

Novel Understanding of Electron States Architecture and Its Dimensionality in Semiconductors

Xiaomin Ren

State Key Laboratory of Information Photonics and Optical Communications
(Beijing University of Posts and Telecommunications), and Zh. I. Alferov Russian-Chinese Joint
Laboratory of Information Optoelectronics and Nanoheterostructures, Beijing, China
Email: xmren@bupt.edu.cn

Received February 27, 2013; revised March 27, 2013; accepted May 8, 2013

ABSTRACT

Some important insights into the electron-states-architecture (ESA) and its dimensionality (from 3 to 0) in a semiconductor (or generally crystalline) material are obtained. The self-consistency of the set of density of states (DOS) expressions with different dimensionalities is remediated through the clarification and rearrangement of the wave-function boundary conditions for working out the eigenvalues in the wave vector space. The actually too roughly observed and theoretically unpredicted critical points for the dimensionality transitions referring to the integer ones are revealed upon an unusual assumption of the intrinsic energy-level dispersion (ELD). The ELD based quantitative physical model had been established on an immediate instinct at the very beginning and has been properly modified afterwards. The uncertainty regarding the relationship between the de Broglie wavelength of electrons and the dimensionality transitions, seeming somewhat mysterious before, is consequentially eliminated. The effect of the material dimensions on the ELD width is also predicted and has been included in the model. The continuous evolution of the ESA dimensionality is convincingly and comprehensively interpreted and thus the area of the fractional ESA dimensionalities is opened. Another new assumption of the spatial extension shrinkage (SES) closely related to the ELD has also been made and thus the understanding of the behavior of an electron or, in a general sense, a particle has become more comprehensive. This work would manifest itself a new basis for further development of nanoheterostructures (or low dimensional heterostructures including the quantum wells, quantum wires, quantum dots and especially the hetero-dimensional structures). Expected should also be the possible inventions of some novel electronic and optoelectronic devices. More basically, it leads to a new quantum mechanical picture, the essential modifications of Schrödinger equation and Newtonian equation that give rise to a full cosmic-scope picture, and a super-low-speed relativity assumption.

Keywords: Energy-level Dispersion; Spatial Extension Shrinkage; Electron-states-architecture; Density of States; Dimensionality; Quantum Wells, Wires and Dots; Schrödinger Equation; Newtonian Equation; Relativity

1. Introduction

Both the classical and the nano- heterostructures of semiconductors have developed dramatically based on the well-recognized theories, especially the theory on the density of states (DOS) of electrons in a semiconductor (or generally crystalline) material and the relevant electron-states-architecture (ESA) dimensionality [1]. However, some problems exist in the previous theory and it had been suspected that some fundamental theoretical changes would be needed for the improvement. Both unusual assumptions of the energy-level dispersion (ELD) [2-3] and the spatial extension shrinkage (SES) have been proposed and it is shown that they have acted quite well for the improvement. And, in proceeding with this investigation, the quantum mechanical picture and the relevant fundamental theories have really been changed.

2. Important Insights into the Previous Theory and a Summary of the Problems

In this paper, without the loss of generality, the investigations on the ESA and its dimensionality will be restricted only on the case of a conduction band (or a band with a lowest minimum). The DOS expressions given by the previous theory in this case for a bulk material (3D), a quantum well (2D), a quantum wire (1D) and a quantum dot (0D) are as follows, respectively:

$$\rho_c(E) = \frac{(m_n^*)^{\frac{3}{2}} \sqrt{2(E-E_c)}}{\pi^2 \hbar^3} \quad (1)$$

$$\rho_c(E, L_z) = \sum_n \frac{m_n^*}{\pi \hbar^2 L_z} \Theta(E - E_{cn}) \quad (2)$$

where

$$E_{cn} = E_c + \frac{\hbar^2}{2m_n^*} \left(\frac{n\pi}{L_z} \right)^2, n = \pm 1, \pm 2, \pm 3, \dots$$

$$\rho_c(E, L_y, L_z) = \sum_{m,n} \frac{\sqrt{m_n^*} \Theta(E - E_{c,mm})}{L_y L_z \sqrt{2\pi\hbar} \sqrt{E - E_{c,mm}}} \quad (3)$$

where, $E_{c,mm} = E_c + \frac{\hbar^2}{2m_n^*} \left[\left(\frac{m\pi}{L_y} \right)^2 + \left(\frac{n\pi}{L_z} \right)^2 \right]$,

$$m, n = \pm 1, \pm 2, \pm 3, \dots$$

$$\rho_c(E, L_x, L_y, L_z) = \sum_{l,m,n} \frac{2}{L_x L_y L_z} \delta(E - E_{c,lmn}) \quad (4)$$

where, $E_{c,lmn} = E_c + \frac{\hbar^2}{2m_n^*} \left[\left(\frac{l\pi}{L_x} \right)^2 + \left(\frac{m\pi}{L_y} \right)^2 + \left(\frac{n\pi}{L_z} \right)^2 \right]$,

$$l, m, n = \pm 1, \pm 2, \pm 3, \dots$$

It is noted that all the material structures under the investigations here are assumed of the cuboidal shapes. The material sizes in the three directions are denoted by L_x, L_y , and L_z , respectively. In addition, the symbols E, E_c and m_n^* are the electron energy with an arbitrary value, the electron energy at the bottom of the conduction band (or at the minimum of a specified band) and the effective mass of an electron in the band, respectively;

$\hbar = \frac{h}{2\pi}$ is a quantity related uniquely to the Planck

constant h ; and, $\Theta(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$ is a Heaviside function.

Although almost every one of us is quite familiar with these expressions, it is still necessary for us to gain some essential insights into them:

1) In a rigorous sense, Eq. (1), the so-called “3D expression”, is true only when L_x, L_y , and L_z are all infinitely long. Otherwise, the value of the \vec{k} -space volume occupied by a single state cannot be identical everywhere and the derivation of this equation cannot be rigorously valid. So, at the most, it could be applicable only approximately to the case of a large enough material volume.

