REVIEWS OF MODERN PHYSICS VOLUME 40, NUMBER 4 OCTOBER 1968

Model for the Metal–Nonmetal Transition in Impure Semiconductors

NOBUO MIKOSHIBA
Electrotechnical Laboratory, Tanashi, Tokyo, Japan

It is convenient to classify the impurity states in semiconductors into three groups: low, intermediate, and high. Even at high concentrations anomalous (i.e., nonmetallic) properties are observed. A simple "inhomogeneity" model is proposed in which the impurity states are regarded as spatial mixtures of metallic and nonmetallic regions. The model provides an explanation of the anomalous properties, at least qualitatively.

I. INTRODUCTION

As is well known, it is convenient to classify the impurity states in impure semiconductors into three groups: the low, intermediate, and high concentrations of impurities. In the low-concentration range at low temperatures, the dc Hall effect cannot be observed and the electrons localized in the impurity site can carry the current only via a hopping motion. In the intermediate-concentration range, the Hall effect is observed, but the carrier concentration estimated is lower than the concentration of impurities. The resistivity in this range shows an activation energy ϵ_2 which is strongly dependent on the concentration of impurities, N, and the overlap of impurity wave functions. In the high-concentration range, the Hall effect becomes nearly equal to that at room temperature. The resistivity is nearly independent of temperature, so that the electronic state can be regarded as a degenerate Fermi gas. The metal–nonmetal transition seems to occur at the concentration between the intermediate and high concentration ranges.

According to the NMR experiment[1] in P-doped Si, the high-concentration range can be classified into two regions, i.e., the transition and metallic regions. In the transition region the Knight shift of ^{29}Si is observed, but does not obey the Korringa relation. The Korringa relation holds in the metallic region, where the so-called impurity band seems to merge into the conduction band. The numerical values of various concentration ranges and some characteristic properties are given in Table I for Sb-doped Ge and P-doped Si.

Recent experiments have disclosed the following anomalous electric and magnetic properties in impure semiconductors:

(i) The negative magnetoresistance[2] and the logarithmic temperature dependence of resistivity[3] are observed in the high-concentration ranges.

(ii) The spin susceptibility obtained from ESR[4] shows no abrupt change at the metal–nonmetal transition.

(iii) The spin susceptibility obtained from ESR[4] in the high-concentration range is not proportional to $N^{1/3}$ and has a temperature dependence below $T=4.2°\mathrm{K}$.

(iv) The static susceptibility in the transition region[5,6] shows a temperature-independent diamagnetism, but does not obey the Landau–Peierls–Pauli formula for the degenerate Fermi gas.

In order to explain the activation energy ϵ_2 in the intermediate-concentration range, the D^- band model has been discussed by some authors.[7,8] However, if we take into account the overlap integral of the D^- wave functions, the bottom of the D^- band is not so lowered that the model cannot explain the properties of ϵ_2.[9] On the other hand, the localized spin model[10] has been proposed for explaining the anomalous properties, (i), mentioned above. This model is based on an analogy with the localized magnetic moment in dilute alloys. The model is not convincing, however, because the correlation effect which leads to the localized moment is not taken into account at the beginning in constructing the impurity band.

In this paper we propose a simple "inhomogeneity" model, in which the impurity states are regarded as spacial mixtures of the metallic and nonmetallic regions. The relative numbers of impurities and the relative volumes of metallic and nonmetallic regions are estimated by using the Poisson distribution of impurities and the critical distance between impurities in the Mott transition.[11] It is shown that all anomalous properties mentioned above can be explained at least qualitatively by this simple model.

---

[1] R. K. Sundfors and D. F. Holcomb, Phys. Rev. **136**, A810 (1964).

[2] W. Sasaki, J. Phys. Soc. Japan Suppl. **21**, 543 (1966).

[3] W. Sasaki, J. Phys. Soc. Japan **20**, 2293 (1965).

[4] S. Maekawa, J. Phys. Soc. Japan Suppl. **21**, 574 (1966).

[5] D. H. Damon and A. N. Gerritsen, Phys. Rev. **127**, 405 (1962).

[6] W. Sasaki, S. Maekawa, and J. Kinoshita, J. Phys. Soc. Japan **22**, 928 (1967).

[7] C. Yamanouchi, J. Phys. Soc. Japan **18**, 1775 (1963).

[8] H. Nishimura, Phys. Rev. **138**, A815 (1965).

