \frac{dp}{d\rho} = c_0^2 \left(\frac{\rho}{\rho_0}\right)^{\gamma-1} \text{ with } c_0^2 = \frac{\gamma p_0}{\rho_0}. \tag{1}$$

(Here, subscript zero denotes undisturbed values.) On linear theory sound is propagated at speed c_0, whereas on nonlinear theory higher pressures travel at an increased sound speed c, amounting on a first approximation to

$$c = c_0 + \frac{\gamma-1}{2}\left(\frac{p-p_0}{\rho_0 c_0}\right) = c_0 + \frac{\gamma-1}{2}u; \tag{2}$$

where u, the gas velocity in the direction of propagation, assumes on a linear approximation the familiar bracketed form. But this propagation speed (2), relative to a fluid which itself moves at velocity u, implies an absolute velocity

$$c + u = c_0 + \frac{\gamma+1}{2}u \tag{3}$$

of wave travel. (For air, with $\gamma = 1.4$, the excess wave speed is $1.2u$, out of which just one-sixth arises from the increase (2) in c while five-sixths is due to convection of sound at the air velocity u.)

Although expression (3) for the wave speed is introduced above by crude approximate arguments, Riemann's subtle mathematical analysis of 1859 had already proved it to be absolutely accurate [5] for plane sound waves of any amplitude propagated one-dimensionally into undisturbed fluid under adiabatic conditions. Briefly, the relationship $c = c_0 + \frac{1}{2}(\gamma-1)u$ is exact; the expressions for pressure and density may be derived from c by equations (1); and, most important of all, each value of u is propagated at precisely the speed (3).

These conclusions were well known to Rayleigh, who recognised also their sensational implications. Figure 1 shows these in the case of a single pulse of positive excess pressure,

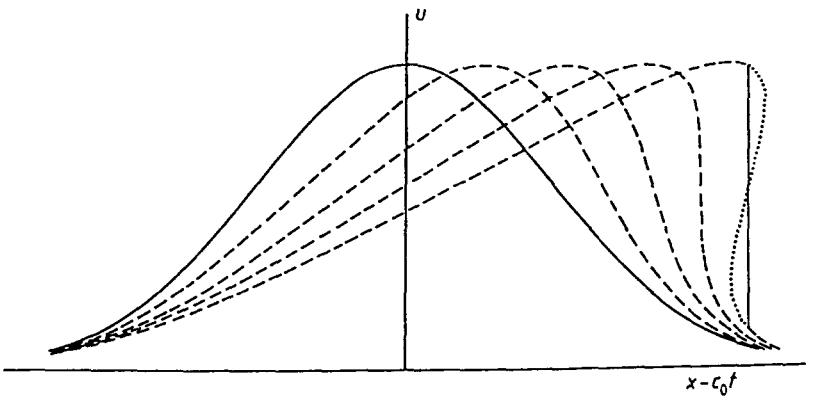

Figure 1. The enigma of nonlinear wave steepening.

represented as an initial graph (solid line) of fluid velocity u against distance. On a linear theory of one-dimensional sound waves, each value of u would be propagated at speed c_0, so that the shape of the pulse would remain unchanged when plotted (as here) against $x - c_0 t$. On the exact nonlinear theory, however, each value of u is propagated at speed $c_0 + \frac{1}{2}(\gamma+1)u$; accordingly, after time t, that value when plotted against $x - c_0 t$ as in Figure 1 has been shifted a distance $\frac{1}{2}(\gamma+1)ut$ to the right. Small values of u have hardly moved at all, while large values have moved much more – allowing them, remarkably, to 'catch up' with smaller values.

These distorted pulse shapes are shown in Figure 1 (broken lines) for a sequence of values of t until a time has been reached when the pulse shape has a vertical tangent. Pulse shapes at still later times continue to be predicted by the theory; however, parts of such a shape are shown in Figure 1 as dotted lines in recognition of the clear impossibility of the fluid velocity u taking three different values at one and the same point!

A tempting idea for resolving this enigma is to suppose that the solution develops a discontinuity. Riemann himself noticed that a discontinuity – indicated in Figure 1 by a vertical solid line – could be inserted in place of the dotted line in such a way that total mass and momentum are conserved, while all continuous parts of the curve (broken lines) still satisfy exact equations for sound propagation under adiabatic conditions. Rayleigh, however, objected that this idea not only (i) left unexplained how a discontinuity would arise but also, still more seriously, (ii) failed to satisfy overall conservation of energy (indeed, the energy content in dotted parts of the waveform is lost).

Early in the 20th century, however, the enigma was resolved in two independent studies, published by Rayleigh himself [6] and by the young G.I. Taylor [7] in a single 1910 number of the Proceedings of the Royal Society. Both studies showed how waveform steepening (see Figure 1) produces in the midst of the fluid a region with gradients large enough for diffusion of heat and momentum to become important. It is because the effects of diffusion (being proportional to gradients) can, if the region of rapid change is thin enough, attain any required level that they are able to cancel out those effects of excess propagation speed which tend to produce the unrealistic 'overturning' of the waveform represented by the dotted curve in Figure 1. Instead, there is formed something close to the discontinuous solution shown

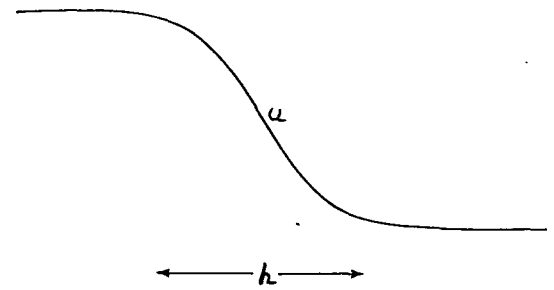

Figure 2. Distribution of velocity u inside a shock wave of Taylor thickness h.

as a solid vertical line; where, however, the discontinuity possesses a definite thickness h ('the Taylor thickness') and an associated internal structure (Figure 2) which allow those 'overturning' and 'diffusion' effects to be in precise balance.

While countering Rayleigh's objection (i) by explaining how this nearly discontinuous wave (the shock wave) arises, the above indication of important diffusion effects within it also demolishes the idea that the process is adiabatic. On the contrary, the fluid traversed by a shock wave experiences (from conduction of heat and from viscous dissipation) a departure from adiabatic conditions; which, in accordance with the second law of thermodynamics, involves an increase in entropy (so that p/p_0 rises above the value given by (1) for given ρ/ρ_0). This counters objection (ii) because the shock waves achieves energy conservation by abandoning entropy conservation – instead of the other way round.

For practical engineering purposes, of course, it often suffices to treat shock waves as sharp discontinuities satisfying conservation equations for mass, momentum and energy. These equations had actually been written down in 1889 – for fluids in general – by Hugoniot [8] who, on the other hand, had not been in a position to appreciate why, in most fluids including air, the physics of shock wave formation permits only the appearance of compressive discontinuities with a (calculable) entropy increase across them.

This survey continues with Section 2, on the dynamics of weak shock waves, which shows how closely the original Riemann idea can follow their behaviour provided it is recognized that the waveform's loss of mechanical energy is a 'hidden' loss, taking the form of that dissipation inside the shock wave whose magnitude is reflected in the entropy change. Then Whitham's extensions of the theory beyond the case of plane waves are described; while, in Section 3, some aeronautical applications are indicated.

By contrast, the dynamics of strong shock waves (Section 4) calls for radically different approaches because specific entropy, although conserved for any fluid particle behind a shock wave, varies steeply with the shock wave's strength at the time when that particle crossed it. These new approaches, making fruitful use of self-similar solutions, were again pioneered by G.I. Taylor.

Yet, despite all the theoretical advances, shock wave dynamics has been able to make progress only through an intimate cooperation between gifted experimentalists, making use

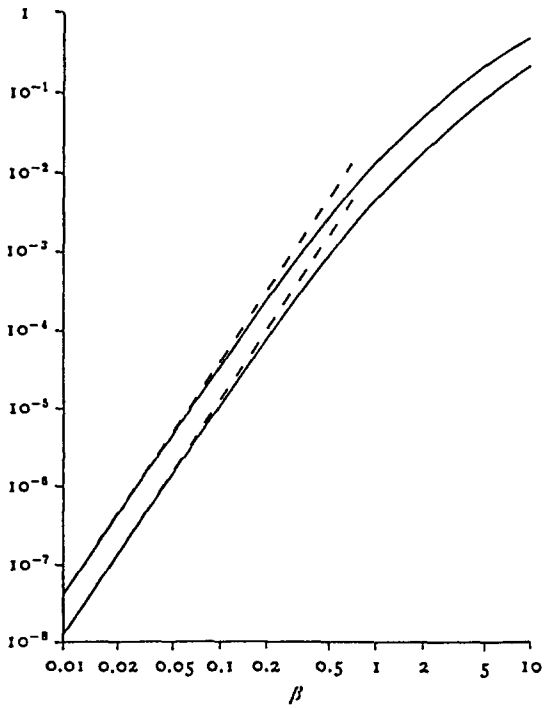

Figure 3. Changes in $\ln(p/\rho^\gamma)$ and $\ln\left[c - \frac{1}{2}(\gamma-1)u\right]$ across a shock wave of strength β.

of refined optical techniques in wind tunnels and shock tubes, with experts in applied mathematical analysis and, more recently, in computational fluid dynamics (CFD). Section 5 outlines accordingly the shadowgraph, schlieren and interferometric techniques that have been the main sources of experimental data, and then concludes by showing how CFD schemes which discretize equations of motion in ways that maintain the precision of overall conservation laws can yield valuable descriptions of shock wave dynamics.

