J. Fluid Mech. (1970), *vol.* 44, *part* 2, *pp.* 315–346
Printed in Great Britain

On the Gunn effect and other physical examples of perturbed conservation equations

By J. D. MURRAY†
Courant Institute, New York University

(Received 10 January 1970)

Many situations of practical importance both in fluid mechanics and elsewhere are governed by perturbed forms of conservation laws. Generally the perturbations are in the nature of *positive* dissipation terms in the sense that any initial disturbance from a uniform state ultimately decays to that state. Diverse examples of these are discussed briefly.

A situation in which the perturbation results naturally in a *negative* dissipation term, in the sense that initial disturbances grow, although not necessarily indefinitely, arises in what has been accepted for a model for the Gunn (1963) effect and other so-called bulk negative resistance effects in semiconductors. The Gunn effect, which is of immense importance in electron-device technology (comparable with transistors), is the appearance of coherent microwave current oscillations in the crystals of a suitable semiconductor, in particular Gallium Arsenide, when they are subjected to a large electric field generally of the order of several kilovolts per centimetre. It now seems to be accepted that the effect is a consequence of the negative resistance (that is the electron drift speed *decreases* with *increasing* electric field) properties of the semiconductor crystal.

A typical model with negative resistance properties is described in detail, the resulting perturbed (both singularly and otherwise) non-linear conservation equations ((2.19) and (2.21)) are studied for practical situations of interest and the physical implications discussed in the light of experimental facts. Particular care is given to the shocks or discontinuities that must appear in the solutions when the diffusion is zero. As a result of these comparisons with experiment a simpler model is suggested which should suffice for a large number of practical situations and various quantitative features of this model are given.

1. Introduction and physical examples of perturbed conservation laws

Conservation laws, which govern many physical processes, give rise to divergence-type equations of the form

$$u_t + \sum_{i=1}^{n} v_{x_i} = 0, \tag{1.1}$$

where subscripts denote differentiation and u, v are n-component vectors which are functions of the space variables $\mathbf{x}$, with components x_i ($i = 1, \ldots, n$) and the

† Present address: Mathematical Institute, University of Oxford.

time t. Equation (1.1) expresses in differential form the physical fact that the rate of change of u in any closed domain V is equal to the flux of the vector v through the hypersurface S enclosing V. In many physical situations of interest what are in effect sources and/or sinks are present which may be functions of the dependent and independent variables. In this case we may have, in place of (1.1) a typical lower-order perturbed equation of the form

$$u_t + \sum_{i=1}^{n} v_{x_i} + f(u, v, \mathbf{x}, t) = 0, \tag{1.2}$$

where f is an n-component vector function.

To complete the physical picture a further equation, relating v to u, is required and is frequently of the type which may be described as an *exchange* equation. This is often simply a relation between u and v or more frequently a first-order differential equation like

$$\epsilon_1 u_t + \sum_{i=1}^{n} (\epsilon_2 u_{x_i} + \epsilon_3 v_{x_i}) = g(u, v, \mathbf{x}, t), \tag{1.3}$$

where ϵ_1, ϵ_2, ϵ_3 are exchange parameters which, like the vector function g, may be functions of $u, v, \mathbf{x}$ and t. Although higher order and more complicated exchange equations can occur a remarkably large number of physical situations, in the biochemical sciences in particular, fall into the class governed by (1.1) and (1.3). It is the purpose of this paper to mention briefly several of these and to consider one, namely that governing the Gunn effect in semiconductors, in detail.

The left side of (1.3) is a dissipative mechanism which may be positive or negative in the sense that an initial distribution may decay or grow with time and distance. The function f in (1.2) also crucially effects the growth or decay depending on whether it is a source ($f < 0$) or sink-like ($f > 0$) term.

We describe the equilibrium state in the exchange equation as that in which the ϵ's in (1.3) are zero in which case $g = 0$. Solving this equation for v and substituting it into (1.1) we obtain an equation for u only. The examples we shall consider here are those in which the ϵ's, f and g are functions only of u and v. A practical situation of importance in § 2 has f a function of t also. Let the solution for v, as a function of u in the equilibrium state ($\epsilon_1 = 0 = \epsilon_2 = \epsilon_3$) be

$$v = B(u) \tag{1.4}$$

and (1.1) becomes

$$u_t + C(u) \sum_{i=1}^{n} u_{x_i} + f(u, B(u)) = 0, \tag{1.5}$$

where

$$C(u) = \operatorname{grad}_u B(u).$$

If the matrix $C(u)$ in (1.5) has real and distinct eigenvalues for all u the system is hyperbolic. It is well known that if $C_u(u) \not\equiv 0$ continuous solutions in the large cannot exist if the initial data is constant outside of a finite range, and shocks or discontinuities must appear (see, for example, Courant & Hilbert (1962) or the general discussion by Gel'fand (1959)).

Let u be discontinuous across the shock (or shocks). Applying the divergence

theorem to the first of (1.5) using an elemental surface enclosing part of the discontinuity we get the speed of propagation, λ, of the shock as

$$\lambda[u] = [B(u)], \tag{1.6}$$

where $[u]$, for example, denotes the jump in u across the shock. Relations (1.6) are the generalized Rankine–Hugoniot relations. The shock relations (1.6) are not unique for the system (1.5). In any practical situation they *must* (and usually can) be justified on physical grounds. Another approach is described in § 2.5 which is more generally physically realistic. We consider (1.6) to be the shock relations which hold for (1.5).

There has been very little work done on systems of the form (1.5) in general. There have been several studies when u and x_i ($i = 1, \ldots, n$) have only one component each. Oleĭnik (1957) in such a situation gives a general discussion with an x, t dependence included in f. Murray (1970) considers a single equation for $u(x,t)$ when f is a positive, hence dissipative, function of u only and the single component of $C(u)$ is monotonic.

Dunwoody (1968) discusses the propagation of plane high-frequency sound waves in an ideal gas with internal dissipation. The one-dimensional system of equations in such a situation are

$$Pu_t + Qu_x + R = 0, \tag{1.7}$$

where the matrices

$$u^T = [w, \rho, p, \alpha], \quad R = [0, 0, 0, h(p, \rho, \alpha)],$$

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & H_\rho & (H_p - \rho^{-1}) & H_\alpha \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad Q = \begin{bmatrix} w & 0 & \rho^{-1} & 0 \\ \rho & w & 0 & 0 \\ 0 & H & w(H_p - \rho^{-1}) & wH \\ 0 & 0 & 0 & w \end{bmatrix},$$

where w, ρ, p, H, α and h are respectively the velocity, density, pressure, specific enthalpy, an internal state variable and the dissipation function. Multiplication of (1.7) by P^{-1} reduces it to the form (1.5). The problem is reduced to a specific single scalar equation and the method suggested by Varley & Cumberbatch (1966) for mildly non-linear wave propagation of high-frequency waves is used to study the resulting weak shock propagation for the single equation.

An example which can be set in the coupled ((1.2) and (1.3)) equation class is a model equation for turbulence suggested by Burgers (1939) and recently studied more generally by Case & Chu (1969). Here we have

$$\left.\begin{aligned} u_t + v_x - \lambda u &= 0, \\ \epsilon u_x &= \tfrac{1}{2}u^2 - v, \end{aligned}\right\} \tag{1.8}$$

where $\lambda > 0$ is some integral of the mean flow (or a solution of a related equation) and ϵ is proportional to the inverse of the Reynolds number. The single scalar equation from (1.8) is

$$u_t + uu_x - \lambda u = \epsilon u_{xx}, \tag{1.9}$$

which is a perturbed conservation equation with a *positive* dissipative term ϵu_{xx} and a negative dissipative source-like term $-\lambda u$. If $\epsilon = 0$ then initial-value

solutions of the resulting (1.9), the 'equilibrium' state of (1.8), simply grow exponentially with shocks present. The role of a non-zero $\epsilon \ll 1$ is primarily to smooth out such shocks in the usual singular perturbation manner. Equation (1.9) is a particularly simple case of the class of equations which arise in the Gunn effect in § 2 below.

Whitham (1967) suggested an integro-differential equation for water waves which is an extension of the Korteweg–de Vries equation in the sense that it includes the Korteweg–de Vries equation. It is of the form

$$u_t + \alpha u u_x + \int_{-\infty}^{\infty} K(x-\xi)\, u_\xi(\xi, t)\, d\xi = 0, \tag{1.10}$$

where u is the height of the surface above the undisturbed depth h_0, $\alpha = \frac{3}{2}(g/h_0)^{\frac{1}{2}}$ and the kernel K is

$$\left.\begin{aligned} K(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} c(K)\, e^{iKx}\, dK, \\ c(K) &= \left(\frac{g}{K}\tanh Kh_0\right)^{\frac{1}{2}}. \end{aligned}\right\} \tag{1.11}$$

If $Kh_0 \ll 1$ the asymptotic form of (1.10) with (1.11) gives the Korteweg–de Vries equation, namely (1.10) with γu_{xxx} in place of the integral, where $\gamma = \frac{1}{6}(gh_0^5)^{\frac{1}{2}}$. The integral perturbated conservation equation (1.10) has been studied by Seliger (1968) who demonstrates that waves, under certain circumstances, can form a bore (or shock) unlike the Korteweg–de Vries equation. Thus as pointed out by Whitham (1967) (1.10) is more analogous to the equation

$$u_t + \alpha u u_x + \beta u = 0, \tag{1.12}$$

which can have discontinuous solutions.

Seymour & Varley (1970) have considered finite-amplitude wave propagation in systems which are dissipative and where responses are rate sensitive. They discuss high-frequency periodic disturbances and one equation they study is, in effect (1.12).

In finite-amplitude wave motion for a certain class of non-linear Maxwell rods with viscosity the governing one-dimensional equation for the stress, σ, is

$$\sigma_{xx} = [1 + \epsilon g_\sigma(\sigma)\, \sigma_t]_t + \lambda h_\sigma(\sigma)\, \sigma_t,$$

where g is a positive monotonic function and $\lambda h_\sigma \sigma_t$ is the non-linear positive dissipation viscosity term. By considering $0 < \epsilon \ll 1$ and stretching the leading semi-characteristic co-ordinate Murray (1970) developed an asymptotic solution for the boundary-value problem for this equation. The first asymptotic term required the solution of a single scalar first-order non-linear wave equation of the form (1.5) where $C(u)$ and $f(u)$ are simply proportional to the functions g and h. Discontinuities in the stress derivatives appeared and the effect of the viscous dissipation on the decay of such solutions was considered.

Perturbed conservation equations which arise physically as coupled equations of the single scalar forms of (1.2) and (1.3) were studied by Goldstein (1953)

and Goldstein & Murray (1959) in a model for ion exchange processes which occur in fixed columns. Their equations are of the form

$$\left.\begin{aligned} u_t + v_x &= 0, \\ \epsilon v_x &= u - rv + (r-1)\,uv, \end{aligned}\right\} \tag{1.13}$$

where ϵ, r are positive constants. The first of (1.13) essentially represents a conservation of ion concentration while the second is a fairly common ion-exchange equation. When $\epsilon \ll 1$ the effect of the dissipative ϵ-term is to smooth out the shocks, and alter their position by $O(\epsilon)$, which appear in the equilibrium ($\epsilon = 0$) situation. A more general class of conservation—exchange equations, encompassing (1.13), and, for example, a generalized Burgers equation has been considered in detail by Murray (1968).

In the above examples the perturbations of the conservation equations have generally been dissipative in the positive sense with the exception of the $-\lambda u$ in Burgers's turbulence model. In the latter there has been interest in solutions where the negative ($-\lambda u$) and positive (ϵu_{xx}) dissipations balance. In the next section we discuss in detail a model for the Gunn effect in semi-conductors which has a *lower*-order *negative* dissipation and a *higher*-order singular perturbation *positive* dissipation. The negative dissipation is the dominant one and the energy for the growth is supplied naturally by an external electric circuit. The reduced ($\epsilon = 0$) problem which should suffice for experimental comparison reduces to solving a scalar equation of the type (1.5) in which the characteristic speed $C(u)$ is *not* monotonic. The model, with appropriate background material, is discussed in §2 and the solutions compared with experiment. The comparison with experiment results in a considerable simplification of the model which has to be considered for practical applications. The results from this simplification seem to be in agreement with experiment and certain quantitative results are given in §2.8.

2. A non-linear analysis of the Gunn effect

2.1. *Introduction*

Experimentally Gunn (1963, 1964) found that coherent microwave current oscillations† appeared when he subjected a crystal sample ($O(10^{-3}\,\mathrm{cm})$) of Gallium Arsenide (GaAs) to a constant electric field larger than a critical field of several (≈ 3) kilovolts per centimetre. This phenomenon is known as the Gunn effect and is a simple way to generate microwave power. Kroemer (1964) suggested that the Gunn effect was consistent with what is called a bulk negative resistance first suggested as a practical proposition by Ridley & Watkins (1961) and independently by Hilsum (1962). That such is the case is now fairly well accepted: it is also clear from the analysis below.

