A.H. Wilson's theory on electronic semiconductor

Ziyi Zhao

November 28, 2018

1 Introduction

Semiconductors, first discovered at the end of nineteenth century, have gained enormous interest from academia and industry alike. This paper aims to discuss the first successful explanation based on then newly developed quantum mechanics [1]. Wilson's *The Theory of Electronic Semi-Conductors. II* was built on Bloch's work and accounted impurity as the reason why semiconductors are possible. Wilson included in this paper an oversimplified calculation of several properties, which matches some of the experimental results available at that time surprisingly well. Although these calculations are proven inaccurate later, the central idea of band overlapping remains to be a useful qualitative tool for understanding experiments results, for instance, nonmetal-to-metal transitions. This paper is organized by a short summery of Wilson's life, his original paper, the breakdown of the original theory, how it is used in nonmetal-to-metal transition and some other ways impurity is used, beyond what Wilson could have predicted.

2 Life and contribution of A.H. Wilson

After receiving a BA degree of Mathematics from Cambridge University, Wilson worked on application of quantum mechanics to electron conduction which led to the development of the theory of semiconductor. This won him the 1932 Adam Prize. He continued to work on *rectification* and *photoelectric effect*, before he turned to nuclear physics from around 1936 till WW2. After war he left academia and joined industry. He worked at Courtaulds, a textile company, and Glaxo, a pharmacy, with short interruption in between until retirement.

3 The Original Paper

By the time of this paper it was already recognized that the Drude model and Sommerfeld model are not satisfactory. Bloch had calculated the temperature dependent conductivity which works well in both high and low temperatures, but at 0. The motivation of this paper stems from the observation that Bloch's theory predicts infinite conductivity at $T = 0\ K$ for all solids, leaving no room for the existence of semiconductors or insulators. It was also incapable of explaining the dependence of conductivity on impurity [2]. These facts suggest that semiconductor differs from metals not only quantitatively, but rather qualitatively. Refraining the scope of semi-conductor on those whose conductivity reduces as impurity being removed, Wilson proposed that nonmetals are characterized by the energy gap between the highest occupied and lowest vacant levels. He further asserts that semiconductor is special since the impurity introduces energy levels close to the vacant conduction band, and under thermal or other excitement, bound electrons from the impurity can jump to the vacant band and become free ones. In his original drawing (Figure 1), band 1 is completely filled and band 2 is empty, while $A - B$ is the impurity energy level.
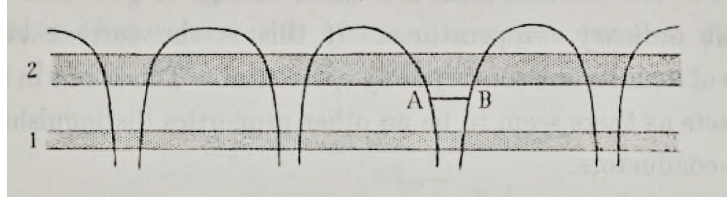
Figure 1: Schematic drawing of energy bands. Adapted from [1].

To verify his theory, Wilson included a calculation of the equilibrium distribution, conductivity and Hall effect coefficient and compared the result with experimental values and other known facts. Purely as a demonstration, Wilson did the simplest version of the problem. He assumed a simple cubic lattice, tightly bound electrons, no electron interaction in Hamiltonian and only one possible transition ($A - B$ to 2), among many others.

3.1 Equilibrium distribution

The derivation starts from finding the allowed energy of the second band. Since it should be close to the p orbitals, use the following ansatz:

$$E = E_2 + \epsilon_2 \quad \psi = \sum_{i,j,k} b_{i,j,k} \chi_{i,j,k}$$

where i, j, k are index for lattice points and χ_{ijk} is the solution of an isolated atom with energy E_2. To find b_{ijk}, consider the first order approximation,

$$\sum_{i,j,k} b_{i,j,k} \int d^3r \, (\epsilon_2 - U'_{i,j,k}) \chi_{i,j,k} \chi_{i',j',k'} = 0$$