2. DYNAMICS OF WEAK SHOCK WAVES

If fluid pressure is abruptly raised by a factor $(1+\beta)$ on the passage of a shock wave, its 'strength' may be defined as the proportional increase β. Commonly, a shock wave with $\beta < 0.5$ is called 'weak' [9].

Such weak shock waves disturb only slightly the accuracy of the adiabatic relationship (1), since the change in $\ln(p/\rho^\gamma)$ across the shock wave (proportional to entropy change) remains less than 0.0055 for any $\gamma > 1$; while for air, with $\gamma = 1.4$, it is at most 0.0027. The upper curve in Figure 3 is a log-log plot of this change for air against the strength β, which confirms the cubic limiting form as $\beta \to 0$,

$$\Delta\left[\ln(p/\rho^\gamma)\right] \sim (\gamma^2-1)\beta^3/12\gamma^2 \ (= 0.041\beta^3 \text{ for air}), \tag{4}$$

although showing that when $\beta = 0.5$ the actual change, 0.0027, is only about half as much as is implied by the limiting expression (4). These very low values suggest that errors in using the constant-entropy relationship (1) to study gas motions which incorporate weak shock waves may be extremely small, even though the entropy change across a shock wave results from dissipation processes within it that are essential (Section 1) to its very existence. Such a suggestion, for a plane wave entering still air, would imply rather good accuracy for

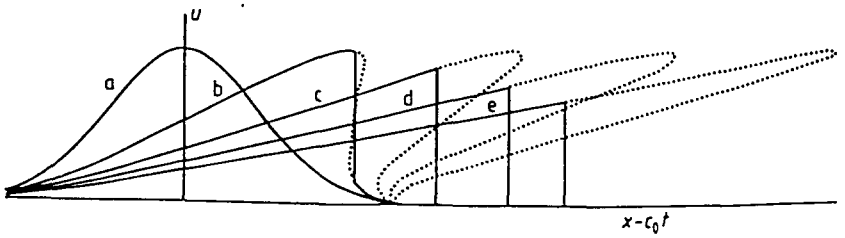

Figure 4. Steepening continued (a and b are first and last waveforms in Figure 1, while c, d, e are later waveforms).

the relationship (2) between c and u which Riemann derived exactly on the assumption of constant entropy; and the lower curve in Figure 3 confirms, indeed, that the difference $c - \frac{1}{2}(\gamma - 1)u$ changes across a weak shock wave by at most a factor 1.0009.

Where shock waves are weak, then, all continuous parts of the waveform obey closely the propagation law (3) used to construct Figure 1; namely, that each value of u, when plotted against $x - c_0 t$, has after time t been shifted a distance $\frac{1}{2}(\gamma+1)ut$ to the right. That construction is continued in Figure 4 to still later times t with, at each stage, the necessary discontinuity placed so that overall mass conservation is retained. This condition, the 'equal area' law of Whitham [10], requires that the discontinuity leaves unaltered the area under the curve (strictly, area under the corresponding curve of density ρ against distance — although a close-to-linear relationship of density variation to u allows the law to be applied, with quite good approximation, to the graph of u itself).

Briefly, then, sound propagation outside any weak shock wave still satisfies constant-entropy laws because entropy inhomogeneities behind it have negligible effects on propagation. By contrast, the wave's overall mechanical energy is progressively diminished (Figure 4) by the accumulated effects of energy losses 'hidden' inside the apparent discontinuity.

Not only energy, but also information, disappears in this process! — as Figure 4 also demonstrates. Absolutely all information about the original shape of the compression pulse, except about the area Q under the curve, has disappeared in the later stages of this process, when the waveform has become a right-angled triangle of area Q with a hypotenuse of slope $\left[\frac{1}{2}(\gamma+1)t\right]^{-1}$ (the reciprocal slope $\delta x/\delta u$ increases at a rate $\frac{1}{2}(\gamma+1)$, because a value $u+\delta u$ propagates faster than a value u by a signal-speed excess of $\frac{1}{2}(\gamma+1)\delta u$). Such a triangular waveform has a height $[4Q/(\gamma+1)t]^{1/2}$ — the jump in u at the shock wave — and a length $[(\gamma+1)Qt]^{1/2}$. Moreover the energy of the wave, proportional to length and to amplitude squared, diminishes as $t^{-1/2}$, while its rate of decrease, proportional to $t^{-3/2}$, is found to agree exactly with the rate of energy loss 'hidden' inside the discontinuity as implied by the cubic form (4) of dependence of entropy change on shock wave strength.

The corresponding conclusion for an initial waveform including negative as well as positive values of u is that it develops asymptotically into an 'N-wave' (Figure 5) which consists of two right-angled triangles, with areas Q_+ above the $u = 0$ axis and Q_- below it and with aligned hypotenuses of slope $\left[\frac{1}{2}(\gamma+1)t\right]^{-1}$. The jumps in u at the shock waves are $[4Q_+/(\gamma+1)t]^{1/2}$ and $[4Q_-/(\gamma+1)t]^{1/2}$ and the overall length is $[(\gamma+1)Q_+t]^{1/2} + [(\gamma+1)Q_-t]^{1/2}$; while, yet

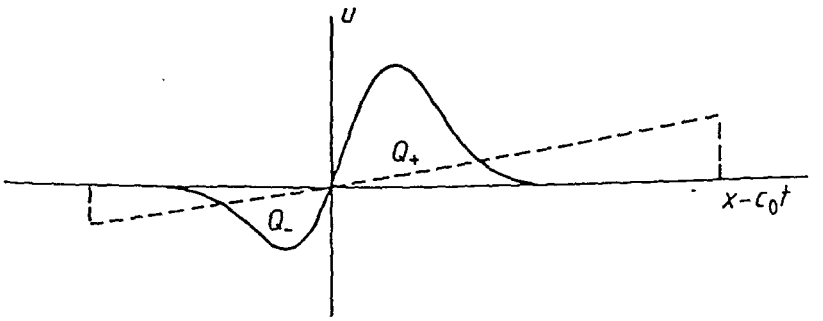

Figure 5. Case of an initial waveform (solid line) with positive lobe of area Q_+ and negative lobe of area Q_-. Broken line: subsequently developing N-wave with unchanged area for positive and negative lobes.

again, mechanical energy diminishes as $t^{-1/2}$ with all energy dissipation 'hidden' inside the two shock waves.

It is not just in plane waves that such behaviour is found; as may be illustrated (again following Whitham) by analysing [11] the 'exploding wire' phenomenon that results from instantaneous discharge of a condenser through a very thin wire. A cylindrically expanding sound wave is generated in the surrounding air by the wire's sudden vaporization; however, an interesting difference from the plane-wave case emerges already from classical linear analysis: cylindrical wave propagation, on such an analysis, converts the sound source (the outward-pushing vapour) into a wave with both positive and negative values of the outward velocity u. Thus, at radial distances r beyond a certain value r_1,

$$u = \left(\frac{r_1}{r}\right)^{1/2} u_1\left(t - c_0^{-1}r\right), \tag{5}$$

where the waveform $u_1(t)$ comprises a positive phase followed immediately by a negative phase of identical area Q (making $\int u_1(t)dt = 0$, and so reconciling the necessary r^{-1} dependence of energy flux to the impossibility of any indefinite growth with r in the total outward displacement $2\pi r \int u dt$).

This means that, on nonlinear theory, an N-wave (and indeed, a 'balanced' N-wave with equal areas $Q_+ = Q_- = Q$) is necessarily formed. Whitham's analysis showed how a value $u_1 = (r/r_1)^{1/2}u$ is again propagated (see (3) above) at a signal speed $dr/dt = c_0 + \frac{1}{2}(\gamma+1)u$ which replaces the simple c_0 value of equation (5). To a close approximation, this gives

$$\frac{dt}{dr} = c_0^{-1} - \tfrac{1}{2}(\gamma+1)\left(\frac{r_1}{r}\right)^{1/2} u_1 c_0^{-2}, \tag{6}$$

which can be integrated as

$$t - c_0^{-1}r + \tfrac{1}{4}(\gamma+1)(r_1 r)^{1/2}u_1 c_0^{-2} = \text{ constant}, \tag{7}$$

correcting the time t when a given value of u_1 is found. Nonlinear theory replaces, then, the argument of u_1 in equation (5) by the left-hand side of equation (7).

Thus the temporal waveform (graph of u_1 as a function of t) is now sheared (backwards) in such a way that the reciprocal slope $\delta t/\delta u_1$ takes an asymptotic value

$$-\tfrac{1}{4}(\gamma+1)(r_1 r)^{1/2}c_0^{-2}. \tag{8}$$

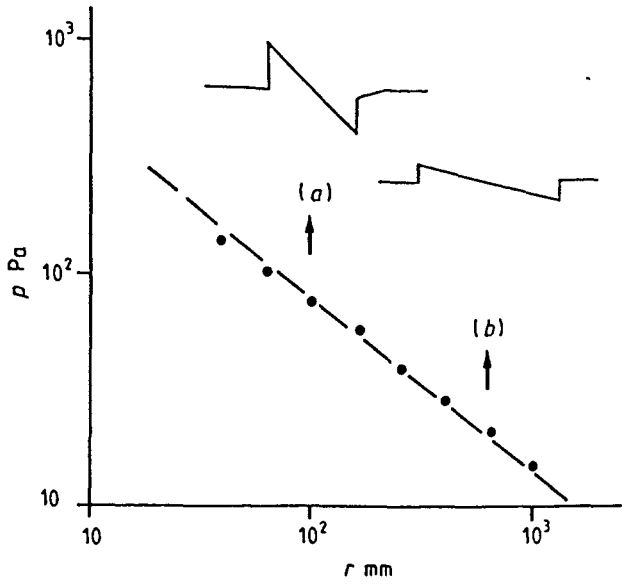

Figure 6. Illustrating how an exploding wire generates pressure-time curves of N-wave shape, (a) and (b) at different radial distances r, and comparing the pressure rise at the initial shock wave with an $r^{-3/4}$ law (broken line).

