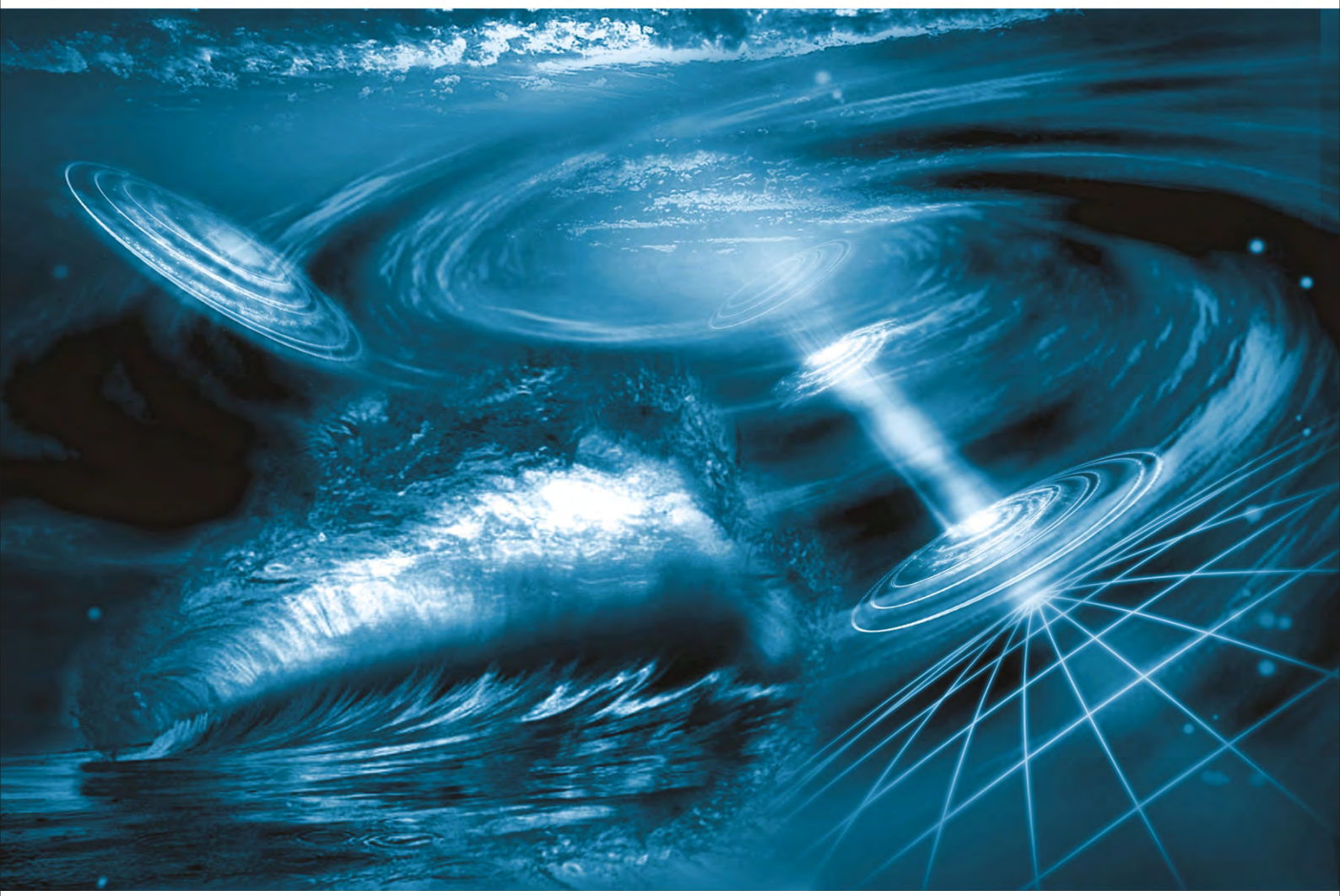
ISSN: 2150-4091 (Print), 2150-4105 (Online) Volume 2, Number 3, March 2010

Scientific Research

Natural Science



www.scirp.org/journal/ns/

Journal Editorial Board

ISSN: 2150-4091 (Print) ISSN: 2150-4105 (Online)
http://www.scirp.org/journal/ns/

Vol.2, No.3, 131-245 (2010) Natural Science

TABLE OF CONTENTS

Volume 2, Number 3, March 2010

Natural Science

Journal Information

SUBSCRIPTIONS

The *Natural Science* (Online at Scientific Research Publishing, www.SciRP.org) is published monthly by Scientific Research Publishing, Inc., USA.

Subscription rates:

Print: $50 per copy.
To subscribe, please contact Journals Subscriptions Department, E-mail: sub@scirp.org

SERVICES

Advertisements
Advertisement Sales Department, E-mail: service@scirp.org

Reprints (minimum quantity 100 copies)
Reprints Co-ordinator, Scientific Research Publishing, Inc., USA.
E-mail: sub@scirp.org

COPYRIGHT

PRODUCTION INFORMATION

For manuscripts that have been accepted for publication, please contact:
E-mail: ns@scirp.org

Vol.2, No.3, 131-138 (2010)
doi:10.4236/ns.2010.23022

Natural Science

Investigation of nonlinear temperature distribution in biological tissues by using bioheat transfer equation of Pennes' type

Ahmed Lakhssassi[1], Emmanuel Kengne[1*], Hicham Semmaoui[2]

[1]LIMA – Laboratoire d'Ingénierie des Microsystèmes Avancés, Département d'informatique et d'ingénierie, Université du Québec en Outaouais, Gatineau, Canada; ekengne6@yahoo.fr
[2]Department of Electrical Engineering, Polytechnique, University of Montrea, Montreal, Canada

Received 14 December 2009; revised 8 January 2010; accepted 30 January 2010.

ABSTRACT

In this paper, a two level finite difference scheme of Crank-Nicholson type is constructed and used to numerically investigate nonlinear temperature distribution in biological tissues described by bioheat transfer equation of Pennes' type. For the equation under consideration, the thermal conductivity is either depth-dependent or temperature-dependent, while blood perfusion is temperature-dependent. In both cases of depth-dependent and temperature-dependent thermal conductivity, it is shown that blood perfusion decreases the temperature of the living tissue. Our numerical simulations show that neither the localization nor the magnitude of peak temperature is affected by surface temperature; however, the width of peak temperature increases with surface temperature.

Keywords: Bioheat Transfer Equation of Pennes Type; Pennes' Bioheat Transfer Equation; Modified Crank-Nicholson Method; Two-Level Finite Difference Scheme

1. INTRODUCTION

The evaluation of thermal conductivities in living tissues is a very complex process which involves several phenomenological mechanisms, such as heat transfer due to perfusion of the arterial-venous blood through the pores of the tissue (blood convection), heat conduction in tissues, metabolic heat generation and external interactions, such as electromagnetic radiation emitted from cell phones, evaporation, metabolism, *etc*. The heat transfer mechanism in biological tissues is important for therapeutic practices, such as cancer hyperthermia, burn injury, brain hypothermia resuscitation, disease diagnostics, cryosurgery, *etc*.

Many of the bioheat transfer problems have been modeled using Pennes' equation [1], which accounts for the ability of tissue to remove heat by both passive conduction (diffusion) and perfusion of tissue by blood. Perfusion is defined as the nonvectorial volumetric blood flow per tissue volume in a region that contains sufficient capillaries that an average flow description is considered reasonable. Pennes' model was adapted by many biologists for the analysis of various heat transfer phenomena in a living body [2-7]. Others, after evaluations of the Pennes model in specifical situations, have concluded that many of the hypotheses (foundational to the model) are not valid. Then these latter modified and generalized the model to adequate systems [8-11]. To analyze nonlinear temperature distribution in living tissues, it is sometime useful to modify the one-dimensional (1D) Pennes' bioheat transfer equation by adding nonlinear terms, which generally account for temperature-dependent variability in tissue perfusion (see for example [12]).

To treat the system of motion in living bodies, we have written the transient 1D bioheat transfer type model in a generalized form as follows [13],

$$\rho c\frac{\partial u^*}{\partial t}=\frac{\partial}{\partial x}\left(k\frac{\partial u^*}{\partial x}\right)-c_b\omega_m(T)\rho_b u^*+Q_m+p(x,t),\qquad 0<x<L, \tag{1}$$

where x is the distance from the surface to the body core (in m), t is the time (in s), ρ, c, k are the density (in gm/m^3), specific heat (in $cal/gm°C$) and thermal conductivity of tissue (in $cal/m\times s\times °C$), respectively and c_b, is the specific heat of blood (in $cal/gm\ °C$), ρ_b is the density of blood (in gm/m^3), Q_m is the metabolic heat production per volume, ω_m is the nondirectional mass flow (in $gm\times s/m^3$) associated with perfusion so that $\omega(T)=c_b\omega_m(T)\rho_b/\rho c$ is the perfusion coefficient, and $p(x,t)$ is the heat deposited per volume due

*E. Kengne dedicates this work to his son Kengneson Fred Jake Sado

to spatially distributed heating. $u^* = T - T_s$ is the elevated temperature (in °*C*) in the x-direction, where T represents the temperature distribution (in °*C*) in the x-direction and T_s represents the skin's steady state temperature (in °*C*). L is the distance (in *m*) between the skin surface and the body core. In this general form, ω_m is a function of temperature to include the specific case of temperature dependent perfusion. We assume that the thermal conductivity k satisfies $\nu \geq k \geq \mu$, where ν and μ are two positive constants.

The form of the perfusion coefficient $\omega(T)$, which incorporates the heat capacity and density of blood and tissue, depends on the characteristics of the response. For a temperature response without an overshoot, the perfusion is defined as a spatially averaged constant, *i.e.*, $\omega(T) = \omega_0$, where ω_0 is the steady-state perfusion at a constant heat flux. With an overshoot response, we assume that the perfusion increases with time because of vasodilation and capillary recruitment with a higher temperature.

For Pennes' model $p(x,t)$ is constant and $\omega_m = \omega_b / \rho_b$, where ω_b is the blood perfusion rate. In this case the second term on the right-hand side of the state **Eq.1** describes the heat transport between the tissue and the microcirculatory blood perfusion. In biological modeling, the nondirectional mass flow ω_m associated with perfusion is of form $\omega_m = \omega_b / \rho_b + F(T) / u^*$ and $p(x,t)$ is null, where $F(T)$ is a function which can be chosen as a polynomial function of the temperature.