A bulk negative resistance or bulk negative conductance or bulk negative differential resistivity is exhibited by certain semi-conductors and in particular

† In the literature it is sometimes referred to as unstable domain formation or propagation.

by n-type† GaAs. Negative resistance is said to occur when the current density *decreases* as the electric field *increases* for some range of the field larger than some finite critical value. It is a bulk effect if at every point inside the material, rather than at the junction between different types of semi-conductors, the negative resistance is caused by the local electric field when it is in the appropriate range. In this paper we shall be concerned with materials which exhibit this bulk negative resistance effect. As mentioned above, the possible existence of bulk negative resistance effects was first mooted by Ridley & Watkins (1961) and Hilsum (1962) using what is called a two-conduction band model, which is briefly described below in § 2.2.

The purpose of this part of the paper is (i) to review briefly the two-conduction band model for negative resistance with a view to (ii) setting up a model equation for the Gunn effect, (iii) to study analytically asymptotic solutions of the resulting general non-linear equation under realistic initial conditions, (iv) to interpret the results physically, keeping in mind the model's limitations, and to compare them with experiment and (v) to suggest, as a result of (iv), a simpler model which although of less intrinsic mathematical interest, is more generally in keeping with experiment.

The Gunn effect has been found in semiconductors other than GaAs but because of its particularly well-developed technology the vast amount of work, which has been carried out since Gunn's (1963) paper, has been on the possible commercial exploitation of the effect in GaAs. In general the possible commercial use, which is extensive, is in making inexpensive semiconductor microwave frequency devices and in creating new applications of microwave technology. The collection of papers in the Special issue of the IEEE Transactions and Electronic Devices (1966) give a survey of the work on the Gunn effect up to that time. An interesting and very readable review is given by Bott & Fawcett (1968). Kroemer (1968) gives a general description of negative conductance. A discussion of bulk-effect devices is given by Sze (1969).

2.2. *Two-conduction band model*

In a pure semiconductor crystal there are practically no electrons in the conduction band (that is an energy band in which the electrons can move under the influence of an electric field). However, if some appropriate impurity, referred to as doping, is added, electrons are released which can move in the lattice. In the case of GaAs if sulphur (a group VI element) is added it replaces arsenic atoms in the crystal lattice and the sulphur atom acts like an arsenic atom with one too many electrons. This electron can then move through the lattice in the presence of a field.

Electrons move in the conduction band of such doped GaAs with an effective mass which depends on the energy. If there are more than one conduction band there is the possibility of having electrons with more than one effective mass as shown below. From a wave-mechanic point of view the energy $\mathscr{E}$ of an electron

† Due to impurities in semiconductors there are immobile and mobile charges. If the mobile charges are negative it is called an n-type semiconductor. The negative mobile charges are balanced by the positive immobile charges in this case.

depends on its wave vector **k**. If the electron were free then **k** is equal to $\mathbf{p}/\hbar$, where **p** is its momentum and $\hbar$ is Planck's constant, in which case $\mathscr{E} = (1/2m)\hbar^2\mathbf{k}^2$, where m is its mass, and the $\mathscr{E}-\mathbf{k}$ curve is a parabola with a minimum or valley at $\mathbf{k} = 0$. When the electron is in a solid such a simple state with a parabolic dependence of $\mathscr{E}$ on **k** does not obtain. In semiconductors there are certain allowed energy bands separated by regions where real **k** do not exist. Such energy bands are usually completely filled or completely empty in pure semiconductor crystals. With doping, however, the spare electrons, as in the GaAs we consider, can exist in a normally empty energy band which is a *conduction* band. We now assume (Ridley & Watkins 1961 and Hilsum 1962) that this conduction consists,

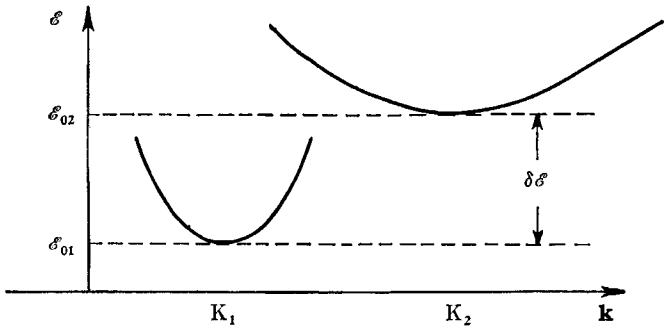

FIGURE 1. Schematic two-conduction band structure.

in effect, of two conduction bands as in figure 1 with two minima representing a central and a satellite valley. We assume further that in the valleys the energy of an electron can be approximated by

$$\mathscr{E}_i(\mathbf{k}) = \frac{\hbar^2}{2m_i}(\mathbf{k}-\mathbf{K}_i)^2+\mathscr{E}_{0i} \quad (i = 1, 2), \tag{2.1}$$

where $\mathscr{E}_{0i}$ and $\mathbf{K}_i$ are respectively the energy and wave vector at the valley minimum and m_i is the effective mass here assumed to be a scalar.

We consider $\delta\mathscr{E} = \mathscr{E}_{02}-\mathscr{E}_{01}$ to be small compared with the energy difference between the central valley and the valence (in effect where the electrons are tied to the ions) energy band, which has an energy band below the central valley: the reason for this is made clear below. For GaAs, for example, $\delta\mathscr{E} \fallingdotseq 0{\cdot}36$ electron volts (eV) and the energy gap between the lower valley and the valence band is approximately 1·43 electron volts (eV).

From (2.1) the effective mass m_i in the valley is inversely proportional to the second derivative of the energy with respect to k, namely $d^2\mathscr{E}/dk^2$ there: that is the effective mass is proportional to the curvature at the valley bottom. Here $m_2 > m_1$ and if we denote by μ_1 and μ_2 the electron mobilities† in the respective valleys then $\mu_1 > \mu_2$. Let n_1 and n_2 be the number densities of the electrons in the two states.

† Mobility is effectively the average velocity of the electrons, resulting from an applied electric field, divided by the electric field.

If the doped crystal is subjected to an increasing electric field E the distribution of electrons between the two valleys will change. This is based on the assumption that if the field is high enough to supply sufficient energy to the electrons a transfer of electrons from the lower to the upper valley occurs but not from the valence band to the lower conduction band. In GaAs the difference in energy between the two valleys (approximately 0·36 eV) is approximately 15 times larger than the room temperature energy of electrons in the central valley and $O(10^{-1})$ times smaller than the energy gap between the central valley and the valence band below it. The electrons which move from the central or lower to the upper or satellite valley increase their effective mass (since the curvature in the valley there is larger) and hence their mobility decreases. Within each state the actual mobilities (as well as n_1 and n_2) can also change as a function of the applied electric field, for example (see the curves in Bott & Fawcett (1968)). The lattice temperature in general also increases when transfer occurs. In our model we do not include the temperature effects *per se*. Since the Gunn phenomenon occurs at room temperatures the inclusion of temperature effects is not perhaps necessary for a basic understanding at this stage.

In a one-dimensional model, which, as is generally accepted, suffices for practical applications, the electrons, in the presence of an electric field pointing in the negative x direction, move in the positive x direction and give rise to a drift electric current, $\mathbf{J}_1$ say. The drift current density carried by the semiconductor is taken to be

$$\mathbf{J}_1(E) = e(n_1\mu_1 + n_2\mu_2)\,\mathbf{E}, \tag{2.2}$$

where $-e$† ($e > 0$) is the charge on the electron and $\mathbf{E} = -\mathbf{i}E$, $\mathbf{J}_1 = -\mathbf{i}J_1$, with $\mathbf{i}$ a unit vector in the x direction. In (2.2) all of n_1, n_2, μ_1 and μ_2 may be functions of E. We introduce the average velocity, $\mathbf{v}(E) = \mathbf{i}v(E)$, of the electrons by writing (2.2) in the form

$$\mathbf{J}_1(E) = -en\mathbf{v}(E), \tag{2.3}$$

where the total electron density n and $v(E)$ in (2.3) are given from (2.2) by

$$n = n_1 + n_2, \quad v(E) = (E/n)\,(n_1\mu_1 + n_2\mu_2). \tag{2.4}$$

The form of $v(E)$ as a function of E has been the subject of considerable study and controversy. Mathematically it requires a study of Boltzmann's equation for, in this case, a two-band structure. Butcher & Fawcett (1965, 1966) (see also Butcher 1967 and Bott & Fawcett 1968) have obtained an approximate solution which is illustrated in figure 2: they also give a review of the state at that time. As seen their theory compares well with the experimental curve of Ruch & Kino (1967) also given in figure 2, a comparison given by Kroemer (1968).

A commonly suggested (but not by Butcher (1967) as figure 2 shows) drift velocity *versus* field curve is illustrated schematically in figure 3. This is obtained from (2.3) and (2.4) on the following basis. At low fields (and normal temperatures) all of the electrons are in the central valley and the drift velocity is given by $v(E) \fallingdotseq \mu_1 E$ since $n \fallingdotseq n_1$, $n_2 \fallingdotseq 0$. As E increases the mean electron energy increases and an increasing number of electrons are transferred from the

† The reason for $-e$ instead of e for the electron charge is that practically all the literature on the Gunn effect uses this notation.

lower to the upper valley with its correspondingly higher effective electron mass. When sufficient electrons have been transferred by virtue of E having reached some threshold E_m, say, the average velocity will actually start to decrease as E increases further: this is the negative resistance part of the curve. For very high fields it might be expected that most of the electrons will have transferred to the upper valley and hence in this range $n_1 \fallingdotseq 0$, $n_2 \fallingdotseq n$ and from (2.4), $v(E) \fallingdotseq \mu_2 E$. The actual mobilities μ_1 and μ_2 themselves can also depend on E

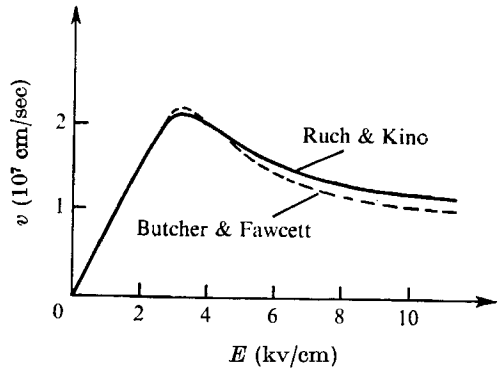

FIGURE 2. Theoretical (after Butcher & Fawcett 1966) and experimental (after Ruch & Kino 1967) curves for the drift velocity $v(E)$ *versus* the electric field E.

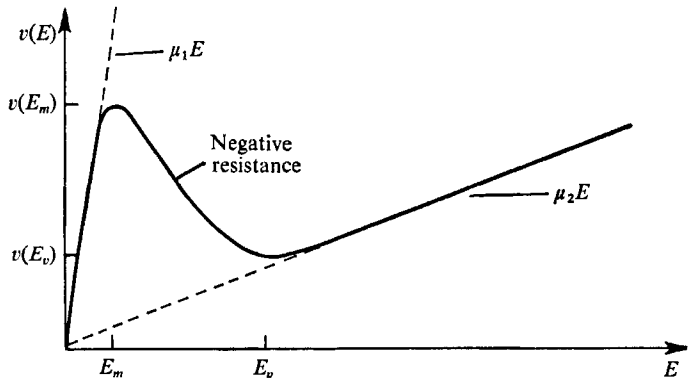

FIGURE 3. Schematic $v(E)$–E curve with negative resistance region using a two-conduction band model.

(see Bott & Fawcett 1968) but such changes are not crucial from a qualitative point of view regarding negative resistance. It is now clear that if the valence band were not separated from the lower central valley by a much larger energy difference than $\delta\mathscr{E}$, that between the central and satellite valleys, it is feasible that electrons could move from it into the central valley. Under these circumstances a negative resistance would not be possible.

In the analysis below it is seen that a negative resistance in the $v(E)$–E curve is necessary for the Gunn effect. An accurate form of it is not so important at this stage but as seen below the assumption that $dv(E)/dE > 0$ for sufficiently

large E is important. It is possible that more complicated situations than the two-valley model can also give rise to a negative resistance region where $dv/dE < 0$. However, Hutson *et al.* (1965) showed experimentally, that if a hydrostatic pressure was applied to GaAs the energy gap $\delta\mathscr{E}$ between the upper and lower valleys was decreased. When the pressure was high enough the Gunn effect did not occur at all. This tends to substantiate the conjecture that a two-conduction band model is a basic physical mechanism for the Gunn effect.

It has been assumed, above and below, in the model that the *static* velocity-field curve (or characteristic) is appropriate to use in the *dynamic* state under which the Gunn effect is observed. The time constant associated with the inter-valley electron transfer is $O(10^{-13}\,\text{sec})$ (Conwell & Vassell 1966) and so the electrons are considered to follow the static $v(\text{E})$–E curve as long as the time scale of the problem is large compared with times $O(10^{-13}\,\text{sec})$. This may not always be an allowable assumption. With a typical sample used in experiment the length is $O(10^{-3}\,\text{cm})$ and the velocity of the waves (or domains as used in the literature) is $O(10^{7}\,\text{cm/sec})$ and so a typical time is $O(10^{-10}\,\text{sec})$. This is small compared with $10^{-13}\,\text{sec}$, but another and perhaps more crucially important time scale is that for the passage of the wave past a point in the sample. If typical wave widths (although they continually vary as seen below in §2.3 *et seq.*) were as small as $O(10^{-6}\,\text{cm})$ with a corresponding typical time of passage of $O(10^{-13}\,\text{sec})$ then there could be some doubt as to the validity of using the static $v(E)$–E curve. Although in our analysis we shall use the static curve it should be kept in mind that a dynamic $v(E)$–E curve with, for example, some dispersion effects included, might be required in a more sophisticated model. Of course other effects, such as an explicit temperature dependence might also have to be included.

