

Zur Theorie der thermoelektrischen Effekte in Halbleitern

VON OTFRIED MADELUNG

Aus dem Forschungslaboratorium der Siemens-Schuckert-Werke AG., Erlangen

(Z. Naturforsch. 13 a, 22—25 [1958]; eingegangen am 9. November 1957)

Die Grundlagen der Theorie der thermoelektrischen Effekte in Halbleitern werden diskutiert. Dabei wird auf einige in der Literatur bestehende Unklarheiten in der Definition der Energiestromdichte und der in der BOLTZMANNschen Stationaritätsbedingung auftretenden Kraft $m \dot{v}$ hingewiesen.

Die Leitungsprozesse in einem Festkörper unter dem Einfluß angelegter Felder und Temperaturgradienten sind schon seit langem Gegenstand zahlreicher Untersuchungen gewesen. Die Theorie dieser Prozesse geht aus von der Annahme, daß der Körper einteilbar ist in makroskopisch kleine Bereiche, in denen lokales Gleichgewicht herrscht, denen also eine definierte Temperatur und ein elektrochemisches Potential zugeordnet werden kann. In diesen Bereichen gelten die bekannten thermodynamischen Relationen zwischen Entropie, innerer Energie, Teilchenzahl usw. Die Thermodynamik irreversibler Prozesse liefert dann, ausgehend von diesen Grundvorstellungen, Beziehungen zwischen den Teilchen- und Energiestromdichten und den treibenden Kräften des elektrochemischen Potentials und des Temperaturgradienten mit Koeffizienten, die unter bestimmten Modellvorstellungen berechnet werden können.

Bei der Ableitung der thermoelektrischen Effekte aus diesen Beziehungen sind nun in der Literatur eine Reihe von Vernachlässigungen vorgenommen worden, die es zweckmäßig erscheinen lassen, die Grundgedanken der Theorie nochmals zusammenzustellen und kritisch zu diskutieren.

I. Thermodynamische Grundlagen, Definition der Energiestromdichten

Wir gehen aus von der Gleichung

$$T ds = du + p dv - \sum_i \zeta_i dn_i, \quad (1)$$

wo s , u , v , n die Entropie, innere Energie, das Volumen und die Teilchenzahl (jeweils bezogen auf die Volumeneinheit), T die Temperatur und ζ_i das elektrochemische Potential der i -ten Teilchensorte bedeuten. Vernachlässigen wir Volumenänderungen und

beschränken wir uns auf eine Teilchensorte (die Erweiterung auf mehrere Teilchensorten bereitet keine Schwierigkeiten), so läßt sich (1) in folgender Form schreiben:

$$T \frac{\partial s}{\partial t} = \frac{\partial u}{\partial t} - \zeta \frac{\partial n}{\partial t}. \quad (2)$$

Unter Einführung einer Entropiestromdichte σ , einer Energiestromdichte w und einer Teilchenstromdichte j erhält man dann mit Benutzung der Kontinuitätsgleichungen

$$\frac{\partial u}{\partial t} + \text{div } w = 0, \quad \frac{\partial n}{\partial t} + \text{div } j = 0, \quad (3)$$

$$T \frac{\partial s}{\partial t} = -\text{div } w + \zeta \text{div } j \quad (4)$$

$$\text{ode. } \frac{\partial s}{\partial t} + \text{div } \sigma = w \cdot \text{grad } \frac{1}{T} - j \cdot \text{grad } \frac{\zeta}{T}. \quad (5)$$

Hier bedeutet das zweite Glied links die Divergenz der Entropiestromdichte $\sigma = 1/T(w - \zeta j)$, während die rechte Seite die lokale Entropieerzeugung ist¹. Da die rechte Seite von (5) eine lineare Funktion in den „Flüssen“ w und $-j$ ist, können wir diese als lineare Funktionen ihrer „Kräfte“ $\text{grad}(1/T)$ und $\text{grad}(\zeta/T)$ ausdrücken:

$$\begin{aligned} -j &= L_{11} \text{grad } \frac{\zeta}{T} + L_{12} \text{grad } \frac{1}{T}, \\ w &= L_{21} \text{grad } \frac{\zeta}{T} + L_{22} \text{grad } \frac{1}{T}. \end{aligned} \quad (6)$$

Eine Anwendung der ONSAGERSchen Beziehungen auf diese Gleichung ergibt schließlich noch $L_{12} = L_{21}$.