Consider only nearest neighbor overlap of the wave function and use the convention:

$$\int d^3r \, U' \chi_{i,j,k} \chi_{i,j,k} = \alpha, \quad \int d^3r \, U' \chi_{i,j,k} \chi_{i,j+1,k} = \beta$$

The allowed energy can be written as:

$$E_{nlm} = W_2 + 6\beta - 2\beta[\cos(n) + \cos(l) + \cos(m)]$$

where we assume W_2 collects the α and E_2 term. $n, l, m = 2\pi G * (0, \pm1, \pm2, ...)$ is relabel of i, j, k. G^3 is total number of pure atoms in this simple cubic crystal. Finally, Taylor expand gives:

$$E_{nlm} = W_2 + \beta(n^2 + l^2 + m^2)$$

Label the energy level of impurity atom W_1 and zero point energy W_0. Applying Fermi-Dirac distribution $n_0(n, l, m) = \dfrac{1}{e^{\frac{E_{nlm} - W_0}{k_B T}} + 1}$ on the impurity atom, Wilson concluded that n_0 satisfy Maxwell distribution:

$$n_0(n, l, m) = \sqrt{4\pi^3 \frac{N_0}{G^3} \left(\frac{\beta}{\pi k T}\right)^{3/2}} e^{-\frac{W_2 - W_1}{2kT}} e^{-\frac{\beta \rho^2}{kT}} \tag{1}$$

where $\rho^2 = n^2 + l^2 + m^2$. And the total number of electron on the second band:

$$N = \sqrt{\frac{N_0 G^3}{4\pi^2} \left(\frac{\pi k T}{\beta}\right)^{3/2}} e^{-\frac{W_2 - W_1}{2kT}} \tag{2}$$

2

which takes the expected form. This is itself a victory. It is to be noticed that $\frac{W_2-W_1}{2}$ is predicted as the "energy gap" rather than $W_2 - W_1$. But using the conductivity derived from this equation (see 3.2), surprisingly, the former expression matches better with the experimental value by Vogt on cuprous oxide. They both lead to the conclusion that $\frac{W_2-W_1}{2}$ in cuprous oxide is $0.3\ eV$. This approximation turns out to be quite lucky.

3.2 Conductivity

The method was established in Wilson's previous paper and the reference therein [2]. The first order approximation of the change in $n_0(n, l, m)$ is found by considering the applied electric field and the elastic lattice vibration. The result is:

$$\sigma = A\frac{N}{G^3}T^{-3/2} = B\sqrt{\frac{N_0}{G^3}}T^{-3/4}e^{-\frac{W_2-W_1}{2kT}} \tag{3}$$

where A, B are constants. We notice that the $\frac{1}{2}$ factor in the energy reappears in these expressions. Wilson compared this result with the classical result:

$$\sigma = \frac{4}{3}\frac{e^2 l}{\sqrt{2\pi mkT}}\frac{N}{(Ga)^3} \tag{4}$$

By recognizing $l \propto T^{-1}$, Wilson argued that his result has the same N and T dependence as the classical one, which is another victory of his theory.

4 Breakdown

As a result of oversimplification, some of the ignored aspects in Wilson's paper turns out to be important enough to change the qualitative picture. To address the limit of Wilson's theory, Mott insulator and effective mass are presented as examples.

4.1 Mott insulator

Electron-electron interaction open all sorts of complexity and leads to lengthy theoretical development. While these theories are out of the scope of this discussion, it is good to know that electron-electron interaction can be fairly important [3]. One of the first examples, NiO is nonmetallic, despite it has odd number of electrons in its unit cell. A simple expression for the energy gap is:

$$E_{gap} = U - 2zt$$

When $U \gg t$, the solid is insulator, which is exactly the case for Mott insulators.