The ultimate N-wave therefore takes the form of two right-angled triangles of area Q whose aligned hypotenuses have the reciprocal slope (8). Each triangle has a height (discontinuity in u_1) equal to

$$\left[\frac{8Qc_0^2}{(\gamma+1)(r_1 r)^{1/2}}\right]^{1/2} \tag{9}$$

which varies as $r^{-1/4}$; but, since $u = (r_1/r)^{1/2}u_1$, this yields Whitham's remarkable prediction of an inverse-three-quarters-power dependence on r for the strengths of both shock waves — which has been accurately verified in experiments on exploding wires (Figure 6). The time interval separating the shock waves takes a value

$$\left[2(\gamma+1)Q(r_1 r)^{1/2}c_0^{-2}\right]^{1/2} \tag{10}$$

which increases as the fourth root $r^{1/4}$ of the distance travelled, while the wave energy decays as $r^{-1/4}$; with all the energy dissipation hidden, once more, inside the shock waves.

It is of course to be understood that the treatment of shock waves as discontinuities throughout Section 2 can be seen as justifiable because each shock wave has a Taylor thickness (Figure 2) very small indeed compared with all other relevant distances. As with singular perturbation problems in general, the true behaviour of u is obtained by matching the outer solution involving a discontinuity with the inner solution depicted in figure 2.

Exclusively in the case of plane waves is it possible, as Hopf [12] first showed, to obtain a uniformly valid approximation to u from the asymptotics of solutions of the Burgers equation. Such an approach is especially valuable [13] for studying shock waves of extremely low strength; and for investigating flow details where, for example, two shock waves unite to form a single stronger shock wave.

3. AERONAUTICAL SHOCK WAVES

Interactions between waves and flows are evident in all the studies outlined above, from equation (3) onwards. Such interactions become even closer, however, in a supersonic flow; that is, one with speed U which exceeds the sound speed c_0 in the undisturbed flow. In an aeronautical application, for example, this flow speed U means the speed of the air relative to a moving aircraft. Then, not only do 'waves and flows interact', but the flow may consist almost entirely of waves whose propagation has been annulled (that is, brought to rest) by the flow.

In particular, propagation at an angle Θ to the wind of very weak sound waves at the undisturbed sound speed c_0 can be annulled by the opposing component $U\cos\Theta$ of airflow if

$$U\cos\Theta = c_0. \tag{11}$$

Provided that $U > c_0$, an angle Θ satisfying (11) exists, and such waves can be a permanent feature of the flow.

For waves that are not so weak [14], the true signal velocity (3) replaces c_0 in equation (11), so that Θ is reduced for positive u (or positive excess pressure) and increased for negative u. Moreover, if a shock wave appears, its own speed of propagation replaces c_0 in (11).

Extra aerodynamic drag is associated with the 'hidden' energy loss arising in any shock waves, such lost energy needing to be restored from additional work done by thrust to overcome this component of drag. So 'aeronautical shock waves' need to be kept as weak as possible.

The above principles may first be illustrated by the two-dimensional flow around a supersonic aerofoil (Figure 7). Pioneers of supersonic aerodynamics considered using straight wings, with spans at right angles to the oncoming flow, and with an 'aerofoil' cross-section uniform along the span, and yet appreciated that aerofoils suitable for supersonic speeds needed to be quite different from low-speed wing sections. They must be thinner, and must also have a sharp leading edge, so that disturbances to be oncoming supersonic stream remain small and the shock waves generated are weak. Then the whole visible flow pattern takes the form of stationary waves propagating at angles Θ satisfying either equation (11) or similar equations with modified right-hand sides.

Each wave arises from a disturbance to the incident airflow by the aerofoil surface. The point A where such disturbance is zero ($u = 0$) emits a wave in accordance with equation (11). But the direction of the surface at B is associated, as Figure 7(b) shows, with propagation of a positive value of u at an increased speed $c_0 + 1.2u$ giving a reduction in Θ; while the direction at C is shown in Figure 7(c) to be compatible with $u < 0$ and an increased value of Θ. The excess pressures, positive at B and negative at C, produce a resultant drag, related rather precisely to the hidden energy loss in the N-waves (compare Figures 7(d) and 5) generated by such nonlinear propagation. But practically no waves arise in the region behind the rear shock wave, where the incident airflow is undisturbed.

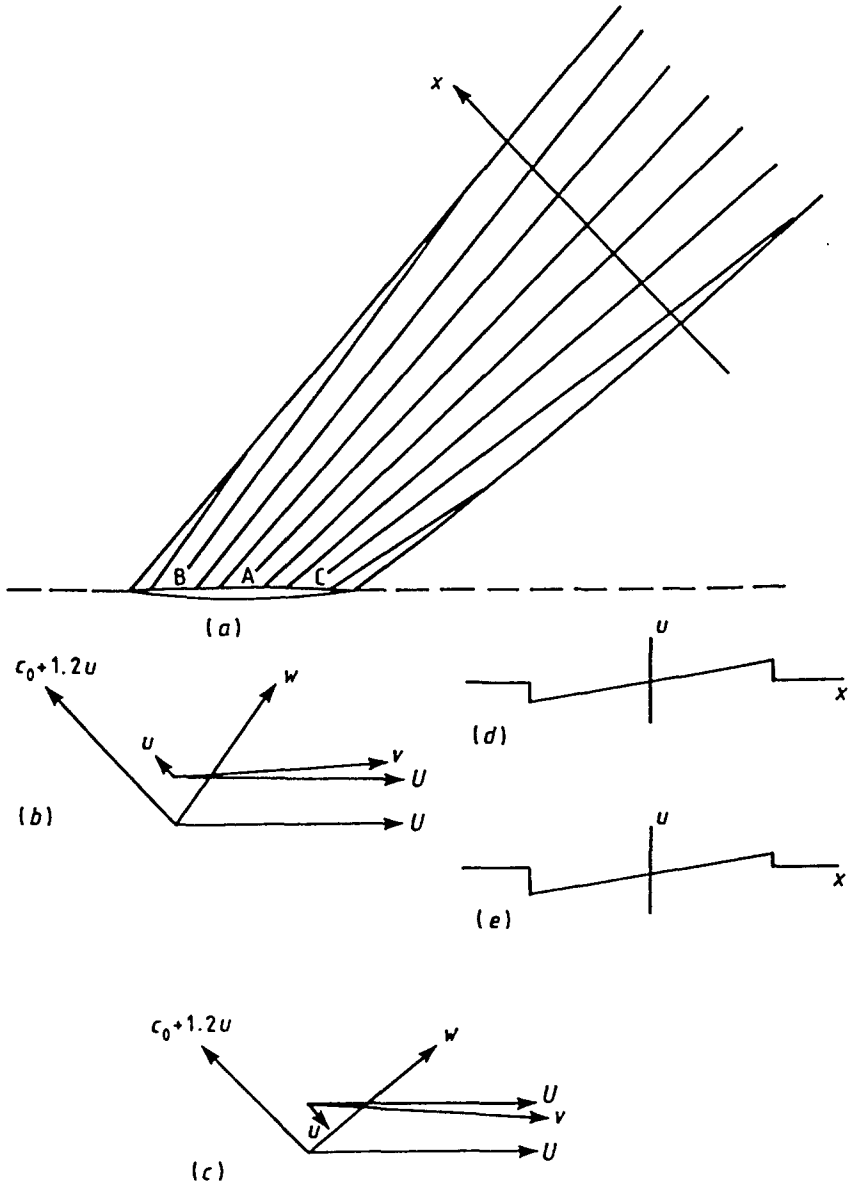

Figure 7. (a) Stationary waves generated in supersonic flow (symmetrical about the broken line) about a thin aerofoil. (b) At B, the wind velocity U and a velocity u carried by the waves have a resultant v that slopes upwards (along the tangent to the aerofoil surface). The stationary wave is in the direction w (resultant of U and the signal velocity). (c) At C, the negative value of u causes the resultant v to slope downwards (again along the tangent) and the stationary wave now has a reduced angle to the wind. (d) Form of the balanced N-wave at the section shown. (e) At positive angle of attack, this becomes an unbalanced N-wave, with downward momentum.

The symmetrical flow of Figure 7(a) changes, at a positive angle of attack, into a flow that gives lift; briefly, because each value of u is decreased on the upper surface while values on the lower surface are increased — so that all pressures on the lower surface exceed the corresponding upper-surface pressures. Needless to say, an equal and opposite force is exerted by the aerofoil, conferring downward momentum on the fluid. Yet this momentum appears, not in any wake region as in low-speed flow, but almost exclusively in the flow between the shock waves! — where the balanced N-waves of Figure 7(d) become unbalanced, as Figure 7(e) shows, and include a downward component of momentum (see also Figure 8).