Gl. (6) stellt den Ausgangspunkt zur Berechnung der Leitungsprozesse dar. Bei der Anwendung dieser Gleichung ist jedoch zu berücksichtigen, daß die in (5) und (6) vorgenommene Aufteilung in „Flüsse“ und „Kräfte“ nicht eindeutig ist^{2,3}. Um die verschiedenen Möglichkeiten zu diskutieren, müssen wir

¹ vgl. etwa die Diskussion bei A. SOMMERFELD, Vorlesungen über theoretische Physik, Bd. V, § 21, Dieterichsche Verlagsbuchhandlung, Wiesbaden 1952.

² S. R. DE GROOT, Thermodynamics of Irreversible Processes, North-Holland Publishing Company, Amsterdam 1951.

³ H. B. CALLEN, Phys. Rev. 73, 1349 [1948].



uns zunächst über den Begriff der Energie eines Elektrons in einem Festkörper klar werden. Dabei beziehen wir uns auf die Verhältnisse in einem Halbleiter. Führt man neben der Energie E des Elektrons noch in das Energiediagramm des Halbleiters die drei charakteristischen Energieparameter $-e\psi$ (elektrostatisches Potential), ζ (elektrochemisches Potential) und E_L (Energie der unteren Kante des Leitungsbandes) ein, so läßt sich E aufteilen in:

$$E = -e\psi + E_0 \text{ (Abspaltung des elektrostatischen Anteils),}$$

$$E = \zeta + E_{\zeta} \text{ (Abspaltung des elektrochemischen Anteils),}$$

$$E = E_L + E_n \text{ (Aufteilung in „potentielle“ Energie } E_L \text{ und „kinetische“ Energie } E_n = m v^2/2).$$

Für die verwendeten Bezeichnungen vgl. Abb. 1.

Zu diesen drei neu eingeführten charakteristischen Energien E_0 , E_z und E_n lassen sich nun weitere Energiestromdichten w_0 , w_z und w_n definieren, die mit der gesamten Energiestromdichte w offensichtlich in folgender Weise zusammenhängen:

$$\begin{aligned} w_0 &= w - (-e\psi)j, & w_\zeta &= w - \zeta j, \\ w_n &= w - E_L j. \end{aligned} \quad (7)$$

$$\frac{\partial u}{\partial t} = -\operatorname{div} \mathfrak{w} \quad (9)$$

$$= -\operatorname{div} \mathfrak{w}_0 + e \, \mathbf{j} \cdot \operatorname{grad} \psi = -\operatorname{div} \mathfrak{w}_0 + \mathbf{i} \cdot \mathfrak{E} \quad (10)$$

$$= -\operatorname{div} \mathfrak{w}_s - \mathbf{j} \cdot \operatorname{grad} E_L = -\operatorname{div} \mathfrak{w}_s + \mathbf{i} \cdot \mathfrak{G} + \mathbf{i} \cdot \operatorname{grad} \frac{\zeta_0}{e} \quad (11)$$

$$= -\operatorname{div} \mathfrak{w}_n - \mathbf{j} \cdot \operatorname{grad} \zeta = -\operatorname{div} \mathfrak{w}_n + \mathbf{i} \cdot \mathfrak{C} + \mathbf{i} \cdot \operatorname{grad} \frac{E_{L_0}}{e}. \quad (12)$$

Nimmt man schließlich statt w als „Fluß“ in (5) und (6) eine der drei neuen Energiestromdichten, so erhält man als „Kräfte“:

<i>Flüsse</i>	<i>Kräfte</i>
$-j, w$	$\text{grad } \frac{\zeta}{T}, \text{ grad } \frac{1}{T} \quad (13)$

$$-j, w_0 \quad \frac{e}{T} \left(\mathfrak{E} + T \operatorname{grad} \frac{\zeta_0}{eT} \right), \operatorname{grad} \frac{1}{T} \quad (14)$$

$$-j, m_\zeta \quad \frac{1}{T} \text{grad } \zeta, \text{ grad } \frac{1}{T} \quad (15)$$

$$-j, w_n; \frac{e}{T} \left(\mathfrak{E} + T \operatorname{grad} \frac{\zeta_n}{eT} + \operatorname{grad} \frac{E_{Lo}}{e} \right), \operatorname{grad} \frac{1}{T} \quad (16)$$

Aus jeder dieser Kombinationen läßt sich dann ein lineares Gleichungssystem der Form (6) mit (wegen der ONSAGER-Beziehungen) symmetrischer Koeffizientenmatrix aufstellen. Die aus diesen verschiedenen

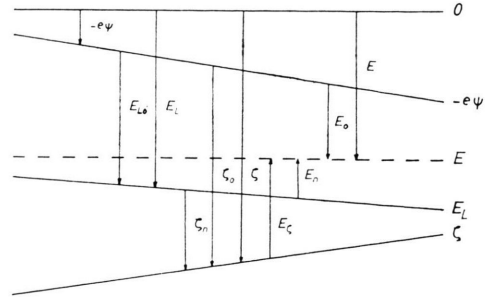


Abb. 1. Darstellung der in einem Halbleiter wichtigen Energien und Energiedifferenzen.