4.2 Effect mass

Predicting semiconductors is all about obtaining the correct band structure. Another line of critique against Wilson's paper is his adoption of tight binding model. Semiconductors usually do not have gap greater than $2\ eV$, while a tight binding model can only be expected to work for deep levels [4]. In contrast, effective mass is a good example for shallow bands. For the simplest case, the final result is easy to understand. The energy of the bottom of the simple band is :

$$E_k = E_c + \frac{\hbar^2 k^2}{2m_e^*}$$

It takes the same form as the solution of free electron, but with $m_0 \mapsto m_e^*$. The shift can qualitatively understood by the "dielectric effect" of electron face when being "put back" into the medium, when Coulomb potential undergoes $r \mapsto \epsilon r$ and $m_e^* = m_0 \epsilon^2$. Without being further involved into derivation, experiment results (in Figure 2) showed good agreement of this approach in Silicon and Germanium donors. This provides motivation of building theories around shallower energy gaps.

TABLE I. Spacings of selected excited states of shallow donors in Si and Ge as calculated by Faulkner (1969) compared with experimental[a] values. All energies are in meV.

States	Theory	P	As	Sb	Bi	Li	S^0	$S^+/4$
Si								
$2p^{\pm}-2p^0$	5.11	5.06	5.12	5.06	4.94	5.13	5.2	5.15
$3p^0-2p^{\pm}$	0.92	0.93	0.86	0.95	0.93	0.88	0.7	1.08
$4p^0-2p^{\pm}$	3.07	3.11	2.6	$\cdots$	2.61	$\cdots$	$\cdots$	$\cdots$
$3p^{\pm}-2p^{\pm}$	3.28	3.27	3.25	3.34	3.31	3.28	3.1	3.45
$4p^0-2p^{\pm}$	4.07	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$4p^0-2p^{\pm}$	4.17	4.21	4.3	4.33	4.34	4.19	$\cdots$	4.35
$4p^{\pm}-2p^{\pm}$	4.21	4.21	4.3	4.33	4.35	4.19	4.35	$\cdots$
$5p^{\pm}-2p^{\pm}$	4.97	4.95	4.9	$\cdots$	5.26	4.93	$\cdots$	$\cdots$
Ge								
$2p^{\pm}-2p^0$	3.02	3.03	3.02	3.02	3.02	3.08	3.04	
$2p^{\pm}-3p^0$	0.83	0.83	0.83	0.84	0.88	0.84	$\cdots$	
$3p^{\pm}-2p^{\pm}$	0.69	0.69	0.70	0.69	0.66	0.67	$\cdots$	

[a]Experimental values from Aggarwal and Ramdas (1965) for Si:P, As, Sb, Bi; from Reuszer and Fisher (1964) for Ge:P, As, Sb, Bi; from Aggarwal, Fisher, Mourzine and Ramdas (1965) for Si:Li₁ and Ge:Li₁; and from Krag and Zeiger (1962) for Si:S.

Figure 2: Theory and experiment results for Si and Ge. Adapted from [4].

5 Wilson Nonmetal-to-metal Transition

Although previous analysis has shown that Wilson's model is not ideal for quantitative understanding of either Mott insulator or shallow levels that dominates conductivity in semiconductors, it still provides a handy qualitative understanding for nonmetal-to-metal transition of group II elements as a response to change in density, whereas band overlapping plays a fundamental role. This phenomenon is to be distinguished from Mott-Hubbard, Peierls, Wigner-Verwey, Slater and Anderson transitions transition, for they each governs different elements and assumes different mechanism [5].

5.1 traditional Wilson transition

This process is most evident if we inspect the energy band as a function of interatomic distance in Figure 3.

The normally separated $n(s)$ and $n(p)$ orbitals can be brought close together as distance decrease. Therefore, after mixing, the original insulator transits to a metal.