The model used in this paper is the following nonlinear bioheat transfer equation

$$\rho c \frac{\partial u^*}{\partial t} = \frac{\partial}{\partial x}\left(k \frac{\partial u^*}{\partial x}\right) - c_b \omega_b \rho_b u^* - \omega_1 \left(T - T_s - \frac{Q_m}{c_b \omega_b}\right)^2 + Q_m, \quad 0 < x < L, \tag{2}$$

which is a special case of **Eq.1** In this equation ω_b and ω_1 are referred to as the temperature independent (basal) perfusion component and the temperature dependent (vasodilation and angiogenesis) perfusion component, respectively. This model is based on a one-dimensional Pennes bioheat transfer equation and is modified to account for temperature-dependent variability in tissue perfusion [12]. Assuming the metabolic heat production Q_m to be constant and denoting $u = u^* - Q_m / c_b \omega_b$, we obtain the following simplified form of **Eq.2**.

$$\frac{\partial u}{\partial t} = \alpha \frac{\partial}{\partial x}\left(k \frac{\partial u}{\partial x}\right) - \beta u - \gamma u^2, \quad 0 < x < L, \tag{3}$$

where $\alpha = 1/\rho c$, $\beta = c_b \omega_b \rho_b / \rho c$, and $\gamma = \omega_1 / \rho c > 0$. Then the perfusion coefficient is $\omega(u) = \beta + \gamma u$. Beside the differential equation, initial and boundary conditions determine the temperature distribution. **Eq.3** can be written, in steady-state form, as $\alpha \frac{\partial}{\partial x}\left(k \frac{\partial u}{\partial x}\right) = \beta u + \gamma u^2$, which, in the case of constant thermal conductivity k, reads

$$\frac{d^2 u}{dx^2} = \frac{\beta}{\alpha k} u + \frac{\gamma}{\alpha k} u^2. \tag{4}$$

The first integral of **Eq.4** is given by

$$\left(\frac{du}{dx}\right)^2 = \frac{2\gamma}{3\alpha k} u^3 + \frac{\beta}{\alpha k} u^2 + \frac{\delta}{6\alpha k}, \tag{5}$$

where δ is a constant of integration. **Eq.5** is an elliptic ordinary differential equation. This equation can be solved using Weierstrass' elliptic function method [14]. For either $\delta = -2\beta^3 / \lambda^2$ or $\delta = -200\beta^3 / 343\gamma^2$, its particular solutions are

$$u(x) = \frac{\beta\left(3 - 2\cos^2\sqrt{\frac{\beta}{4\alpha k}}x\right)}{2\gamma \cos^2\sqrt{\frac{\beta}{4\alpha k}}x}$$

and

$$u(x) = \frac{\beta\left(9 - 5cn^2\left(\sqrt{\frac{2\beta}{7\alpha k}}x, \frac{5}{8}\right)\right)}{14\gamma cn^2\left(\sqrt{\frac{2\beta}{7\alpha k}}x, \frac{5}{8}\right)},$$

respectively, where $cn(x,m)$ denotes the Jacobi ellitic cosine function with modulus *m*.

Mathematically, solving bioheat transfer equation requires knowledge of both the initial and boundary conditions applicable to the case under study. The initial conditions describe the state of the biological system at time, $t = 0$ while the boundary conditions give information both on the surface $x = 0$ and at depth $x = L$. In this work, we use the following initial and boundary conditions. The initial condition is that a biological tissue has a uniform temperature at the steady-state temperature of the biological tissue as in

$$t = 0; \quad u = u_b(x). \tag{6}$$

Boundary conditions include a constant temperature

on the tissue surface whereas no heat flow is assumed at the boundary $x = L$:

$$u(0,t) = u_{surf}, \quad \frac{\partial u(L,t)}{\partial x} = 0, \quad t > 0 \tag{7}$$

u_{surf} being the surface temperature of the biological tissue.

Restricting ourselves to nonnegative $u(x,t)$, we follow Zhao *et al.* [15] and construct a two level finite difference scheme for (3) which has the same order of accuracy. As in [15] our scheme requires only one initial condition and is also unconditionally stable and convergent. The rest of the paper is organized as follows: in Section 2 we construct a finite difference scheme for the modified Pennes **Eq.3** and prove its solvability. Numerical experiments are reported in Section 3, while a brief summary of the results is done in Section 4.

2. NUMERICAL SCHEME

The main difficulties in numerical solving the bioheat transfer **Eq.3** are the nonlinearity due to the perfusion term and the different material properties of the tissue. Using a modified Crank-Nicholson method (see the description below), we are able to integrate the heat equation efficiently. Within this approach, the approximate elevated temperature u^n at time t^n is constructed by a combination of previous elevated temperature u^{n-1} at time t^{n-1}.

For the numerical computation of the nonlinear heat transfer **Eq. 3** with initial and boundary conditions (6) and (7) we use h to represent the space mesh and τ to represent the time step so that x will be incremented by h and t by τ. We choose an integer N and h such that $Nh = L$. We construct a finite difference scheme for (3) by using the second-order central difference scheme in space and a scheme of Crank-Nicholson [16] type in time. In what follows, u_j^n denotes the approximate elevated temperature at time t^n at depth x_j: $u_j^n \approx u(x_j, t^n)$. If δ_x denotes the central difference operator, the discretization scheme is then

$$\frac{u_j^n - u_j^{n-1}}{\tau} = -\frac{\beta}{2}\left(u_j^n - u_j^{n-1}\right) - \frac{\gamma}{2}u_j^{n-1}\left(u_j^n - u_j^{n-1}\right) + \frac{\alpha}{2}$$

$$\times\left[\delta_x\left(k\left(x_{j+1/2}, t_n\right)\delta_x u_j^n\right) + \delta_x\left(k\left(x_{j+1/2}, t_n\right)\delta_x u_j^{n-1}\right)\right],$$

$$0 < j < N,$$

$$u_0^n = u_0, \quad \frac{u_N^n - u_{N-1}^n}{h} = 0.$$

The difference scheme then takes the form

$$\frac{-\alpha\tau k_j^{n-1}}{2h^2}u_j^{n-1} + \left(1 + \frac{\beta\tau}{2} + \frac{\gamma\tau}{2}u_j^{n-1} + \frac{\alpha\tau\left(k_j^{n-1} + k_{j-1}^{n-1}\right)}{2h^2}\right)$$
$$-\frac{\alpha\tau k_{j+1}^{n-1}}{2h^2}u_{j+1}^n = \frac{\alpha\tau k_j^{n-1}}{2h^2}u_{j-1}^{n-1} + \left(1 - \frac{\beta\tau}{2} - \frac{\gamma\tau}{2}u_j^n\right. \tag{8}$$
$$\left.\times u_j^{n-1} - \frac{\alpha\tau\left(k_j^{n-1} + k_{j-1}^{n-1}\right)}{2h^2}\right)u_j^{n-1} + \frac{\alpha\tau k_{j+1}^{n-1}}{2h^2}u_{j-1}^{n-1},$$

$$u_0^n = u_0, \; u_N^n = u_{N-1}^{n-1},$$

where $k_j^n = k\left(x_j = jh, t^n = n\tau\right)$. The truncation error of this difference schemes in of order $O\left(h^2 + \tau^2\right)$ for each interior grid point $\left(x_j, t^n\right), n \geq 1$ and $0 < j < N$. In matrix form, the difference scheme (8) reads

$$Q_{L_n}\mathrm{u}^n = Q_{R_n}\mathrm{u}^{n-1}, \quad \mathrm{n} = 1, 2, ..., \tag{9}$$

where the vector $\mathbf{u}^n = \left(u_0^n, u_1^n, u_2^n, ..., u_N^n\right)^T$ represents the numerical solutions at the time level t^n $(u_j^n \approx (x_j, t^n))$, Q_{L_n} (left-hand side matrix depending on n) and Q_{R_n} (right-hand side matrix depending on n) are both square tridiagonal matrices of dimension N+1 of the form

$$Q_{L_n} = \begin{cases} 1, \text{ if } p = q \in \{1, N+1\}, \\ -\alpha\tau/(2h^2)\,k_{p-1}^n, \text{ if } q = p-1, \\ 1 + \dfrac{\beta\tau}{2} + \dfrac{\gamma\tau}{2}u_{p-1}^{n-1} + \dfrac{\alpha\tau\left(k_{p-1}^n + k_p^n\right)}{2h^2}, \\ -\alpha\tau/(2h^2)\,k_p^n, \text{ if } q = p+1, \\ 0, \text{ for other } p \text{ and } q, \end{cases}$$

$$Q_{R_n} = \begin{cases} 0, \text{ if } p = q = 1, \\ 1, \text{ if } p = q = N+1, \\ \alpha\tau/(2h^2)\,k_{p-1}^{n-1}, \text{ if } q = p-1, \\ 1 - \dfrac{\beta\tau}{2} - \dfrac{\gamma\tau}{2}u_{p-1}^{n-1} - \dfrac{\alpha\tau\left(k_{p-1}^{n-1} + k_p^{n-1}\right)}{2h^2}, \text{ if } q = p, \\ \alpha\tau/(2h^2)\,k_p^{n-1}, \text{ if } q = p+1, \\ 0, \text{ for other } p \text{ and } q. \end{cases}$$

In other words,

$$Q_{L_n} = \begin{pmatrix} 1 & 0 & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots \\ 0 & \cdots & -\dfrac{\alpha\tau}{2h^2}k_{j-1}^n & 1 + \dfrac{\beta\tau}{2} + \dfrac{\gamma\tau}{2}u_{j-1}^{n-1} + \\ \vdots & \vdots & \ddots & \ddots \\ 0 & \cdots & \cdots & \cdots \end{pmatrix}$$

$$\begin{pmatrix} \cdots & 0 \\ \vdots & \vdots \\ \frac{\alpha\tau\left(k_{j-1}^{n}+k_{j}^{n}\right)}{2h^2} & -\frac{\alpha\tau}{2h^2}k_{j-1}^{n} \quad \cdots \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix},$$

$$Q_{R_n}=\begin{pmatrix} 0 & 0 & \cdots & & \cdots \\ \vdots & \vdots & \ddots & & \ddots \\ 0 & \cdots & \frac{\alpha\tau}{2h^2}k_{j-1}^{n-1} & 1-\frac{\beta\tau}{2}-\frac{\gamma\tau}{2}u_{j-1}^{n-1}- & \\ \vdots & \vdots & \ddots & & \ddots \\ 0 & \cdots & \cdots & & \cdots \end{pmatrix}$$

$$\begin{pmatrix} \cdots & 0 \\ \vdots & \vdots \\ \frac{\alpha\tau\left(k_{j-1}^{n-1}+k_{j}^{n-1}\right)}{2h^2} & \frac{\alpha\tau}{2h^2}k_{j-1}^{n-1} \quad \cdots \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix}.$$

Theorem (on the solvability of system (9)). *System* (9) *is unconditionally solvable for each time step* n.

Proof. It is sufficient to show that matrix Q_{L_n} is inversible. From the expression for Q_{L_n} we have

$$Q_{L_n}^{11}=1>0=\sum_{q=2}^{N+1}\left|Q_{L_n}^{1q}\right|;\quad Q_{L_n}^{N+1N+1}=1>0=\sum_{q=1}^{N}\left|Q_{L_n}^{N+!q}\right|;$$

$$Q_{L_n}^{pp}-\sum_{q=1,q\neq p}^{N+1}\left|Q_{L_n}^{pq}\right|=1+\frac{\beta\tau}{2}+\frac{\gamma\tau}{2}u_{p-1}^{n-1}>0,$$

So that matrix Q_{L_n} is diagonally dominant. By Gershgorin's theorem [17] we conclude that the matrix Q_{L_n} is invertible. This proves the theorem.