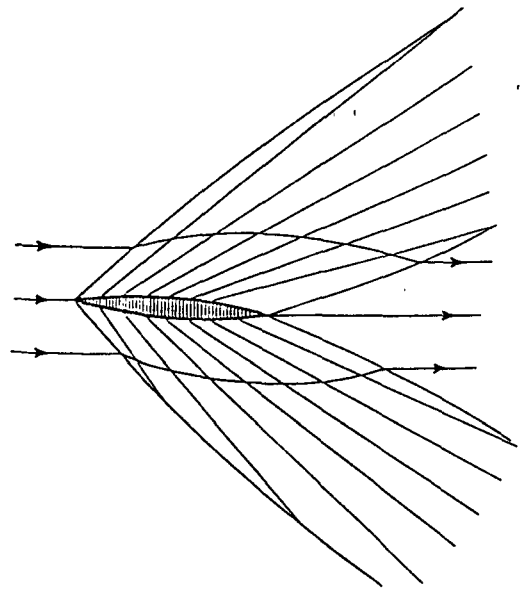

Figure 8. The complete flow field [14] around a lifting aerofoil (actually, one twice as thick as that in Figure 7).

Nonetheless, designers of supersonic aircraft in the twentieth century were increasingly drawn away from wings with such aerofoil sections by hard facts about aeronautical shock waves. The aim of keeping these weak, so as to reduce shock-wave drag (and, also, the supersonic boom heard at ground level), is achieved best, not by shapes that are thin in just one dimension (Figure 7), but by shapes that are slender in both dimensions perpendicular to the flow. Accordingly, the rather simple concepts of supersonic aerofoil theory (Figure 7) required careful extension to involve three-dimensional wave propagation if essential features of the 'slender-body aerodynamics' needed for $M > 1$ were to be usefully exploited.

A first suggestion of what may be called for is offered by the analysis (Section 2) of shock waves produced in air by an 'exploding wire'. Here, the source of sound is the appearance in the fluid of foreign matter, the outward-pushing vapour, whose cross-sectional area S increases very suddenly with time; generating, on a linear analysis, the air motions (5), which are converted by nonlinear effects into an N-wave. Flight at supersonic speeds through a mass of air similarly introduces foreign matter (the aircraft), whose overall cross-section S in contact with that air mass increases very suddenly with time — though it then falls abruptly to zero. For a slender shape at supersonic speeds, this suggests the possible importance of a function $S(x)$, describing variation in the wind-direction x of the body's cross-sectional area S normal to that direction.

Slender-body aerodynamics showed [14] how shapes where the function $S(x)$ varies rather smoothly and gradually along the aircraft's length have two advantages: (i) shock-wave strengths are kept relatively low; and (ii) they are quite well estimated, as for an exploding wire, in a two-stage analysis — linear and then nonlinear — with an acoustic source strength depending on the total area $S(x)$ of each cross-section rather than on its detailed shape. Of course the waves spread, not cylindrically but conically at the angle Θ defined by equation (11); as the linear part of the analysis makes clear already because equation (11) represents a 'stationary phase' condition.

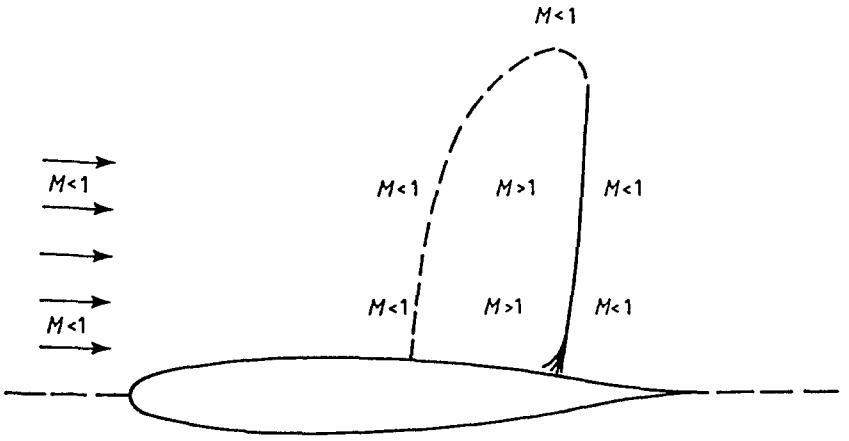

Figure 9. Illustrating a flow (symmetrical about the broken line) around an aerofoil at high subsonic Mach number, when there appears in the flow a limited supersonic region terminated by a 'discontinuity' that incorporates a shock-wave/boundary-layer interaction.

This initial, linear analysis can be further improved if the nondirectional simple-source radiation is modified [15] by adding a distribution of downward-pointing acoustic dipoles of strength $L(x)$, representing forces on the fluid equal and opposite to the lift forces on aircraft cross-sections. This improvement, after the necessary nonlinear adaptation, produces modest increases of N-wave strength below, and decreases above, the aircraft.

During the 'century of shock-wave dynamics' a significant first step in extending air transportation to supersonic speeds was made with Concorde [16], which offers busy passengers the benefit of journey times of only 3 hours over ranges of around 6000 km. In spite of substantial shock-wave drag, its slender shape achieves at $U/c_0 = 2$ a lift-drag ratio of around 10, and the shock waves produced below it are found, after propagation into the much denser air at ground level, to involve pressure jumps of only about a millibar.

Even the passage over a wing of a subsonic airstream may accelerate it locally to supersonic speeds and allow shock-wave formation [17]. Figure 9 illustrates this for symmetrical flow over a typical subsonic-speed aerofoil section, showing how a supersonic region can form downstream of the point of maximum thickness, with readjustment to subsonic flow possible only at a discontinuity. (For a lifting aerofoil, it is especially on the upper surface that such a discontinuity tends to form.) Moreover, this discontinuity involves not just a simple shock wave but a complex 'shock-wave/boundary-layer interaction', including upstream influence [18-21] of the shock wave on the boundary layer (that is, an influence extending much farther upstream than might have been expected).

In the design of transport aircraft it is important to avoid such effects, above all by the use of sweptback wings. These allow aeronautical shock waves to be eliminated up to quite high subsonic speeds ($U/c_0 = 0.85$); essentially, because just the component of airflow perpendicular to the wing is accelerated by passage over it, while the parallel component remains unaltered.

Other concepts valuable in design for Mach numbers $M = U/c_0$ close to 1 include Whitcomb's famous 'transonic area rule'[22]. Briefly, an aircraft for which $S(x)$, the distribution

of cross-section area defined above, varies smoothly and gradually with x will experience aerodynamic forces which, to an even greater extent at transonic than at supersonic speeds, are determined almost entirely by that distribution. For such a determination, however, the earlier two-stage analysis — first linear and then nonlinear — needs at transonic speeds to be replaced by a single-stage, fully nonlinear analysis.

But, beyond any such conceptual aids to aerodynamic design, it is mainly on advanced computational fluid dynamics (CFD) that modern designers rely for detailed analysis of flows over an aircraft configuration in complicated conditions like those of transonic flight. Section 5 outlines techniques which ensure that modern CFD programs [23] can closely predict locations and strengths of shock waves appearing where continuous flow has become impossible.

4. DYNAMICS OF STRONG SHOCK WAVES

Propagation laws behind a shock wave of large, yet variable, strength cannot correctly be simplified by assuming closely uniform entropy; nevertheless, a fluid particle's entropy per unit mass stays at the level — different for each particle — to which passage through the shock wave raised it. Thus equation (1) is replaced by an equation

$$\frac{\partial}{\partial t}\left(\frac{p}{\rho^\gamma}\right) + \mathbf{u}\cdot\nabla\left(\frac{p}{\rho^\gamma}\right) = 0, \tag{12}$$

stating that the ratio p/ρ^γ (with logarithm proportional to entropy) is unchanging for any particle with the fluid velocity $\mathbf{u}$.

Formidable difficulties oppose attempts to solve equation (12) together with the ordinary Euler equations of continuity and momentum for a compressible fluid. Yet, from 1933 onwards, important successes in overcoming those difficulties were won by G.I. Taylor: the pioneer earlier responsible, with Lord Rayleigh, for resolving the enigma described in Section 1. These successes came from the computation of 'self-similar' solutions to the equations described above; that is, solutions where nondimensional forms of p, ρ and $\mathbf{u}$ are just functions of a single nondimensional quantity ζ formed from the space and time coordinates — and, therefore, can be computed by solving ordinary differential equations with respect to ζ.

Actually, in Taylor's first such study, made jointly with J.W. Maccoll, no time coordinate arises since all the waves remain stationary in a steady supersonic stream, as in Section 3 above. In this study [24] of axisymmetrical flow past a cone, ζ may be taken as a spherical polar coordinate ψ, and the flow equations are solved in the interval $\psi_c < \psi < \psi_s$ where ψ_c and ψ_s are the semi-angles of the solid cone and of the attached conical shock wave, respectively.