5.1.1 Mercury

One of the prominent examples is the transition occurs in low-density liquid mercury or mercury cluster [6, 7]. In liquid, the experimental result is shown in Figure 4.
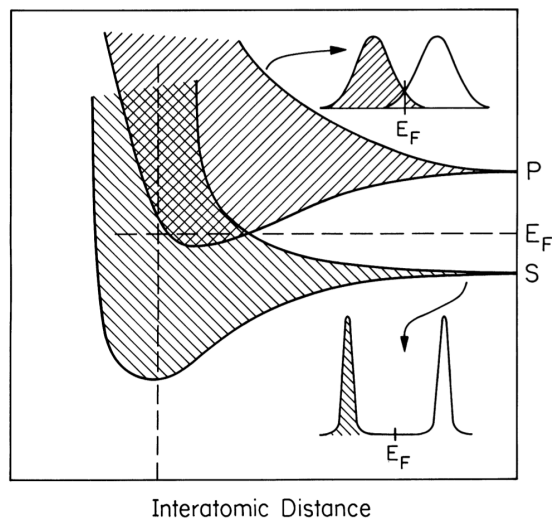
Figure 3: Schematic drawing of the band formation in a solid composed of divalent atoms, such as the alkaline earths, as a function of lattice separation. Adapted from [5].
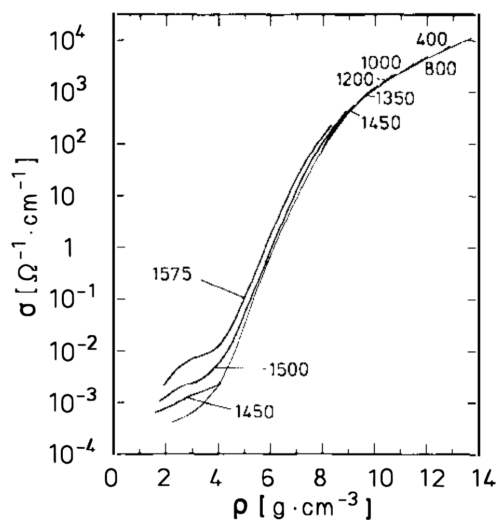

Figure 4: Conductivity of mercury with respect to density. Adapted from [6].

Wilson transition is a suitable description since isolated a Hg atom has a filled $6s$ orbital and an empty $6p$ orbital. When in form of a cluster, the bands broaden, leaving a gap Δ as shown in Figure 5a, which decreases as increasing the average number n of a cluster. Authors report that the gap closes at around $n = 100$, where the cluster become metallic, as shown in Figure 5b. In this plot the ionization potential is used as indication of the bond type. Note that only ■ labelled data is produced in the cited paper, the ○ and × labels are from other sources. The solid lines are represents the bulk work function. The important message is that as the distance $R^{-1} \propto n^{-1/3}$ decreases, the system becomes metallic.

5.1.2 Zinc cluster

Kostko *etc.* discuss the electronic properties of Zinc cluster, whose size ranges from $n = 3$ to $n = 107$. Authors assert that, when the system size is small $n \leq 18$, the metal to insulator of the system matches

5

(a) Band structure of Mercury of different form.
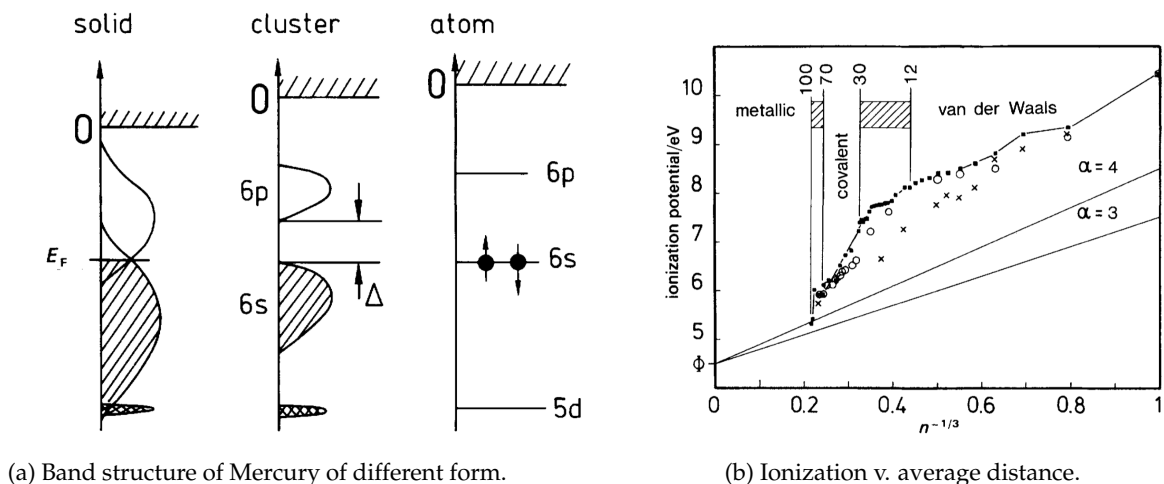
(b) Ionization v. average distance.