2) Attention should also be paid to another issue regarding the derivation of Eq. (1). In the derivation, the number of states per unit \vec{k} -space volume is actually (or should be) taken as

$$\begin{aligned} \frac{dZ}{d\Omega^*} &= 2 \times \lim_{\substack{L_j \rightarrow \infty \\ j=x,y,z}} \frac{1}{\left(\frac{2\pi}{L_x} \right) \left(\frac{2\pi}{L_y} \right) \left(\frac{2\pi}{L_z} \right)} \\ &= \lim_{\substack{L_j \rightarrow \infty \\ j=x,y,z}} \frac{L_x L_y L_z}{4\pi^3} = \frac{\infty^3}{4\pi^3} \end{aligned} \quad (5)$$

where, the intervals between the adjacent wave-vector eigenvalues corresponding to the specified pairs of the wave-function solutions are taken as $\lim_{L_x \rightarrow \infty} \frac{2\pi}{L_x}, \lim_{L_y \rightarrow \infty} \frac{2\pi}{L_y}$

and $\lim_{L_z \rightarrow \infty} \frac{2\pi}{L_z}$ in the three directions, k_x, k_y , and k_z , re-

spectively, by adopting the periodic boundary conditions. This kind of conditions are truly valid upon a homogenous-material assumption which means that the materials within and outside the volume defined by L_x, L_y , and L_z are exactly the same. This is an absolutely-homogenous ideal (infinite) 3D model and thus, in principle, the use of Eq. (1) should be excluded from the modeling of a heterostructure system.

For the completeness of description, it should be stated here additionally that, based on this ideal 3D model with the homogenous-material assumption, Eq. (1) can be immediately obtained by using the three more equations as follows (the material volume, the number of states per unit \vec{k} -space volume and per unit material volume, and the \vec{k} -space volume per unit energy interval, respectively):

$$V = \lim_{\substack{L_j \rightarrow \infty \\ j=x,y,z}} (L_x L_y L_z) = \infty^3 \quad (6)$$

$$\frac{1}{V} \cdot \frac{dZ}{d\Omega^*} = \frac{1}{\infty^3} \cdot \frac{\infty^3}{4\pi^3} = \frac{1}{4\pi^3} \quad (7)$$

$$\frac{d\Omega^*}{dE} = \frac{4\pi(m_n^*)^{\frac{3}{2}} \sqrt{2(E-E_c)}}{\hbar^3} \quad (8)$$

However, a heterogenous-interface between the materials within and outside the volume defined by L_x, L_y , and L_z is a more suitable assumption for the DOS modeling because most cases we are interested in refer to heterostructure systems. With this assumption, we can immediately get another ideal (infinite) 3D model which is applicable to the heterostructure systems with infinite or at least large enough material dimensions.

Nevertheless, it should be noted at the moment that this heterogenous-interface assumption has not been completely excluded in the previous theory. Actually, it has been adopted partially in the derivations of both Eq. (2) and Eq.(3) and even completely in obtaining Eq.(4).

With this assumption, the intervals between the adjacent wave-vector eigenvalues may become $\lim_{L_x \rightarrow \infty} \frac{\pi}{L_x}$,

$\lim_{L_y \rightarrow \infty} \frac{\pi}{L_y}$ and $\lim_{L_z \rightarrow \infty} \frac{\pi}{L_z}$ in the three directions, respec-

tively, by adopting a commonly used typical boundary condition called “a potential well of infinite depth” or briefly “an infinite well”. Then, Eq.(1) should be re-

placed by

$$\rho_c(E) = 8 \times \frac{(m_n^*)^3 \sqrt{2(E-E_c)}}{\pi^2 \hbar^3} \quad (1^*)$$

If adopted are a set of hybrid boundary conditions, say, in the k_x and k_y directions using the periodic boundary conditions and in the k_z direction using the one of “an infinite-well”, Eq.(1) should be replaced by another one as follows:

$$\rho_c(E) = 2 \times \frac{(m_n^*)^3 \sqrt{2(E-E_c)}}{\pi^2 \hbar^3} \quad (1')$$

These clarifications are very essential for ensuring the self-consistency of the DOS expressions as a whole set for different ESA dimensionalities. Actually and unfortunately, due to improper adoptions of the boundary conditions, the equations from Eq.(1) to Eq.(4) in their well-known current forms cannot rigorously meet the requirement of the self-consistency. This observation will be further discussed in the following text.

3) Similarly, Eq. (2), the so-called “2D expression”, is true only when L_x and L_y are infinitely long. Moreover, as to L_z , as long as it is of a finite value, Eq.(2) is always obviously valid, even when L_z is very large and the structure cannot be regarded as a “well”; when L_z reaches its upward limit, i.e. the infinity, Eq.(2) cannot be degenerated to Eq.(1) due to the mismatch in the boundary condition choices, but can be degenerated to Eq. (1') with a perfect consistency owing to the modification of the boundary conditions made hereinabove.

4) Also, Eq. (3), the so-called “1D expression”, is true only when L_x is infinitely long. As to L_y and L_z , as long as they are of finite values, Eq. (3) is always obviously valid, even when L_y and L_z are (or, when each of them is) very large and the structure cannot be regarded as a “wire”, or even neither as a “well”. When L_y becomes infinity, it can also not be degenerated to Eq. (2) but can be degenerated to a new equation as follows:

$$\rho_c(E, L_z) = 2 \times \sum_n \frac{m_n^*}{\pi \hbar^2 L_z} \Theta(E - E_{cn}) \quad (2')$$

The modification adopted for the derivation of Eq.(2') refers to a set of hybrid boundary conditions including the periodic boundary condition in the k_x direction and the “infinite-well” conditions in the k_y and k_z directions.