Die Entropiestromdichte wird dann mit diesen neuen Definitionen:

$$\sigma = \frac{1}{T} (w - \zeta j) = \frac{1}{T} (w_0 - \zeta_0 j) = \frac{1}{T} (w_n - \zeta_n j); \quad (8)$$

w_n ist nach Definition die Stromdichte der kinetischen Energie der Elektronen, während w_z wegen der Analogie der Beziehungen $w_z = \sigma T$ und $q = sT$ als Wärmestromdichte bezeichnet werden kann³.

Die Kontinuitätsgleichung (3) wird dann (mit $-e \mathbf{j} = \mathbf{i} = \text{elektr. Stromdichte}$, $-\text{grad } \psi = \mathbf{E} = \text{elektr. Feldstärke}$ und mit $\text{div } \mathbf{j} = 0$):

Möglichkeiten folgenden physikalischen Aussagen sind natürlich identisch. Man muß nur beachten, daß zum System (13), (14), (15) bzw. (16) die Kontinuitätsgleichung (9), (10), (11) bzw. (12) gehört. Dies ist in der Literatur häufig nicht genügend beachtet worden. So benutzt etwa JOHNSON⁴ das System (16) mit der Kontinuitätsgleichung (10), TER HAAR und NEAVES⁵ (15) mit (10). Dies hat im letzteren Fall z. B. zur Folge, daß in der Gleichung für die Wärmetönung neben der JOULEschen Wärme, der THOMSONschen Wärme und der Wärmeleitung noch ein Glied $-\dot{i} \cdot \text{grad}(\zeta_0/e)$ auftritt, das keinen physikalischen Inhalt besitzt.

⁴ V. A. JOHNSON, Progress in Semiconductors, Vol. I, S. 63, Heywood and Comp. Ltd., London 1956.

⁵ D. TER HAAR u. A. NEAVES, *Adv. Phys.* **5**, 241 [1956].

Mit Hilfe einer der vier möglichen Gleichungen des Typs (6) und einer der Gln. (9) bis (12) lassen sich dann die spezifische Leitfähigkeit und die verschiedenen thermoelektrischen Koeffizienten bestimmen. Dazu bringt man das gewählte Gleichungssystem zweckmäßig in die Form

$$\frac{1}{e} \text{grad } \zeta = \frac{1}{\sigma} \dot{\mathbf{i}} + \varphi \text{grad } T, \quad \mathbf{w}_\zeta = II \dot{\mathbf{i}} - \kappa \text{grad } T, \\ \frac{\partial u}{\partial t} = \frac{i^2}{\sigma} + \text{div}(\kappa \text{grad } T) - \mu_E \dot{\mathbf{i}} \cdot \text{grad } T, \quad (17)$$

wo σ , φ , II , κ und μ_E direkt die spez. Leitfähigkeit, die differentielle Thermospannung, den Peltier-Koeffizienten, die spez. Wärmeleitfähigkeit und den THOMSON-Koeffizienten bedeuten. Die aus den ONSAGER-Beziehungen geforderte Symmetrie der Koeffizientenmatrix der Gleichungen des Typs (6) sorgt dann automatisch für die Erfüllung der THOMSON-Beziehungen

$$\mu_E = T \frac{\partial}{\partial T} \varphi = T \frac{\partial}{\partial T} \left(\frac{II}{T} \right). \quad (18)$$

II. Explizite Bestimmung der Leitungskoeffizienten

Zur expliziten Bestimmung der in (17) enthaltenen Leitungskoeffizienten, d. h. zur Bestimmung der in (6) bzw. einer äquivalenten Gleichung auftretenden L_{ik} wird allgemein die BOLTZMANNsche Stationaritätsbedingung herangezogen. Diese liefert für die gestörte Verteilungsfunktion der Elektronen im isothermen Halbleiter in erster Näherung⁶

$$f = f_0 - \frac{l}{v} (\mathbf{v} \cdot \text{grad}_r f_0 + \dot{\mathbf{v}} \cdot \text{grad}_v f_0), \quad (19)$$