Figure 5: Adapted from [7].

the picture of Wilson transition, where the band gap closes. When the system is large, $n \geq 32$, the system behaves like free electron model since the band overlap allows electrons to be fully delocalized [8]. The free electron behavior is irrelevant to our discussion and we focus only on the part where n is small.

Contrary to previous example where the transition is triggered by the change in density, this metal to insulator transition in clusters is caused by the change in number nearest neighbors. "Band gap closure" in these systems is defined to be the situation where the distance between bands is equal to the average distance between energy levels within one band. The authors use a photoelectron spectroscopy (PES) setup for the experiment, to determine the binding energy of different sizes. The result is presented in Figure 6.
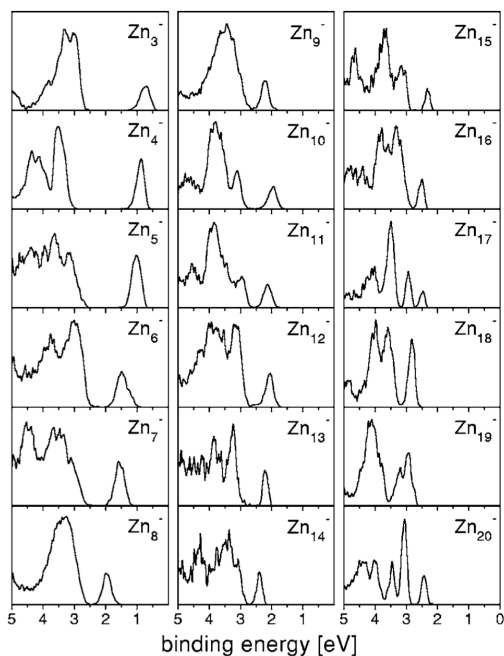

Figure 6: Photoelectron spectra of zinc cluster anions, adapted from [8].

6

The way to read this graph is to notice that all plots corresponding to $n \leq 18$ has a single peak at the lower energy side (right) and multiple peaks at the higher energy side. The single peak is supposed to be the single electron in the highest occupied state, a p-band state in the case of mercury and zinc. This feature is not present in the larger clusters. Authors interpret the disappearance of the single peak as evidence of closure of band gap since the top band then contains more than one electron.

The authors also compare the band gap of zinc cluster and that of mercury. They note that, as expected from Wilson's model where band gap is proportional to the energy difference s orbital and p orbital, the gap of zinc should be always smaller than that of mercury, which matches the result shown in Figure 7.
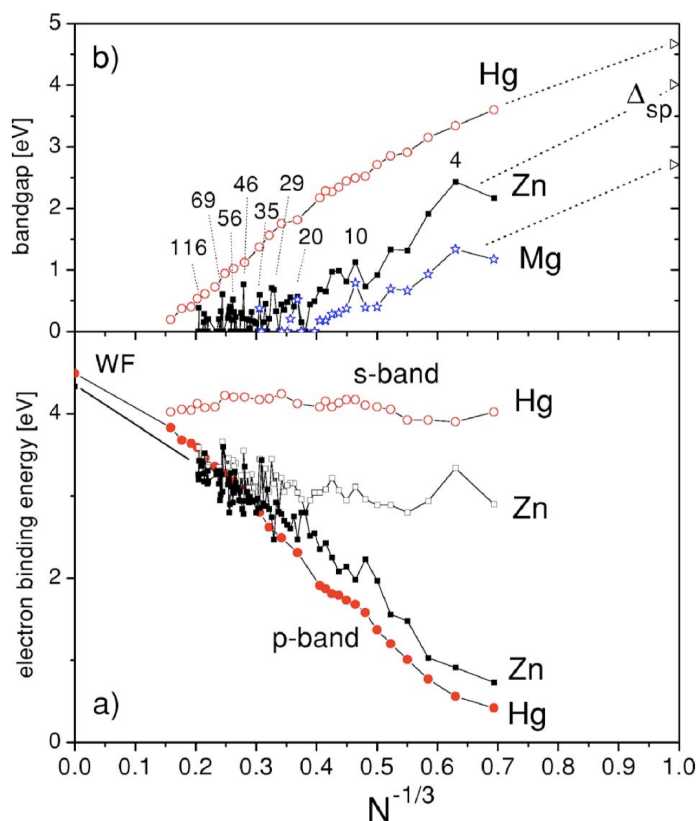