In addition, when the “infinite-well” boundary conditions are adopted in all the three directions, both Eq.(2) and Eq. (2') are no longer valid and should be replaced by the following expression:

$$\rho_c(E, L_z) = 4 \times \sum_n \frac{m_n^*}{\pi \hbar^2 L_z} \Theta(E - E_{cn}) \quad (2^*)$$

5) Finally, Eq.(4), the so-called “0D expression”, is

always obviously true as long as L_x , L_y , and L_z are of finite values, even when L_x , L_y , and L_z (or, when each two or each one of them) are (or, is) very large and the structure cannot be regarded as a “dot” (actually, it could be a wire, a well, or even a bulk one). When L_x becomes infinity, the degeneration from Eq.(4) to Eq.(3) fails unexceptionally but to a new one, as shown below, it works:

$$\begin{aligned} \rho_c(E, L_y, L_z) &= 2 \times \sum_{m,n} \frac{\sqrt{m_n^*} \Theta(E - E_{c,mn})}{L_y L_z \sqrt{2\pi \hbar} \sqrt{E - E_{c,mn}}} \\ &= \sum_{m,n} \frac{\sqrt{2m_n^*} \Theta(E - E_{c,mn})}{L_y L_z \pi \hbar \sqrt{E - E_{c,mn}}} \end{aligned} \quad (3'^*)$$

In the derivation of Eq.(3'*) , a modification on the boundary conditions has been done: the “infinite-well” boundary condition is also adopted in the k_x direction and consequently the same kind of conditions has been adopted in all the three directions.

It is noted that the boundary conditions in all-directions adopted in the derivation of Eq.(4) are straightforwardly those of the “infinite-wells”. So, for the convenience of descriptions hereinafter, Eq.(4) can be directly re-denoted as Eq.(4*).

In summary, two conclusions can be reached. First, the true significations of all these expressions do not exactly match with what they are called by names and, also unfortunately, we cannot find any accurate criterion from the expressions themselves to define the critical midway points (either L_x , L_y , or L_z with a certain finite value) for the transitions between different dimensionality regions, although such transitions actually happen as a fact of our common knowledge. These transitions cannot be convincingly interpreted even with the help of the concept of de Broglie wavelength of electrons which lacks an objective tie to the transition phenomena for doing a scientific reasoning other than a mysterious one. In addition and more basically, it is surely unsatisfactory to explain the formation of energy bands in such a way: the adjacent levels are so near that they can be treated APPROXIMATELY as a continuous band. This explanation works as merely a mathematical trick other than an intrinsic physical settlement. So, the previous theory looks quite casual and the truthfulness of the dispersion-free picture of the energy level in the previous theory has to be suspected. The author believes that something actually happens to originate a dimensionality transition. Secondly, to make the theory validly applicable to the heterostructure systems and ensuring the self-consistency of the whole set of DOS expressions for different ESA dimensionalities, the heterogeneous-interface assumption should be followed for all the cases and in all the directions unexceptionally. In this paper, the boundary conditions of the “infinite-wells” are adopted due to its typicality and thus

a consistent group of equations consisting of Eqs. (1^{*}), (2^{*}), (3^{*}) and (4^{*}, identical to 4), as a whole, becomes the basis for the modeling and discussions throughout the text hereinafter.

3. The Concepts of the ELD and the SES

To solve the above mentioned problems regarding the ESA dimensionality transitions and evolutions, the concept of ELD was proposed [2] and it is regarded as an “intrinsic” effect. It is believed that, in general, any electron energy level could not be an ideal line with zero-linewidth. Instead, all the energy levels might be dispersive, at least slightly sometime but unexceptionally. The position of each level in an energy band defined by the previous theory can be regarded as the peak of a certain energy distribution (so it will be referred to as “peak energy level” hereinafter). It is this kind of dispersion that makes different energy levels merging together and TRULY (not APPROXIMATELY) form a continuous energy band when the intervals between adjacent energy levels are sufficiently small. A preliminary model had been established in [2] where the ELD and the associated wave vector dispersion (WVD) were related by the original dependence of electron energy versus wave vector, which itself did not take into account the ELD effect. Quite soon afterwards, it was found that such a model needs to be modified. The biggest problem of the model is that it does not permit any energy level either in the conduction band or in the valence band to disperse to the energy region of the forbidden band, otherwise the wave vector should have to take imaginary values. To overcome this theoretical discomfiture, a better model was proposed [3] except that the normalization condition was not fully proper yet. In the model, the ELD is expressed as a function of the continuous independent variable E , by using a symmetric decayed exponential lineshape with its peak at each of the peak energy levels E^* . Now, a proper normalization condition $\int_{-\infty}^{\infty} |F_D(E, E^*)|^2 dE = 1$ is adopted and then a new normalized ELD function of this type can be defined more rigorously as follows:

$$F_D(E, E^*) = \begin{cases} \frac{1}{\sqrt{D}} e^{-\frac{E-E^*}{D}} & E \geq E^* \\ \frac{1}{\sqrt{D}} e^{\frac{E-E^*}{D}} & E \leq E^* \end{cases} \quad (9)$$

where, D is the width of the ELD which is surely a parameter of the most importance in this model.

Generally speaking, to use this type of expression is just a probing choice and there might be some other optional lineshapes such as a Lorentzian one. However, recent advancement of the investigations related to the modification of Schrödinger equation which will be

briefly mentioned hereinafter has made the author more convincing for the choice of the exponential one.

To make a further introduction to the ELD concept, it needs to be noted that, for a material with finite volume, the envelope of the dependence of the discrete peak energy levels E^* on the wave vector eigenvalues \vec{k}^* along a specified \vec{k} -space direction can be approximately regarded as a parabolic curve either at the bottom of conduction band or at the top of valence band. Here, as a reminder, it should be repeated that our investigations will be restricted only on the case of a conduction band. Being aware of the fact that the dispersions happen to all the peak energy levels “located” within such an envelope, and as shown in the figure of E - \vec{k} curves (**Figure 1**, the electron energy versus the wave vector), a series of (actually an infinite series of) dispersion-resulted new “envelopes” appear and all look similar to the original one. In an ideal case with infinitely long material size(s), each of the above mentioned discrete E - \vec{k} dependences becomes a continuous curve while all the other stories are the same.

Consequently, the associated WVD can be observed by referring to the ELD-resulted shifts of the specified E - \vec{k} envelope or curve (also shown in **Figure 1**).

In an essential point of view, the ELD function $F_D(E, E^*)$ shows a spectrally varied intrinsic permissibility for a continuous energy deviation ΔE from E^* ($\Delta E = E - E^*$) caused by possible external forces (including all kinds of completely deterministic forces and the thermo effect, etc.), while D implies a permitted E^* -centric energy scope in which the continuous energy deviation mentioned above could manifestly happen.