3. NUMERICAL EXPERIMENTS

For the numerical experiments we use $L=0.01208\,m$ and $\alpha=1/4200000$ [18]. Because our main aim in this work is the impact of the nonlinear term on the temperature distribution, we will work with different choices of nonlinear coefficient γ. As initial elevated temperature $u_B(x)$, we use one of the functions $u_B(x)=u^0[1+\exp(-x/x_u)]$ and $u_B(x)=u^0/[1+\exp(-x/x_u)]$, where u^0 and x_u are two constants. The thermal conductivity of tissue k is taken to be either a function of x, namely, $k(x)=-10000x^2+112.1x+0.9$ or a linear function of the elevated temperature, namely, $k=k_0+k_1u$, with $k_0=0.4574$ and $k_1=0.001403$ (see [19]). To analyze the effects of the blood perfusion, we used the following three couples (β,γ) of parameters $(0.0005,0.001),(0.009,0.01)$, and $(0.005,0.02)$.

3.1. Effects of Blood Perfusion, Thermal Conductivity, and Initial Elevated Temperature

To analyze the effects of the blood perfusion, we used the following three couples (β,γ) of parameters $(0.0005,0.001),(0.009,0.01)$, and $(0.005,0.02)$. The calculation results of the influences of the blood perfusion, the thermal conductivity, and the initial elevated temperature on the temperature distribution are shown in **Figures l-4**. **Figures 1** and **3** and **Figures 2** and **4** are obtained with the initial conditions $u_B(x)=u^{01}\left[1+\exp\left(x/x_u\right)\right]$, respectively, with $x_u=1/200$, $u^{01}=6$, and $u^{02}=12.928$. For all these figures, we used $u_{surf}=u_B(0)$. Plot (a) gives the elevated temperature profile along the x direction at time $t=5.2\,s$, while plot (b) shows the elevated temperature profile along the t direction at depth $x=0.01148\,m$. Plots (c) and (d) are obtained with $(\beta,\gamma)=(0.005,0.02)$ and show the elevated temperature along the x-direction for different times t and along the t-direction for different depths x, respectively. Plots (e) and (f) show the elevated temperature for temperature-dependent (plot (1)) and depth-dependent (plot (2)) thermal conductivity k. Here, we used the parameter $(\beta,\gamma)=(0.005,0.02)$

Figures 1 and **2** give the elevated temperature when the thermal conductivity of tissue k is a function of distance *x*. In **Figure 1(a)** and **(b)**, as well as in **Figure 2(a)** and **(b)**, the results show that the higher the blood perfusion is, the lower is the elevated temperature. In other words, to decrease the temperature in the living tissue, it is sufficient to increase the blood perfusion.

As in the case of x-dependent thermal conductivity, plots (a) and (b) of **Figures 3** and **4** show that, in the case of temperature-dependent thermal conductivity, the higher the blood perfusion is, the lower is the elevated temperature. We may then conclude from **Figures 1-4** that the effect of the blood perfusion is to decrease the temperature in the living tissue.

Figures 1(c) and **(d)** and **Figure 3(c)** and **(d)** show that the peak temperature is a decreasing function of both depth and time. When the initial temperature of the living tissue is a decreasing function of depth x (**Figures 1** and **3)**, a phenomenon of heating (in the sense that the temperature is almost above the initial temperature) is observed both in depth and time, and the peak temperature at any depth is above the initial temperature (see plots (c) and (d) of **Figures 1** and **3**). To the contrary, a

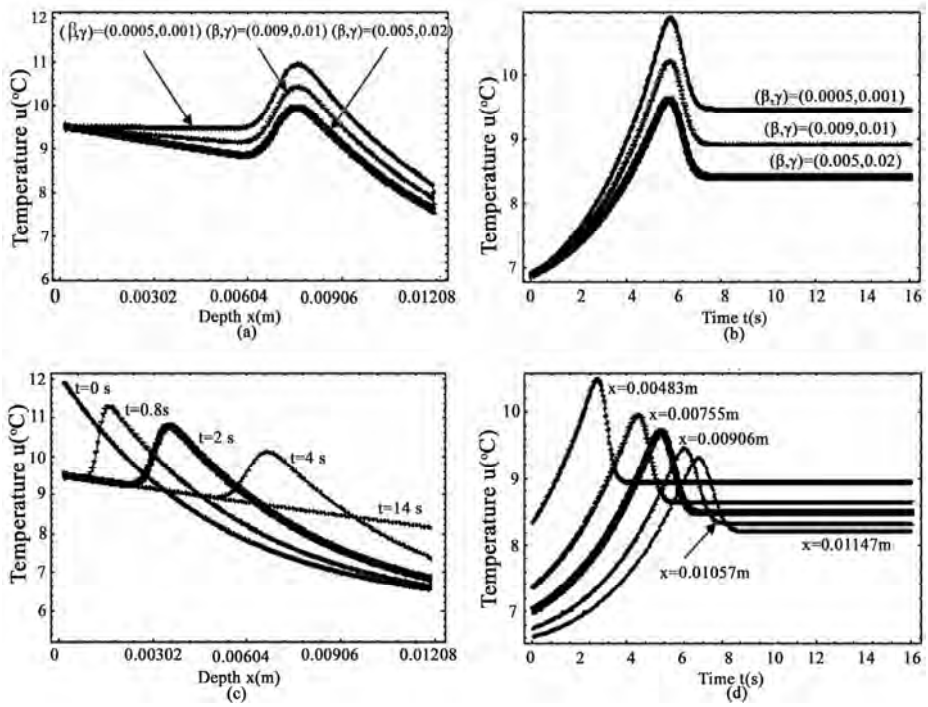
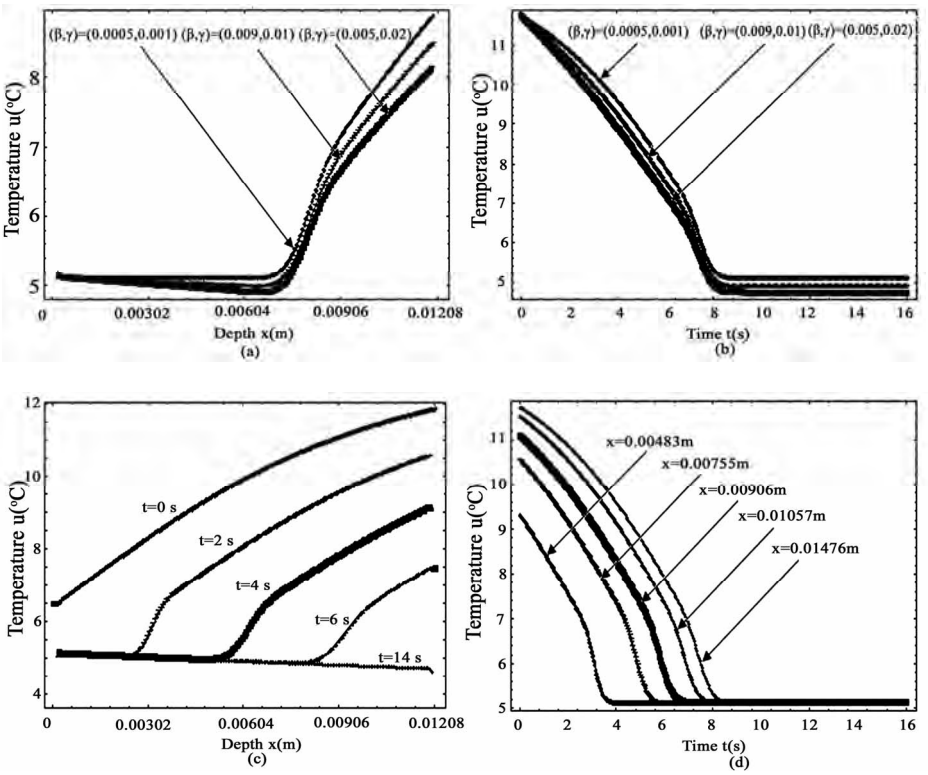


Figure 1. Effect of blood perfusion on the temperature in the case of depth-dependent thermal conductivity of tissue $u(x,0) = u_B(x) = 6[1+\exp(-200\,x)]°C$, with decreasing initial temperature. Plots of the first row show the temperature elevation for different blood perfusion while plots of the second give the temperature elevation for the same blood perfusion at different time t (c) and at different depth x (d).



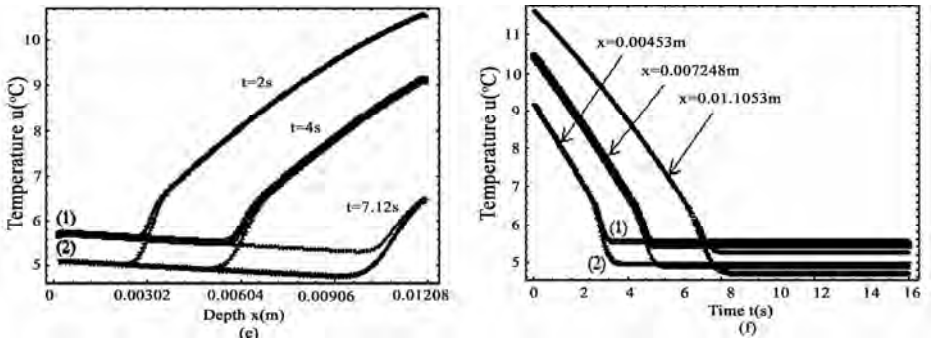


Figure 2. Effect of blood perfusion on the temperature in the case of depth-dependent thermal conductivity of tissue with increasing initial temperature $u(x,0)=u_B(x)=12.928\times[1+\exp(-200x)]^{-1}\,°C$. Plots of the first row show the temperature elevation for different blood perfusion while plots of the second and third row give the temperature elevation for the same blood perfusion at different time t ((c) and (e)) and at different depth$x$ ((d) and (f)). Plots (e) and (f) of the third row indicate the elevated temperature obtained when using temperature-dependent and depth-dependent thermal conductivity, respectively.