[9] T. Kasuya and N. Mikoshiba (private communication).

[10] Y. Toyozawa, J. Phys. Soc. Japan **17**, 986 (1962).

[11] N. F. Mott, Phil. Mag. **6**, 278 (1961).

II. MODEL

Let us consider one of the impurities distributed at random in semiconductors. The probability of no neighbor closer than radius r is given by the Poisson distribution,

$$P=\exp[-(4\pi/3)Nr^3]. \tag{1}$$

We assume that the impurity which has no neighbor closer than a critical radius r_c belongs to the nonmetallic region. The concentration of nonmetallic impurities is then

$$N_n=N\exp(-t_c), \qquad t_c=(4\pi/3)Nr_c^3. \tag{2}$$

The concentration of metallic impurities is therefore given by

$$N_m=N[1-\exp(-t_c)]. \tag{3}$$

The relative volume of nonmetallic region can be estimated as follows. The probability of no neighbor closer than radius r and of finding one neighbor in a shell of thickness dr at r is

$$dP=4\pi Nr^2\,dr\exp(-t). \tag{4}$$

Let us assign a volume of $(4\pi/3)r^3$ to this impurity. The relative volume of nonmetallic region is then given by

$$V_n=\int_{t_c}^{\infty} t\,dt\exp(-t)=(1+t_c)\exp(-t_c). \tag{5}$$

Hence, the relative volume of metallic region becomes

$$V_m=1-(1+t_c)\exp(-t_c). \tag{6}$$

The average impurity concentrations in the both regions are given by

$$\langle N_n\rangle=N_n/V_n, \qquad \langle N_m\rangle=N_m/V_m, \tag{7}$$

respectively (see Fig. 1).

Let us now estimate the critical radius r_c by using the simplest method in the Mott transition.[11] In the degenerate Fermi gas with concentration N, the Coulomb potential of an impurity is screened as

$$V(r)=-(e^2/\kappa r)\exp(-qr),$$

$$q^2=(4\cdot 3^{1/3}/\pi^{1/3})(N^{1/3}/a). \tag{8}$$

If q is larger than $1/a$, there is no bound state. Inserting $(4\pi r^3/3)^{-1}$ into N and $1/a$ into q in Eq. (8), we obtain

$$r_c=(144/\pi^2)^{1/3}a, \tag{9}$$

where a is an appropriate Bohr radius.

For simplicity we assume that the impurity states in the nonmetallic region can be approximated by isolated impurity states, while in the metallic region the electronic state can be described by a degenerate Fermi gas in a band with the same band parameters as the conduction band.

III. MAGNETIC SUSCEPTIBILITY

The contributions from the nonmetallic region to the paramagnetic and diamagnetic susceptibilities are given by

$$\chi_{np}=N_n\mu_B^2/\rho kT, \tag{10}$$

$$\chi_{nd}\cong-\frac{N_ne^2a^2}{6\rho mc^2}\left[m_l^{-1}+\frac{1+(b/a)^2}{m_t}\right], \tag{11}$$

respectively.[12] Here μ_B is the Bohr magneton, ρ is the density of materials, a and b are the transverse and longitudinal radii of the ground-state wave function, m_t and m_l are the transverse and longitudinal effective masses in the conduction band.

In the metallic region we have

$$\chi_{mp}=3N_m\mu_B^2/2\rho E_f, \tag{12}$$

$$\chi_{md}=-(N_m\mu_B^2/2\rho E_f)[(2m_t+m_l)/3m_t^2m_l], \tag{13}$$

for the paramagnetic and diamagnetic components of susceptibility, where E_f is the Fermi energy.

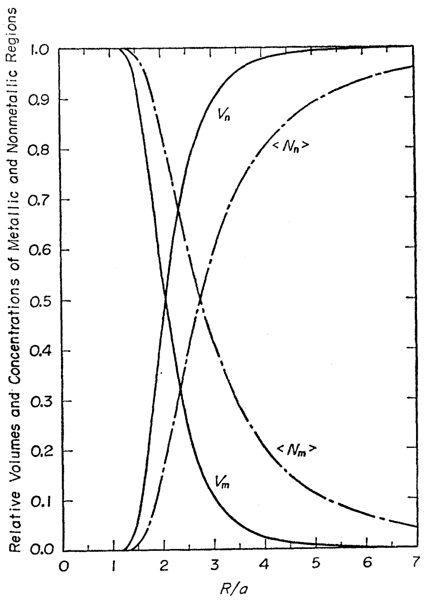

FIG. 1. Relative volumes of nonmetallic and metallic regions and average concentrations in each region.