The ELD actually leads to an essential change regarding the wave function in Schrödinger equation. The original wave function Ψ should be replaced by a re-defined one, i.e.

$$\Psi_D = F_D(E, E^*) \Psi \quad (10)$$

which is no longer a function with a fixed $E = E^*$, but a wave-spectrum function varying continuously with E . However, it can be easily found that the well-known Schrödinger equation is still valid for this new wave function, i.e.

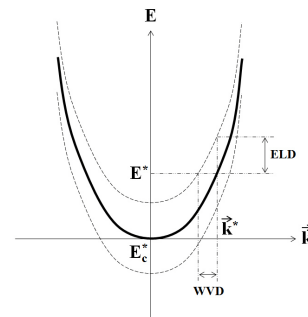


Figure 1.

$$i\hbar \frac{\partial \Psi_D}{\partial t} = -\frac{\hbar^2}{2\mu} \nabla^2 \Psi_D + U(\vec{r}) \Psi_D \quad (11)$$

where, t is time, μ is the electron mass, $U(\vec{r})$ is the potential energy of an electron at a specified location \vec{r} , and $i = \sqrt{-1}$ is the imaginary unit.

In crystalline materials, it is assumed that the intrinsic D could be extended temperature-dependently and dimension-dependently to D_{ext} due to the lattice-vibration and the multi-electrons effects, respectively. The expression of D_{ext} can be written as

$$D_{ext} = D + D_T + D_l = D + D_{LV} (1 - e^{-\alpha_T T}) + D_M [M_l(L_x, L_y, L_z) - 1], \quad T < T_b \quad (12)$$

where, D_T plus D_l forms the extended portion of D_{ext} ; D_{LV} , α_T and D_M are three assumed constant coefficients; T_b is the material-broken temperature; and $[M_l(L_x, L_y, L_z) - 1]$ is an assumed dimension-dependent factor (please ignore the corresponding formulation $D = M_l(L_x, L_y, L_z) D_e$ in [3]).

When the material dimensions are not too small, $M_l(L_x, L_y, L_z)$ could be assumed almost as a constant (a saturated value). Otherwise, in case of the sufficiently small dimensions, it is assumed to firstly increase with the enlargement of each of the three dimensions L_x , L_y and L_z remarkably and then get to becoming saturated. Furthermore, $M_l(L_x, L_y, L_z)$ can be expressed as a product of three individual factors, $M_{lx}(L_x)$, $M_{ly}(L_y)$ and $M_{lz}(L_z)$, referring to the dimensions L_x , L_y and L_z , respectively, i.e.,

$$M_l(L_x, L_y, L_z) = M_{lx}(L_x) M_{ly}(L_y) M_{lz}(L_z) \quad (13)$$

Each of the factors, in turn, can be initially assumed as follows:

$$M_{lj}(L_j) = \left[1 + \frac{(M_{ls} - 1)}{3} \cdot \left(\frac{L_j^2}{L_t^2} \right) \right] \Theta(L_t - L_j) + M_{ls} \left[1 - \frac{2(M_{ls} - 1)}{3M_{ls}} \left(\frac{L_t}{L_j} \right) \right] \Theta(L_j - L_t) \quad (14)$$

where, $j = x, y, z$ again, and L_t is the turning point where $M_{lj}(L_j)$ begins to approach the saturation limit M_{ls} which as well as L_t , respectively, are identical for all the three factors. Obviously, we have $1 \leq M_l \leq M_{ls}^3$.

So, in crystalline materials, the normalized ELD function can be redefined with the extended width D_{ext} as follows:

$$F_{D, ext}(E, E^*) = \begin{cases} \frac{1}{\sqrt{D_{ext}}} e^{-\left(\frac{E-E^*}{D_{ext}}\right)} & E \geq E^* \\ \frac{1}{\sqrt{D_{ext}}} e^{\left(\frac{E-E^*}{D_{ext}}\right)} & E \leq E^* \end{cases} \quad (15)$$

To see how a dimensionality transition happens with the existence of the ELD, we may take the case of the

dimensionality transition from the ideal 2D ($L_x = \infty, L_y = \infty; L_z = 0$) to the ideal 3D ($L_x = \infty, L_y = \infty; L_z = \infty$) as an example. As a reasonable approximation when the material dimensions are not too small, it can be assumed that D_l takes its saturated value and D_{ext} becomes dimension-independent. The derivations and discussions hereinafter regarding the dimensionality transitions will be made with this assumption. It is believed that at a certain big value of L_z , the energy intervals between the adjacent separated $E-\vec{k}$ surfaces will be so small that at least some of the surfaces will get to merging together remarkably duo to ELD. It seems quite reasonable to designate the critical energy interval for such a merging probably as the full width at half maximum of the redefined ELD function Eq. (15), which can be calculated as

$$\Delta E = 2 \times \ln 2 \times D_{ext} \quad (16)$$

The bottoms of the separated $E-\vec{k}$ surfaces can be expressed as:

$$E_{cn}^* = E_c^* + \frac{\hbar^2 k_{zn}^2}{2m_n^*}, \quad k_{zn} = \frac{n\pi}{L_z}, \quad n = 1, 2, 3, \dots \quad (17)$$

It in turn leads to the following two equations:

$$E_{cn}^* - E_c^* = \frac{n^2 h^2}{8m_n^* L_z^2} \quad (18)$$

$$E_{c,n+1}^* - E_{cn}^* = \frac{(2n+1)h^2}{8m_n^* L_z^2} \quad (19)$$

By taking the criterion of $E_{c,n+1}^* - E_{cn}^* = \Delta E$, we can determine the critical value of $L_z = L_{z,2 \rightarrow 3(n)}$ for the merging of the two peak energy levels, E_{cn}^* and $E_{c,n+1}^*$, as follows:

$$L_{z,2 \rightarrow 3(n)} = \frac{\sqrt{2(2n+1)}}{4\sqrt{\ln 2}} \cdot \frac{h}{\sqrt{2m_n^* D_{ext}}} \quad (20)$$