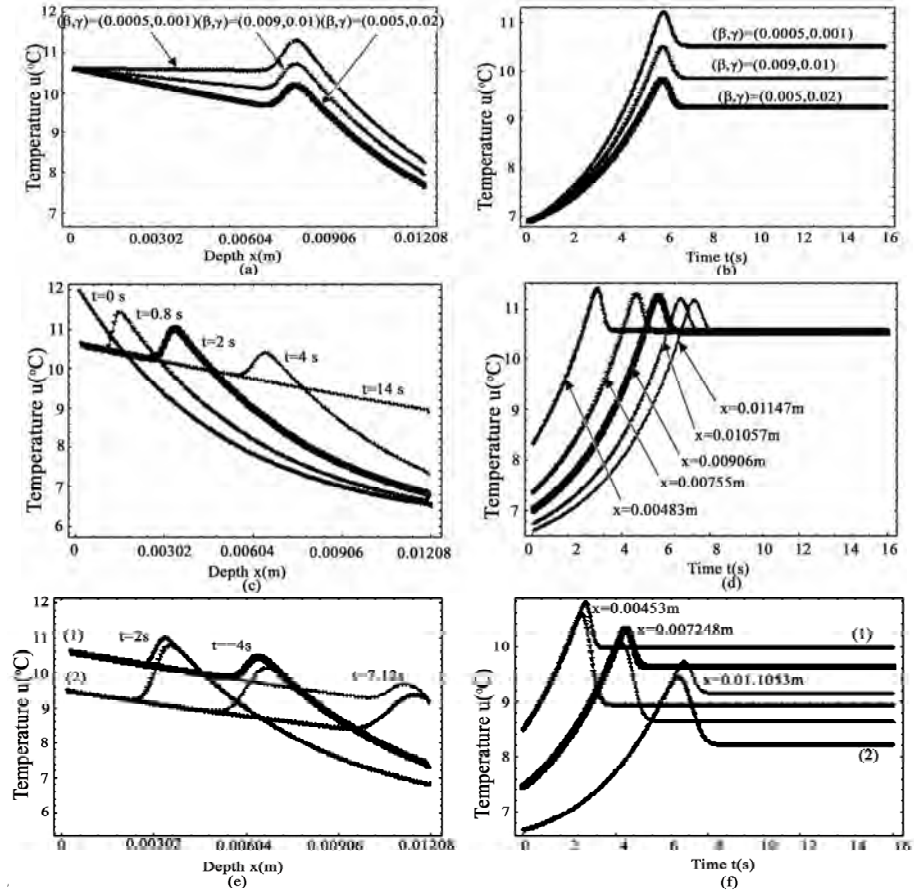


Figure 3. Effect of blood perfusion on the temperature in the case of temperature dependent thermal conductivity of tissue with decreasing initial temperature $u(x,0)=u_B(x)=12.928[1+\exp(-200x)]\,°C$. Plots of the first row show the temperature elevation for different blood perfusion while plots of the second and third row give the temperature elevation for the same blood perfusion at different time t ((c) and (e)) and at different depth x ((d) and (f)). Plots (e) and (f) of the third row indicate the elevated temperature obtained when using temperature-dependent and depth-dependent thermal conductivity, respectively.

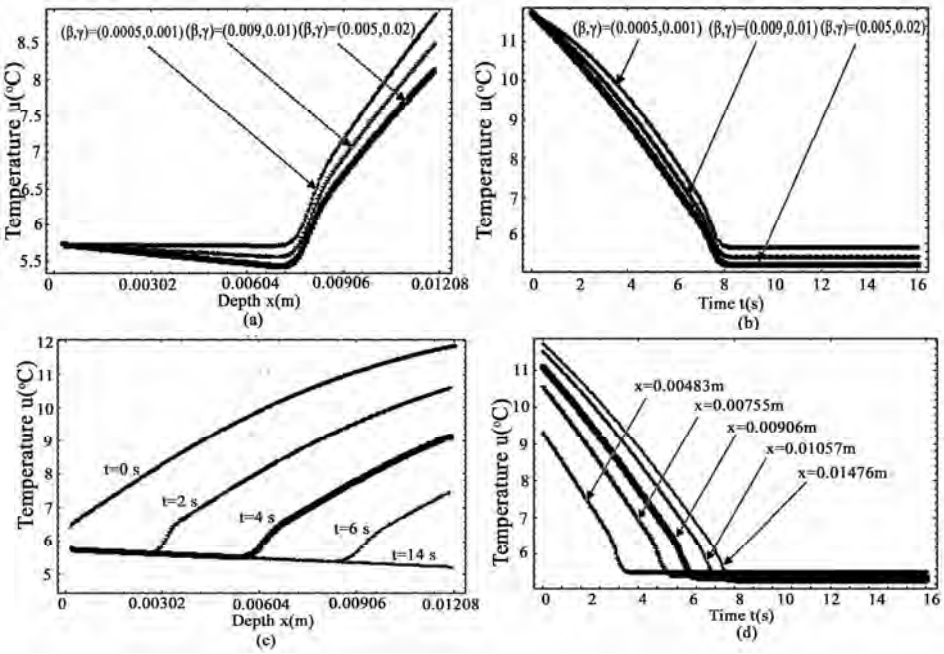


Figure 4. Effect of blood perfusion on the temperature in the case of temperature-dependent thermal conductivity of tissue with increasing initial temperature $u_B(x) = 12.928/[1+\exp(-200x)]°C$. Plots of the first row show the temperature elevation for different blood perfusion while plots of the second row give the temperature elevation for the same blood perfusion at different time t (c) and at different depth x (d).

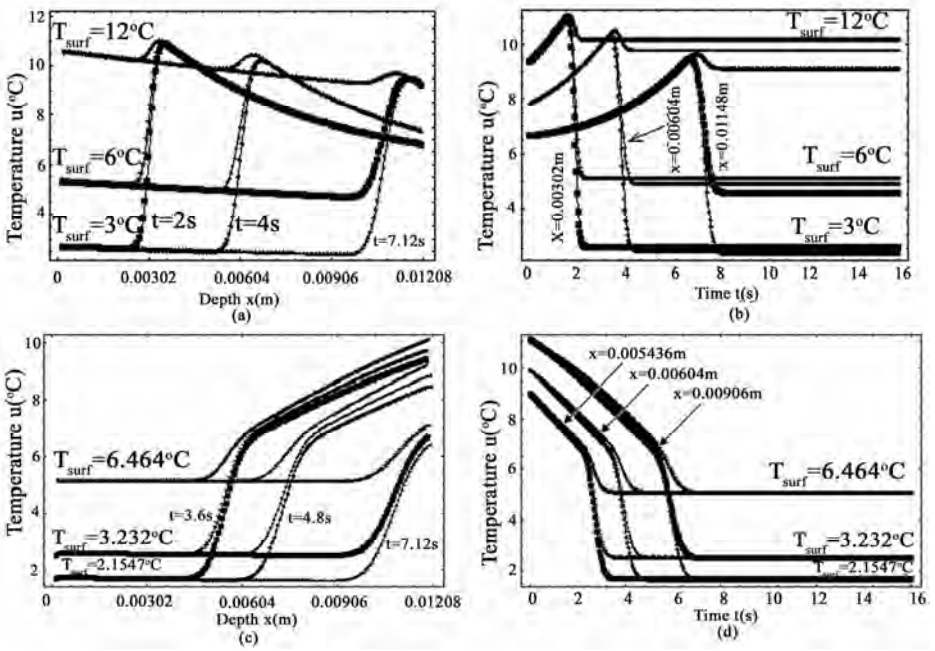


Figure 5. Effect of surface temperature on the temperature distribution, location, magnitude, and width of peak temperature.

cooling phenomenon (in the sense that the temperature is almost below the initial temperature) is observed when the initial temperature increases as a function of depth x (**Figures 2** and **4**).

Plots (e) and (f) of **Figure 2** and **Figure 3** show that the elevated temperature, numerically obtained in the case of temperature-dependent thermal conductivity (plots (1)) is larger than the one obtained in the case of

depth-dependent thermal conductivity (plots (2)).

3.2. Effect of Surface Heating

The effects of surface heating (when the surface of the living tissue is maintained at different constant temperature) on the temperature distribution and on the location of the peak temperature have been studies and observed when the initial temperature is either a decreasing function of depth x (**Figure 5(a)** and **(b)**) or an increasing function of depth x (**Figure 5(c)** and **(d)**). The plots of this figure are obtained with $\alpha = 1/4200000$ and $(\beta,\gamma) = (0.005, 0.02)$. **Figure 5(a)** gives the depth at which the temperature peak occurs at a given time while **Figure 5(b)** gives the time at which the temperature peak at a given depth occurs. **Figure 5(a)** shows that the temperature peak does not depend on the surface temperature. These two plots also show that the magnitude of peak temperature does not depend on the surface temperature: the magnitude of peak temperature is the same for all three surface temperatures. As we can see from **Figures 5(c)** and **(d)**, the depth (**Figure 5(c)** and the time (**Figure 5(d)** from which the variation of the temperature suddenly changes do not depend on the surface temperature. Many other information about the temperature distribution in the living tissue can be obtained from **Figure 5**. For example, it is seen from plots (a) and (b) that the width of the peak temperature increases with the surface temperature: the higher the surface temperature is, the higher is the width of the peak temperature.

4. CONCLUSIONS

In this work, we present a numerical investigation of a 1D bioheat transfer equation with either depth-dependent or temperature-dependent thermal conductivity and with temperature-dependent blood perfusion. An implicit unconditional numerical scheme of the Crank-Nicholson type is constructed and used to solve the nonlinear bioheat transfer equation with given initial and boundary conditions. We found that blood perfusion decreases the temperature in living tissue. It is also shown that the localization and the magnitude of peak temperature do not depend on the surface temperature, while its width increases with surface temperature. The computation presented in this paper can be used to predict the temperature distribution in living tissue during many bioheat transfer processes. The method used in this paper has potential to provide the temperature distribution in tissue in the absent of heat flux.

REFERENCES

[1] Pennes, H.H. (1948) (1998) Analysis of tissue and arterial blood temperatures in the resting human forearm. *Journal of Applied Physics*, **1**, 93-122; **85**, 5-34.

[2] Meyer, C., Philip, P. and Troltzsch, F. (2004) Optimal control of semilinear PDE with nonlocal radiation interface conditions. *IMA Preprint Series*, **2002.**

[3] Yamamoto, M. and Zou, J. (2001) Simultaneous reconstruction of the initial temperature and heat radiative coefficient. *Investment Problems*, **17**, 1181-1202.

[4] Ganzler, T., Volkwein, S. and Weiser, M. (2006) SQP methods for parameter identification problem arising in hyperthermia. *Optimization Methods and Software*, **21**, 869-887.

[5] Arkin, H., *et al.* (1986) Thermal pulse decay method for simultaneous measurement of local thermal conductivity and blood perfusion: A theoritical snalysis. *Journal of Biomechanical Engineering*, **108**, 208-214.

[6] Deuflhard, P. and Seebass, M. (1998) Adaptive multilevel FEM as decisive tools in clinical cancer therapy hyperthermia. Konrad-Zuse-Zentrum für Information stechn, Berlin Takustr, Berlin, 7.

[7] Hill, J. and Pincombe, A. (1992) Some similarity temperature profiles for the microwave heating of a half-space. *Journal of the Australian Mathematical Society, Series B*, **33**, 290-320.

[8] Chato, J.C. (1980) Heat transfer to blood vessels. *Journal of Biomechanical Engineering*, **102**, 110-118.

[9] Weinbaum, S. and Jiji, L.M. (1985) A two simplified bioheat equation for the effect of blood flow on average tissue temperature. *Journal of Biomechanical Engineering*, **107**, 131-139.

[10] Chen, M.M. and Holmes, K.R. (1980) Microvascular contributions in tissue heat transfer. *Annals of the New York Academy of Sciences*, **335**, 137-150.

[11] Charney, C.K. (1992) Mathematical models of bioheat transfer. *Advanced Heat Transfer*, **22**, 19-155.

[12] Davies, C.R., Saidel, G.M. and Harasaki, H. (1997) Sensitivity analysis of 1-D heat transfer in tissue with temperature-dependent perfusion. *Journal of Biomechanical Engineering*, **119**, 77.

[13] Lang, J., Erdmann, B. and Seebass, M. (1999) Impact of nonlinear heat transfer on temperature control in regional hyperthermia. *IEEE Transactions on Biomedical Engineering*, **46**, 1129-1138.