Such a solution can be found [25] for all cone angles ψ_c up to a certain maximum value $(\psi_c)_{\max}$, which is plotted against the Mach number M of the undisturbed stream in Figure 10. For this maximum cone angle, ψ_s takes a critical value $(\psi_s)_{\text{crit}}$, also shown, which can be defined as that angle ψ_s between a stationary shock wave and the oncoming supersonic stream for which streamlines are deflected (abruptly) through the greatest possible angle $\delta_{\max}$. Such streamlines, on the other hand, proceed to curve away from the cone after

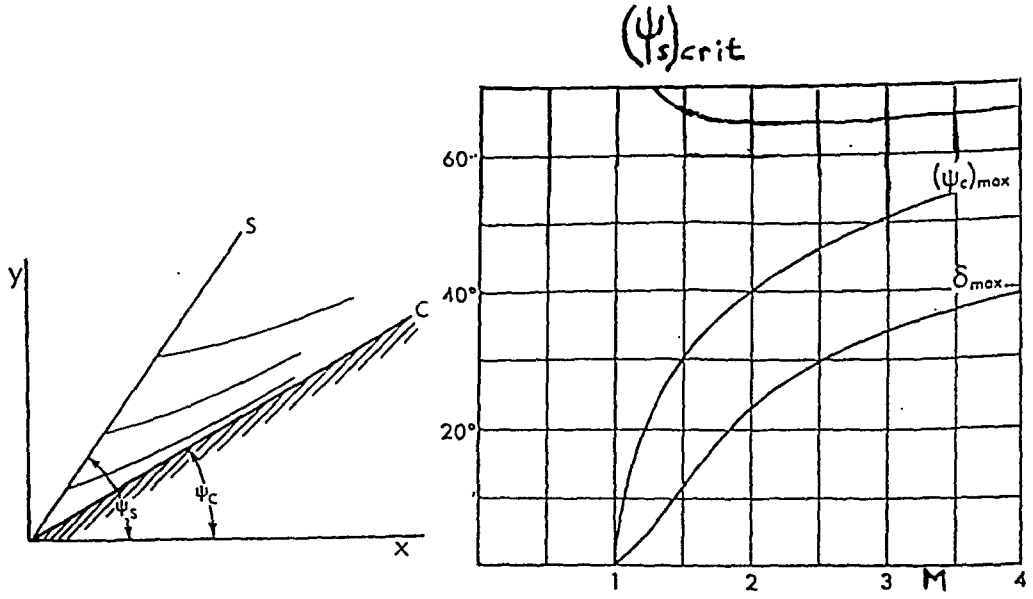

Figure 10. Defining the cone angle ψ_c and the shock wave angle ψ_s , and plotting $(\psi_s)_{\text{crit}}, (\psi_c)_{\max}$ and $\delta_{\max}$ against M.

crossing the shock wave, which is why this maximum deflection $\delta_{\max}$ is less (Figure 10) than the corresponding cone angle $(\psi_c)_{\max}$.

A surprising property of the equations is that they possess two solutions for each $\psi_c < (\psi_c)_{\max}$: one with $\psi_s < (\psi_s)_{\text{crit}}$ and one with $\psi_s > (\psi_s)_{\text{crit}}$. Only the first of these is found to be of practical importance, representing with good accuracy the axisymmetrical supersonic flow over cone-shaped noses of missiles. Yet early conjectures that the second solution might fail to arise owing to instability were incorrect; that solution, involving a stronger shock wave, has in fact been found to represent locally the fluid flow near the tip of a cone which protrudes forwards out of a wide enough bluff body in a supersonic stream.

Taylor also studied how strong shock waves develop with time. Here, an important limiting case is that of highly intense shock waves travelling at speed U into an atmosphere at rest with density ρ_0. Then the pressure p_s, density ρ_s and gas velocity v_s behind the shock wave satisfy simple laws

$$p_s(U - v_s) = \rho_0 U^2, \quad \rho_0 U v_s = p_s, \quad \rho_0 U \left[\tfrac{1}{2} v_s^2 + \frac{p_s}{(\gamma - 1)\rho_s} \right] = p_s v_s \tag{13}$$

which (i) equate mass flows of gas entering and leaving unit area of shock wave while (ii) equating rates of gain of momentum and energy for that gas to force and its rate of working, respectively. Here, γ (again assumed constant) is a ratio of specific heats for the hot gas behind the shock wave; while, on the intense-shock-wave approximation, both pressure and internal energy in the undisturbed gas can be neglected. Equations (13) have the simple solutions

$$p_s = \frac{2}{\gamma + 1} \rho_0 U^2, \quad \rho_s = \frac{\gamma + 1}{\gamma - 1} \rho_0, \quad v_s = \frac{2}{\gamma + 1} U. \tag{14}$$

Rather as in 1910 (see section 1), it was two independent investigations in 1941 by G.I. Taylor and John von Neumann (see [26] and [27]) which simultaneously uncovered how a

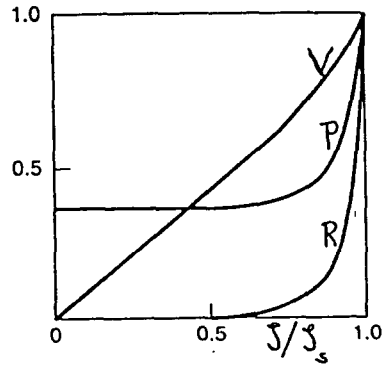

Figure 11. The function $P(\zeta)$, $R(\zeta)$ and $V(\zeta)$ plotted against ζ/ζ_s.

highly intense shock wave spreads spherically after the release, very close to a central point, of an extremely large amount of energy E. With r as radial distance from that point, they solved equations of continuity and momentum,

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho v)}{\partial r} + 2\frac{\rho v}{r} = 0, \quad \frac{\partial v}{\partial t} + v\frac{\partial v}{\partial t} + \frac{1}{r}\frac{\partial p}{\partial r} = 0 \tag{15}$$

together with the entropy equation (12), subject to boundary conditions (14) at the shock wave and the constancy of total energy E. Self-similar solutions of equations (15) and (12) take the form

$$p = \rho_0 \frac{r^2}{t^2} P(\zeta), \quad \rho = \rho_0 R(\zeta), \quad v = \frac{r}{t} V(\zeta), \tag{16}$$

where $P(\zeta)$, $R(\zeta)$ and $V(\zeta)$ are nondimensional functions of an appropriate nondimensional variable ζ.

In the Taylor-von-Neumann case, dimensional analysis yields

$$\zeta = r\left(\frac{\rho_0}{Et^2}\right)^{1/5} \tag{17}$$

since the bracketed fraction has dimensions mass times (length)$^{-3}$ divided by mass times (length)^2. Then, if $\zeta = \zeta_s$ at the shock wave, its radius grows as

$$r = \zeta_s \left(\frac{Et^2}{\rho_0}\right)^{1/5} \text{ with velocity } U = \frac{2}{5}\frac{r}{t} \tag{18}$$

so that equations (14) and (16) yield simple boundary conditions for the functions $P(\zeta)$, $R(\zeta)$ and $V(\zeta)$ at $\zeta = \zeta_s$; from which they can be calculated (see Figure 11 for the case $\gamma = 1.4$) in terms of ζ/ζ_s. Finally, ζ_s is obtained from an equation

$$4\pi \int_0^{\zeta_s} \left[\frac{1}{2} R V^2 + (\gamma - 1)^{-1} P\right] \zeta^4 d\zeta = 1; \tag{19}$$

which states that the total energy (integral within the spherical shock wave of the energy per unit volume $\frac{1}{2}\rho v^2 + (\gamma - 1)^{-1} p$) is equal to E, and which gives $\zeta_s = 1.033$ for the case $\gamma = 1.4$ treated in Figure 11. It is perhaps noteworthy that the velocity field shown there implies a huge rate of expansion, able to reduce density behind the shock wave to very low

values indeed — yet without this fall being mimicked by the pressure because fluid traversed relatively early by the shock wave has relatively high entropy per unit mass.

Taylor [26] demonstrated good agreement with observation for the two-fifths-power law (18) derived from dimensional analysis. By contrast, two other important calculations have used solutions of equations (15) and (12) in the form (16) but with a choice of ζ different from (17) because dimensional analysis was not available.

As a first example, Barenblatt and Sivashinsky [28] studied certain problems where the energy E is itself variable, because gas crossing the shock wave either loses energy per unit mass (by radiation) or gains it (by chemical reaction). Those problems may be treated by modifying γ in equations (13) to a value γ_1 different from the true ratio of specific heats γ(here, $\gamma_1 < \gamma$ for extraction, but $\gamma_1 > \gamma$ for addition, of energy); the boundary conditions then take the form (14) with γ replaced by γ_1 — even though γ is unchanged in the partial differential equation (12). For this 'mixed' problem, they used equations (16) with

$$\zeta = r/(At^\alpha), \text{ and with } \zeta = 1 \text{ at the shock wave;} \tag{20}$$

then the shock wave velocity is $\alpha r/t$ and, when this value is inserted for U in the modified boundary conditions (14), boundary values $P(1), R(1)$ and $V(1)$ may be determined. However, a solution of the partial differential equations behaving regularly at the central point $r = 0$ can attain these boundary values only for one value of α; which is found to be rather less than two-fifths when energy is extracted but greater when it is added.

A second example of solutions (16) with a different choice of ζ from (17) had appeared already in 1942, when K.G. Guderley [29] studied an 'implosion'; that is, an intense spherical shock wave which converges on a point, reaching it at time $t = t_0$. For $t < t_0$, the use of

$$\zeta = \frac{r}{B(t_0 - t)^\beta}, \text{ with } \zeta = 1 \text{ at the shock wave,} \tag{21}$$

again allows determination of $P(1)$, $R(1)$ and $V(1)$; and Guderley showed when $\gamma = 1.4$ that, among solutions of equations (15) and (12) for $\zeta \geqslant 1$ attaining these boundary values, the one with finite energy has $\beta = 0.717$.