The shortest $L_{z,2 \rightarrow 3(n)}$ should be that with the minimum sequence number $n = 1$, i.e.,

$$L_{z,2 \rightarrow 3(n=1)} = \frac{\sqrt{6}}{4\sqrt{\ln 2}} \cdot \frac{h}{\sqrt{2m_n^* D_{ext}}} = 0.736 \frac{h}{\sqrt{2m_n^* D_{ext}}} = 0.736 \lambda_{de,ELD} \quad (21)$$

where

$$\lambda_{de,ELD} = \frac{h}{\sqrt{2m_n^* D_{ext}}} = \sqrt{\frac{\mu E_{k,de}^*}{m_n^* D_{ext}}} \cdot \lambda_{de} \quad (22)$$

$$\lambda_{de} = \frac{h}{\sqrt{2\mu E_{k,de}^*}} = \frac{h}{\sqrt{2\mu(E_{de}^* - E_c^*)}} \quad (23)$$

In the above equations, λ_{de} is the well-known de

Broglie wavelength with E_{de}^* and $E_{k,de}^*$ being the average total energy and the average kinetic energy of an electron, respectively, in a specified system, while $\lambda_{de,ELD}$ is such a “golden quantity” that it times a fixed proportional constant (0.736) simply equals the critical size $L_{z,2 \rightarrow 3(n=1)}$. We may name $\lambda_{de,ELD}$ as “the nominal de Broglie wavelength” referring to the dimensionality transition due to its above mentioned feature as well as its similarity to λ_{de} in mathematical expression.

We may also make $L_{z,2 \rightarrow 3(n=1)}$ explicitly related with λ_{de} by writing it as follows:

$$L_{z,2 \rightarrow 3(n=1)} = 0.736 \sqrt{\frac{\mu E_{k,de}^*}{m_n^* D_{ext}}} \cdot \lambda_{de} \quad (24)$$

In this equation, it has been shown so clearly how the extended ELD width D_{ext} serves as an objective tie between de Broglie wavelength and the critical point of the dimensionality transition and how de Broglie wavelength plays its role in the quantum size effect.

Relevant experiments are needed to judge the truthfulness of the above mentioned theoretical assumptions and predictions both qualitatively and quantitatively.

In addition, owing to the above described effort of inquiring into the relationship between the concept of de Broglie wavelength of electrons and the dimensionality transition phenomena, the author has been further suspecting the possibility if D would increase somehow with the increasing of the peak energy level E^* . If this assumption would be true, it might be found that we had approached the bridge between the microcosm and the macrocosm. Actually, if we deal with not only an electron but also a bigger particle or an object in a general sense with arbitrary mass and energy, we may do a further reasoning to assume the increasing of D with the increases of both the mass and the energy of the object, for example, $D \propto (\mu E^*)^n$ with n being a positive rational number. Then, as a result, an object with a big enough mass and a high enough energy would manifest a huge ELD and there no longer exists a states-architecture of discrete energy levels. Instead, the states-architecture of this kind of objects should be always continuous. It refers therefore to a non-quantum world or a Newtonian world.

Even more additionally, as we comprehensively think about the difference between the mechanical behaviors of an object in the quantum regime and in the Newtonian regime, it seems that another assumption in association with the ELD should also be made: there might exist an effect of SES, i.e. the spatial extension shrinkage as mentioned hereinabove. As we know, the Newtonian mechanics is deterministic while the quantum mechanics is statistic. Therefore, the confinement tendency of the spatial extension of an object should be enhanced during the evolution from the quantum mechanics to the Newtonian

mechanics. Similarly to the above given descriptions of $F_D(E, E^*)$ and D , the nomorlized SES function could be assumed as

$$F_S(\vec{r} - \vec{r}^*) = \frac{(2/\pi)^{\frac{3}{4}}}{S^{\frac{3}{2}}} e^{-\frac{|\vec{r} - \vec{r}^*|^2}{S^2}} \quad (25)$$

and the SES parameter S as the measure of the confined extension could also be conceived as a function of μE^* and probingly could be $S \propto (\mu E^*)^{-n}$. Where, \vec{r}^* is the spatial position with the peak probability of appearance.

If this assumption of SES would be true, the wave function should be changed further and, eventually, Schrödinger equation should no longer be kept unchanged; instead, it should have to be modified. At the same time, Newtonian equation surely also needs to be modified. Actually, the author has fulfilled both modifications and the relevant details will be described later.

Here given are the modified Schrödinger-Newtonian equations themselves as follows:

$$i\hbar \frac{\partial \Psi_M}{\partial t} - DC_{DS}^2 \left[\tilde{\Theta}(\tilde{E} - 1) - \tilde{\Theta}(1 - \tilde{E}) \right] \frac{\partial \Psi_M}{\partial \tilde{E}} \quad (26-a)$$

$$= \frac{\left(-\frac{\hbar^2}{2\mu} \right) \nabla^2 \Psi_M}{\left\{ 1 + \frac{2}{(1 + C_{DS}^2)} \left[i\sqrt{2} C_{DS} \tilde{p} \cdot \left(\frac{\vec{r} - \vec{r}^*}{S} \right) - C_{DS}^2 \frac{|\vec{r} - \vec{r}^*|^2}{S^2} \right] \right\}} + U(\vec{r}) \Psi_M$$

$$\left(\iiint_{\Omega \rightarrow \infty} |\Psi_M|^2 dV \right)^{-1} \cdot \iiint_{\Omega \rightarrow \infty} |\Psi_M|^2 \vec{F}(\vec{r}) dV = \xi_\psi \cdot \frac{d(\mu v)}{dt} (\vec{v}/v) \quad (26-b)$$

$$+ \mu v \frac{d(\vec{v}/v)}{dt}, \quad \xi_\psi = \left[1 + C_E \frac{\partial}{\partial E} \left(\iiint_{\Omega \rightarrow \infty} |\Psi_M|^2 dV \right) \right]^\gamma$$

where, Ψ_M is the modified wave function; $\tilde{E} = E/E^*$ the normalized energy; \tilde{p} the unit vector along the dominant momentum direction; $\vec{F}(\vec{r})$ the field of force; \vec{v} the velocity vector; $\tilde{\Theta}(x)$ a modified Heaviside function of x with $\tilde{\Theta}(0_+) = 1/2$; C_{DS} and C_E are two constants; and, $\gamma = -1, 0, 1$ (case-dependent).