[14] Weierstrass, K. (1915) Mathematische Werke V, New York, Johnson, 4-16; Whittaker, E.T. and Watson, G.N. (1927) A Course of Modern Analysis. Cambridge University Press, Cambridge, 454.

[15] Zhao, J.J., Zhang, J., Kang, N. and Yang, F. (2005) A two level finite difference scheme for one dimensional Pennes' bioheat equation. *Applied Mathematics and Computation*, **171**, 320-331.

[16] Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vallerling, W.T. (1992) Numerical Recipes in Fortran. Cambridge University Press, Cambridge.

[17] Heath, M.T. (2002) Scientific Computing, an Introductory Survey, second edtion, McGraw-Hill, New York.

[18] Liu, J., Chen, X. and Xu, L.X. (1999) New thermal wave aspects on burn evaluation of skin subjected to instantaneous heating. *IEEE Transactions on Biomedical Engineering*, **46**, 420-428.

[19] Tunç, M., Çamdali, Ü., Parmaksizoglu, C. and Cikrik çi, S. (2006) The bioheat transfer equation and its applications in hyperthermia treatments. *Engineering Computations*, **23**, 451-463.

Vol.2, No.3, 139-144 (2010)
doi:10.4236/ns.2010.23023

Natural Science

Student-level numerical simulation of conditions inside an exploding fission-bomb core

Bruce Cameron Reed

Department of Physics Alma College, Alma, USA; reed@alma.edu

Received 2 December 2009; revised 13 January 2010; accepted 30 January 2010.

ABSTRACT

This paper describes a freely-available spreadsheet that has been developed to simulate the conditions of reaction rate, core acceleration and velocity, energy generation, and pressure within a detonating fission-bomb core. When applied to a model of the Hiroshima *Little Boy* bomb, the spreadsheet predicts a yield of 12.7 kilotons, a figure in reasonable agreement with published values.

Keywords: Nuclear Weapons; Numerical Simulation; Fission; Little Boy

1. INTRODUCTION

The discipline of computational physics is now regarded as possessing importance equal to the traditional areas of experimental and theoretical studies. As such, it is important that our students be introduced early in their careers to the power of modern desktop computational tools that can be used to model physically interesting systems. Computational models can facilitate study of physical systems where the theory cannot be solved purely analytically and/or that are not easily realizable experimentally. Sometimes a single graph can serve to dramatically make a point about the behavior of a system that is "latent" in the mathematics but whose magnitude is not immediately apparent. This paper describes and makes freely available an Excel spreadsheet to carry out a student-level simulation of a system that exemplifies all of these points: the conditions of energy release, pressure, fission rate, and expansion inside the core of a detonating fission bomb. The physics of nuclear weapons has been and will remain a fascinating and timely subject. These devices are forbidding and mysterious to students and the general public alike; by helping our students to understand some of the details of their functioning we can equip them to play constructive roles in furthering public understanding of them.

The basic physics of criticality conditions in both "bare" (untamped) and tamped fission cores is described in two papers previously published by this author [1,2]. Because the time-dependence of conditions within a fissioning core is highly non-linear, approximate analytic expressionss had to be developed in those papers in order to arrive at estimates of bomb yields and efficiencies. The purpose of this paper is to utilize the theoretical foundations established in those papers to build an approximate but easy-to-use numerical simulation of a core whose mass, nuclear properties, and tamper properties are set by the user. The resulting spreadsheet is made freely downloadable to interested readers.

The structure of this paper is as follows. In Section 2 I lay out the theoretical background; while this is adopted directly from References [1] and [2] it is presented here for sake of a self-contained discussion. The programming of the simulation itself is described in Section 3, and in Section 4 I present the results of a simulation of the Hiroshima *Little Boy* uranium bomb. A brief summary is presented in Section 5.

2. FISSION-CORE PHYSICS

The essential physical quantity within a fissioning bomb core is the number density of neutrons, N. As described in References [1] and [2], diffusion theory leads to an approximate expression for the space and time- dependence of N within a spherically symmetric core of the form

$$N_{core}(r,t) = N_o e^{(\alpha/\tau)t}\left[\frac{\sin(r/d_{core})}{r}\right] \qquad (1)$$

In this expression, N_o is the initial neutron number density at the center of the core, and is set by whatever initiating device starts the chain reaction. τ is the average time that a neutron will travel before causing a fission,

$$\tau = \lambda_{fiss}^{core} / v_{neut} \qquad (2)$$

where v_{neut} is the average neutron speed and λ_{fiss}^{core} is the mean free path for neutrons between fissions in the

core,

$$\lambda_{fiss}^{core} = \frac{1}{\sigma_f n} \tag{3}$$

where n is the nuclear number density and σ_f the fission cross-section of the fissile core material. d_{core} in **Eq.1** has units of length and can be thought of as fundamentally setting the size of the critical radius; it is given by

$$d_{core} = \sqrt{\frac{\lambda_{fiss}^{core}\lambda_{trans}^{core}}{3(-\alpha+\nu-1)}} \tag{4}$$

where ν is the number of neutrons emitted per fission (so-called secondary neutrons) and λ_{trans}^{core} is the transport mean free path for neutrons,

$$\lambda_{trans}^{core} = \frac{1}{\sigma_t n} \tag{5}$$

σ_t is the so-called transport cross-section. If non-fission neutron capture can be ignored (which should be the case for any sensibly pure fissile material), this is given by the sum of the fission and elastic-scattering cross-sections:

$$\sigma_t = \sigma_f + \sigma_{el} \tag{6}$$

The parameter α in **Eqs.1** and **4** arises in a separation of variables in solving the diffusion equation for the neutron density, and is itself time-dependent as it depends on the core radius as described in what follows. First consider the case of an untamped core, also known as a *bare* core. As described in References [1] and [2], application of appropriate boundary conditions to **Eq.1** to ensure that there is no way for neutrons which have passed out through the surface of the core to be reflected back inside it from the external world leads to the following constraint for the core radius R_C:

$$(R_C/d_{core})\cot(R_C/d_{core}) + (3d_{core}/2\lambda_{trans}^{core})(R_C/d_{core}) - 1 = 0 \tag{7}$$

This is the *criticality* condition for a bare core. Satisfaction of this constraint depends on α through its appearance in d_{core}. If $\alpha = 0$, then the neutron number density is neither growing nor diminishing in time, a condition known as *threshold* criticality. (In an implosion weapon, the moment when this state first occurs during compression of a subcritical core into a supercritical mass is known as *first criticality*. The present work is not inteneded to model implosion scenarios, which are much more complex on account of the inward motion of the core at the moment when fissions are initiated.) In this case d_{core} is completely determined by the fissility parameters, and the value of R_C which satisfies **Eq.7** is the bare threshold critical radius, R_{bare}^{thresh}; for pure U-235 this is about 8.4 cm, equivalent to about 46 kg.

As explained in Reference [1], if one starts with a core of specified radius $R_C > R_{bare}^{thresh}$, then **Eq.7** can be solved for α (through d_{core}); one will find that $\alpha > 0$, which means that the neutron density is growing exponentially in time, a condition known as "supercriticality." As the core rapidly (within a microsecond) expands due to the extreme rate of release of energy by fissions, α will decline as a function of time until it reaches zero, at which point the chain reaction will rapidly shut down; this situation is known as "second criticality" and effectively marks the end of the detonation phase.

In Reference [2], it is shown that if the core is surrounded by a snugly-fitting tamper of non-fissile and non-neutron-absorbing material, diffusion theory leads to the following expressions for the neutron density within the tamper material:

$$N_{tamp} = e^{(\alpha/\tau)t}\begin{cases}\left[\frac{A}{r}+B\right] & (\alpha=0)\\ \left[A\frac{e^{r/d_{tamp}}}{r}+B\frac{e^{-r/d_{tamp}}}{r}\right] & (\alpha>0),\end{cases} \tag{8}$$

where α is as above and τ is again the mean travel time for neutrons between fissions in the *core*, A and B are constants of integration to be determined by boundary conditions, and d_{tamp} is given by

$$d_{tamp} = \sqrt{\frac{\lambda_{trans}^{tamp}\lambda_{fiss}^{core}}{3\alpha}} \tag{9}$$

where λ_{trans}^{tamp} is the mean free path for neutron transport in the tamper material, analogous to **Eq.5** except that the tamper material is presumed to have no fission cross-section:

$$\lambda_{trans}^{tamp} = \frac{1}{\sigma_{el}^{tamp} n_{tamp}} \tag{10}$$

If the core and tamper have outer radii R_{core} and R_{tamp}, then demanding the continuity of neutron density and flux at the core/tamper interface and again requiring that that no neutrons which escape from the tamper to the external world can be reflected back, one finds that the criticality conditions emerge as

$$\left[1 + \frac{2R_{core}^{thresh}\lambda_{trans}^{tamp}}{3R_{tamp}^2} - \frac{R_{core}^{thresh}}{R_{tamp}}\right]\left[\left(\frac{R_{core}^{thresh}}{d_{core}}\right)cot\left(\frac{R_{core}^{thresh}}{d_{core}}\right) - 1\right] + \frac{\lambda_{trans}^{tamp}}{\lambda_{trans}^{core}} = 0, \quad (\alpha = 0) \tag{11}$$

and

$$e^{2(x_{ct}-x_t)}\left[\frac{x_c \cot x_c - 1 - \lambda(x_{ct}-1)}{R_{tamp} + 2\lambda_{trans}^{tamp}(x_t-1)/3}\right]$$

$$= \left[\frac{x_c \cot x_c - 1 + \lambda(x_{ct}+1)}{R_{tamp} - 2\lambda_{trans}^{tamp}(x_t+1)/3} \right] \quad (\alpha > 0) \qquad (12)$$

where

$$\left.\begin{aligned} x_{ct} &= R_{core}/d_{tamp} \\ x_c &= R_{core}/d_{core} \\ x_t &= R_{tamp}/d_{tamp} \\ \lambda &= \lambda_{trans}^{tamp}/\lambda_{trans}^{core} \end{aligned}\right\} \qquad (13)$$

Eq.11 corresponds to tamped threshold (α = 0) criticality. Once values for the d's and λ's are given the only unknown is R_{core}^{thresh}, the core radius for tamped threshold criticality. In using **Eqs.12** and **13** the idea is that one again *specifies* the mass and hence radius of the core, R_{core} ($> R_{core}^{thresh}$), and solves for α. This would presumably be the value of α when fissions are initiated at t = 0; as the core expands α will subsequently decline until second criticality is reached. A tamper serves to increase the efficiency (and hence the yield) of a weapon through two effects. First, by briefly retarding the expansion of the core, the tamper causes α to remain greater for a longer time than it would have otherwise; this is beneficial as the rate of fissions – and hence the energy release – depends exponentially on α. Second, the tamper serves to reflect neutrons back into the core, effectively decreasing the loss of fission-causing neutrons from within the core to the outside world. In modern weapons – engineering parlance a tamper is known as a reflector; I retain the historical terminology. The retardation effect is difficult to model analytically at this level and so is not accounted for in **Eqs.12** and **13**; they do, however, include the reflection effect in the boundary conditions used to establish them. The retardation effect is treated approximately in the simulation as described following **Eq.17** below.