In view of the theoretical complexity and variety of conditions under which experiments can be conducted there still seems to be some disagreement as to the most appropriate form to take for the $v(E)$–E curve, and in particular for the negative resistance section. There is a divergence in the experimental curves as well as the theoretical ones although there is some agreement as in figure 2. However, all of them display *at least* the general form with a negative resistance section. For the general discussion we do not require anything more specific than a form for $v(E)$ which is similar to that in figure 3. It seems that, at this stage, the $v(E)$–E curves obtained experimentally and by some detailed transport calculations (see figure 2) have not yet shown that beyond some E, E_v in figure 3, the velocity increases again: some theoretical models do predict it, however. If the $v(E)$–E curve does *not* possess a relative minimum for some $E > 0$ the physical situation from the point of view of wave propagation is very different as discussed below in §2.8: the algebra is considerably easier, and is the basis for the simplified model which agrees more generally with experiment. We shall consider curves at this stage where $dv(E)/dE > 0$ for some E sufficiently large. For specific details the analytical solution is complicated for all but the simplest feasible forms for $v(E)$. For simplicity and for some quantitative description of the non-linear wave propagation which might occur in GaAs when the Gunn effect occurs, we

shall use a simplified form as in figure 4 which is continuous, piece-wise linear and, following Butcher, Fawcett & Hilsum (1966), is given by

$$\left.\begin{aligned} v(E) &= \mu_1 E \quad (0 \leqslant E \leqslant E_m) \\ &= (\mu_1+\mu_3)E_m - \mu_3 E \quad (E_m \leqslant E \leqslant E_v) \\ &= \mu_2 E \quad (E_v \leqslant E), \\ \text{with} \qquad \mu_3 &= (\mu_1 E_m - \mu_2 E_v)/(E_v - E_m), \end{aligned}\right\} \tag{2.5}$$

where μ_1, μ_2 are the lower and upper valley mobilities, μ_3 given by (2.5) ensures continuity at E_v and E_m, E_m is the threshold electric field for which larger values

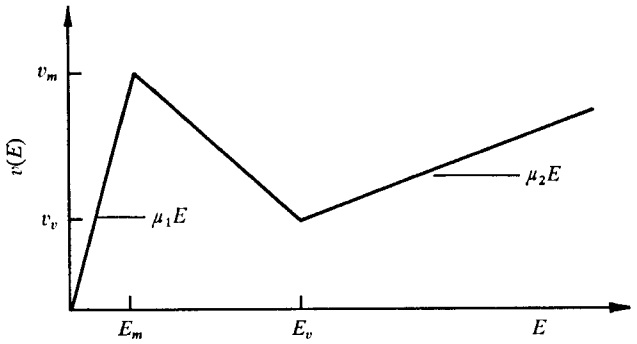

FIGURE 4. Piece-wise linear approximation for the $v(E)$–E curve (see (2.36) and (2.37) for details and typical values of the parameters).

of E give negative resistance, and E_v is the value of E at the relative minimum of $v(E)$. Typical values for these parameters are (Bott & Fawcett 1968)

$$\left.\begin{aligned} \mu_1 &\fallingdotseq 1{\cdot}1 \times 10^4 \text{ cm}^2/\text{volt sec}, \quad \mu_2 \fallingdotseq 50 \text{ cm}^2/\text{volt sec}, \\ E_m &\fallingdotseq 3{\cdot}3 \times 10^3 \text{ volt/cm}, \qquad E_v \fallingdotseq 10^5 \text{ volt/cm}, \end{aligned}\right\} \tag{2.6}$$

which from (2.5) give

$$\left.\begin{aligned} v_m &= v(E_m) \fallingdotseq 2{\cdot}7 \times 10^7 \text{ cm/sec}, \\ v_v &= v(E_v) \fallingdotseq 5 \times 10^6 \text{ cm/sec}, \\ \mu_3 &\fallingdotseq 2{\cdot}6 \times 10^2 \text{ cm}^2/\text{volt sec}. \end{aligned}\right\} \tag{2.7}$$

An experimentally more realistic $v(E)$–E curve is obtained at least for $E < E_v$ from (2.5) by simply rounding off the slope discontinuities as they stand which results in a lower $v(E_m)$ closer to the accepted experimental one of approximately $2{\cdot}2 \times 10^7$ cm/sec. The reason for considering a piecewise linear form is that the integrations involved in the solution can be carried out simply. Until more general agreement is found for the $V(E)$–E curve there is no point in using a more complicated form: quantitative differences would not in any case be major. From a theoretical and practical point of view what is more important is whether $dv/dE > 0$ for any realistic range of values of $E > E_m$. Some current views (for example, Butcher 1967) exclude such a possibility. These views are substantiated by the results below.

2.3. *One-dimensional model equation*

When a sample of n-type GaAs, or other suitable semiconductor, is subjected to an electric field in the vicinity of the critical field E_m, or a field in the region where $dv(E)/dE < 0$, the microwave current oscillations observed depend on the various parameters and characteristics of the sample and the external circuit. A schematic form of the device is given in figure 5. It is clear from experiment and it is generally accepted that a one-dimensional model will suffice for most practical situations and so all dependent quantities are taken to be functions of a single space co-ordinate x and the time t. As far as the author is aware most of the models and subsequent analyses have been linear and/or numerical studies or those in which the wave (or domain) does not change in shape. Here we wish to consider the non-linear constantly-changing wave problem analytically. Relevent to a non-linear approach is the work of Knight & Peterson (1966): it is discussed below at the appropriate place.

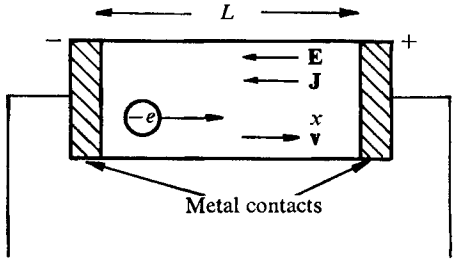

FIGURE 5. Schematic Gunn effect device.

As a first step we isolate the effect by considering the situation which occurs *in* the sample only and we shall not consider the external circuit problem. We require equations therefore which will govern the total current $\mathbf{J}$ and field $\mathbf{E}$ in the semiconductor sample. From Maxwell's equation $\mathbf{J} = \operatorname{curl} \mathbf{H}$, where $\mathbf{H}$ is the magnetic field, $\mathbf{J}$, with a single component, J, in the negative x direction, is a function of t only: this is supplied by the external circuit. This current in the crystal is made up of the drift current $\mathbf{J}_1$ of the form (2.3) plus the displacement current plus the diffusion current. The current conservation equation is thus

$$\mathbf{J} = -ne\mathbf{v}(E) + k\epsilon_0 \mathbf{E}_t + \mathbf{i}e\partial[nD(E)]/\partial x, \tag{2.8}$$

where subscripts denote partial derivatives and $k\epsilon_0$, assumed to be constant, is the semiconductor dielectric constant, $-e$ is the electron charge and $D(E)$ is the field-dependent diffusion coefficient. A form for $D(E)$–E has been given by Bott & Fawcett (1968).

Let the impurity or doping density in the crystal be n_0 which we shall take to be uniform and independent of x: (it is easy to include an x dependence in the model equation). For low fields ($E < E_m$) any group of excess mobile electrons or space charges simply disperses due to Coulomb forces and the total electron density n is then equal to n_0 and the electric field, because of Poisson's equation ((2.9) below), is uniform throughout the sample. In this situation Ohm's law

holds and the electron drift velocity is proportional to E as in figures 3 and 4. For values of $E > E_m$ the excess mobile electrons do not disperse but in a sense can pile up due to the negative differential resistance and give rise to inhomogeneous regions of electron density. The equation relating **E** to the charge density n in this case is Poisson's equation, namely,

$$k\epsilon_0 E_x = e(n - n_0). \tag{2.9}$$

In (2.9) it should be remembered that **E** points in the negative x direction.

If the field is low enough then $n \fallingdotseq n_0$ and the electrons drift with a uniform velocity for a given E. When the field is high enough however this is not the case and inhomogenities in electron density appear and $n \neq n_0$. From (2.9) the integral of $n - n_0$ over the sample length must still be zero of course. Since $n \geqslant 0$ physically, (2.9) puts a limitation on allowable E_x, namely $E_x \geqslant -en_0/k\epsilon_0$.

Equations (2.8) and (2.9), or with slight variations, with $v(E)$ such as in figure 3, are those which seem to be fairly generally accepted as the temperature-independent model equations for the Gunn effect.

Substituting n from (2.9) into (2.8) in scalar form ($\mathbf{E} = -\mathbf{i}E$, $\mathbf{J} = -\mathbf{i}J$, $\mathbf{v} = \mathbf{i}v(E)$) gives

$$k\epsilon_0 E_t + k\epsilon_0 v(E) E_x + [en_0 v(E) - J] = \partial[D(E)(en_0 + k\epsilon_0 E_x)]/\partial x. \tag{2.10}$$

Equation (2.10) is in the form of a single scalar conservation equation (1.5) with a *higher*- and *lower*-order perturbation included.

If we consider short samples then $J(t)$ (in (2.10)) is intimately connected with the external circuit. However, if the sample is long enough so that the wavelength of a disturbance in E (or n) is small compared with the sample length, J can be obtained from (2.10), when E is constant, and is

$$J = en_0 v(E_e), \tag{2.11}$$

where $E = E_e$, say, is the uniform constant field far from the disturbance wave. When $E = E_e$, $n(x, t) = n_0$, of course.

We wish to consider the initial-value problem for (2.10) in the situation when the sample is infinitely long (that is J from (2.11) is a given constant) and so we consider $0 \leqslant x \leqslant \infty$. We thus seek a solution of (2.10) with J from (2.11), $v(E)$ and $D(E)$ known functions of E, and

$$E(x, 0) = \phi(x), \quad E(0, t) = E_e = E(\infty, t), \tag{2.12}$$

where a typical $\phi(x)$ is as shown in figure 6. The $\phi(x)$ is a form of disturbance field the history of which we wish to study. If such a solution, or an approximate one, can be found certain pertinent information regarding the Gunn effect would be available. The basic idea behind this model for the Gunn effect is that when the uniform field $E_e > E_m$ (or as we shall see if $\max \phi(x) > E_m$) an initial disturbance or wave travels across the sample until it reaches the anode where it is immediately absorbed and another with profile $\phi(x)$ appears at the cathode.

There seems to be some question as to how a disturbance wave appears at all. The fact that there appears to be only one wave in the sample at any given time suggests that it may be a property of the external circuit. Another possibility

is a special type of instability property of the crystal: we discuss this briefly in §2.8.

Here we study the wave-type propagation problem posed by (2.10) with (2.11) and (2.12). If a realistic solution can be found then such quantities as wave speed, wave shape, shock generation in the zero diffusion case can be given and an estimate of the frequency of the current pulses using the length, L, of the sample for distance of travel. After E has been found an approximation to the excess voltage, V, which is required in a finite-length sample is given by

$$V = \int_0^L (E - E_e)\, dx. \tag{2.13}$$

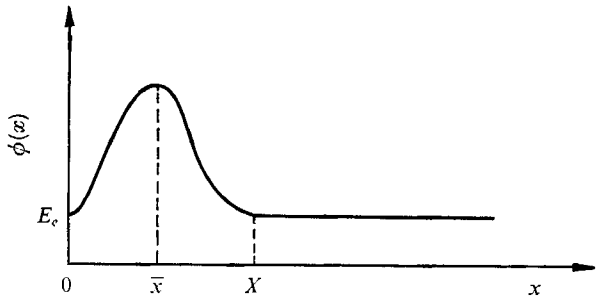

FIGURE 6. Initial electric field profile.