Beyond treatments of the dynamics of strong shock waves by use of a single similarity variable ζ, two other analytical approaches may be described. When a shock wave's strength, though large, varies relatively slightly, Lighthill showed in a series of papers [30-33] that the difficulties posed by equation (12) disappear. Thus, although entropy inhomogeneity generates vorticity and so rules out the acoustician's favourite use of a velocity potential, nonetheless the pressure can be used as a sort of 'acceleration potential' satisfying the usual acoustic equations relative to the flow behind the shock wave. Briefly, the (small) flow velocities $\mathbf{u}$ in this frame of reference satisfy linearised equations

$$\nabla p = -\rho_1 \frac{\partial \mathbf{u}}{\partial t} \text{ and } \nabla \cdot \mathbf{u} = -\frac{1}{\rho_1}\frac{\partial \rho}{\partial t} = -\frac{1}{\rho_1 c_1^2}\frac{\partial p}{\partial t}, \tag{22}$$

where ρ_1 and c_1 are the density and sound speed behind the undisturbed strong shock wave, and where a linearised form of equation (12) equates the last two quantities (since pressure

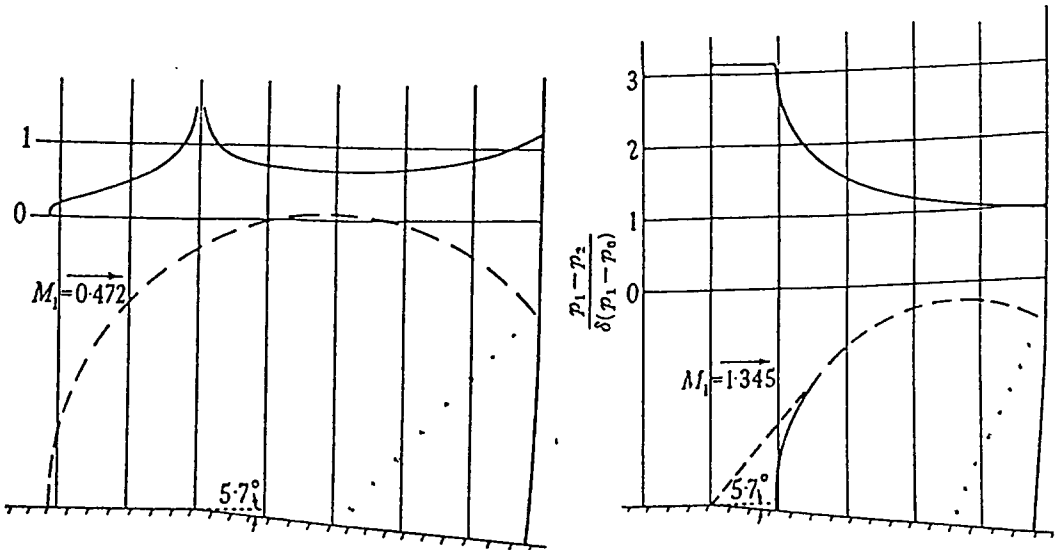

Figure 12. Diffraction of shock waves of pressure ratio $p_1/p_0 = 2$ (on left) and 10 (on right) at a corner with angle $\delta = 0.1$ radian. The plots of $(p_1 - p_2)/\delta(p_1 - p_0)$ show distributions of wall pressure p_2.

and density are adiabatically related for each fluid particle). Thus the linear wave equation $\partial^2 p/\partial t^2 = c_1^2 \nabla^2 p$ is satisfied by the pressure.

Just one interesting problem solved by this approach is sketched here (see the book by Srivastava [34] for a general review). It may be described as 'diffraction of a strong shock wave' by an oblique-angled corner where a straight wall abruptly turns, away from the flow, through a small angle δ.

The absence of any characteristic lengths in the definition of this problem allows solution [32] in terms of two similarity variables $x = X/c_1 t$ and $y = Y/c_1 t$, where (X, Y) are space coordinates in the frame of reference defined above. Their origin is at a distance $c_1 M_1 t$ to the right of the corner where M_1 is the Mach number of the flow (from left to right)behind the shock wave. A circular arc of radius $c_1 t$ with this origin as centre bounds the acoustic disturbances; however, because any shock wave travels subsonically relative to the flow behind it, that arc is but part of a semicircle, the other part being cut off by the disturbed shock wave — where a linearized boundary condition on p has to be carefully determined so that the whole pressure field can be calculated.

Figure 12 shows for $\gamma = 1.4$ the shape of that field, and the distribution of p along the wall, in two interesting cases: a shock wave of strength $\beta = 1$ (pressure-ratio 2) for which $M_1 = 0.472$ so that the corner lies inside the circular arc, and one with $\beta = 9$ (pressure-ratio 10) for which $M_1 = 1.345$ and the flow around the corner, outside the circular arc, is just a steady supersonic flow at this Mach number. Pressure reductions on the wall beyond the corner are seen to be about 3 times bigger for the stronger shock wave.

In both cases, because entropy stays constant for a particle, with X and Y roughly constant, isentropic lines in the (x, y) plane are approximately radii $y/x =$ constant through the circular arc's centre. It follows that entropy variations appear in triangular regions,

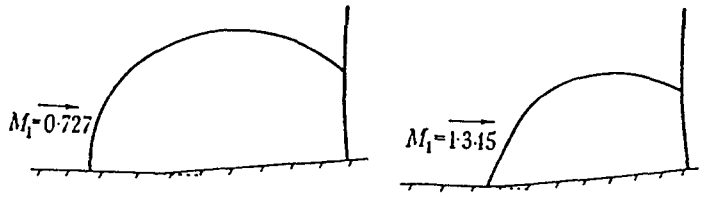

Figure 13. Shock waves of pressure ratios 3 and 10 incident on a concave corner.

to the right of the dotted lines in Figure 12, comprising all such radii that pass through disturbed parts of the shock wave.

When the angle δ is negative, the previously convex corner becomes concave to the flow, and pressure reductions change to pressure excesses; while expansion waves, shown as broken lines in Figure 12, are converted into shock waves. The altered configuration was illustrated by Lighthill [32] for incident shock waves of strengths $\beta = 2$ and 9 as in Figure 13.

From the days of Mach himself [35], reflexions of plane shock waves had been observed (see Section 5) to take 'regular' forms at plane walls making large angles to the flow behind the shock wave, while changing for small angles into 'Mach reflexion' — comprising an intersection of three shock waves and of a surface separating regions of uniform and nonuniform entropy — as in Figure 13. The analysis that produced this Figure offered a first indication of the dynamical processes underlying the formation of such intersections.

The dynamics, and also the physics, of strong shock waves were hugely developed in the 1950s to meet the demands of reentry-vehicle design [36]. In this survey of just the former field I highlight a simple piece of fluid dynamics [37] which shows how a reentry-vehicle nose having locally the form of a spherical cap generates at high hypersonic speed V an intense shock wave in the form of a concentric spherical cap, whose (larger) radius depends on K: that density ratio at an intense shock wave which conditions (14) equate to 6 for $\gamma = 1.4$ — although various physical effects in air (primarily, dissociation of molecular oxygen) raise K to larger values for reentry-vehicle shock waves. The analysis, adapted to effectively incompressible flow fields near the nose, proceeds inversely by showing that a spherical-cap shock wave of radius c gives rise to a flow with a spherical-cap streamsurface of a certain radius $a < c$.

In spherical polar coordinates, the Stokes stream function ψ satisfies

$$\psi = \frac{Vc^2 \sin^2\theta}{2K}, \quad \frac{\partial\psi}{\partial r} = Vc\sin^2\theta \text{ on the shock wave } r = c, \tag{23}$$

while the azimuthal vorticity component ω is

$$\omega = \frac{1}{r\sin\theta}\frac{\partial^2\psi}{\partial r^2} + \frac{1}{r^3}\frac{\partial}{\partial\theta}\left(\frac{1}{\sin\theta}\frac{\partial\psi}{\partial\theta}\right). \tag{24}$$

Now the vorticity produced by gradients of entropy immediately behind the shock wave may be shown to satisfy the equation

$$\frac{\omega}{r\sin\theta} = \frac{(K-1)^2 V}{Kc^2}; \tag{25}$$

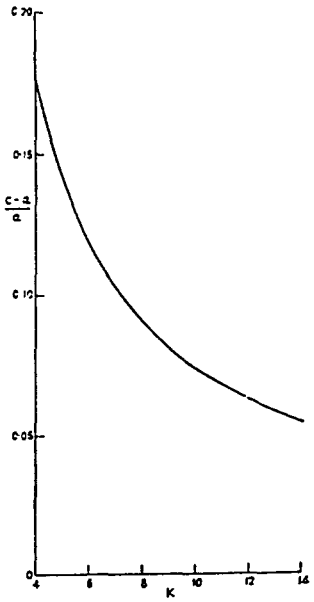

Figure 14. Stand-off ratio $(c-a)/a$ plotted against the density ratio K across an intense shock wave.

moreover, the left-hand side is a quantity known to remain constant along every streamline of an axisymmetrical flow. Therefore equation (25) for ω may be used everywhere in equation (24) to give a partial differential equation for ψ, which can be solved under the boundary conditions (23) as

$$\psi = \frac{Vc^2 \sin\theta}{30K}\left[3(K-1)^2\left(\frac{r}{c}\right)^4 - 5K(K-4)\left(\frac{r}{c}\right)^2 + 2(K-1)(K-6)\left(\frac{r}{c}\right)^{-1}\right]; \qquad (26)$$

the spherical streamsurface $r = a$ being determined as the greatest value of $r < c$ for which the quantity in square brackets vanishes. Figure 14 confirms that the 'stand-off ratio' $(c-a)/a$ for the strong shock wave falls steeply as the density ratio K increases (cramming into less and less space the flow between shock wave and solid surface). Lighthill shows how this solution helps to validate a certain more comprehensive hypersonic approximation (often called 'Newtonian plus centrifugal') to flow around reentry vehicles.