wo f die Verteilungsfunktion im Lagen-Geschwindigkeits-Raum (Anzahl der Elektronen im Volumenelement $(\mathbf{r}, d^3r, \mathbf{v}, d^3v)$, l die freie Weglänge und f_0 die ungestörte Verteilungsfunktion (Zustandsdichte mal FERMI-Verteilung)

$$f_0 = 2 \left(\frac{m^*}{h} \right)^3 \left(1 + \exp \left\{ \frac{E - \zeta}{kT} \right\} \right)^{-1} \quad (20)$$

bedeuten. In (19) ist die Gradientenbildung im ersten Glied rechts bei konstantem \mathbf{v} , im zweiten Glied bei konstantem \mathbf{r} vorzunehmen. Da in E nur der Anteil $E_n = E - E_L$ von \mathbf{v} abhängt ($E_n = m^* v^2/2$), setzt man in (20) zweckmäßig $E - \zeta = E_n - \zeta_n$ und

erhält

$$\text{grad}_r f_0 = - \frac{\partial f_0}{\partial E_n} \left(\frac{E_n - \zeta_n}{T} \text{grad } T - \text{grad } \zeta_n \right), \quad (21)$$

$$\text{grad}_v f_0 = \frac{\partial f_0}{\partial E_n} m^* \mathbf{v}. \quad (22)$$

Für $\dot{\mathbf{v}}$ wird in der Literatur allgemein $-(e/m^*) \mathfrak{E}$ angesetzt. Dies ist jedoch inkorrekt. Die auf das Elektron wirkende „äußere“ Kraft $\mathfrak{K} = m^* \dot{\mathbf{v}}$ ist nicht allein der Gradient des elektrostatischen Potentials, sondern der Gradient der gesamten potentiellen Energie des Elektrons. Diese ist aber gleich $E - E_n = E_L$. Im isothermen Fall ist diese Unterscheidung gegenstandslos, da dann $\text{grad } E_L = \text{grad}(-e\psi)$ ist. Bei Gegenwart eines Temperaturgradienten ist dagegen die Temperaturabhängigkeit von E_L mit zu berücksichtigen. Man hat also zu setzen:

$$\dot{\mathbf{v}} = - \frac{1}{m^*} \text{grad } E_L = - \frac{e}{m^*} \mathfrak{E} - \frac{1}{m^*} \frac{\partial E_{L0}}{\partial T} \text{grad } T. \quad (23)$$

Hiermit wird (19):

$$f = f_0 + \frac{l}{v} \frac{\partial f_0}{\partial E_n} T \mathbf{v} \cdot \left(\frac{1}{T} \text{grad } \zeta - (E_n - \zeta_n) \text{grad } \frac{1}{T} \right) \quad (24)$$

$$= f_0 + \frac{l}{v} \frac{\partial f_0}{\partial E_n} \mathbf{v} \cdot$$

$$\left\{ e \mathfrak{E} + \left[\frac{E_n}{T} + T \frac{\partial}{\partial T} \left(\frac{\zeta_n}{T} \right) + \frac{\partial E_{L0}}{\partial T} \right] \text{grad } T \right\}. \quad (24a)$$

Durch Einsetzen in die Gleichungen für den elektronischen Anteil der Stromdichten (unter Vernachlässigung aller *phonon-drag*-Effekte und der Gitterwärmeleitung)

$$\dot{\mathbf{j}} = \int \mathbf{v} f d^3 v, \quad (25)$$

$$\mathbf{w} = \int E \mathbf{v} f d^3 v, \quad (26)$$

$$\mathbf{w}_i = \int E_i \mathbf{v} f d^3 v \quad (i=0, \zeta, n) \quad (27)$$

folgen dann Gleichungen vom Typ (6), z. B. bei Verwendung von (24) in $\dot{\mathbf{j}}$ und \mathbf{w}_ζ direkt das Gleichungssystem (15) und damit die gesuchten Koeffizienten.

III. Diskussion

Trotz der oben erwähnten Inkonsistenz bei der Wahl der richtigen Kraft in der BOLTZMANNschen Stationaritätsbedingung sind die in der Literatur mitgeteilten Ergebnisse der Theorie fast durchweg richtig. Dies hat verschiedene Gründe.