We can now consider the time-dependence of various quantities in order to begin formulating a simulation. At a moment when the core has volume V_{core}, the rate of fissions $R(t)$ (fission/sec) is given by

$$R(t) = \left(\frac{N_o V_{core}}{\tau} \right) e^{(\alpha/\tau)t} \qquad (14)$$

If each fission releases energy E_{fiss} (typically ~ 180 MeV), then the rate of energy release within the core is

$$\frac{dE}{dt} = \left(\frac{N_o V_{core} E_{fiss}}{\tau} \right) e^{(\alpha/\tau)t} \qquad (15)$$

The total energy liberated to a given time can be tracked by numerically integrating **Eq.15**; this determines the pressure within the core as a function of time. This follows from the thermodynamic pressure – energy density relationship

$$P_{core}(t) = \frac{\gamma E(t)}{V_{core}(t)} \qquad (16)$$

The choice of the parameter γ depends on whether gas pressure (γ = 2/3) or radiation pressure (γ = 1/3) is dominant; the latter dominates for per-particle energies greater than about 2 keV and will presumably be the case for the later, more energetic stages of the reaction. I use the core volume in **Eq.16** on the rationale that the fission products which cause the gas/radiation pressure will likely largely remain within the core.

Following Reference [1], I model the core as an expanding sphere of radius $r(t)$ with all parts of the sphere moving at speed $v(t)$, driven by the energy release from fissions. Do not confuse this velocity with the average neutron speed, which does not directly come into this part of the development (it does enter implicitly, however, through τ). Invoking the work-energy theorem in its thermodynamic formulation, $W = P(t)\,dV$, I equate the work done by the gas (or radiation) pressure in changing the core volume by dV over time dt to the change in the core's kinetic energy over that time:

$$P\frac{dV}{dt} = \frac{dK}{dt} \qquad (17)$$

For simplicity in developing the simulation, I treat the tamper as remaining of constant density. Now, it is desirable to make some effort to account for the retarding effect of the tamper on the core. To do this, I treat the dK/dt term in **Eq.17** as involving the velocity of the core expansion but with the mass involved being that of the core *plus* that of the tamper. The dV/dt term is taken to apply to the core. With r as the radius and v the expansion velocity of the core, we have

$$\frac{\gamma E(t)}{V_{core}(t)}\left(\frac{dV_{core}}{dt}\right) = \frac{dK_{total}}{dt} \Rightarrow$$

$$\frac{\gamma E(t)}{V_{core}(t)}\left(4\pi r^2 \frac{dr}{dt}\right) = \frac{1}{2} M_{c+t}\left(2v\frac{dv}{dt}\right)$$

With $dr/dt = v$

$$\frac{dv}{dt} = \frac{4\pi r^2 \gamma E(t)}{V_{core} M_{c+t}} \qquad (18)$$

Numerically integrating this result will give $v(t) = dr/dt$, which can be integrated to give the core radius as a function of time.

3.THE SIMULATION

I have developed an Excel spreadsheet to carry out the calculations described above. This is freely available at http://othello.alma.edu/~reed/FissionCore.xls. In this Section I describe the general layout of this spreadsheet; some results are described in Section 4 below.

This spreadsheet consists of three interlinked sheets. On the first, the user inputs fundamental data such as core and tamper material densities, atomic weights, cross-sections, the secondary neutron number, the average secondary-neutron energy, values for E_f and γ, the desired core mass, the outer radius of the tamper, and the number of "initial neutrons" in the core at $t = 0$. These are entered in convenient units such as g/cm^3, barns, and MeV; the spreadsheet subsequently carries out all calculations in MKS units. The Excel "Goal Seek" function is then run three times, to establish values for 1) the bare threshold critical radius, 2) the tamped threshold critical radius, and 3) the value of α corresponding to the chosen core mass. The masses in 1) and 2) are computed for reference and for the fact that they are needed for some calculations involving the expansion of the core as described below. The chosen core mass should exceed that corresponding to R_{tamp}^{thresh}.

A significant complexity in carrying out this simulation is that one apparently needs to solve **Eq.12** for the value of α corresponding to each time-stepped core radius between first and second criticality: the fission rate, energy generation rate, and pressure all depend on α as a function of time. I have found, however, that α is usually quite linear as a function of core radius. This behavior greatly simplifies the actual time-dependent simulation. Sheet 2 of the spreadsheet allows one to establish parameters for this linear behavior for the values of the various parameters that the user inputs on Sheet 1. Here, the user solves (again using the Goal Seek function) for the value of α for 25 values of the radius. These start at the initial core radius and proceed to 1.25 times the value of the second-criticality radius for a *bare* core of the mass chosen by the user on Sheet 1; this range appears to be suitable to establish the behavior of α. The rationale for this arrangement is as follows. As shown in Reference [2], if the chosen core mass is equal to C bare threshold critical masses, criticality will hold over a range of radii given by

$$\Delta r = \left(C^{1/2} - C^{1/3}\right) R_{bare}^{thresh} \tag{19}$$

The presence of a tamper means that the core will expand somewhat beyond Δr before second criticality is reached, but **Eq.19** sets the essential length scale of the expansion. For convenience, Sheet 2 utilizes a "normalized" radius defined as

$$r_{norm} = \frac{r - C^{1/3} R_{tamp}^{thresh}}{\left(C^{1/2} - C^{1/3}\right) R_{tamp}^{thresh}} \tag{20}$$

where C is now defined as the number of tamped threshold critical masses. $r_{norm} = 1$ corresponds to the second criticality radius one would compute from **Eq.19** if it applied as well to a tamped core. Sheet 2 tracks the changing mass density, nuclear number densities, and mean-free-paths within the core as a function of r. By running the Goal Seek function on each of the 25 radii, the user adjusts α in each case to render **Eq.12** equal to zero. The behavior of $\alpha(r)$ is then displayed in an automatically-generated graph. On a separate line with α fixed to a value very near zero (10^{-10} is built-in), the user adjusts the radius to once again render **Eq.12** equal to zero, thus establishing the radius of second criticality for his or her parameters. The slope and intercept of a linear $\alpha(r)$ fit are then automatically computed in preparation for the next step.

The actual time-dependent simulation occurs on Sheet 3. The simulation is set up to involve 500 timesteps, one per row. The initial core radius is transferred from Sheet 1 for $t = 0$. Because much of the energy release in a nuclear weapon occurs during the last few generation of fissions before second criticality, this Sheet allows the user to set up two different timescales: an "initial" one (dtinit) intended for use in the first few rows of the Sheet when a larger timestep can be tolerated without much loss of accuracy, and a later one (dtlate), to be chosen considerably smaller and used for the majority of the rows. In this way a user can optimize the 500 rows to both capture sufficient accuracy in the last few fission generations and arrange that $\alpha(r)$ is just approaching zero at the last steps of the process. Typical choices for dtinit and dtlate might be a few tenths of a microsecond and a few tenths of a nanosecond, respectively. At each radius, the Sheet computes the value of $\alpha(r)$ from the linear approximation of Sheet 2, the core volume, mass density, nuclear number densities and mean free paths within the core, τ, rates of fission and energy generation, pressure, and total energy liberated to that time. The acceleration of the core is computed from **Eq.18**, and the core velocity and radius are updated depending upon the timestep in play; the new radius is transferred to the subsequent row to seed the next step. The user is automatically presented with graphs of $\alpha(r)$, the fission rate, pressure, and total energy liberated (in kilotons equivalent) as functions of time.

4. A SIMULATION OF THE HIROSHIMA *LITTLE BOY* BOMB

As described in References [2] and particularly [3], the Hiroshima *Little Boy* core comprised about 64 kg of enriched U-235 in a cylindrical configuration surrounded by a cylindrical tungsten-carbide tamper of diameter and length 13 inches, mass approximately 310 kg, and density 14.8 g /cm^3. Values for the various core and tamper parameters are given in **Table 1**; these are adopted from Reference [2]. Assuming these values and taking the core to be spherical (radius 9.35 cm at a density of 18.71 g/cm^3; this figure is 235/238 times the density of natural uranium, 18.95 g/cm^3) and surrounded by a spherical

tungsten-carbide tamper of outer radius 18 cm (mass 311 kg), Sheet 1 of the author's spreadsheet indicates that the tamped threshold critical mass of U-235 in this configuration is 18.4 kg, about a 60% reduction from the bare threshold critical mass of 45.9 kg. **Figure 1** shows that the run of $\alpha(r)$ for this situation is quite linear out to the computed second criticality radius of 12.04 cm. Upon

Table 1. Parameters for U-235 core/tungsten-carbide tamper model.

Quantity	Meaning	Unit	Core value	Tamper value
ρ	Mass·density	g·cm^{-3}	18.71	14.8
σ_f	Fission·cross-section	bn	1.235	---
σ_{si}	Scattering·cross-section	bn	4.566	6.857
ν	Secondary·neutron·number	---	2.637	---
A	Atomic·weight	g·mol^{-1}	235.04	195.84
E_f	Energy·per·fission	MeV	180	---
K_n	Fission·neutron·average·energy	MeV	2	---

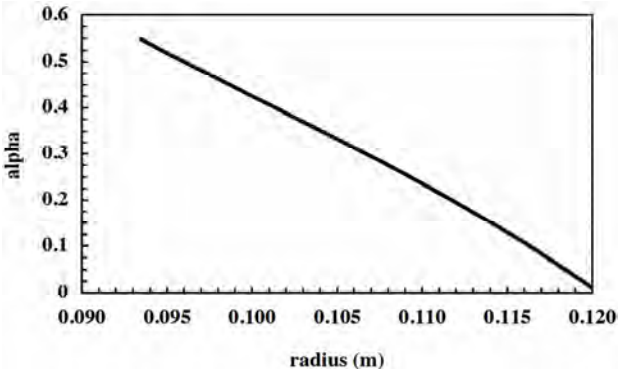


Figure 1. Criticality parameter α as a function of radius for the simulation of the *Little Boy* 64-kg U-235 core plus 311-kg tungsten-carbide tamper. $\alpha(r)$ is approximately linear, with slope -20.35 m^{-1} and intercept 2.45 m.

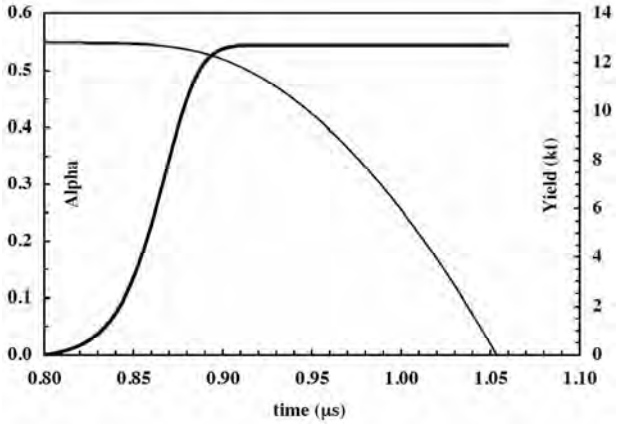


Figure 2. Criticality parameter α (thin line, left scale) and cumulative energy yield in kilotons (thick ascending solid line, right scale) for the *Little Boy* simulation.