5. OPTICAL AND NUMERICAL EXPERIMENTS IN SHOCK WAVE DYNAMICS.

During most of the 20th century, crucial contributions to the development of shock wave dynamics were made by optical observation. The essential techniques had been invented much earlier; the schlieren method by A. Toepler [38] in 1867 and the interferometer method by L. Mach [39] in 1889 — while direct shadow photography of shock waves had been successfully used in 1887 by his famous father E. Mach [3]. Accordingly, it may suffice in a history devoted to the past 100 years to stress that all these inventions were aimed at acquiring data on a density field from the essentially linear relationship between density and refractive index. The interferometer method yields the best quantitative data, since it can compare the integrated density along each optical path through a flowing gas against integrated density along an equal path through undisturbed gas. For identifying shock waves and other regions of sharp density gradient, however, the schlieren method, which effectively

integrates along an optical path the density gradient in one particular direction perpendicular to that path, is specially appropriate. By contrast, direct shadow photography is of much less quantitative value; although, admittedly, it is easier to use in non-laboratory conditions.

Progress during the century of shock wave dynamics was due above all [40] to close cooperation between (on the one hand) optical experimenters, using schlieren and interferometric techniques in supersonic wind tunnels — and later in shock tubes — and (on the other hand) theoretical analysts applying methods like those described already to interpret existing experiments or to suggest new ones. A lot later in the 20th century, significant advances were made in numerical experimentation on flows incorporating shock waves; that is, on techniques within computational fluid dynamics (CFD) specially adapted to capturing the appearance of effective discontinuities within the flows; and, once again, continued progress has depended on good cooperation of numerical with theoretical analysts, and of both with optical experimenters.

Although all the excellent visual aids available in lecture-theatre conditions will allow participants in the IIAV Congress to be shown many schlieren and interferometric records of shock wave behaviour, together with impressive results of numerical experiments, their inclusion in the preprinted written version of this survey is very sharply restricted by space limitations. Accordingly, the remainder of Section 5 is devoted (i) to sketching the development of the shock tube, as a preeminent specialised device for studying shock wave dynamics, and (ii) to outlining the essential ideas underlying accurate shock-capturing methods in CFD.

History repeated itself yet once more in 1946, when accounts of two independent inventions of essentially the same shock-tube device were published simultaneously (a) in the Proceedings of the Royal Society [41] and (b) in Physical Reviews [42], with G.I. Taylor involved in one of them! — as theoretical analyst for (a) with the experimental team of William Payman and W.C.F. Shepherd — while the fine Princeton physicist Walker Bleakney led those responsible for (b), including A.H. Taub as theoretical analyst. An immediate benefit from both experimental programmes was a comprehensive verification of theories of nonlinear acoustics and shock wave dynamics.

In a shock tube, gases at pressure p_0 and at a much greater pressure p_1 are separated by a diaphragm, whose sudden rupture emits a shock wave into the former gas and an expansion wave in the latter; but, because the waves travel in opposite directions, an intermediate region of uniform pressure p_s and velocity v_s behind the shock wave separates it from the expansion-wave region. For shock waves of arbitrary strength the relationship between p_s and v_s (calculated from a fuller form of equations (13), without neglect of the pressure p_0 or the internal energy $p_0/(\gamma-1)\rho_0$ in the undisturbed gas) is

$$\frac{p_s}{p_0} = 1 + \tfrac{1}{4}\gamma(\gamma+1)\left(\frac{v_s}{c_0}\right)^2 + \gamma\frac{v_s}{c_0}\left[1 + \frac{(\gamma+1)}{16}\left(\frac{v_s}{c_0}\right)^2\right]^{1/2}. \tag{27}$$

On the other hand, the expansion wave satisfies the exact equations (Section 1) of nonlinear acoustics so that the local sound speed is $c_0 + \frac{1}{2}(\gamma-1)u$ (here, both gases are taken to have the same undisturbed sound speed $c_1 = c_0$), with u as the gas velocity in the direction of propagation. However, u as so defined takes the value $(-v_s)$ in the intermediate region;

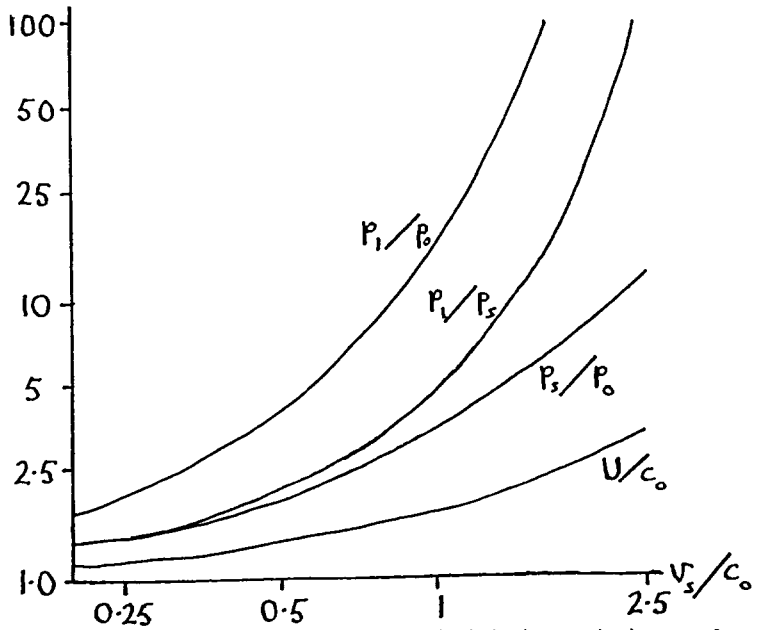

Figure 15. Log-log plots against v_s/c_0 of expressions (27),(28) and (29), together with the overall pressure ratio p_1/p_0 (which, as the product of p_1/p_s and p_s/p_0, is on a log-log plot obtained from them by adding).

where, therefore, $c_s = c_0 - \frac{1}{2}(\gamma-1)v_s$ and, by equations (1),

$$\frac{p_1}{p_s} = \left(\frac{c_0}{c_s}\right)^{2\gamma/(\gamma-1)} = \left[1 - \tfrac{1}{2}(\gamma-1)\frac{v_s}{c_0}\right]^{-2\gamma(\gamma-1)}. \tag{28}$$

Figure 15, for $\gamma = 1.4$, plots on a log-log basis expressions (27) and (28) as functions of v_s/c_0, together with their product p_1/p_0 which is the initial pressure ratio across the diaphragm. For any p_1/p_0, therefore, Figure 15 determines both v_s/c_0 and also the pressure ratio p_s/p_0 across the shock wave, together with its velocity U where

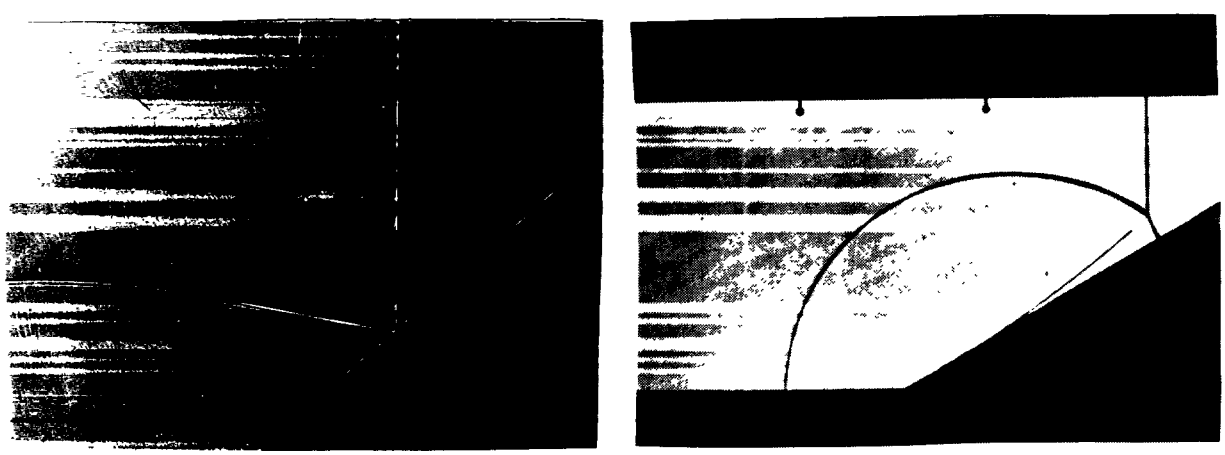

Figure 16. Shadow photograph of regular reflexion of shock wave travelling from left to right, and schlieren photograph of Mach reflexion.

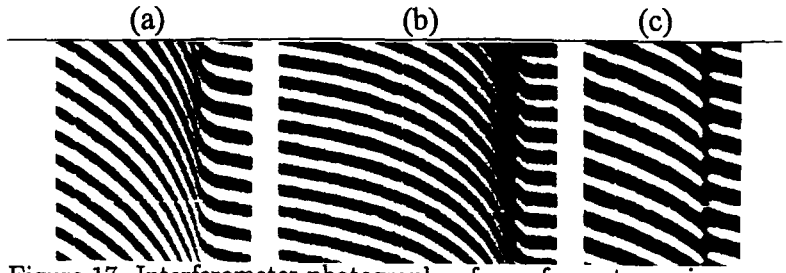

Figure 17. Interferometer photographs of waveform steepening.

$$\frac{U}{c_0} = \tfrac{1}{4}(\gamma+1)\frac{v_s}{c_0} + \left[1 + \frac{(\gamma+1)^2}{16}\left(\frac{v_s}{c_0}\right)^2\right]^{1/2} . \tag{29}$$

Both the papers published in 1946 obtained excellent agreement with this simple theory.