So wird in vielen (vor allem frühen) Arbeiten über die thermoelektrischen Effekte in Halbleitern

⁶ vgl. etwa O. MADELUNG, Halbleiter, in Handbuch der Physik, Bd. 20, Springer-Verlag, Berlin-Göttingen-Heidelberg 1957.

als Energienullpunkt die Bandkante E_L selbst gewählt. Dies hat zur Folge, daß auf der rechten Seite von (23) das Glied $\text{grad } E_{L0}$ fehlt, also

$$\dot{v} = - \frac{e}{m^*} \mathfrak{E}$$

wird. Außerdem wird dann $v = v_n$, $E = E_n$ und $\zeta = \zeta_n$. Alle E_L enthaltenden Glieder der oben angegebenen Gleichungen fallen fort und diese Vernachlässigungen heben sich gerade bei der Berechnung der thermoelektrischen Effekte eines Störstellenleiters heraus. Berechnet man jedoch auf diese Weise die Beiträge der beiden Bänder getrennt und kombiniert sie in der richtigen Weise, so liefern die Ergebnisse bekanntlich nicht die von der ambipolaren Diffusion⁷ herrührenden Beiträge. Nur in diesen Beiträgen macht sich die Temperaturabhängigkeit der Bandkanten, also der Breite der verbotenen Zone, bemerkbar.

Andererseits wird häufig in (23) und damit in (24 a) das Glied $\text{grad } E_{L0}$ unberücksichtigt gelassen, das System (16) angewandt, für die Energiebilanz jedoch (10) benützt. Berechnet man dann etwa die Thermospannung aus dem THOMSON-Koeffizienten, so heben sich beide Fehler gerade wieder heraus.

Auch bei Berücksichtigung der Temperaturabhängigkeit der Bandkanten, aber Ersetzen von $m^* \dot{v}$ durch $-e \mathfrak{E}$ kommt man zu richtigen Ergebnissen, wenn man in (19) $\text{grad}_r f_0$ nicht bei festem E_n , sondern festem E bildet. Auf diese inkorrekte Weise kam u. a. der Verfasser in einer früheren Arbeit⁶ zur richtigen Verteilungsfunktion (24 a)⁸.

In vielen Fällen ist die Temperaturabhängigkeit der Bandkanten so klein, daß ihre Vernachlässigung

keinen großen Fehler bringt. Dies ist jedoch nicht immer so. Bei der Berechnung des ambipolaren Diffusionsstromes im nicht-isothermen gemischten Halbleiter⁷ ergibt sich beispielsweise eine Proportionalität dieses Stromes zu $(\Delta E + 4 k T)$. Ohne Berücksichtigung der Temperaturabhängigkeit der Bandkanten ist hier statt $\Delta E(T)$ nur $\Delta E(0)$ zu setzen. Etwa im Falle des InSb reduziert sich hierdurch bei Zimmertemperatur dieser Faktor um fast 30 Prozent⁹.

Ein weiterer Fall, in welchem der Ansatz

$$\dot{v} = - \frac{e}{m^*} \mathfrak{E}$$

versagt, findet sich in der Theorie der inhomogenen Halbleiter. Liegt die Inhomogenität nicht nur in einer ortsabhängigen Störstellenverteilung, sondern in einer variablen Bandstruktur (Mischkristalle örtlich variabler Zusammensetzung), so ist auch m^* und somit der „kinetische“ Anteil der Energie ortsabhängig. In diesem Fall ist

$$\dot{v} = - \frac{e}{m^*} \text{grad } E$$

zu setzen. Man erhält dann für die gestörte Verteilungsfunktion wieder Gl. (24), jetzt aber mit ortsabhängiger freier Weglänge l . Auch die phänomenologischen Gln. (17) bleiben erhalten, nur werden jetzt die Koeffizienten σ , φ , Π , κ und μ_E Variable des Ortes.

Herrn Prof. Dr. W. SCHOTTKY danke ich für eine lehrreiche Diskussion.

⁷ P. J. PRICE, Phil. Mag. 46, 1252 [1955].

⁸ Wir brauchen deshalb hier nicht auf die Ergebnisse der Theorie einzugehen und können für die explizite Angabe der Gleichungen für die thermoelektrischen Koeffizienten auf l. c.⁶ verweisen.

⁹ PRICE vernachlässigt in seiner Arbeit⁷ die Temperaturabhängigkeit von ΔE . Bei Berücksichtigung dieses Einflusses ist in seiner Gl. (11) zwar $\Delta = \Delta E(T)$ zu setzen, in (10) aber $\Delta = \Delta E(0)$ [$n_1 n_2 \sim \exp(-\Delta E(0)/k T)$]. Gleichzeitig tritt dann aber in seiner Gl. (2) ein zusätzliches Glied $(\mu_i/e) \text{grad } E_i$ ($E_i = E_L$ bzw. $-E_V$) auf, so daß im Endresultat der Gl. (12) wieder $\Delta = \Delta E(T)$ ist. Vgl. hierzu l. c.⁶, Gln. (35.3) und (36.4).