Figure 7: Band gap of zinc, mercury and magnesium, adapted from [8].

5.2   type-II

In this paper, Yonezawa *etc.* introduce a "type-II" Wilson transition in the sense that band overlapping remains to be the effective description, while macroscopic properties change in a contradicting way as in normal Wilson transition [9]. The authors report an insulator to metal transition in liquid Se as the density is first reduced, which is the opposite of expectation. Further reducing the density causes the liquid Se to return to an insulating state. A model is constructed as following:

The system is composed of chains with rotation angle to be $120°$ of an indefinite length. The bond length is r_1 while the separation is r_{sep} head-to-head. These parallel line form a triangular prism, as shown in Figure 8a. The energy bands are calculated with numerical methods assuming a psuedo potential.

The authors stress that the most important feature is that, the lowest energy level of the anti-binding levels (denoted as σ^* in Figure 8b) decreases as the separation increases, causing a band overlap to occur at certain
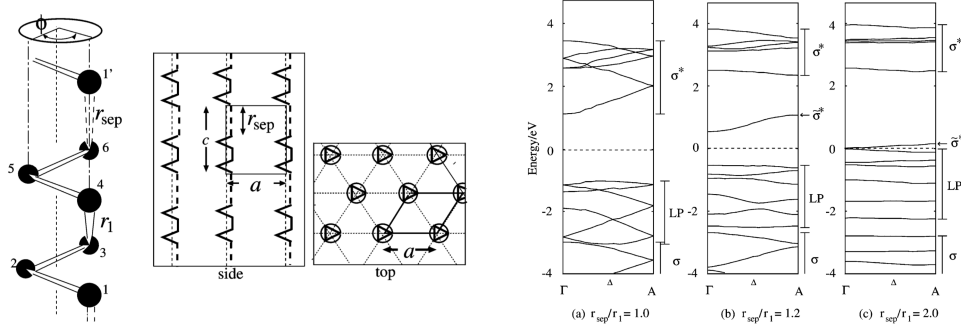
(a) Schematic drawing of model structure.

(b) Calculated energy bands as function of r_{sep}.

Figure 8: Adapted from [9].

point. The lowest energy level in σ^* is denoted as $\tilde{\sigma}^*$. This phenomena can be evidently shown in Figure 9. We can see that the anti-binding band overlaps with the lone pair band (labelled *LP* in the Figure 9).
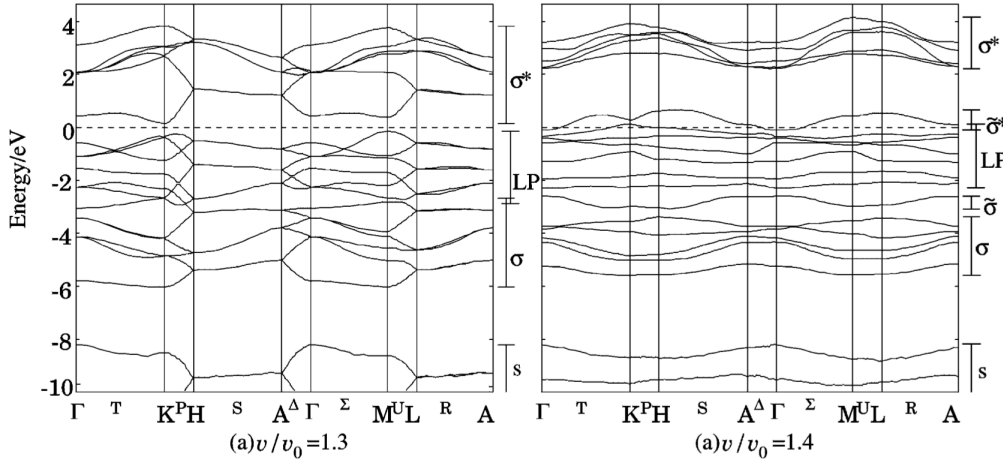