---

[12] W. Kohn, Solid State Phys. 5, 257 (1967).

First of all, we consider the spin susceptibility obtained from ESR[4] in P-doped Si. If the ESR signal originates from the spins in the nonmetallic region, we have a relation

$$\chi(\text{ESR})=\chi_{np}. \qquad (14)$$

Comparing the theoretical curve of (14) with the experimental one, we conclude that the nonmetallic region occupies a more appreciable fraction for $N>10^{19}$ cm^{-3} than that given by Eq. (14). However, the behavior of χ_{np} is in qualitative agreement with the experiment, see Fig. 2.

Next, let us consider the static susceptibility[5] in Sb-doped Ge. The static susceptibility consists of four terms,

$$\chi(\text{static})=\chi_{np}+\chi_{nd}+\chi_{mp}+\chi_{md}. \qquad (15)$$

In our model the contribution of χ_{nd} is cancelled by that of χ_{np} at $T=1.24°\text{K}$. In the intermediate- and high-concentration ranges in Ge, χ_{nd} seems to be nearly cancelled by χ_{np} below $T=4.2°\text{K}$ because of the spin–spin interaction in the nonmetallic region. Thus, we can reduce Eq. (15) to

$$\chi(\text{static})\cong\chi_{mp}+\chi_{md}. \qquad (16)$$

The theoretical curve of (16) explains (a) the temperature-independent diamagnetism and (b) the deviation from the Landau–Peierls–Pauli formula, see Fig. 3.

The Knight shift of ^{29}Si in P-doped Si in the transition region has been shown to fall below the extrapolation of the $N^{1/3}$ line from the metallic region.[1] This indicates that not all of the impurities contribute electrons to the free-electron system. In our model the Knight shift is given by

$$K=(8\pi/3)\chi_{mp}P_f \qquad (17)$$

for ^{29}Si in the metallic region and $K\cong 0$ for ^{29}Si in the nonmetallic region. Here P_f is the mean wave-function probability density at the Fermi surface normalized in unit volume. Assuming that P_f is independent of N, we have compared the theoretical values of (17) with the experimental ones, see Fig. 4. Although the deviation from the $N^{1/3}$ line in the experiment is much larger than that given by Eq. (17), the concentration, $N=3\times10^{19}$ cm^{-3}, at which the deviation occurs is in good agreement with the experiment.

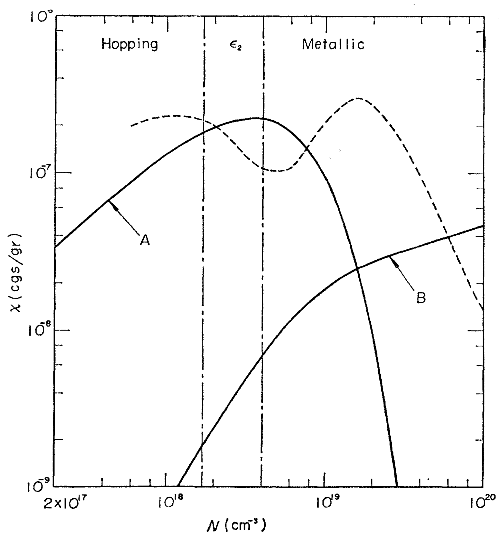

FIG. 2. The spin susceptibility versus the impurity concentration at $T=1.5°\text{K}$ in P-doped Si. ---: The experimental values obtained from ESR (see Ref. 4). A: The contribution from the nonmetallic region. B: The contribution from the metallic region.

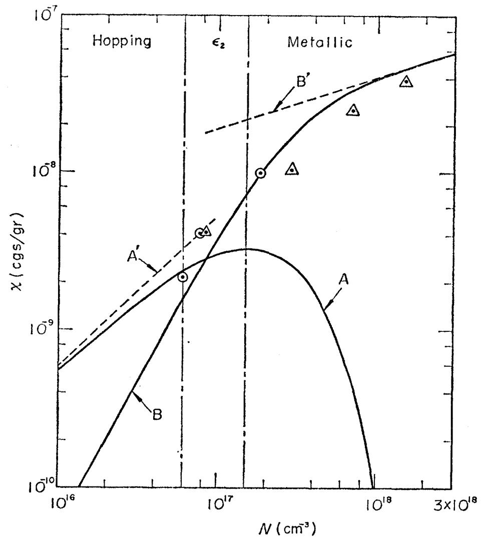

FIG. 3. The diamagnetic susceptibility in Sb-doped Ge. A: The diamagnetic contribution from the nonmetallic region. B: The Landau–Perierls–Pauli formula in the metallic region. △: After Hedgcock. ⊙: After Damon and Gerritsen (see Ref. 5).