We now non-dimensionalize (2.10) and (2.11) by using the following reference quantities giving, where appropriate, typical practical values for GaAs:

$$\left.\begin{aligned}
&L, \text{ length of the sample} \approx 10^{-3}\,\text{cm};\\
&E_0, \text{ electric field} \approx 3{\cdot}3 \times 10^3\,\text{volt/cm};\\
&v_0 = \mu_0 E_0, \text{ drift velocity of electrons} \approx 2 \times 10^7\,\text{cm/sec};\\
&\mu_0, \text{ electron mobility} \approx 5 \times 10^3\,\text{cm/volt sec};\\
&\omega = v_0/L, \text{ drift velocity frequency} \approx 10^{10}/\text{sec};\\
&D_0, \text{ diffusion coefficient (when } E = E_0) \approx 4 \times 10^2\,\text{cm}^2/\text{sec};\\
&n_0 = \sigma_0 E_0/e v_0, \text{ doping density};\\
&\sigma_0, \text{ static conductivity} \approx 1\,\text{mho/cm};\\
&k\epsilon_0 = 13{\cdot}5\epsilon_{\text{air}}, \text{ dielectric constant};\\
&\omega_c = \sigma_0/k\epsilon_0, \text{ low field dielectric relaxation frequency} \approx 10^{12}/\text{sec};\\
&\omega_d = v_0^2/D_0, \text{ diffusion frequency} \approx 10^{12}/\text{sec}.
\end{aligned}\right\} \tag{2.14}$$

Write

$$\left.\begin{aligned}
x = x'L, \quad t = (L/v_0)t' = t'/\omega, \quad E = E_0 u,\\
v(E) = v_0 v'(u), \quad D(E) = D_0 D'(u), \quad E_e = E_0 u_e,
\end{aligned}\right\} \tag{2.15}$$

and so u, and all primed quantities in (2.15) are non-dimensional. Using (2.14) and (2.15) the governing equation (2.10) with (2.11) takes the non-dimensional form

$$u_t + v(u)\,u_x + \frac{\omega_c}{\omega}[v(u) - v(u_e)] = \frac{\partial}{\partial x}\left[D(u)\left(\frac{\omega_c}{\omega_d} + \frac{\omega}{\omega_d}u_x\right)\right], \tag{2.16}$$

where for convenience in (2.16) and below we have dropped the primes on x, t, v and D. A solution of the parabolic equation (2.16) is sought subject to (2.12) which we take in dimensionless form as

$$u(x,0) = \phi(x), \quad u(0,t) = u_e = u(\infty,t). \tag{2.17}$$

Figure 6 in dimensionless form illustrates a typical disturbance $\phi(x)$ we shall consider.

For GaAs in particular and for Gunn effect semiconductors in general $\omega/\omega_d \ll 1 (O(10^{-2})$ from (2.14)). For *high* conductivity semi-conductors $\omega/\omega_c \ll 1$ and ω_c/ω_d is $O(1)$ as in (2.14) while for *low* conductivity ones ω/ω_c is $O(1)$ and $\omega_c/\omega_d \ll 1$. Writing

$$\epsilon = \omega/\omega_d, \quad \lambda = \omega_c/\omega, \tag{2.18}$$

(2.16) becomes

$$u_t + [v(u) + \epsilon\lambda D_u(u)]\, u_x + \lambda[v(u) - v(u_e)] = \epsilon\partial[D(u)\, u_x]/\partial x, \tag{2.19}$$

where

$$\left.\begin{array}{l} \epsilon \ll 1, \quad \lambda = O(1), \quad \text{low} \\ \epsilon \ll 1, \quad \lambda \gg 1, \quad \epsilon\lambda = O(1), \quad \text{high} \end{array}\right\} \text{conductivity semiconductors.}$$

The problem is thus a singular perturbation one, for *all* λ, with the small parameter ϵ which multiplies the highest derivative being the ratio of a typical drift current frequency to the static diffusion frequency or, what is the same thing, the ratio of the diffusion velocity (D_0/L_0) to the drift velocity v_0.

The problem posed by (2.19) with (2.17) and given non-monotonic $v(u)$ and $D(u)$ (as is the case here) is interesting mathematically. For some range of u, $D(u)$ may also have large derivatives. However, keeping in mind the physical background to the governing equation (2.19) we must put limitations on the minimum value for u_x for any practical Gunn phenomenon application. The non-dimensional form of (2.9) is

$$\lambda u_x = (n/n_0) - 1, \tag{2.20}$$

and so $u_x \geqslant -1/\lambda$ is a physical restriction since $n \geqslant 0$. For $D(u)$ and its derivatives $O(1)$ it is shown below in § 2.5 that realistic solutions for a long (infinite) sample are never possible for all time for ϵ small. This restriction on u_x, also limits the type of allowable $\phi(x)$ in (2.17) in the $\phi_x(x) < 0$ region.

From a mathematical and physically general point of view if we consider only (2.19) the term on the right is dissipative and it tends to make the initial field wave $\phi(x)$ decay. However, since $\lambda[v(u) - v(u_e)] < 0$ for some range of $u > u_e > u_m$ $(= E_m/E_0)$ it is, for such u, a source term with a *negative* dissipative effect which tends to make $\phi(x)$ grow. With $\epsilon = 0$, which is the same as zero diffusion, the problem reduces to that of non-linear wave propagation governed by a first-order non-linear hyperbolic equation in the form of a perturbed conservation equation such as (1.5). Discontinuities or shocks in u will form in general and the role of the higher-order ϵ-dissipation term with $\epsilon \ll 1$ will simply be to alter the position of these shocks asymptotically by $O(\epsilon)$ and to smooth them out over a distance $O(\epsilon)$ in which the first-order derivatives are $O(\epsilon^{-1})$. A class of such equations when $\lambda = 0$, $\epsilon \ll 1$ has been studied by Murray (1968) and when $\lambda \neq 0$, $\epsilon = 0$ by Murray (1970) in both cases with $v(u)$ a monotonic function of u.

Propagation of wave-like solutions of, in effect, (2.19) which do not change in shape have been studied by Butcher, Fawcett & Hilsum (1966) by writing the dependent variables E and n in (2.8) and (2.9) as functions of the single variable $x-Ut$ where U, the constant wave speed of propagation, is to be determined. They deduce from the ordinary differential equations, which are obtained from (2.8) with $J(t)$ constant and (2.9) on this assumption, that $U = v(E_e)$, the drift speed in the undisturbed state where $E = E_e$. The maximum value of E in the wave is then found. They then proceed to the limit as the diffusion coefficient D tends to zero and so from (2.18) and (2.14) they have, in effect, $\epsilon = 0$. However, the limit $D \to 0$ is *not* uniform since the resulting equation is a first-order non-linear hyperbolic wave equation with wave solutions which *cannot* propagate without change in shape nor without the appearance of discontinuities; such an assumption would imply that the characteristic speeds for the equation, namely $v(u)$ would have to be constant. This can happen only when the λ-multiple is also zero. With $\epsilon \neq 0$ the existence of such constant shape wave-like solutions implies a cancelling out of the two competing dissipations mentioned above. This is physically unlikely since in practical situations $\epsilon \ll 1$ with all the consequences it implies from a singular perturbation viewpoint.

Thus since $\epsilon \ll 1$ for both high and low conductivity semiconductors the *right* side of (2.19) does not affect the quantitative features to $O(1)$ as regards shape and propagation speed. It smooths out the discontinuities which must appear if $\phi(x) - u_e$ is continuous and is zero outside a finite range in x. Thus, as a first and physically realistic step we shall consider the wave propagation when the diffusion coefficient is zero: that is $\epsilon = 0$. It is suggested that the detailed physical features obtained from solutions of the reduced ($\epsilon = 0$) problem, for whatever is the appropriate $v(u)$ preserve the important quantitative features of the phenomenon governed by (2.19) with conditions (2.17) for the *low* conductivity case. In the high conductivity case $\epsilon\lambda = O(1)$ and from (2.19) the importance of $D(u)$ is evident. If $D(u)$ is taken to be constant then the reduced problem obtained on letting $\epsilon \to 0$ even with $\epsilon\lambda = O(1)$ is the same as the low conductivity case. If $D_u(u) \neq 0$ the situation is quite different. It can be studied in exactly the same way as the following although the results could be markedly different since the characteristic speeds are $v(u) + \epsilon\lambda D_u(u)$ in place of $v(u)$. We shall not consider this specific case further.

2.4. *Non-linear wave propagation problem*

With $\lambda \neq 0$ the mathematical problem reduces to finding solutions of

$$\left.\begin{aligned} u_t + v(u)\,u_x + \lambda[v(u) - v(u_e)] &= 0, \\ u(x,0) = \phi(x), \quad u(0,t) &= u_e, \end{aligned}\right\} \tag{2.21}$$

obtained from (2.19) with $\epsilon = 0$ and (2.17). In dropping the order of (2.19) on setting $\epsilon = 0$ we removed one boundary condition, namely the one at infinity, from (2.17). Equation (2.21) is a perturbed conservation equation of the form (1.5).

It should be mentioned again at this stage that once u (the electric field) has

been found the electron density $n(x,t)$ is given immediately by (2.20). If we integrate (2.20) with respect to x from zero to infinity (or a sufficiently large x so that $u = u_e$) the left side is zero and so, as mentioned above $\int_0^\infty (n-n_0)\,dx = 0$. Thus if n is not simply a constant the $n(x,t)-x$ curve must have a solution with $n > n_0$ in one part, called an accumulation layer, and $n < n_0$ in the other part, called a depletion layer. We shall not discuss the specific solution for n but it should be noted that it can be obtained, once u has been found, on taking appropriate care when the u solution contains shocks discussed in detail below. The restriction on $u_x \geqslant -1/\lambda$ from (2.20) should be kept in mind for any practical application.

In (2.21) λ (> 0) can take any value covering both the high ($\lambda \gg 1$) and low ($\lambda = O(1)$) conductivity cases, although for the former we require $D(u)$ to be constant here, and $v(u)$ will be of the form illustrated dimensionally in figures 3 (see also figure 7) and 4. We further consider $\phi(x)$ of the form

$$\left.\begin{aligned} \phi(x) &= u_e \quad (x \leqslant 0) \\ &= \phi(x) \quad (0 \leqslant x \leqslant X) \\ &= u_e \quad (X \leqslant x), \end{aligned}\right\} \tag{2.22}$$

with the maximum value $\overline{\phi}$ defined as the value of $\overline{\phi}(x)$ at $x = \bar{x}$, that is

$$\overline{\phi} = \max_{0 \leqslant x \leqslant X} \phi(x) = \phi(\bar{x}), \tag{2.23}$$

and where $\phi(x)$ is monotonic for $0 \leqslant x \leqslant \bar{x}$ and for $\bar{x} \leqslant x \leqslant X$ as in figure 6 (see also figure 8).

There are several cases to consider which result in different wave configurations and which are of varying practical importance. The two main classes are (I) $u_e > u_m$ and (II) $u_e < u_m$ (see figures 7) of which the former is of more importance practically for the Gunn effect. Within each class the solution depends on the form of $\phi(x)$ of course but more importantly on the maximum of $\phi(x)$, $\overline{\phi}$, given in (2.23).

In case I, for example, if $u_e < \overline{\phi} < u_v$ then $\lambda[v(u)-v(u_e)] < 0$ is a source or type of negative dissipation term and $u(x,t)$ increases from its initial form $u(x,0)$. In fact, omitting for the moment shock discontinuity considerations, u would increase until u reaches u_v and finally (see figure 7) u_s defined by

$$v(u_s) = v(u_e) \quad (u_s > u_e). \tag{2.24}$$

Clearly the *larger* the background field (but $u_e < u_v$) the *smaller* is the maximum field u_s. At this point the λ-term tends to zero, as is also the case when $u \to u_e$. For that section of the $u(x,t)-x$ curve where u is increasing (similar to the $0 \leqslant x \leqslant \bar{x}$ section in figure 8) *and* $u_e \leqslant u \leqslant u_v$, $v(u)$ is a monotonic decreasing function of u and wave steepening occurs and the appearance of discontinuities of shocks is inevitable. For the section where u is increasing and $u_v \leqslant u \leqslant u_s$, $v(u)$ is a monotonic increasing function and so this section of the wave tends to flatten.

On the other hand in that section where u is decreasing (similar to $\bar{x} \leqslant x \leqslant X$, in figure 8) the wave tends to steepen for $u_v \leqslant u \leqslant u_s$ and to flatten for $u_e \leqslant u \leqslant u_v$.

As the flattening and steepening occurs the wave continually grows with a *maximum* $u = u_s$ and at the same time the area under the curve $(u-u_e)$–x grows, as it must in view of the source like λ term.

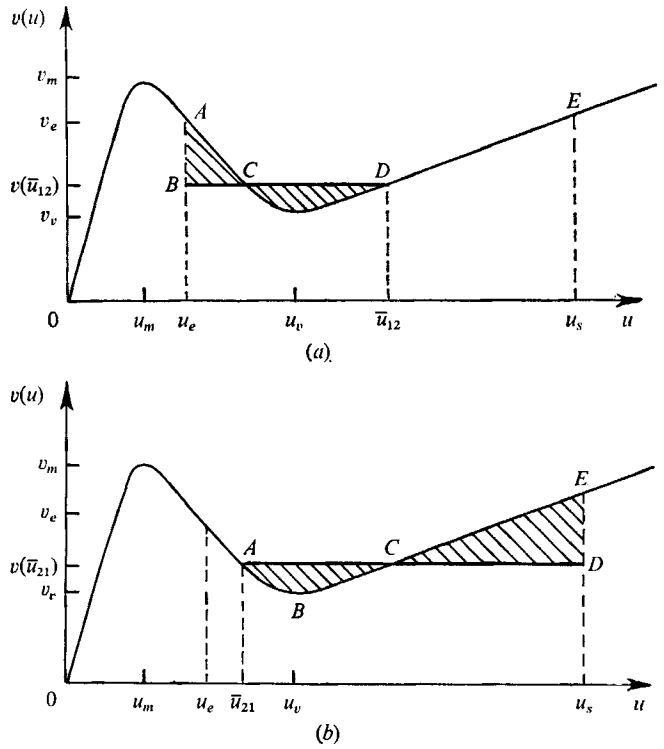

FIGURE 7. Equal area rule for determining the constant shock properties (see §2.5).