As a fully natural consequence of the formulation of the modified Schrödinger equation, i.e. Eq.(26-a), one essential relation, $S = C_S / \sqrt{\mu E^*} = \hbar C_{DS}^{-1} / \sqrt{\mu E^*}$, has been obtained and the other one, $D = C_D \sqrt{\mu E^*}$, has also been assumed tentatively (both C_S and C_D are constants). The two relations make the concepts of ELD/SES quantitatively defined and lead to an important idea that $\sqrt{\mu E^*}$ could be taken as the measure of the cosmo-level (the author suggests to call it the “cosmicity”).

The above given modified Schrödinger-Newtonian equations should be well valid in the mid-cosmic regime

(between the microcosm and the macrocosm) where no relativistic effect would need to be taken into account. However, in the super-high speed regime, Einstein’s theory of relativity featuring an upper limit of velocity c should be adopted to make further modification of this pair of equations. In association with this consideration, a more unusual idea came out and had greatly surprised the author: there must appear a similar theory of relativity for super-low speed regime featuring a lower limit of velocity (LLV) \hat{c} ! This idea is necessary not only for the further modification of the equations in the super-low speed regime and the unveiling of a philosophical beauty of the cosmic symmetry but also for the essential causal interpretation of the statistical spatial extension of a particle, or the so-called “probability wave” phenomenon. It might be this lower-limit effect that intrinsically makes any particle keeping un-static or in continuous motion directed randomly. The author suggests to call the matter moving with the velocity \hat{c} , if it would exist, as “symight” (it is similar to “light” but the first three letters are newly introduced to replace “l” and to signalize that it is a symmetrical existence for “light”; and, the word regarding its particle-property corresponding to “photon” could be “symiton”) and to call \hat{c} itself the velocity of symight. Consequentially, Einstein’s theory of relativity can be reasonably modified with an extended Lorentz transformation as follows:

$$x' = (x - vt) \sqrt{\frac{1 - \left(\frac{\hat{c}}{v}\right)^2}{1 - \left(\frac{v}{c}\right)^2}} \tag{27-a}$$

$$t' = t \frac{\sqrt{1 - \left(\frac{\hat{c}}{v}\right)^2}}{\sqrt{1 - \left(\frac{v}{c}\right)^2}} - \frac{x \left[\left(\frac{v}{c}\right)^2 - \left(\frac{\hat{c}}{v}\right)^2 \right]}{\sqrt{1 - \left(\frac{v}{c}\right)^2} \cdot \sqrt{1 - \left(\frac{\hat{c}}{v}\right)^2}} \tag{27-b}$$

And then, the mass μ should be expressed as:

$$\mu = \mu_{mid} \sqrt{\frac{1 - \left(\frac{\hat{c}}{v}\right)^2}{1 - \left(\frac{v}{c}\right)^2}} \tag{28}$$

In these relativistic equations, v is the velocity and μ_{mid} is the mass at the critical mid-cosmic velocity $v_{mid} = \sqrt{c\hat{c}}$ and is used to replace the static mass μ_0 (or m_0) in Einstein’s theory of relativity.

As we can imagine, it should be difficult to directly determine the lower limit velocity \hat{c} . However, an indirect method for the determination of \hat{c} with assistance of the critical mid-cosmic velocity v_{mid} could be

somewhat easy. Actually, \hat{c} can be determined from $\hat{c} = c^{-1} (v_{mid})^2$ when v_{mid} is known. In turn, v_{mid} can be determined by measuring S or D as precisely as possible in a certain range of v around v_{mid} .

Combining this essentially extended theory of relativity, or more explicitly the full velocity-scope theory of relativity, with the modified Schrödinger-Newtonian equations given as Eq.(26) is so perfective and enlightening that it makes the latter (a) being more rigorously valid for all the possible cases especially in the super-high and super-low speed regimes regarding the particles or objects in a common sense, and (b) probably even being valid in the velocity regimes upwards beyond c and downwards beyond \hat{c} where, as the author suspected, a new matter featuring an imaginary mass might exist in a certain variety of intriguing forms. The properties and behaviors of the new matter could be interpreted with the aid of properly modified ELD and SES functions and we may find that the imaginary mass could actually make sense. This “new-matter” prediction could hopefully conclude or at least share the century-long challenges to Einstein’s prediction regarding the upper-limit velocity. It is believed that any common particle or object cannot surpass or just reach the velocity limits c and \hat{c} indeed, while for the new matter, it is no problem at all to manifest a velocity beyond c or \hat{c} , or furthermore, there are actually no other choices for the new matter except moving essentially with a velocity either higher than c or lower than \hat{c} . The key point here is that different matters behave differently.

4. The DOS Expressions in the Dispersive Cases

To make it convenient, we may refer the two cases, before and after the introduction of ELD model, as “the dispersion-free case” and “the dispersive case”, respectively. And, the values of the electron energy in the two cases need to be denoted with different symbols, E^* and E , also respectively. Then, in the dispersive case, Eqs. (1*), (2*), (3*) and (4*, identical to 4) should be replaced respectively by the following new expressions (just taking $F_{D, ext}^2(E, E^*)$ as the typical convolute integration factor):

$$\begin{aligned} \rho_{c,D}(E) &= \int_{E_c^*}^{\infty} F_{D, ext}^2(E, E^*) \rho_c(E^*) dE^* \\ &= \frac{8\sqrt{2}(m_n^*)^2}{D_{ext}\pi^2\hbar^3} \{ \Theta(E - E_c^*) [\int_{E_c^*}^E e^{-2\left(\frac{E-E^*}{D_{ext}}\right)} \sqrt{E^* - E_c^*} dE^* \\ &\quad + \int_E^{\infty} e^{2\left(\frac{E-E^*}{D_{ext}}\right)} \sqrt{E^* - E_c^*} dE^*] \\ &\quad + \Theta(E_c^* - E) \int_{E_c^*}^{\infty} e^{2\left(\frac{E-E^*}{D_{ext}}\right)} \sqrt{E^* - E_c^*} dE^* \} E \end{aligned} \tag{29}$$