IV. ACTIVATION ENERGY ϵ_2

In the framework of our simple model it is impossible to get information about the distribution of the metallic and nonmetallic regions, so that further assumptions must be made in discussing the transport phenomena. We make the following two assumptions (see Fig. 5):

(a) There is no metallic path network through the sample in the intermediate-concentration range, while in the high-concentration range there is such a path for carrying the current without activation energy.

(b) At finite temperatures in the intermediate concentration range, there is a probability that some electrons in the nonmetallic region are in the excited states which form a band due to the large overlap of wave functions. Because of a series connection of the metallic

TABLE I. Various concentration ranges and some characteristic properties in Sb-doped Ge and P-doped Si.

Concentration range	Low	Intermediate	High	
			Transition	Metallic
Conduction	ϵ_1, ϵ_3 No Hall effect $(\Delta\rho/\rho)_H \propto R^3H^2$	$\epsilon_1, \epsilon_2, \epsilon_3$ $1/R_Hec<N$ $\epsilon_2(H)=\epsilon_2(0)+\gamma H^2$	Anomalous T dependence $1/R_Hec=N$ $(\Delta\rho/\rho)_H<0$	
Magnetic susceptibility (static)	Deviation from the Curie law	Deviation from the Curie law (Si), T-independent diamagnetism (Ge)	Sharp increase of diamagnetism with increasing N (Ge)	Anomalous paramagnetic component
Magnetic susceptibility (ESR)	Various clusters	Line narrowing	Line broadening	
	Deviation from the Curie law		Anomalous T and N dependence	
Knight shift (^{29}Si)			Observable; the Korringa relation does not hold	The Korringa relation holds
Concentration (cm^{-3})	(P-doped Si) 1.7×10^{18} $(R/a=3.1)$	4×10^{18} $(R/a=2.3)$		3×10^{19} $(R/a=1.2)$
	(Sb-doped Ge) 6×10^{16} $(R/a=3.4)$	1.5×10^{17} $(R/a=2.5)$		1×10^{18} $(R/a=1.3)$

and nonmetallic regions, the resistivity is determined by that in the nonmetallic region and has the activation energy.

Let us make a rough estimation of the energy gap between the bottom of the band of excited states and the ground level. Since the overlap of wavefunctions of excited states is so large that we cannot use the tight-binding-like approximation, as in the discussion of the D^- band.[9] In the nearly free-electron approximation, the lowered energy of the conduction band due to impurities is given by

$$\bar{V}=\left(\frac{4\pi R^3}{3}\right)^{-1}\int_0^R\left(-\frac{e^2}{\kappa r}\right)4\pi r^2\,dr=-3V_0\left(\frac{a}{R}\right), \qquad 4\pi R^3/3=N^{-1}, \tag{18}$$

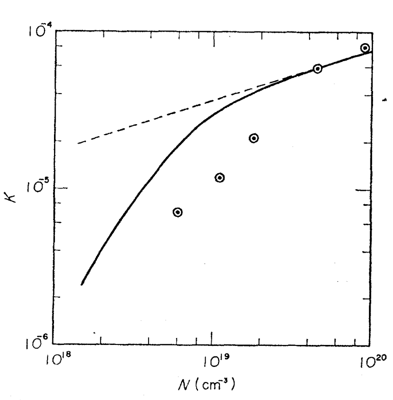

FIG. 4. The Knight shift of ^{29}Si in P-doped Si at $T=1.6$°K. ——: Calculated curve, ⊙: After Sundfors and Holcomb (see Ref. 1).

where $V_0=(e^2/2\kappa a)$ is the activation energy of the ground level. We can therefore define $2\epsilon_2$ as

$$2\epsilon_2=V_0-3V_0(a/R), \tag{19}$$

which leads to $\epsilon_2=0$ for $R/a=3$ in agreement with the experiment.