Figure 9: The band structure with respect to volume.

This concludes that band overlapping as a suitable qualitative way of thinking such nonmetal-to-metal transitions. Authors note that, besides Group II elements, Group IV elements, including *C*, *Si*, *Ge* and $\alpha - Sn$ are subject to this description.

6 Other Roles of Impurity

Wilson correctly predicted that density of the impurity can be used to control resistivity by changing temperature, but there are various other ways that impurity can be important. Beyond Wilson's time, current transistor can control $p - n$ junctions' various properties by varying density of the impurity. Impurity can catalyst electron and hole recombination since it provides an intermediate energy level. This enables impurity to control the lifetime of carriers. Examples include photocells and a junction laser [4].

References

[1] AH Wilson. The theory of electronic semi-conductors. - II. *PROCEEDINGS OF THE ROYAL SOCI-ETY OF LONDON SERIES A-CONTAINING PAPERS OF A MATHEMATICAL AND PHYSICAL CHAR-ACTER*, 134(823):277–287, NOV 1931. Also available as `http://rspa.royalsocietypublishing.org/content/royprsa/134/823/277.full.pdf`.

[2] AH Wilson. The theory of electronic semi-conductors. *PROCEEDINGS OF THE ROYAL SOCIETY OF LONDON SERIES A-CONTAINING PAPERS OF A MATHEMATICAL AND PHYSICAL CHARACTER*, 133(822):458–491, OCT 1931. Also available as `http://rspa.royalsocietypublishing.org/content/royprsa/133/822/458.full.pdf`.

[3] N F Mott and Z Zinamon. The metal-nonmetal transition. *Reports on Progress in Physics*, 33(3):881, 1970.

[4] Sokrates T. Pantelides. The electronic structure of impurities and other point defects in semiconductors. *Rev. Mod. Phys.*, 50:797–858, Oct 1978.

[5] P.A. Dowben. The metallicity of thin films and overlayers. *Surface Science Reports*, 40(6):151 – 247, 2000. Also available as `https://www.sciencedirect.com/science/article/pii/S0167572900000108`.

[6] F Hensel and H Uchtmann. The metal-insulator transition in expanded fluid metals. *Annual Review of Physical Chemistry*, 40(1):61–83, 1989.

[7] Hellmut Haberland, Hans Kornmeier, Helge Langosch, Michael Oschwald, and Gregor Tanner. Experimental study of the transition from van der waals, over covalent to metallic bonding in mercury clusters. *J. Chem. Soc., Faraday Trans.*, 86:2473–2481, 1990. Also available as `http://dx.doi.org/10.1039/FT9908602473`.

[8] Oleg Kostko, Gert Wrigge, Ori Cheshnovsky, and Bernd v Issendorff. Transition from a bloch-wilson to a free-electron density of states in zn n- clusters, 2005.

[9] Fumiko Yonezawa, Hiroaki Ohtani, and Toshio Yamaguchi. A completely new type of bloch–wilson nonmetal-to-metal transition. *Physica B: Condensed Matter*, 296(1):289 – 300, 2001. Proceedings of the Symposium on Wave Propagation and Electronic Structure in Disordered Systems. Also available as `https://www.sciencedirect.com/science/article/pii/S0921452600008127`.