$$\begin{aligned} \rho_{c,D}(E, L_z) &= \int_{E_c^*}^{\infty} F_{D, ext}^2(E, E^*) \rho_c(E^*, L_z) dE^* \\ &= \frac{4m_n^*}{\pi \hbar^2 L_z} \sum_n \left\{ \left[1 - \frac{e^{-2\left(\frac{E-E_{cn}^*}{D_{ext}}\right)}}{2} \right] \Theta(E - E_{cn}^*) \right. \\ &\quad \left. + \frac{1}{2} e^{2\left(\frac{E-E_{cn}^*}{D_{ext}}\right)} \Theta(E_{cn}^* - E) \right\} \end{aligned} \quad (30)$$

$$\begin{aligned} \rho_{c,D}(E, L_y, L_z) &= \int_{E_c^*}^{\infty} F_{D, ext}^2(E, E^*) \rho_c(E^*, L_y, L_z) dE^* \\ &= \frac{\sqrt{2m_n^*}}{D_{ext} L_y L_z \sqrt{\pi \hbar}} \sum_{m,n} \left\{ \int_{E_{c,mn}^*}^E \frac{e^{-2\left(\frac{E-E^*}{D_{ext}}\right)} \Theta(E - E_{c,mn}^*)}{\sqrt{E^* - E_{c,mn}^*}} dE^* \right. \\ &\quad \left. + \int_E^{\infty} \frac{e^{2\left(\frac{E-E^*}{D_{ext}}\right)} \Theta(E - E_{c,mn}^*)}{\sqrt{E^* - E_{c,mn}^*}} dE^* \right. \\ &\quad \left. + \int_{E_{c,mn}^*}^{\infty} \frac{e^{2\left(\frac{E-E^*}{D_{ext}}\right)} \Theta(E_{c,mn}^* - E)}{\sqrt{E^* - E_{c,mn}^*}} dE^* \right\} \end{aligned} \quad (31)$$

$$\begin{aligned} \rho_{c,D}(E, L_x, L_y, L_z) &= \int_{E_c^*}^{\infty} F_{D, ext}^2(E, E^*) \rho_c(E^*, L_x, L_y, L_z) dE^* \\ &= \frac{2}{D_{ext} L_x L_y L_z} \sum_{l,m,n} \left[e^{-2\left(\frac{E-E_{c,lmn}^*}{D_{ext}}\right)} \Theta(E - E_{c,lmn}^*) \right. \\ &\quad \left. + e^{2\left(\frac{E-E_{c,lmn}^*}{D_{ext}}\right)} \Theta(E_{c,lmn}^* - E) \right] \end{aligned} \quad (32)$$

Among all these four expressions, Eq.(32) is uniquely a general one which is applicable to all the cases including the bulk, quantum well, wire and dot materials; while, the others, in a rigorous sense, are applicable only to those specified ideal cases where there should be three dimensions, two dimensions or at least one dimension reaching infinity (please ignore Eqs.(6) and (7) in [3]).

5. The Accurate and Comprehensive Understanding of the Dimensionality Evolution

1) In the ideal 3D case corresponding to Eq.(29), the size of the material along each direction reaches its upward

limit, so the individual contribution from each of the three sizes to the identification number of the dimensionality (dimensionality ID-number) should be unity and thus the dimensionality ID-number itself as the sum of those contributions is “3”. As an alternative denotation, we may also define a triplet of dimensionality indices (DI-triplet) as [1,1,1] to separately specify the 3 fold contributions.

2) As to Eq.(30), it is the case that the dimensionality ID-number varies from 2 (the ideal 2D, $L_z = 0$) to 3 (the ideal 3D, $L_z = \infty$). A rigorous critical point $L_{z,2 \rightarrow 3}$, which should be near but not exactly identical to the value of $L_{z,2 \rightarrow 3(n=1)}$ given in Section 3, can be determined by a method of similarity calculation [2-3]. The definition of the Similarity Function for $\rho_{c,D}(E, L_z)$ is given as:

$$S_\rho[\rho_{c,D}(E, L_z), \rho_r(E)] = \frac{\int_{-\infty}^{\infty} \rho_{c,D}(E, L_z) \rho_r(E) dE}{\int_{-\infty}^{\infty} \rho_r^2(E) dE} \quad (33)$$

where, $\rho_r(E)$ is a proper reference function case by case.

Then, the criterion for the rigorous determination of $L_{z,2 \rightarrow 3}$ should be as

$$\begin{aligned} S_\rho[\rho_{c,D}(E, L_z \neq \infty), \rho_c(E, L_z \neq \infty)] \\ = S_\rho[\rho_{c,D}(E, L_z \neq \infty), \rho_c(E)] \end{aligned} \quad (34)$$

A fractional dimensionality ID-number corresponding to the critical point $L_{z,2 \rightarrow 3}$ should be naturally the midpoint between the numbers of 2 and 3, i.e. should be equal to 2.5. Then, the dimensionality ID-numbers within the 2D and 3D regions should be (2 - 2.5) and (2.5 - 3), respectively. We can also use the relevant DI-triplets, written as [1,1, d_z], to denote the dimensionalities mentioned above, where d_z has been defined in [3] as follows:

$$d_z (\leq 0.5) = \frac{\int_{-\infty}^{\infty} \rho_{c,D}(E, L_z) \rho_{c,D}(E, L_{z,2 \rightarrow 3}) dE}{2 \int_{-\infty}^{\infty} \rho_{c,D}^2(E, L_{z,2 \rightarrow 3}) dE} \quad (35)$$

$$d_z (\geq 0.5) = 0.5 + \frac{\int_{-\infty}^{\infty} \rho_{c,D}(E, L_z) \rho_{c,D}(E) dE - I_{z,0}}{2 \left[\int_{-\infty}^{\infty} \rho_{c,D}^2(E) dE - I_{z,0} \right]} \quad (36)$$

where,

$$I_{z,0} = \int_{-\infty}^{\infty} \rho_{c,D}(E, L_{z,2 \rightarrow 3}) \rho_{c,D}(E) dE \quad (37)$$

The dimensionality ID-numbers can be derived from the corresponding DI-triplets by simply taking the sum of the indices, i.e. $2 + d_z$.