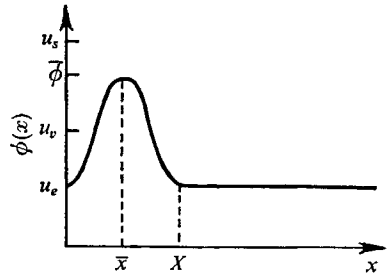

FIGURE 8. Dimensionless initial field profile.

In the next three sections we proceed with the mathematical solution, since certain interesting facts are found about solutions of such equations in which $v(u)$ is not monotonic. The method may also be of practical interest in the reduced $(\epsilon \to 0)$ situation in which $D(u)$ is not constant. In certain practical situations the solution demonstrates the inadequacy of the model for explaining the physical phenomenon we had in mind since the wave can steepen and eventually a shock may, but not always, appear on the wave side in which $u_x < 0$ and there

$u_x \to -\infty$. In this case the condition from (2.20) that $u_x \geqslant -1/\lambda$ is violated for some $t > 0$ even if the initial value $\phi_x(x) > -1/\lambda$. A specific example is worked out to provide typical numbers. That shocks can eventually appear on the $u_x < 0$ side does *not* necessarily mean the model solution is not valid for the range of t of practical interest.

If a shock does appear on the $u_x < 0$ side in the range of t found in experiment the singular perturbation solution, in which $\epsilon \ll 1$, *cannot* compensate for the violation of the condition $u_x \geqslant -1/\lambda$ since near this shock it gives

$$u_x = O(-1/\epsilon) \ll -1/\lambda.$$

In such a situation the model (with $J(t)$ constant in (2.10)) can still be used for the Gunn effect if $v(u)$ has a different form from that assumed above and is a form frequently found experimentally. It must have an effective unlimited negative resistance region: that is $v(u) < 0$ for all $u > u_m$ in the *practical rangeof interest.*

The point about the possible existence of shocks on the $u_x < 0$ side has also been noted by Knight & Peterson (1966) for an equation like (2.21). However, they proceed by excluding this shock and by suggesting a resultant constant wave solution. Such a solution cannot exist for (2.21): the growth of any wave is calculated in § 2.6 and it is not zero. The growth is directly related to the voltage across the sample.

2.5. *Shock formation and speed of propagation*

The characteristics form of (2.21) is

$$x_\sigma = v(u(x(\sigma), t(\sigma))), \quad t_\sigma = 1, \; u_\sigma + \lambda[v(u) - v(u_e)] = 0, \tag{2.25}$$

where σ is a parameter measured along the characteristic. From (2.25)

$$x_\sigma = v(u) = v(u_e) - (1/\lambda)\, u_\sigma.$$

On choosing $t = 0$ when $\sigma = 0$ and integrating the last equation, (2.25) becomes

$$\left.\begin{aligned} x &= x_0 + v(u_e)\,\sigma + (1/\lambda)\,[\phi(x_0) - u], \\ t &= \sigma, \\ \int_{\phi(x_0)}^{u} \frac{ds}{v(s) - v(u_e)} &= -\lambda\sigma, \end{aligned}\right\} \tag{2.26}$$

where x_0 is the value of x at $t = 0$. The solution obtained from (2.26) ceases to be single-valued as soon as two or more characteristics intersect which, from (2.26) is when x ceases to be a single-valued function of x_0: that is, when $\partial x/\partial x_0 = 0$. From (2.26)

$$\left.\begin{aligned} x_{x_0} &= 0 = 1 + (1/\lambda)\,[\phi'(x_0) - u_{x_0}], \\ u_{x_0} &= \phi'(x_0)\left[\frac{v(u) - v(u_e)}{v(\phi(x_0)) - v(u_e)}\right], \end{aligned}\right\} \tag{2.27}$$

where primes denote differentiation with respect to the single argument. Thus if t_c is the smallest time t when x first ceases to be single-valued then from (2.27) it is the least t satisfying

$$\mathrm{X}(x_0, t) = 1 + \frac{\phi'(x_0)}{\lambda}\left[1 - \frac{v(u(x_0, t)) - v(u_e)}{v(\phi(x_0)) - v(u_e)}\right] = 0, \tag{2.28}$$

where X is defined by (2.28), $0 \leqslant x_0 \leqslant X$ and u as a function of x_0 and t ($= \sigma$) is obtained from the integral in (2.26).

Case I: $u_v > u_e > u_m$.† Since $\phi(x_0)$ is monotonically increasing and decreasing according as $0 \leqslant x_0 \leqslant \bar{x}$ and $\bar{x} \leqslant x_0 \leqslant X$ respectively, $\phi'(x_0 > 0$ until $x_0 = \bar{x}$ where $\phi = \overline{\phi}$ and $\phi'(x_0) = 0$ at which point, from (2.28), $\mathrm{X}(\bar{x}, t) = 1$: $\mathrm{X}(x_0, 0) = 1$ also. At a time $t > 0$ with $0 \leqslant x_0 \leqslant \bar{x}$ any $u(x_0, t)$ in the range $u_e \leqslant u \leqslant u_v$ (see figure 7) has increased from its initial value $u(x_0, 0) = \phi(x_0)$ and so

$$v(\phi(x_0)) > v(u(x_0, t))$$

and since $\phi'(x_0) > 0$, $\mathrm{X}(x_0, t) < 1$. Further for $t > 0$ with $0 \leqslant x_0 \leqslant \bar{x}$ and $u(x_0, t)$ in the range $u_v \leqslant u \leqslant u_s$ by a similar argument $\mathrm{X}(x_0, t) > 1$. Thus a critical time $t = {}_1t_c$, say, occurs for a critical $x_0 = {}_1x_{0c}$, say, such that $u_e \leqslant u({}_1x_{0c}, {}_1t_c) \leqslant u_v$ and $\phi({}_1x_{0c}) \leqslant u_v$. On the other hand when $\bar{x} \leqslant x_0 \leqslant X$ and so $\phi'(x_0) < 0$ there may be another critical time $t = {}_2t_c$ which occurs for a critical $x_0 = {}_2x_{0c}$ such that $u_v \leqslant u({}_2x_{0c}, {}_2t_c) \leqslant u_s$ and $\phi({}_2x_{0c}) \geqslant u_v$. If λ is large enough this latter critical time ${}_2t_c$, and hence a shock in this region, may *not* exist (see (2.42) below, for example). Alternatively ${}_2t_c$ may be sufficiently large to be outside of the range of practical interest and this solution could *still* be of use for experimental comparison.

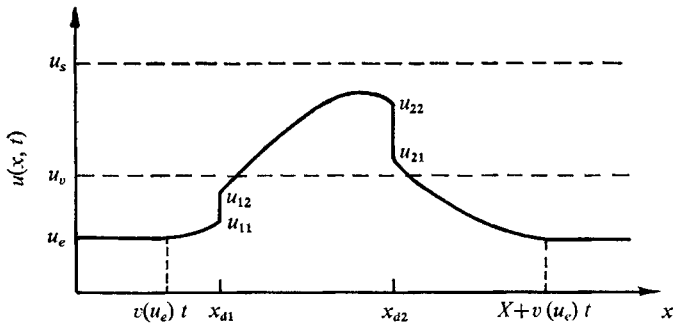

FIGURE 9. Typical wave with two shocks.

As soon as the solution ceases to be single-valued we must have a shock in the usual way. The first shock to form depends on $\phi(x_0)$ and λ. For $t > \max[{}_1t_c, {}_2t_c]$ a typical situation is as illustrated in figure 9 where two shocks exist at $x = x_{d1}$ and $x = x_{d2}$. Such a situation is not allowed physically if ${}_2t_c$ is in the time range of experiment.

Case II: $u_e < u_m$. In this case if $\overline{\phi} \leqslant u_q$, say where u_q is defined as the least $u > u_e$ satisfying $v(u_q) = v(u_e)$ (and so $u_m < u_q < u_s$: see also figure 7), then $\lambda[v(u) - v(u_e)] \geqslant 0$ for all $u_q \geqslant u \geqslant u_e$. Thus the λ term is a dissipative sink-like term and the initial wave $\phi(x)$ simply decays to u_e with time. If $\overline{\phi}$ is large enough ($\overline{\phi} > u_q$) then part of the wave will grow while the rest decays and spike like solutions appear. For $\overline{\phi} < u_q$ the specific manner of the decay and the appearance or otherwise of shocks depends on the form of $\lambda[v(u) - v(u_e)]$ as $u \to u_e$: such

† If $e_e > u_v$ this case is similar to case II as seen below.

situations have been considered in detail by Murray (1970). From this it is clear that, in this case II, for any Gunn phenomenon to exist $\phi(x)$ must be such that $\overline{\phi} > u_q$ for some $0 \leqslant x \leqslant X$ so that $\lambda[v(u)-v(u_e)] < 0$ for some u, namely $u_q < u \leqslant u_s$. But this case is in essence similar to that in case I considered in detail below.

If $u_e > u_v$ then $\lambda[v(u)-v(u_e)] > 0$ which is again like case II and so will not be discussed further here. We are thus left with case I as that containing the essential features of a model which gives solutions related to the Gunn phenomenon.

Case I: $u_v > u_e > u_m$. Consider $\phi(x)$ to be as illustrated in figure 8 in which $u_s \geqslant \phi(x) > u_v$ for some $0 \leqslant x \leqslant X$ and λ and $\phi(x)$ such that ${}_1t_c$ and ${}_2t_c$ exist. At $t > \max[{}_1t_c, {}_2t_c]$ let the shock which started at $t = {}_1t_c$ have position $x = x_{d1}(t)$ and let u increase discontinuously from $u(x_{d1}-,t) = u_{11}(t)$ to $u(x_{d1}+,t) = u_{12}(t)$ across it. Across the leading shock at $x = x_{d2}(t)$ let u decrease from $u(x_{d2}-,t) = u_{22}(t)$ to $u(x_{d2}+,t) = u_{21}(t)$ as in figure 9. From (2.26), incidentally, the positions of those parts of the wave which at $t = 0$ had $u = u_e$ were at $x_0 = 0$, X are, for $t > 0$, now at $x = v(u_e)t$ and $x = X + v(u_e)t$ respectively: the former of these holds as long as $u_{11}(t) > u_e$ which for large enough t is not the case as seen below.

The relations which hold across shocks in solutions of (2.21) are not unique unless they are constrained by some prescribed condition which must really come from the physics or is obtained from the limit of a higher-order equation, (2.19), in this case, as $\epsilon \to 0$. In situations of practical interest the shock conditions are generally the same as those obtained simply from applying Gauss's theorem (or as in Courant & Hilbert 1962) to (2.21) across each shock in turn. These are the same, of course, as those obtained from the condition that requires all solutions of (2.21), *with or without* shocks, to satisfy the integrated form of (2.21): we recall this below where such an integral is required for the voltage. This integration approach was used, in effect, by Goldstein (1953), Goldstein & Murray (1959) and Murray (1968, 1970). Thus, for the two shocks at x_{d1} and x_{d2} we obtain from (2.21) the appropriate shock propagation speeds as

$$\frac{dx_{d1}}{dt} = \frac{1}{u_{12}-u_{11}}\int_{u_{11}}^{u_{12}} v(u)\,du, \quad \frac{dx_{d2}}{dt} = \frac{1}{u_{22}-u_{21}}\int_{u_{21}}^{u_{22}} v(u)\,du. \tag{2.29}$$

When the shocks are weak and the changes in u are small (2.29) give the shock speeds as the usual average of the speeds of the characteristics meeting on the shock, namely, $\frac{1}{2}[v(u_{11})+v(u_{12})]$ and $\frac{1}{2}[v(u_{22})+v(u_{21})]$ respectively.

We now consider the characteristic solution (2.26) to hold in the regions $0 \leqslant x \leqslant x_{d1}-$, $x_{d1}+ \leqslant x \leqslant x_{d2}-$, $x_{d2}+ \leqslant x$. Thus, for the shock at $x_{d1}(t)$, for example, we have from (2.26)

$$\left.\begin{aligned} x_{d1} &= x_0 + v(u_e)\,t + (1/\lambda)\,[\phi(x_0(x_{d1})) - u_{11}] \\ \int_{\phi(x_0)}^{u_{11}} &\frac{ds}{v(s)-v(u_e)} = -\lambda t, \end{aligned}\right\} \tag{2.30}$$

which on eliminating x_0 gives u_{11} as a function of x_{d1} and t. A similar equation gives u_{21} as a function of x_{d1} and t. These together with the first of (2.29) determine $x_{d1}(t)$ with the boundary condition at $t = {}_1t_c$ where $u_{11}({}_1t_c) = u_{12}({}_1t_c)$. Similarly $x_{d2}(t)$ can be obtained in principle. Once x_{d1}, x_{d2} have been found $u_{11}(t)$, $u_{12}(t)$, $u_{21}(t)$ and $u_{22}(t)$ can be obtained from (2.30) and similar equations. Note that (it was shown above) $u_e \leqslant u_{11}({}_1t_c) \leqslant u_v$ and $u_v \leqslant u_{21}({}_2t_c) \leqslant u_s$. The shocks, once formed, continually change in strength and, certainly for t close to the critical times, grow.