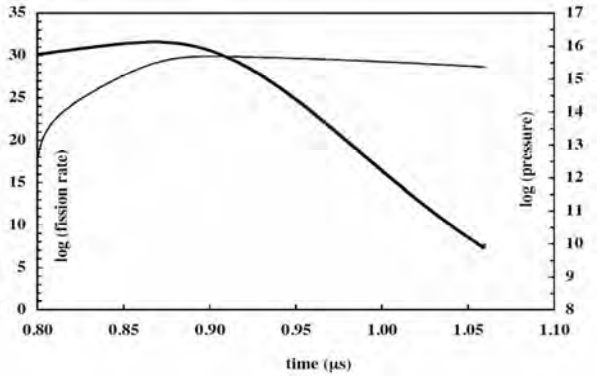


Figure 3. Logarithmic plots of pressure in Pa (thin line, right scale) and rate of fissions per second (thick descending solid line, left scale) for the *Little Boy* simulation.

adopting $\gamma = 1/3$, an equivalence of 4.2×10^{12} Joule/kt and a single initial neutron to start the reaction ($N_o = 292$ m^{-3}), **Figures 2** and **3** show the results of the simulation.

The brevity and violence of the detonation are astonishing. The vast majority of the energy is liberated within an interval of about 0.1 μs. The pressure peaks at close to 5×10^{15} Pa, or about 50 *billion* atmospheres, equivalent to about one-fifth of that at the center of the Sun, and the fission rate peaks at about 4×10^{31} per second. The core acceleration peaks at about 1.5×10^{12} m/s^2 at $t - 0.9$ μs, and second criticality occurs at $t - 1.05$ μs, at which time the core expansion velocity is about 270 km/s. These graphs dramatically illustrate what Robert Serber wrote in *The Los Alamos Primer* : "Since only the last few generations will release enough energy to produce much expansion, it is just possible for the reaction to occur to an interesting extent before it is stopped by the spreading of the active material" [4]. The predicted yield of *Little Boy* from the present model is 12.7 kt. This result is in surprisingly good agreement with the estimated – 12 kt yield published by Penney, *et al.* [5]. At a fission yield of 17.59 kt per kg of pure U-235, this represents an efficiency of only about 1.1% for the 64 kg core. Some of this agreement must be fortuitous, however, in view of the approximations incorporated in the present model. That the yield estimate needs to be taken with some skepticism is demonstrated by the fact that increasing the initial number of neutrons to 10 increases the yield to 18.7 kt. However, this change does not much affect the timescale or the peak pressure and fission rates. Users who download FissionCore.xls will find this example pre-loaded.

5. SUMMARY

This paper describes the development of a spreadsheet for simulating the conditions within a detonating fission-bomb core. The simulation is straightforward enough to be used with students, and for a simulation of

the Hiroshima *Little Boy* bomb predicts a yield in reasonable accord with published values. This type of simulation can help students grasp some of the underlying physics of and get a sense of the extreme physical conditions that briefly occur within such devices.

REFERENCES

[1] Reed, C. (2007) Arthur Compton's 1941 report on explosive fission of U-235: A look at the physics. *American Journal of Physics*, **75(12)**, 1065-1072.

[2] Reed, C. (2009) A brief primer on tamped fission-bomb cores. *American Journal of Physics*, **77(8)**, 730-733.

[3] http://nuclearweaponarchive.org/Nwfaq/Nfaq8.html

[4] Serber, R. (1992) The Los Alamos primer: The first lectures on how to build an atomic bomb. University of California, Berkeley, CA, 12.

[5] Penney, L., Samuels, D.E.J. and Scorgie, G.C. (1970) The Nuclear Explosive Yields at Hiroshima and Nagasaki. *Philosophical Transactions of the Royal Society of London*, **A266**, 357-424.

Vol.2, No.3, 145-154 (2010)
doi:10.4236/ns.2010.23024

Natural Science

Signature of chaos in the semi quantum behavior of a classically regular triple well heterostructure

Tiokeng Olivier Lekeufack[1], Serge Bruno Yamgoue[2], Timoleon Crepin Kofane[1,3]

[1]Laboratory of Mechanics, Department of Physics, Faculty of Science, University of Yaounde I, Yaounde, Cameroon; lekeufackolivier@gmail.com
[2]Department of Physics Higher, Teacher's Training College, Bamenda, Cameroon
[3]The Abdus Salam International Center for Theoretical Physics, Trieste, Italy

Received 28 December 2009; revised 14 January 2010; accepted 25 January 2010.

ABSTRACT

We analyze the phenomenon of semiquantum chaos in the classically regular triple well model from classical to quantum. His dynamics is very rich because it provides areas of regular behavior, chaotic ones and multiple quantum tunneling depending on the energy of the system as the Planck's constant $\hbar$ varies from 0 to 1. The Time Dependent Variational Principle TDVP using generalized Gaussian trial wave function, which, in many-body theory leads to the Hartree Fock Approximation TDHF, is added to the techniques of Gaussian effective potentials and both are used to study the system. The extended classical system with fluctuation variables nonlinearly coupled to the average variables exhibit energy dependent transitions between regular behavior and semi quantum chaos monitored by bifurcation diagram together with some numerical indicators.

Keywords: Nonlinear Dynamics; Semi Quantum Chaos; Effective Potential

1. INTRODUCTION

The quantum computer science, the quantal dynamics of hetero-structures, the mesoscopic behavior of some systems, the chaotic entropy production in open quantum systems, the zero momentum (long wavelength) part of the problem of pair production of charged scalar particles by a strong external electric field [1], the quantum suppression of diffusion (dynamic localization) [2] and the quantum unique ergodicity in statistical thermodynamics are good candidates for a wide range of the study and the application of semiquantum chaos in experimental physics, nuclear physics and quantum chemistry as well. The definition and observation of chaotic behavior in classical systems are familiar and well understood [3,4]. However, the proper definition of chaos for quantum systems and its experimental manifestations are still unclear [5-8]. We use the term *semiquantum chaos* to refer to the study of the quantum dynamics of systems whose classical limit is regular (restriction to Hamiltonian systems).

Over the last years, different approaches of studying chaos in classical and quantum systems have attracted increasing attention. For example, we have the problem of pair production of particles by strong external electric field, the two particles interaction through a biquadratic coupling, *i.e.*, a two-degrees of freedom system of which one is classical and the other purely quantum non linearly coupled and which exhibit chaos [9-11]. In addition, great attention is focused on a system of two particles non-linearly coupled, whose classical limit is chaotic, involving quantum properties. This is the case for the authors of [9], who studied the duality wave/particle to take into account the quantum dynamics, then combined Quantum Theory of Motion (**QTM**) with Quantum Fluid Dynamics (**QFD**) in classical chaos, and found quantum parameters being chaotic. For the full quantum dynamics, only experimental studies have been done up till date on hetero-structures undergoing multiple tunneling resonance, and leading to different approaches of building quantum computers [12,13]. In addition to all the aforementioned routes of analyzing non-classical chaos, many theories and formalisms are always used. One of the most important of these routes is the time dependent variational approach, which, in the many body theory, leads to the **TDHF** using a Gaussian trial wave function [14-20]. The appearance of this wave function has provided great and interesting results not only in general universe physics field, but also in nuclear physics and quantum chemistry [14,16]. This usage generally integrates the mean field theory [21-24]. The signature of the Gaussian Effective potential **GEP** [1,17,25,26] is also a good indicator of non-classical chaos. In fact, effective potentials [26] are used to assess the impact of quantum

effects such as zero point fluctuation and tunneling on the magnitude and the geometry of classical potentials for which they are an extension because carrying quantum corrections. Since some of the diagnostics [17,27,28] for chaos are based on the geometry of the potential, the effective potential techniques are especially powerful in combination with such method.

Our aim in this paper is to study the semiquantal dynamics of a triple-well potential hetero−structure. We introduce additional (fluctuation) degrees of freedom at order of the Planck's constant, representing quantum counterpart. The non-linear coupled system of the first order autonomous flow is obtained and both analytically and numerically studied. The complete dynamics of the coupled quantum and classical oscillators is described by a classical effective Hamiltonian, which is the expectation value of the quantum Hamiltonian or equivalently the Dirac's action. The utilization of the (**GEP**) to draw the scheme of the system evolution provides fixed points, tunneling and multiple resonant. The numerical simulations are sometimes in good agreement, and bear somewhat surprises.

Our paper is organized as follows: in Section 2, we briefly describe our model of triple-well hetero-structure and apply the time dependent variational principle from which we obtain our basic set of equations. Section 3 presents some analytical considerations and introduces the **GEP**. Section 4 contains numerical simulations to confirm assumptions made earlier in the previous sections. Finally, Section 5 summarizes the main results of the paper and provides discussion with perspectives for future works.

2. MODEL AND EQUATIONS

This section is to describe the triple-well hetero-structure model and draw the basic set of motions equations. Hetero-structures, besides offering very interesting new technological perspectives, represent a unique opportunity to study fundamental question of mechanics such as many-body interaction, resonant tunneling, ergodicity and chaos [29-31]. A triple quantum well structure **TQWS**, all like a double quantum well structure **DQWS** constructed by the authors of the reference [13], can be constructed under several practical apparatus with GaAs/AlGaAs. Physically, the structure presents the diagram of a tri-stable potential energy, which is the potential energy including stable and unstable equilibrium positions:

$$V(Q)=\frac{A}{2!}Q^2-\frac{B}{4!}Q^4+\frac{C}{6!}Q^6 \tag{1}$$

where Q is the coordinate and $V(Q)$ the potential energy; $A, B,$ and C are physical parameters on which depends the numerous variety of configurations. This type of potential is called $\lambda\Phi^6$ potential and exhibits several configurations according to the values of the constants A, B, and C. On the one hand, we can have mono- and bi-stable catastrophic potentials *i.e.* one and two potential wells, respectively. It corresponds to cases of beams, flexible or breakable structures. On the other hand, we have mono-, bi- and tri-stable non-catastrophic configurations, and it corresponds for example to the potentials of O-H chemical bound in ice. It can also describe the dynamics of some rigid structures, oscillators, hetero structures as well. We focus our attention on triple well configuration in order to study its various usages, to find out the importance and the rich dynamics offered by the additive potential well considering its symmetry. Note that more extensive works have recently been done on the $\lambda\Phi^6$ potential [32-34] and good results were obtained about its higher precisions brought in the study of the various systems described by it. The graph of **Figure.1** shows the potential energy. Our case belongs to mesoscopic physics, which deals with systems that are macroscopic but retain essential quantum features [30].