The magnetic-field dependence of ϵ_2 in the lower-concentration part of the intermediate-concentration range is also roughly estimated as follows. Let us approximate the excited band by the hydrogenic $2s$ state. The energies of the $2s$ and $1s$ states are changed by the quadratic Zeeman perturbation H_Z:

$$E(2s)=E(2s,0)+\langle\Psi_{2s}, H_Z\Psi_{2s}\rangle$$
$$=-(V_0/4)+(7e^2H^2a^2/2m^*c^2), \tag{20}$$

$$E(1s)=-V_0+(e^2H^2a^2/4m^*c^2), \tag{21}$$

where m^* is an appropriate effective mass. Thus, we have a relation

$$\epsilon_2=(1/2)\{E(2s)-E(1s)\}=\epsilon_2(0)+\gamma H^2, \tag{22}$$

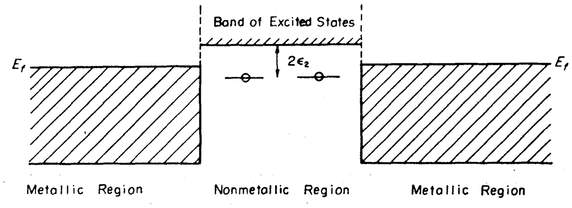

FIG. 5. A model for ϵ_2.

where

$$\gamma=(13/8)(e^2a^2/m^*c^2)$$

and

$$\epsilon_2(0)=(3/8)V_0. \quad (23)$$

The theoretical value $\gamma=3.15\times10^{-13}$ eV G^{-2} for Sb-doped Ge ($m^*=0.2m$, $a=47$ Å) is in reasonably agreement with the experimental values[7] $\gamma_{\text{exptl}}=0.5\sim4\times10^{-13}$ eV G^{-2}.

V. ANOMALOUS ELECTRICAL PROPERTIES AT HIGH-CONCENTRATION RANGE

In our model there is no difficulty to consider the localized spins in the Fermi sea. The negative magnetoresistance and the logarithmic temperature dependence of resistivity could be explained at least partially by the spin-dependent scattering or tunneling of the metallic electrons on or through the nonmetallic region.

Discussion of Mikoshiba's Paper

R. J. **Sladek** (Purdue University): It was pointed out on the last slide that there are certain regimes where there is negative magnetoresistance and where you can measure a Hall effect and so on. Recently we have been making some measurements on *n*-type gallium arsenide. Originally we thought we would have a very simple system because it has a simple conduction band. What we find is that in samples with relatively low carrier concentration $\sim2\times10^{15}$ (the total impurity concentration is perhaps larger) we can see negative magnetoresistance at low magnetic fields and at high magnetic fields we get an extremely large positive magnetoresistance. So I am adding another dimension to the argument, namely, it looks like we can have a transition between that so-called ϵ_2 region and the transition region as was indicated on Dr. Mikoshiba's slide. Also, with respect to trying to explain this ϵ_2 and the resistivity, our data on gallium arsenide shows that, when we can measure the Hall coefficient, the Hall coefficient has an activation energy that is smaller than the activation energy that the resistivity exhibits. So a couple of questions for Dr. Mikoshiba: Does he have any model for an activated mobility? And finally, his most interesting point that I would have a question on is, how does the negative magnetoresistance come about from the model he has proposed?

N. **Mikoshiba**: Unfortunately I haven't yet developed a theory on the difference of activation energy of the Hall effect and of the resistivity and also the negative magnetoresistance. It is still in the stage of speculation.

H. **Fritzsche** (University of Chicago): I think the Hall effect is a difficult problem in this region. Does the negative and the positive anomalous magnetoresistance come out from your tunnelling model from metallic regions through nonmetallic regions?

N. **Mikoshiba**: I think negative magnetoresistance and also the anomalous temperature dependence can be partially explained by the spin-dependent scattering of metallic electrons through the nonmetallic regions.

E. L. **Wolf** (Eastman Kodak): I would like to raise the question of the possible relevance of the theoretical work of I. M. Lifshitz [Advan. Phys. **13**, 483 (1964)] to our present discussions. It seems to me that his theory probably relates more to an Anderson transition than to a Mott transition, in that I understand him to use a one-electron approach. But he proceeds and takes into account very carefully the effect of the random distribution on the density-of-state spectrum of interacting shallow impurities. One feature of this which I think is striking is a sharp minimum in this density of one-electron states at the energy of the original ground state. He, in fact, finds an approximate symmetry in the density of states around this original energy level. This seems to raise the possibility that this minimum might look like a gap, in that activation processes might occur across the steep minimum and conceivably could have some relevance to the question of the small activation energy ϵ_2.