We now show that the shocks asymptotically grow to a *maximum stable* size and speed of propagation given by a simple area rule.

2.6. *Equal area rule for shock propagation*

Consider in the first instance the shock at $x = x_{d1}(t)$ where $u_e \leqslant u_{11}({}_1t_c) \leqslant u_v$. For u in the range $u_e \leqslant u \leqslant u_v$, $v(u)$ is a monotonic decreasing function of u and so for $u_e \leqslant u_{11} \leqslant u_{12} \leqslant u_v$

$$v(u_{11}) > \frac{dx_{d1}}{dt} = \frac{1}{u_{12}-u_{11}}\int_{u_{11}}^{u_{12}} v(u)\,du > v(u_{12}). \tag{2.31}$$

From (2.31), therefore, the shock speed is less than the characteristic speed $v(u_{11})$ and greater than the characteristic speed $v(u_{12})$. Thus the characteristics for $x < x_{d1}$ catch up with the shock and u_{11} decreases with time until it eventually reaches u_e. On the other hand the shock overtakes the characteristics for $x > x_{d1}$ and u_{12} grows at least until $u_{12} = u_v$. Since u_{11} cannot decrease beyond u_e (since $\lambda[v(u)-v(u_e)]$ is zero for $u = u_e$) u_{12} will continue to increase *beyond* u_v as long as $dx_{d1}/dt > v(u_{12})$. Referring to figure 7(*a*) the right inequality in (2.31) says that the area under the curve $v(u)$ between u_{11} and u_{12} is greater than the area $(u_{12}-u_{11})\,v(u_{12})$. It is clear therefore that the shock will grow until u_{11} decreases to u_e and u_{12} increases to a maximum $\bar{u}_{12}$, say, where $u_s > \bar{u}_{12} > u_v$, given by

$$\frac{1}{\bar{u}_{12}-u_e}\int_{u_e}^{\bar{u}_{12}} v(u)\,du = v(\bar{u}_{12}). \tag{2.32}$$

Equation (2.32) says that $\bar{u}_{12}$ is determined by simply requiring that the areas ABC and CDE in figure 7(*a*) are equal. In this way the maximum shock speed for the shock at x_{d1} equals the characteristic speed at $x = x_{d1}+$, namely $v(\bar{u}_{12})$. For times larger than that required for u_{11} to reach u_e and u_{12} to reach $\bar{u}_{12}$ the shock speed of propagation and its strength remain *constant*. Note that the expressions in (2.32), and (2.33) below, are *independent* of λ. The time to reach the steady state, which is *finite* for x_{d1}, is a function of λ, of course.

Considering now the shock at x_{d2}, its constant strength $u_s-\bar{u}_{21}$ and constant speed of propagation $v(\bar{u}_{21})$ are given by a similar equal area rule where area ABC = area CDE in figure 7(*b*) and $\bar{u}_{21}$ is given by

$$\frac{1}{u_s-\bar{u}_{21}}\int_{\bar{u}_{21}}^{u_s} v(u)\,du = v(\bar{u}_{21}). \tag{2.33}$$

The approach to the constant shock state is less simple, however, since $u_{22} \to u_s$ only after an infinite time and it may not be a simple monotonic growth.

When the constant state has almost been reached so that $dx_{d2}/dt = v(\bar{u}_{21})$ that part of the wave $x > x_{d2}$ flattens and stretches out since $v(\bar{u}_{21}) < v(u)$ for $u_e \leqslant u \leqslant \bar{u}_{21}$.

For that part of the wave where $x_{d1} < x < x_{d2}$, $u_v < \bar{u}_{12} < u < u_s$, u grows continuously with time since $u_t > 0$ from the last of (2.25). By a simple linearization of the last two equations of (2.25) when $u \fallingdotseq u_s$ (and hence $v(u) - v(u_e) \fallingdotseq v(u_s) - (u_s - u)\, v'(u_s) - v(u_e) \fallingdotseq -(u_s - u)\, v'(u_s))\, u$ tends to u_s exponentially in time with exponent $\lambda v'(u_s)\, t$.

Since the magnitude of the slope of that part of the $v(u)$–u curve for $u_e \leqslant u \leqslant u_v$ is almost everywhere larger than the slope where $u_v \leqslant u \leqslant u_s$ it can be shown in general (or demonstrated simply using (2.5) as a basis for $v(u)$) that when the constant shock strengths and speeds have effectively been achieved the equal area rule implies that $dx_{d2}/dt \fallingdotseq v(\bar{u}_{21}) > v(\bar{u}_{12}) = dx_{d1}/dt$. Thus the distance between the shocks increases and the wave spreads out. A typical form is shown in figure 10.

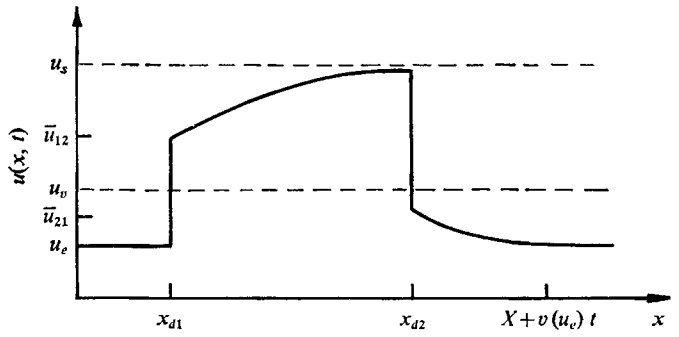

FIGURE 10. Typical solution when the two shocks have reached their steady state.

The area under $u - u_e$ continuously increases (see also figures 11 and 12) by the spreading of the wave since there is a maximum relative amplitude $u_s - u_e$. A measure of the growth is obtained by integrating the differential equation in (2.21) from $x = 0$ to $x = X + v(u_e)\, t$, the leading edge of the wave, to give

$$\int_0^{X+v(u_e)\,t} u_t\, dx + \int_0^{X+v(u_e)\,t} v(u)\, u_x\, dx + \lambda \int_0^{X+v(u_e)\,t} [v(u) - v(u_e)]\, dx = 0.$$

If no shocks are present $\partial/\partial t$ may be taken outside of the integral and the second becomes simply an integral in u. If shocks are present we can do exactly the same thing *only* if we use the specific shock relations (2.29). As mentioned above this is another way of obtaining the shock conditions. In view of the physical significance of the integrated form in this problem (see (2.35) below) this is clearly the physical justification here, namely that the voltage at a given time is the same whether or not shocks are present. Thus, with or without shocks the

second integral in the last equation is zero ($u = u_e$ at both limits) and the integrated equation becomes, using (2.29) when shocks are present,

$$\left.\begin{aligned}\frac{\partial}{\partial t}\int_0^{X+v(u_e)t} u\,dx &= \lambda\int_{v(u_e)t}^{X+v(u_e)t}[v(u_e)-v(u)]\,dx \quad (t\leqslant \tilde{t}),\\ &= \lambda\int_{x_{d1}}^{X+v(u_e)t}[v(u_e)-v(u)]\,dx \quad (t\leqslant \tilde{t}),\end{aligned}\right\} > 0, \tag{2.34}$$

where $\tilde{t}$ is the time when u_{11} has decreased to u_e.

Of practical importance is the excess voltage V (defined slightly differently in (2.13) for the finite sample) over that required to maintain u_e uniformly. Thus

$$\left.\begin{aligned}V(t) &= \int_{v(u_e)t}^{X+v(u_e)t}(u-u_e)\,dx \quad (t\leqslant \tilde{t}),\\ &= \int_{x_{d1}}^{X+v(u_e)t}(u-u_e)\,dx \quad (t\geqslant \tilde{t}).\end{aligned}\right\}$$

From (2.34) it is easily shown that

$$\left.\begin{aligned}V(t) = &\int_0^X \phi(x)\,dx - u_e[X+v(u_e)t]\\ &+\begin{cases}\lambda\int_0^t\int_{v(u_e)\tau}^{X+v(u_e)\tau}[v(u_e)-v(u)]\,dx\,d\tau & (t\leqslant\tilde{t})\\ \lambda\int_0^{\tilde{t}}\int_{v(u_e)\tau}^{X+v(u_e)\tau}[v(u_e)-v(u)]\,dx\,d\tau+\lambda\int_{\tilde{t}}^t\int_{x_{d1}(\tau)}^{X+v(u_e)\tau}[v(u_e)-v(u)]\,dx\,d\tau & (t\geqslant\tilde{t}).\end{cases}\end{aligned}\right\} \tag{2.35}$$

The major qualitative and some quantitative features of the wave growth, form and speed of propagation of the wave motion phenomenon governed by (2.21) are given from the above. For example, if we consider a finite length ($= 1$) sample then the front of the wave will reach the anode after a dimensionless time $1/v(u_e)$. Practically what is important then is whether ${}_2t_c$ is greater than or less than $1/v(u_e)$. The latter situation *cannot* hold physically. On the basis that the wave is immediately absorbed when the front of the wave reaches the anode, with another starting immediately at the cathode, the dimensional frequency $\omega = O(v(E_e)/L) = O(10^{10}/\text{sec})$ from (2.14). On the other hand if we required the trailing edge to reach the anode before another appeared then the typical dimensionless frequency is $O(v(\bar{v}_{12}))$ on the basis that the shock at x_{d1} forms the trailing edge. The frequency on this basis is smaller than $O(10^{10}/\text{sec})$.

The time of passage of a wave is given approximately by

$$[X+v(u_e)t-x_{d1}(t)]/v(\bar{u}_{12})$$

which for a meaningful comparison with the Gunn phenomenon, irrespective of the presence of the leading shock, must be small compared to unity.

Since the actual form for $v(u)$ as a function of u is still not definite we consider in specific detail the solution when the salient features of the model are incorporated in $v(u)$ by considering it to be piece-wise linear as in (2.5). The results are specifically relevant to our simplified Gunn effect model below in §2.8. The

solution also gives typical values in a specific and seemingly relevant situation and demonstrates the limit of possible practical validity of the model by giving specific expressions for ${}_1t_c$ and ${}_2t_c$.

2.7. *Simple illustrative example*

We consider the specific case where $v(u)$ is piece-wise linear and, on dividing both sides of (2.5) by $v_0 (= \mu_0 E_0$, see (2.14)), is given by

$$\left.\begin{aligned} v(u) &= \mu_1 u \quad (0 \leqslant u \leqslant u_m), \\ &= (\mu_1+\mu_3)\, u_m - \mu_3 u \quad (u_m < u \leqslant u_v), \\ &= \mu_2 u \quad (u_v < u), \\ \mu_3 &= (\mu_1 u_m - \mu_2 u_v)/(u_v - u_m), \end{aligned}\right\} \tag{2.36}$$

where the dimensionless μ_1, μ_2, μ_3 are the dimensional mobilities in (2.5) divided by μ_0 ($\fallingdotseq 5 \times 10^3$ cm/volt sec). From (2.6), (2.7) and (2.14)

$$\left.\begin{aligned} &\mu_1 \fallingdotseq 1{\cdot}62 \times 10, \quad \mu_2 \fallingdotseq 10^{-2}, \quad u_m = 1, \\ &u_v \fallingdotseq 3 \times 10, \quad \mu_3 \fallingdotseq 5{\cdot}48 \times 10^{-1}. \end{aligned}\right\} \tag{2.37}$$

Here u_s, defined by (2.24) is given from (2.36) as

$$u_s = [(\mu_1+\mu_3)\, u_m - \mu_3 u_e]/\mu_2,$$

which with (2.37) and a typical u_e gives

$$u_e \fallingdotseq 10, \quad u_s \fallingdotseq 1{\cdot}127 \times 10^3. \tag{2.38}$$

Note that whatever form of $\phi(x)$ is used the maximum field is u_s, defined by (2.24), where typically from (2.38) and E_0 from (2.4) the maximum field $E_s \fallingdotseq 3{\cdot}67 \times 10^3$ kv/cm. The larger the background field u_e the *smaller* is the maximum field as is seen from (2.24) and, for example, figure 7.

With the initial wave $\phi(x)$ such as in figure 6 shocks form from the $\phi'(x) > 0$ side when $t = {}_1t_c$ is the least t satisfying (2.28) with $\phi' > 0$. From (2.26) $u(\phi(x_0), t)$ is obtained on setting $v(u) = (\mu_1+\mu_3)\, u_m - \mu_3 u$ (since the shock forms at $u_e < u < u_v$) and integrating to give

$$u(\phi(x_0), t) = u_e(1 - e^{\lambda\mu_3 t}) + \phi(x_0)\, e^{\lambda\mu_3 t}, \tag{2.39}$$

where x_0, t are so restricted in (2.39) that $u_e \leqslant \phi(x_0) < u_v$ and $u_e \leqslant u \leqslant u_v$. Equation (2.39) with the first of (2.26) gives $u(x, t)$. Substitution of (2.39), with (2.36) from which $v(u) - v(u_e) = \mu_3(u_e - u)$ in the range in which we are concerned, into (2.28) gives ${}_1t_c$ as the least t satisfying

$${}_1t_c = \frac{1}{\lambda\mu_3} \log\left[1 + \frac{\lambda}{\{\phi'(x_0)\}_{\max}}\right], \tag{2.40}$$

where x_0 takes values in the range where $\phi'(x_0) > 0$ and $u_e \leqslant \phi(x_0) > u_v$.