3) As to Eq.(31), it refers to a complicated case. When $L_z = 0$ and L_y varies, similarly to the discussion regarding Eq.(30), the dimensionality ID-number can vary from 1 (the ideal 1D, $L_y = 0$) to 2 (the ideal 2D, $L_y = \infty$). The fractional dimensionality ID-number corresponding to the critical point $L_{y,1 \rightarrow 2}$ ($= L_{z,2 \rightarrow 3}$) for the transition from 1D to 2D regions should be 1.5. Then, the dimensionality ID-numbers within the 1D and 2D regions should be (1 - 1.5) and (1.5 - 2), respectively. The corresponding DI-triplet should be $[1, d_y, 0]$. If $L_z \neq 0$, the situation is quite different. It is impossible to consistently define a single dimensionality ID-number for an arbitrary combination of L_z and L_y . So, the DI-triplets, i.e. $[1, d_y, d_z]$ becomes the unique choice for the dimensionality denotation. All the possible dimensionality regions can be defined as D1: $[1, d_y < 0.5, d_z < 0.5]$; D2: $[1, d_y > 0.5, d_z < 0.5]$ or $[1, d_y < 0.5, d_z > 0.5]$; and D3: $[1, d_y > 0.5, d_z > 0.5]$.

4) Quite similarly but referring to a more complicated case, Eq.(32) can be interpreted as follows: When $L_z = L_y = 0$ and only L_x varies, the dimensionality ID-numbers can vary from 0 (the ideal 0D, $L_x = 0$) to 1 (the ideal 1D, $L_x = \infty$). The fractional dimensionality ID-number corresponding to the critical point $L_{x,0 \rightarrow 1}$ ($= L_{y,1 \rightarrow 2} = L_{z,2 \rightarrow 3}$) for the transition from 0D to 1D regions should be 0.5. Then, the dimensionality ID-numbers within the 0D and 1D regions should be (0 - 0.5) and (0.5 - 1), respectively. The corresponding DI-triplets should be $[d_x, 0, 0]$. If $L_z \neq 0$ and $L_y \neq 0$, the dimensionality should be denoted only by the DI-triplets, i.e. $[d_x, d_y, d_z]$ and the dimensionality regions can be defined as D0: $[d_x < 0.5, d_y < 0.5, d_z < 0.5]$; D1: $[d_x > 0.5, d_y < 0.5, d_z < 0.5]$, $[d_x < 0.5, d_y > 0.5, d_z < 0.5]$ or $[d_x < 0.5, d_y < 0.5, d_z > 0.5]$; D2: $[d_x > 0.5, d_y > 0.5, d_z < 0.5]$, $[d_x < 0.5, d_y > 0.5, d_z > 0.5]$ or $[d_x > 0.5, d_y < 0.5, d_z > 0.5]$; and D3 $[d_x > 0.5, d_y > 0.5, d_z > 0.5]$.

6. Discussions and Conclusions

This work has provided a novel understanding of ESA and its dimensionality in semiconductors or crystalline materials; has established a more reasonable and convincing basis for further development of nanoheterostructure physics, especially the hetero-dimensional structure physics; has made three new assumptions of the ELD, the SES and the LLV; has led to a new quantum-and furthermore a full cosmic-scope mechanical picture, a pair of the modified Schrödinger-Newtonian equations being valid throughout the evolution from the microcosm regime to the macrocosm regime, an idea and the initial efforts to establish a theory of super-low-speed relativity, and a “new-matter” prediction. Some fundamental phenomena such as the tunneling effect can be modeled

more accurately and becomes more understandable by adopting the ELD assumption. Expected should also be the possible inventions of some novel electronic and optoelectronic devices. For example, it seems quite reasonable to predict a DOS gradient caused electron (or more generally, carrier) diffusion in a hetero-dimensional structure or a structure featuring a gradually changed dimensionality. Then a sort of novel semiconductor electronic devices featuring the mono-carrier transportation (recombination free) and the remarkably reduced power consumption can be conceived and demonstrated, while relevant works have been on its way in the author’s laboratory.

7. Acknowledgements

I would like to express my deep gratitude to Prof. Zh. I. Alferov for his enlightening words. I am also very grateful to Prof. V. G. Dubrovskii for his concentrated discussion on this topic as well as his warm encouragement, to Researcher YE Xiaoyan for her all the way support and helpful suggestions, to my colleagues in my group as well as others in the university, including Associate Prof. WANG Qi, Associate Prof. DUAN Xiaofeng, Prof. HUANG Yongqing, Prof. ZHANG Xia, Associate Prof. WANG Jun, Mr. CAI Shiwei, Mr. HAO Jiaquan, Mr. WANG Jin, Prof. FANG Binxing, and Prof. WANG Yajie for their valued supports and relevant discussions, and to Ms. LUO Fen as well as some of my students, especially Mr. LIU Xiaolong, Mr. JIA Zhigang and Mr. YAN Xin, for their precious helps and necessary assistances. This work has been supported by the National Basic Research Program of China (No. 2010CB327601), the National Natural Science Foundation of China (No. 61020106007, 61108048), International Science & Technology Cooperation Program of China (No. 2011DFR11010) and the 111 Project (No. B07005).

REFERENCES

- [1] Zh. I. Alferov, “The Double Heterostructure: Concept and Its Applications in Physics, Electronics and Technology”, *Nobel Lecture*, December 8, 2000.
- [2] X. M. Ren, “Theoretical Investigation on the Continuous Evolution of the Electron-Gas Dimensionality: From Bulk Materials to Quantum Dots” (postdeadline paper), *2012 Asia Communication and Photonics Conference*, November 7-11, 2012.
- [3] X. M. Ren, “Modification of the Theory on the Energy-Level Dispersion and the Continuity of the Electron-Gas Dimensionality” (invited talk / plenary session; to be delivered), *21st International Symposium on Nanostructures: Physics and Technology*, June 24-28, 2013.