B. I. **Halperin** (Bell Telephone Lab): I would like to comment on the question of Dr. Wolf, with regard to the possible relevance here of Lifshitz's paper. I think first of all, that it is a bit questionable whether the minimum at the impurity energy that Lifshitz calculates is going to actually be there or whether it's an artifact of the theory. He has an expansion in which each term will be zero at the impurity level, but by the time you add up all the terms, it is not clear whether you will close up that minimum, or whether it will still be there. Even within the first few terms of this theory, it should be pointed out that the number of levels below the minimum is not equal to the number of levels above the minimum. In other words, even though a density-of-states minimum occurs at the energy of the isolated state, this is not the center of the impurity band, and if you put in one electron per impurity, you would find that the Fermi level was *above* this minimum. I therefore do not think that Lifshitz's minimum in the density of states could be relevant to the activation energy.

E. L. **Wolf**: I just wanted to reply to one aspect of Dr. Halperin's comment. I think that the physical origin of the density-of-states minimum that Lifshitz finds is easily understood within an approximation. That is, consider that in a random array each impurity has usually a unique nearest neighbor. If one makes the approximation of pairing up impurities, and assumes that the impurities interact as hydrogen atoms interact in forming a hydrogen molecule, the singlet and a triplet states are split nearly symmetrically from the original ground state level. This, of course, gives the minimum. Now there is the question of what about the next-nearest and other neighbors. I can't say anything about this question of convergence, other than that it would seem to warrant serious attention.

F. **Stern** (I.B.M.): There is another experimental system which might show some of the phenomena that we have been talking about at this conference, and that is the inversion layer in a semiconductor. There we can produce a two-dimensional system of electrons near the surface. We can vary electron concentration by varying the electric field that's applied, and (in principle) we can vary the impurity concentrations present near the surface. Now we have in that system an additional degree of freedom, because the electron concentration is not tied to the impurity concentration. The experiments of Fang and Fowler [F. F. Fang and A. B. Fowler, Phys. Rev. (to be published)] may show at low temperatures effects that could be interpreted as metal–nonmetal transitions [F. Stern and W. E. Howard, Phys. Rev. **163**, 816 (1967), pp. 829–830].

M. **Pollak** (University of California, Riverside): I would like to ask about the relation of your calculation of the D^- states to the interesting model you propose. We know from general considerations of the Mott transition, and also from experiments, that for small impurity concentrations there is only one state per impurity atom accessible to the carriers. On the other hand we know that in the high-concentration region there are two states per impurity atom available to the carriers, the additional states presumably being some linear combinations of D^- states. In your model of the intermediate region there are spatially alternating regions of the first type with regions of the second type. There seems to be a contradiction between the accessibility of the D^- states to the carriers in the metallic regions and between your calculations that the energy of such states does not decrease noticeably with concentration. Does that contradiction imply that your calculation must break down for high concentrations?

N. Mikoshiba: The difference between the D^- band model and my model is that there are conduction band (or impurity band) regions and isolated impurities in my model. At the intermediate concentrations some part of the specimen becomes metallic but there are no connected paths through the sample. In some regions, electrons are in the excited levels and form their band. In the D^- band model electrons transfer to atoms already occupied by an electron and in this case there is no lowered energy because of the large overlap of the D^- wave functions and the weak potential due to neutral atoms. In my case because of the large Coulomb interactions, the bottom of the excited energy band can be lowered. Now I would like to answer the question by Dr. Pollak. As to correlations and repulsions between electrons, I have made a drastic assumption that these electron–electron interactions are very important in the nonmetallic regions but can be neglected in the metallic regions. Further investigations will be required on this important point.

R. J. Sladek: Experimentally, we had considered a possibility of there being a metallic and nonmetallic region in our gallium arsenide. Probably the ones we were thinking of were on a larger scale than those which Dr. Mikoshiba has been talking about. The result is the following: We had a mass-spectrographic analysis made as a function of distance along the sample and for a number of different samples of about the same excess donor concentration. Some samples, as far as this could tell, looked very uniform and some didn't. So the question is unanswered, but we have already look for, in a sense, something related to Dr. Mikoshiba's two types of regions.