Then the Princeton team, after naming the new devices 'shock tubes', went on to use them [43] for advanced studies in shock wave dynamics, including accurate measurements of many types of shock wave interaction. Figure 16, for example, includes a shadow photograph of the reflexion of a shock wave at a plane wall inclined at a relatively large angle, giving regular reflexion, along with a schlieren photograph for the case of a smaller angle, giving 'Mach reflexion' (see Section 4) with 3 shock waves and a discontinuity of entropy all meeting at a point in the midst of the fluid.

Again, in a special experiment to verify the theory of waveform steepening (Section 1), three successive interferometer photographs of the development of a plane wave were taken (Figure 17). Here, each dark line is essentially a graph of the density (on a suitable scale) against distance. In (a) they have moderate slope, in (b) they have become steeper, and in (c) discontinuous.

Later developments of the shock tube included some in which still stronger shock waves could be achieved for a given pressure ratio p_1/p_0 by use of a driving gas such as hydrogen with high sound speed c_1. Then the augmented c_1 replaces c_0 in equation (28) for p_1/p_s so that, in Figure 15, its log-log plot as a function of v_s/c_0 is shifted bodily to the right; evidently, this allows the contributions of p_1/p_s and p_s/p_0 to their product p_1/p_0 to become much more unbalanced in favour of the second ratio.... Such further developments helped to add greatly to knowledge of the physics [44], as well as of the dynamics, of intense shock waves.

Another fine Princeton thinker about shock waves was John von Neumann (whose 1941 discovery of a key self-similar solution has been mentioned in Section 4 above). In 1950 he produced with R.D. Richtmyer a far-seeing paper [45] about how numerical schemes for CFD would need to be adapted in order that the appearance of shock waves in the real flow might be properly 'captured' in a numerical representation. Here, it is only the essential idea of such schemes that is outlined.

It stems from the fundamental 1910 discovery (Section 1) that, in a real gas, wave steepening processes only form a discontinuity because they are opposed by certain specific

physical effects — including viscous diffusion of momentum as well as thermal diffusivity. By contrast, any CFD scheme describes the dynamics of an 'artificial' gas, with the 'graininess' that comes from representation on some discrete grid system; although the numerical analysts's object, of course, is to make its behaviour mimic that of the real gas as closely as possible. To this end, it is necessary to ask: can this 'artificial' gas be endowed with some specific 'artificial' effect possessing two essential properties (i) that it will oppose wave steepening 'cleanly', in such a way that the resulting opposition creates rather a sharp, monotone discontinuity (within limitations imposed by mesh size), and yet (ii) that it will produce very little effect in other regions of the flow.

Von Neumann and Richtmyer [45] proposed the name 'artificial viscosity' for any such specific effect in a CFD scheme that would exhibit both properties (i) and (ii). They explained why an attempt at close numerical representation of the viscous and/or other diffusive processes in the real gas would fail to satisfy the conditions. Furthermore, they gave a valuable first suggestion for the mathematical form which a successful artificial viscosity would need to take if numerical representations of compressible-flow 'Euler' equations of continuity, momentum and energy were to be able to represent as closely as possible inviscid compressible flows in which shock waves may possibly appear. In the meantime, the need (stressed in Section 1) for overall conservation of mass, momentum and energy could be satisfied if a 'conservative' CFD scheme were used.

Any full account of later developments in this field would be a very long story indeed; yet it is noteworthy that, at the end of that long story, the essential idea of von Neumann has proved successful — as in several highly efficient CFD schemes for compressible-flow Euler equations in the development of which another fine Princeton innovator, Anthony Jameson, played a crucial role. These advanced schemes, now used throughout the aircraft industry, incorporate an enormous range of different special features. Yet the feature essential for shock capturing is precisely a modern form of artificial viscosity [23], adapted admirably to meet conditions (i) and (ii) above in flow fields of all the diverse types that can arise in particular aeronautical applications.... And, needless to say, such numerical schemes have also become specially important sources of knowledge in all areas of the field at the end of our 'century of shock wave dynamics'.

REFERENCES.

1. Brown, L.M., Pais, A. and Pippard, B. (eds.) 1995 *Twentieth Century Physics.* Bristol: Institute of Physics.
2. Rayleigh, Lord 1896 *The Theory of Sound*, 2nd edn. London: Macmillan.
3. Mach, E. and Salcher, P. 1887 *Sitzungsber. Akad. Wiss. Wien* **95**, 764.
4. Laplace, P.S. 1816 *Ann. Chim. Phys.* **3**, 238.
5. Riemann, B. 1859 *Abh. Göttingen Ges. Wiss.* **8**, 243.
6. Rayleigh, Lord 1910 *Proc. Roy. Soc. A* **84**, 247.
7. Taylor, G.I. 1910 *Proc. Roy. Soc. A* **84**, 371.
8. Hugoniot, A. 1889 *J. de l'Ecole Polytech.* **58**, 1.
9. Lighthill, J. 1978 *Waves in Fluids.* Cambridge Univ. Press.
10. Whitham, G.B. 1956 *J. Fluid Mech.* **1**, 290.
11. Whitham, G.B. 1952 *Commun. Pure Appl. Math.* **5**, 301.

12. Hopf, E. 1950 *Commun. Pure Appl. Math.* **3**, 201.
13. *Lighthill, M.J. 1956 Viscosity effects in sound waves of finite amplitude, pp 250-351, *Surveys in Mechanics* (ed. Batchelor, G.K. and Davies, R.M.). Cambridge Univ. Press.
14. *Lighthill, M.J. 1954 Higher approximations, pp 345-489 of *General Theory of High Speed Aerodynamics* (ed. Sears, W.R.). Princeton Univ. Press.
15. Hayes, W.D. 1971 *Ann. Rev. Fluid Mech.* **3**, 269.
16. Morgan, M.B. 1972 *J. Roy. Aeronaut. Soc.* **76**, 1.
17. Zierep, J. and Oertel, H. (eds.) 1988 *Symposium Transsonicum III.* Berlin: Springer.
18. *Lighthill, M.J. 1953 *Proc. Roy. Soc. A* **217**, 478.
19. Stewartson, K. 1969 *Mathematika* **16**, 106.
20. Neiland, V. Ya. 1969 *Izv. Akad. Nauk SSSR Mekh. Zhidk. Gaz.* **4**, 33.
21. Messiter, A.F. 1970 *SIAM J. Appl. Math.* **18**, 241.
22. Whitcomb, R.T. 1952 *Nat. Adv. Comm. Aeron. Memorandum* RN L52 H08.
23. Jameson, A., Baker, T.J. and Weatherill, N.P., 1986 *Am. Inst. Aeronaut. Astron. Paper* 86-0103.
24. Taylor, G.I. and Maccoll, J.W. 1933 *Proc. Roy. Soc. A* **139**, 278.
25. Ferri, A. 1958 Supersonic flow with shock waves, pp. 670-747 of *Fundamentals of Gas Dynamics* (ed. Emmons, H.W.). Princeton Univ. Press.
26. Taylor, G.I. 1950 *Proc. Roy. Soc. A* **201**, 159.
27. Von Neumann, J. 1963 The point source solution, pp. 219-237 of *Collected Works*, Vol. VI. New York: Pergamon.
28. Barenblatt, G.I. and Sivashinsky, G.I. 1969 *Prikl. Math. Mekh.* **34**, 655.
29. Guderley, K.G. 1942 *Luftfahrtforschung* **19**, 302.
30. *Lighthill, M.J. 1949 *Philos. Mag.* **40**, 214.
31. *Lighthill, M.J. 1949 *Philos. Mag.* **40**, 1201.
32. *Lighthill, M.J. 1950 *Proc. Roy. Soc. A* **198**, 454.
33. *Lighthill, M.J. 1950 *Proc. Roy. Soc. A* **198**, 554.
34. Srivastava, R.S. 1994 *Interaction of Shock Waves.* Dordrecht: Kluwer.
35. Mach, E. 1878 *Sitzungsber. Akad. Wiss. Wien* **77**, 819.
36. Hayes, W.D. and Probstein, R.F. 1966 *Hypersonic Flow Theory.* New York: Academic.
37. *Lighthill, M.J. 1957 *J. Fluid Mech.* **2**, 1.
38. Toepler, A. 1867 *Poggendorfs Ann. Phys. Chem.* **131**, 33.
39. Mach, L. 1889 *Sitzungsber. Akad. Wiss. Wien* **98**, 1318.
40. Polachek, H. and Seeger, R.J. 1958 Shock Wave Interactions, pp. 482-525 of *Fundamentals of Gas Dynamics* (ed. Emmons, H.W.). Princeton Univ. Press.
41. Payman, W. and Shepherd, W.C.F. 1946 *Proc. Roy. Soc. A* **186**, 293.
42. Bleakney, W. 1946 *Phys. Rev.* **69**, 678a.
43. Bleakney, W. and Taub, A.H. 1949 *Rev. Mod. Phys.* **21**, 584.
44. Zeldovich, Ya. B. and Raizer, Yu. P. 1966 *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena.* New York: Academic.
45. Von Neumann, J. and Richtmyer, R.D. 1950 *J. Appl. Phys.* **21**, 232.

*The asterisked references are also available in: Hussaini, M.Y. (ed.) 1996 *Collected Papers of Sir James Lighthill.* Oxford University Press.