In an analogous manner when $u_v \leqslant u \leqslant u_s$ and x_0 is such that $u_v \leqslant \phi(x_0) \leqslant u_s$,

$$u(\phi(x_0), t) = u_s(1 - e^{-\lambda\mu_2 t}) + \phi(x_0)\, e^{-\lambda\mu_2 t}, \tag{2.41}$$

where in (2.26) we used the appropriate $v(u) = \mu_2 u$ from (2.36). Equation (2.41) with the first of (2.26) gives the solution in x, t. With (2.41) the critical shock formation time ${}_2t_c$ is given, using (2.28), with $v(u_s)$ for $v(u_e)$ for convenience in this range, by

$$ {}_2t_c = -\frac{1}{\lambda\mu_2}\log\left[1-\frac{\lambda}{|\phi'(x_0)|_{\max}}\right], \tag{2.42}$$

where in (2.42) x_0 takes values in the range where $\phi'(x_0) < 0$ and $u_v \leqslant \phi(x_0) \leqslant u_s$. Note that in (2.42) if λ is large enough ${}_2t_c$ may not exist and no shock forms on the $\phi' < 0$ side. Thus in the *high* conductivity case with $\lambda \gg 1$ this model and solution could obtain physically since the limit on u_x ($u_x \geqslant -1/\lambda$) could still hold for all t. As mentioned above from a practical point of view if ${}_2t_c$ existed and was found to be sufficiently large that the wave would have passed through the sample in a time $t < {}_2t_c$ the solution could still be relevant: that is we would require ${}_2t_c > 1/v(u_e)$ at the least.

With ${}_1t_c$ and ${}_2t_c$ as in (2.40) and (2.42), which also give the respective critical x_0's, the u's at which shocks begin, namely $u_{11}({}_1t_c)$ $(= u_{12}({}_1t_c))$ and $u_{21}({}_2t_c)$ $(= u_{22}({}_2t_c))$, are then given by (2.39) and (2.41) with the appropriate t from (2.40) and (2.42).

Suppose, for example, that $\phi(x)$ may be approximated simply by

$$\left.\begin{aligned} \phi(x) - u_e &= x/\alpha \quad (0 \leqslant x \leqslant \tfrac{1}{2}X), \\ &= (X-x)/\alpha \quad (\tfrac{1}{2}X \leqslant x \leqslant X), \\ &= 0 \quad (X \leqslant x). \end{aligned}\right\} \tag{2.43}$$

The global physical restriction that $u_x \geqslant -1/\lambda$ requires from (2.43) that $\alpha \geqslant \lambda$. Thus, from (2.40), ${}_1t_c = (1/\lambda\mu_3)\log(1+\alpha\lambda)$ and from (2.42) ${}_2t_c$ is either infinite if $\alpha\lambda = 1$ or does not exist if $\alpha\lambda > 1$.

The complete picture of the wave problem posed by (2.21) with $v(u)$ from (2.36) is given by the first of (2.26) with (2.39) and (2.41) for $t \leqslant \min[{}_1t_c, {}_2t_c]$. For example, if we consider a point on the initial wave which has $u(\phi(x_0), 0) < u_v$ then $u(x,t)$ is given by (2.26) with (2.39) for $t \leqslant t_{0v}$, say, where t_{0v} is defined as the time it takes for that initial $u(\phi(x_0), 0)$ to reach u_v: t_{0v} is a function of x_0. For $t > t_{0v}$ we must use (2.41) with t replaced by $t - t_{0v}$ and $u(\phi(x_0), t_{0v})$ replacing $\phi(x_0)$ as the multiple of $e^{-\lambda\mu_2 t}$.

The first shock starts at $t = {}_1t_c$ and the shock speed, growth and strength can easily be found from the first of (2.29) together with (2.39), (2.41) and the first of (2.26) all taken at $x = x_{d1}$ with $u = u_{11}(t)$ at $x_{d1}-$ and $u = u_{12}(t)$ at $x_{d1}+$. The time for growth to the maximum stable shock can be found. The maximum size is $\bar{u}_{12} - u_e$ where, from the equal area rule (2.32), $\bar{u}_{12}(u_v)$ is given by

$$\int_{u_e}^{u_v}[(\mu_1+\mu_3)u_m - \mu_3 u]\,du + \int_{u_v}^{\bar{u}_{12}}\mu_2 u\,du = (\bar{u}_{12}-u_e)\mu_2\bar{u}_{12},$$

namely,
$$\bar{u}_{12} = u_e + \left[\frac{2}{\mu_2}(\mu_1+\mu_3)u_m(u_v-u_e) - \left(1+\frac{\mu_3}{\mu_2}\right)(u_v^2-u_e^2)\right]^{\frac{1}{2}}. \tag{2.44}$$

With the typical values as in (2.37) and (2.38), $\bar{u}_{12} \fallingdotseq 1{\cdot}6\times10^2$ $(u_v < \bar{u}_{12} < u_s)$.

Note that the shock speed $v(\bar{u}_{12}) < v(u_e)$. A typical solution is illustrated in figure 11 for a $\phi(x)$ which is effectively a rounded off version of (2.43).

The width of the wave continuously increases and the width $W(t)$ between the leading edge and the trailing shock for a $\phi(x)$ such as gives figure 11 is asymptotically $[v(u_e) - v(\bar{u}_{12})]t$, with $v(u)$ given by (2.36).

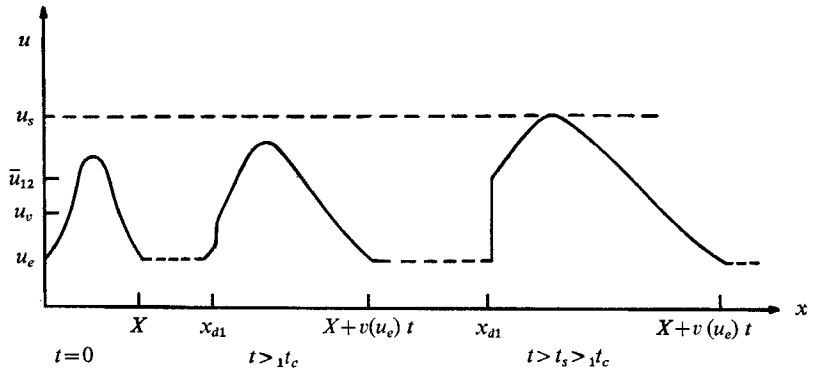

FIGURE 11. Typical solution with one shock where t_s is the time to reach the steady shock state.

2.8. *Discussion of results, experimental relevance and a simplified model for the Gunn effect*

The above phenomenological temperature independent model is primarily based on the assumption of a two-conduction band structure for the doped semiconductor. The above analysis is crucially dependent on the fact that there is a negative resistance region and that it is limited in extent with the ion velocity $v(E)$ having a relative minimum at E_v (see figures 3, 4 and 7). By considering a long sample, and so $J(t)$ can be taken as constant, the governing non-dimensional equation in one space variable is a non-linear second-order parabolic equation in which, for most practical situations, a small parameter ϵ, directly connected with the diffusion multiplies the highest derivative. The reduced, $\epsilon = 0$, equation is a perturbed conservation equation which is a first-order non-linear wave equation, for which wave-like solutions *cannot* propagate without change in shape. Since the effect of ϵ is to change the magnitude of the electric field by terms $O(\epsilon)$ this reduced equation should suffice for comparison with experiment. This change is still $O(\epsilon)$ when shocks are present: the role of ϵ in this case is simply to smooth out the shocks over a distance $O(\epsilon)$.

With the typical $v(E)$–E curves which with exceptions noted above seem to have been widely accepted two shocks, or steep regions, in the electric field can appear in the solution when $\epsilon = 0$ even if the initial wave satisfies the physical restriction on the maximum allowable negative slope of the electric field ($u_x \geqslant -1/\lambda$). (This ensures that the number density does not become negative.) It was shown in §2.7 for a particular case, although it holds generally, that if λ is sufficiently large the second shock at x_{d2} will not exist resulting in a typical wave shape similar to that in figure 11. It should be mentioned at this stage that

if a shock at x_{d2} does appear in the solution the singular perturbation solution cannot smooth it out sufficiently to accommodate the restriction $u_x \geqslant -1/\lambda$. In the vicinity of the shock it makes $u_x = O(-1/\epsilon) \ll -1/\lambda$ for $\epsilon \ll 1$ and $\lambda = O(1)$ or $\lambda \ll 1$. In the high conductivity case the effect of $\epsilon\lambda D_u(u)$ in (2.19) could, of course, substantially alter the situation as to a shock at x_{d2} but for constant $D(u)$ or slowly varying $D(u)$ it certainly could not. We must thus consider the solutions of the reduced equation (2.21) to be those which have to be used for experimental comparison: a singular perturbation analysis is unnecessary.

The role of $\phi(x)$ is very important. That such a wave appears in the sample at all is interesting. It is probably a property of the external circuit. Alternatively (if not the actual cause then perhaps a contribution to it) the appearance could be the result of a basic instability for which there is a preferred wavelength which grows at the expense of the energy from neighbouring disturbance waves. In this situation a $\phi(x)$ would be obtained which would not violate the physical restriction on the negative slope. Here there would be the question of why the wave grew only at one end.

Given $\phi(x)$ the solution depends crucially on whether or not $\max \phi(x) > u_v$ or is sufficiently close to it, or the sample is long enough, so that $u > u_v$ for some $t > 0$. If $u > u_v$ for some $t > 0$ or if $\phi(x) > u_v$ for some x we have the possibility of two shocks appearing. Even if the time ${}_2t_c$ is sufficiently large that the second shock would not appear with typical experimental times the typical shape of the solution in figure 11 is not what is observed experimentally (see Gunn 1967). The experimental curves tend to have a steep region equivalent to a smoothed shock at x_{d1}, in which u_{12} is effectively the maximum u in the solution, and the region $x > x_{d1}$ is monotonically decreasing as in figure 12. The broken line curves in figure 12 for $t \geqslant t_c$ are fairly typical experimental shapes.

We have seen that the possible appearance of the shock at x_{d2} and the shape of the wave for $x > x_{d1}$, as in figure 11, are a direct consequence of the existence of a limit to the negative resistance region, assuming $\phi(x)$ to have a maximum close enough to u_v so that $u > u_v$ for some $t \geqslant 0$. It is now suggested in view of the above analysis and the observed experimental results, as to shape, that for all situations of practical interest the effective $v(E)$–E curve is one in which the negative resistance region for $E > E_m$ is to all intents and purposes *unlimited*. This is in keeping with Butcher (1967) and Butcher & Fawcett (1966) (see figure 2). In this case $\lambda[v(u) - v(u_e)] < 0$ in (2.21) for *all* $u > u_e$ in the range of interest. From §2.4 this implies that only *one* shock at x_{d1} will appear at a time given by ${}_1t_c$. Further, only the first of (2.29) will be required and from the last of (2.25) all parts of the wave in which $u > u_e$ will grow without limit. The magnitude will be determined by the sample length (and λ and $v(u)$, of course).

By way of illustration we choose $v(u)$ from (2.36) as

$$\left.\begin{aligned} v(u) &= \mu_1 u \quad (0 \leqslant u \leqslant u_m) \\ &= (\mu_1 + \mu_3)\, u_m - \mu_3 u, \quad u_m \leqslant u < (\mu_1 + \mu_3)\, u_m/\mu_3, \end{aligned}\right\} \tag{2.45}$$

with the upper limit on u in (2.45) ensuring that $v(u) > 0$. A curve in which $v(u) > 0$, $v_u(u) < 0$ for $u > u_m$ will suffice theoretically. The shock will form at ${}_1t_c$ given by (2.40) and its speed of propagation, position and strength are then given

by the first of (2.29) with (2.30) for u_{11} and a similar set for u_{12}. With (2.45) these give (using (2.39))

$$\left.\begin{aligned}
\frac{dx_{d1}}{dt} &= v(u_e) - \tfrac{1}{2}\mu_3(u_{12}-u_{11}),\\
x_{d1} &= x_0 + v(u_e)\,t + (1/\lambda)\,[\phi(x_0(x_d-)) - u_{11}],\\
x_{d1} &= x_0 + v(u_e)\,t + (1/\lambda)\,[\phi(x_0(x_d+)) - u_{12}],\\
u_{11} &= u_e(1-e^{\lambda\mu_3 t}) + \phi(x_0)\,e^{\lambda\mu_3 t},\\
u_{12} &= u_e(1-e^{\lambda\mu_3 t}) + \phi(x_0)\,e^{\lambda\mu_3 t}.
\end{aligned}\right\} \tag{2.46}$$

for $t > {}_1t_c$. For a given $\phi(x_0)$, (2.46) can be solved.

As an example, consider the initial wave $\phi(x_0)$ to be as in (2.43) (see figure 12) where the physical condition from (2.20), namely $u_x(x,t) \geqslant -1/\lambda$ requires $\alpha \geqslant \lambda$. For $t \leqslant {}_1t_c = (1/\lambda\mu_3)\log(1+\alpha\lambda)$ from (2.40) we have, from (2.26) using (2.45),

$$\left.\begin{aligned}
x &= v(u_e)\,t + (1/\lambda)\,(u-u_e)\,[(1+\alpha\lambda)\,e^{-\lambda\mu_3 t}-1],\\
u &= u_e + (x_0/\alpha)\,e^{\lambda\mu_3 t},
\end{aligned}\right\} 0 \leqslant x_0 \leqslant \tfrac{1}{2}X,$$

$$\left.\begin{aligned}
x &= X + v(u_e)\,t + (1/\lambda)\,(u-u_e)\,[(1-\alpha\lambda)\,e^{-\lambda\mu_3 t}-1],\\
u &= u_e + [(X-x_0)/\alpha]\,e^{\lambda\mu_3 t}.
\end{aligned}\right\} \tfrac{1}{2}X \leqslant x_0 \leqslant X, \tag{2.47}$$

The wave form thus consists at any given t of two straight lines the leading part of which is flattening while the trailing part steepens. At $t = {}_1t_c$ the shock at $x = x_{d1}$ has formed and, as shown in figure 12, at its inception, has $u_{11}({}_1t_c) = u_e$ and from (2.47) with $x_0 = \frac{1}{2}X$ and $t = {}_1t_c = (1/\mu_3)\log(1+\alpha\lambda)$,

$$u_{12}(t_c) = u_e + \tfrac{1}{2}X/\alpha + \tfrac{1}{2}X\lambda = u(\tfrac{1}{2}X, 0) + \tfrac{1}{2}X\lambda. \tag{2.48}$$

For $t > {}_1t_c$ we use (2.46) with $u_{11}(t) = u_e$ for all $t > {}_1t_c$. We thus have only the first, third and fifth of (2.46), with the appropriate $\phi(x_0)$ given by the last of (2.47). The shock problem with these reduces to

$$\left.\begin{aligned}
\frac{dx_{d1}}{dt} &= v(u_e) - \tfrac{1}{2}\mu_3(u_{12}-u_e),\\
x_{d1} &= X_0 + v(u_e)\,t - (u_{12}-u_e)\,(1/\lambda)\,[(\alpha\lambda-1)\,e^{-\lambda\mu_3 t}+1],
\end{aligned}\right\} \tag{2.49}$$

which have to be solved subject to

$$\left.\begin{aligned}
u_{12}({}_1t_c) &= u_e + \tfrac{1}{2}(X/\alpha) + \tfrac{1}{2}X\lambda,\\
x_{d1}(t_c) &= v(u_e)\,{}_1t_c,\\
{}_1t_c &= (1/\lambda\mu_3)\log(1+\alpha\lambda).
\end{aligned}\right\} \tag{2.50}$$

Differentiating the second of (2.49) and using the first of (2.49) an equation for $u_{12}(t)$ is obtained, the solution of which is simply found to be

$$u_{12}(t) = u_e + X(\tfrac{1}{2}\lambda)^{\frac{1}{2}}\,e^{\lambda\mu_3 t}\,[e^{\lambda\mu_3 t} + \alpha\lambda - 1]^{-\frac{1}{2}}, \tag{2.51}$$

which on substitution into the second of (2.49) gives

$$x_{d1}(t) = X + v(u_e)\,t - [X/(2\alpha\lambda)^{\frac{1}{2}}]\,[e^{\lambda\mu_3 t} + \alpha\lambda - 1]^{\frac{1}{2}}. \tag{2.52}$$

For $x > x_{d1}$ the solution for u is, for any given t, a straight line and is given by the third of (2.47). Clearly these solutions for $v(u)$ as in (2.45) cannot hold for all time since all $u(> u_e) \to \infty$ and so $v(u) \to 0$ for some finite time. With the typical values for the constant parameter in (2.47) from §2.7 and $\lambda = O(1)$ times large compared with the time of passage across a typical sample are still allowable in (2.51) and (2.52) which can thus be used for comparison. The solution is illustrated in figure 12. If one compares this solution with the experimental curves in, for example, Gunn (1967) the similarity is very marked. The role of the singular perturbation ϵ diffusion term will be to smooth out the discontinuities as indicated by the dashed lines in figure 12.

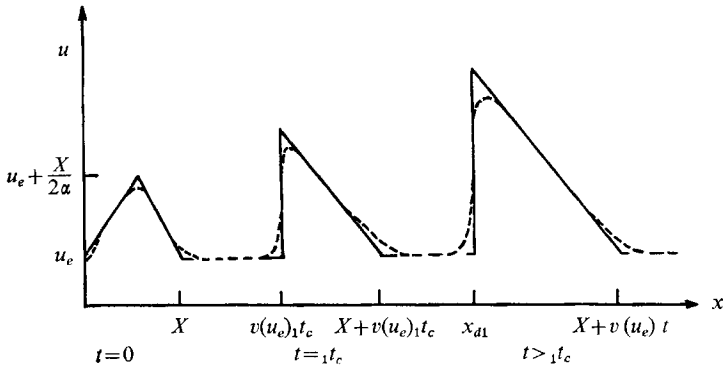

FIGURE 12. Theoretical solution for simplified model (see §2.8) which is similar to the experimentally observed wave motion. Broken line indicates the effect of diffusion in its role as a singular perturbation.

The width of the wave $W(t)$ for $t < {}_1t_c$ is simply X and for $t > t_c$ is given by

$$\begin{aligned} W(t) &= X + v(u_e)\, t - x_{d1}(t) \\ &= [X/(2\alpha\lambda)^{\frac{1}{2}}][e^{\lambda\mu_3 t} + \alpha\lambda - 1]^{\frac{1}{2}}. \end{aligned} \tag{2.53}$$

The excess voltage $V(t)$ from (2.35) is given simply by the area under the wave and bounded below by $u = u_e$. Since u as a function of x for $x_{d1} < x \leqslant X + v(u_e)\, t$ is simply a straight line, $V(t)$ is the area of the triangular wave and so, using (2.51) and (2.52)

$$\begin{aligned} V(t) &= \tfrac{1}{2} W(t)\, (u_{12}(t) - u_e) \\ &= \tfrac{1}{4}(X^2/\alpha)\, e^{\lambda\mu_3 t}. \end{aligned} \tag{2.54}$$

If we consider the number density then from (2.20) the accumulation layer ($n > n_0$) simply becomes more of a spike while the depletion ($n < n_0$) layer becomes wider.

In conclusion then it is suggested that a realistic model is simply that in which $v(u)$ is taken to be a monotonic decreasing function of u for $u > u_m$. The crucial parameters are thus λ and μ_3, the magnitude of $|v_u(u)|$ for $u > u_m$. Equation (2.21) is then very simply solved under any realistic conditions. Of practical importance is the magnitude and rate of growth of the maximum u: these are

given immediately by (2.51) for a piece-wise linear $v(u)$. For comparison with experiment in which the sample length is finite the solution (2.51) may be used for times up to $1/v(u_e)$.

There are clearly other problems of practical interest even with this simple monotonic model which could be considered. Of most importance, perhaps, are studies of the effect of finite length and a non-constant diffusion coefficient.

Other bulk effect phenomena in semiconductors, such as when part of the carrier density gets trapped at impurity sites (known as 'trapping') also give rise to similar non-linear wave equations: Ridley & Wisbey (1967), for example, discuss some aspects of this. Sze (1969) discusses other such phenomena.

It is a pleasure and a privilege to contribute this paper to an anniversary volume for Sydney Golstein with whom I had the good fortune to work for several years. His continual encouragement, insight and influence as a mathematician and kindness as a friend have been for me an enriching and delightful experience.

REFERENCES

BOTT, I. B. & FAWCETT, W. 1968 The Gunn effect in gallium arsenide. *Advances in Microwaves*, Vol. III. (Ed. L. Young). Academic.

BURGERS, J. M. 1939 Mathematical examples illustrating relations occurring in the theory of turbulent fluid motion. *Verhandel. Kon. Nederl. Akad. Wetensch.* **17**, 1–53. (See also 1948 A mathematical model illustrating the theory of turbulence. *Adv. Appl. Mech.* **1**, 171–199.)

BUTCHER, P. N. 1967 The Gunn effect. *Rep. Progress in Phys.* **30**, 97–148.

BUTCHER, P. N. & FAWCETT, W. 1965 The intervalley transfer mechanism of negative resistivity in bulk semiconductors. *Proc. Phys. Soc. Lond.* **86**, 1205–1219.

BUTCHER, P. N. & FAWCETT, W. 1966 Calculation of the velocity-field characteristic for gallium arsenide. *Phys. Lett.* **21**, 489–490.

BUTCHER, P. N., FAWCETT, W. & HILSUM, C. 1966 A simple analysis of stable domain propagation in the Gunn effect. *Brit. J. Appl. Phys.* **17**, 841–850.

CASE, K. M. & CHU, C. S. 1969 Burgers' turbulence models. *Phys. Fluids*, **12**, 1799–1809.

CONWELL, E. M. & VASSELL, M. O. 1966 High field distribution function in GaAs. *Trans Inst. Elect. Electron. Engrs, Electron Devices*, **13**, 22–27.

COURANT, R. & HILBERT, D. 1962 *Methods of Mathematical Physics*, Vol. 2. Interscience.

DUNWOODY, J. 1968 High-frequency sound waves in ideal gases with internal dissipation. *J. Fluid Mech.* **34**, 769.

GEL'FAND, I. M. 1959 Some problems of the theory of quasilinear equations. *Uspekhi Mat. Nauk*, **14**, 87–158. (English translation 1960, Office of Tech. Services, U.S. Dept. Commerce, Washington, D.C.)

GOLDSTEIN, S. 1953 On the mathematics of exchange processes in fixed columns. I and II. *Proc. Roy. Soc.* A **219**, 151–185.

GOLDSTEIN, S. & MURRAY, J. D. 1959 On the mathematics of exchange processes in fixed columns. III, IV, and V. *Proc. Roy. Soc.* A **252**, 334–375.

GUNN, J. B. 1963 Microwave oscillations of current in III–V semiconductors. *Solid State Commun.* **1**, 88–91.

GUNN, J. B. 1964 Instabilities in III–V semiconductors. *IBM J. Res. Develop.* **8**, 141–159.

GUNN, J. B. 1967 Solid-state microwave devices: domain and LSA modes. *Electro-Technology*, August, 69–75.

HILSUM, C. 1962 Transferred electron amplifiers and oscillators. *Proc. IRE* **50**, 185–189.

HUTSON, A. R., JAYARAMAN, A., CHYNOWETH, A. G., CORIELL, A. S. & FELDMAN, W. L. 1965 Mechanism of the Gunn effect from a pressure measurement. *Phys. Rev. Letters*, **14**, 639–641.

KNIGHT, B. W. & PETERSON, G. A. 1966 Nonlinear analysis of the Gunn effect. *Phys. Rev.* **147**, 617–621.

KROEMER, H. 1964 Theory of the Gunn effect. *Proc. IEEE* (*correspondence*) **52**, 1736.

KROEMER, H. 1968 Negative conductance in semiconductors. *IEEE Spectrum*, **5**, 47–56.

MURRAY, J. D. 1968 Singular perturbations of a class of nonlinear hyperbolic and parabolic equations. *J. Math. Phys.* **47**, 111–133.

MURRAY, J. D. 1970 Perturbations of nonlinear first order wave equations and their effect on discontinuous solutions. *S.I.A.M. J. Appl. Math.* (in press).

OLEĬNIK, O. A. 1957 Discontinuous solutions of nonlinear differential equations. *Uspehi Mat. Nauk.* **12**, 3–73. (English Translation. 1963 *AMS Translations*, Series 2, **26**, 95–173.)

RIDLEY, B. K. & WATKINS, T. B. 1961 The possibility of negative resistance effects in semiconductors. *Proc. Phys. Soc. Lond.* **78**, 293–304.

RIDLEY, B. K. & WISBEY, P. H. 1967 Nonlinear theory of electrical domain in the presence of trapping. *Brit. J. Appl. Phys.* **18**, 761–771.

RUCH, J. G. & KINO, G. S. 1967 Measurement of the velocity field characteristic of gallium arsenide. *Appl. Phys. Lett.* **10**, 40–42.

SELIGER, R. L. 1968 A note on the breaking of waves. *Proc. Roy. Soc.* A **303**, 493–496.

SEYMOUR, B. R. & VARLEY, E. 1970 High frequency, periodic disturbances in dissipative systems. Part I. Small amplitude, finite rate theory. *Proc. Roy. Soc.* A (in press).

SZE, S. M. 1969 *Physics of Semiconductor Devices*. Wiley.

VARLEY, E. & CUMBERBATCH, E. 1966 Non-linear, high frequency sound waves. *J. Inst. Math. Applic.* **2**, 133–143.

WHITHAM, G. B. 1967 Variational methods and applications to water waves. *Proc. Roy. Soc.* A **299**, 6–25.