Consider a particle submitted to that potential energy. The classical action is given by:

$$S=\int dt\left\{\frac{1}{2}\left(\frac{\partial Q}{\partial t}\right)^2-V(Q)\right\} \tag{2}$$

From this action, we can derive the Lagrangian and through the Euler-Lagrange equations with respect to the coordinates, we come out with the first order autonomous system flow. This is the classical approach of our problem. It is obvious to realize, by solving the Schrödinger equation, that the corresponding quantum dynamics looks regular no matter whether this classical system behaves regularly or not:

$$i\hbar\frac{\partial}{\partial t}\Psi(Q;t)=\left[-\frac{1}{2}\frac{d^2}{dQ^2}+V(Q)\right]\Psi(Q;t) \tag{3}$$

As explained in the introduction, it is quite an interesting and difficult approach to find whether our system bears quantum features or not. It corresponds to a situation where a classical oscillator interacts with quantum one through bi-cubic non-linear coupling. This necessitates the introduction of additive degrees of freedom, namely

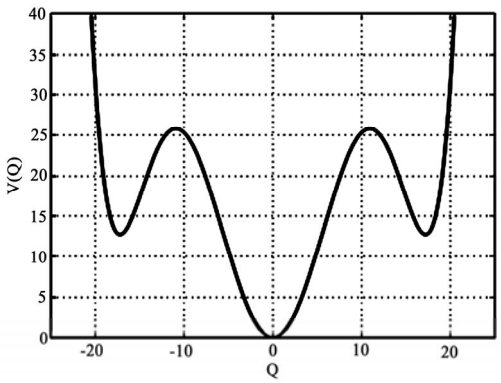


Figure 1. The triple well oscillator potential, **Eq.1** with A=1.0; B=0.0706 and C=0.0034.

fluctuation ones. Their quantum nature belongs to the significance of the Planck's constant $\hbar$. We introduce classical and quantum coordinates [18]

$$q(t)=<Q>\ ;\ \ \hbar G(t)=<Q^2>-<Q>^2 \tag{4}$$

Together with classical and quantum momenta [18]

$$\Pi(t)=<P>\ ;\ \ p(t)=<-i\hbar\frac{\partial}{\partial Q}> \tag{5}$$

The $<\ >$ has the sense of the mean value. We now consider a trial wave function, particularly the one that has been successfully introduced by Gauss [15,18,35] and whose results were in good convenience with the physics of the considered system

$$\Psi(Q,t)=N\exp\left\{-\frac{1}{2\hbar}(Q-q)^2\Omega+\frac{i}{\hbar}P(Q-q)\right\} \tag{6}$$

With $\Omega=\frac{1}{2}G^{-1}(t)-2i\Pi(t)$ from the authors of [15]. This wave function has to satisfy usual quantum requirements such as the normalization condition and the Heisenberg uncertainty principle. The normalization condition gives $N=(2\Pi\hbar G)^{-1/4}$ and the mean values are easily calculated

$$<Q>_\Psi=q(t)\ ;\ \ <-i\hbar\frac{\partial}{\partial Q}>_\Psi=p(t)\ ;$$

$$<Q^2>_\Psi=q^2+\hbar G(t)\ ;\ \ <i\hbar\frac{\partial}{\partial t}>_\Psi=p\dot{q}-\dot{\Pi}G$$

Moreover, the uncertainty principle is obtained:

$$\left(<Q^2>-<Q>^2\right)^{1/2}\left(<P^2>-<P>^2\right)^{1/2}=\hbar G^{1/2}\left(\frac{1}{4}G^{-1}+4\Pi^2G\right)^{1/2}$$
$$=\frac{\hbar}{2}\left[1+(4\Pi G)^2\right]^{1/2}\geq\frac{\hbar}{2}$$

Q, P, G and Π are variational parameters and we demand that their variation vanishes at infinity.

We now turn to the derivation of appropriate semi quantum equations of motion. The **TDHF** or Gaussian variational approximation can easily be performed with the help of Dirac's variational principle [15]. We require the effective action

$$S_{eff}[\Psi]=\int dt\ <\Psi(t)\left|i\hbar\frac{\partial}{\partial t}-H\right|\Psi(t)> \tag{7}$$

to be stationary against arbitrary variation of a normalized wave function which vanishes at $t\to\infty$: $\frac{\delta S_{eff}}{\delta\Psi}=0$ for all Ψ, with $<\Psi|\Psi>=1$. This is equivalent to the exact time dependent Schrödinger equation. With this variational principle, one can solve the quantum mechanical time evolution problem approximately by restricting the variation of the wave function to a sub space of Hilbert space. The effective action is therefore given by:

$$S_{eff}=\int_{-\infty}^{+\infty}dt\left\{p\dot{q}-H_{cl}-\frac{\hbar}{8G}-\hbar G\left[\dot{\Pi}+2\Pi^2+\frac{1}{2}V^{(2)}(q)\right]\right.$$
$$\left.-\frac{\hbar^2G^2}{8}V^{(4)}(q)-\frac{\hbar^3G^3}{48}V^{(6)}(q)\right\} \tag{8}$$

where

$$H_{cl}(q,p)=\frac{p^2}{2}-V(q)\ \ ;\ \ V^{(n)}=\frac{\partial^n V}{\partial Q^n} \tag{9}$$

This contains higher order additive term in $\hbar^3G^3$ symbolizing higher quantum correction as compared to **Eq.6** of [18] and **Eq.18** of [35]. The appearance of the additive term may increase the richness of the studied systems dynamics. The semi quantum variational equations of motion are therefore derived via the Euler Lagrange equation for the effective action, **Eq.9**, by independent variation with respect to Q, P, G and Π:

$$\begin{cases}\dot{q}=p\\ \dot{p}=-V'(q)-\frac{\hbar G}{2}V^{(3)}(q)-\frac{\hbar^2G^2}{16}V^{(5)}(q)\\ \dot{G}=4\Pi G\\ \dot{\Pi}=\frac{1}{8}G^{-2}-2\Pi^2-\frac{1}{2}V^{(2)}(q)-\frac{\hbar G}{4}V^{(4)}(q)-\frac{\hbar^2G^2}{16}V^{(6)}(q)\end{cases} \tag{10}$$

As compared to **Eq.7** of [18] and **Eqs.22-25** of [35], there appear additive non-linear terms. We expect more corrections on the dynamics of the aforementioned studied system. The above equations are the **TDHF** ones because using the Gaussian wave function in Dirac's principle. The validity of this **TDHF** approximation, which has been widely tested [11,18] by being applied to various quantum mechanical problems and drawing good results, is awaited here. Since the equations are highly nonlinear, we expect the trajectories to be regular and irregular providing chaotic behavior. It is important to note that our equations are coupled, showing the link between classical and quantum interactions. At $\hbar\to 0$ classical limit, only the first two equations remain, confirming that the fluctuation variables are responsible for quantum effects. In addition, the Ehrenfest theorem is then verified to confirm the validity of our system [17].

3. GEP AND THEORETICAL ANALYSIS

The purpose of this section is to derive the Gaussian Effective Potential (**GEP**) and to report some analytical considerations, which may help to better understand the dynamics of the semiquantum equations of motion.

3.1. The Static GEP

The variety of techniques used to study dynamical systems comes about because the measures (such as K entropy [36], Lyapunov numbers [36-39]), the diagnostics (the Melnikov functions [38-40]), and the signatures of chaos, which generally lie in phase space, are dynamic and have no direct interpretation in quantum dynamics. We consider a parallel approach, that of using classical techniques of analysis by reducing the problem to that of an effective classical system *i.e.* looking at Hamiltons' equations in a modified potential. Effective potentials are therefore used to assess the impact of quantum ef-

fects such as zero-point fluctuations and tunneling on the magnitude and the geometry of the classical potentials. It was introduced by Stevenson [26] and successfully tested in the Henon Heiles and four leg potentials problem. Most of the diagnostics for chaos are based on the geometry of the potential. The effective potential technique is especially powerful in combination with such method [17]. The idea of Stevenson is to approximate the effective potential of a system by using Gaussian wave function. In general, the effective potential of a system gives us a picture of how the quantum fluctuations modify the classical potential. Following the conserved total energy for our non-dissipative model with its complicated functions of q, p, G and Π, we choose the simplest initial conditions with zero momenta $p(t=0)=\Pi(t=0)=0$; then, evaluating the initial total energy $E = < V(Q) > - \frac{\hbar^2}{2} < \frac{\partial^2}{\partial Q^2} >$ we have:

$$E(q,G)=V(q)+\frac{\hbar}{8G}+\frac{\hbar G}{2}\left[A-\frac{B}{2}q^2+\frac{C}{24}q^4\right]+\frac{\hbar^2G^2}{8}\left[-B+\frac{C}{2}q^2\right]+\frac{\hbar^3G^3}{48}C \tag{11}$$

At the minima of the classical potential, (*i.e.* q = 0 or q = ± 17.2, numerically), we obtain the initial energy for the correspondent minimum q = 0,

$$E_q(G)=\frac{\hbar}{8G}+\frac{\hbar A}{2}G-\frac{\hbar^2 B}{8}G^2+\frac{\hbar^3 C}{48}G^3 \tag{12}$$

which is a function of the initial value of G: this is the relevant control parameter for our model. We now derive the GEP. In analogy to a classical system, we consider the *effective potential* which is defined as the total minus the kinetic energy, from **Eq.8**, the kinetic part being zero according to the initial conditions. Thus,

$$V_{eff}(q,G)=E(q,G)=V(q)+\frac{\hbar}{8G}+\frac{\hbar G}{2}\left[A-\frac{B}{2}q^2+\frac{C}{24}q^4\right]+\cdot$$
$$\frac{\hbar^2G^2}{8}\left[-B+\frac{C}{2}q^2\right]+\frac{\hbar^3G^3}{48}C \tag{13}$$

In **Figure 2**, we present equipotentials for the GEP corresponding to couples of (q,G) varying through

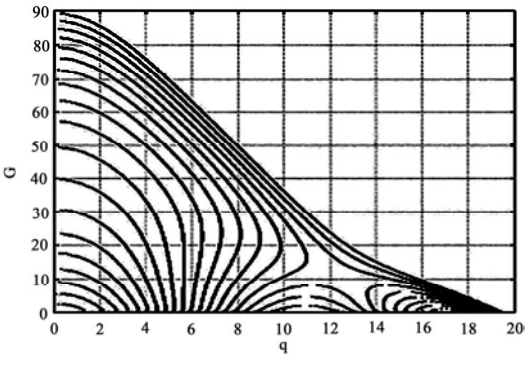


Figure2. The effective semi-quantum potential (GEP), **Eq.13** restricted in the q-G positive plane.

Eq.10 for the same conservative energy. It shows how the classical potential is modified.

In the particular case of a potential well, the Heisenberg uncertainty principle implies that, if the centroid is concentrated in a small region ΔQ then the uncertainty on the conjugate momentum ΔP is very large $\Delta P \geq \hbar/2\Delta Q$; there appears a large kinetic contribution in the total energy. In the triple well, the quantum fluctuations lower the potential barriers such that a particle with classical insufficient energy to spray over the barrier finds it possible: that is quantum tunnel effect [21-24]. The most interesting features are the valleys in the potential energy surface that lead to the saddle point regions separating the three effective potential wells. They are observable at q = 10.9, corresponding to the maximum of the classical potential. The GEP, in its analytical expression, shows the quantum corrections on the classical potential. New phenomena are expected as shown in **Figure 2**. In some situations, this potential is shown to exhibit chaotic behavior for some restricted energy ranges [41]. According to the graph, we expect multiple quantum tunnel effects, as well as energy (medium value) dependent transitions between regular and chaotic motion in the GEP; this needs to be confirmed in the numerical study. Quantum tunneling must be observable in our system, for energy range between the potential wells. Since the potential has symmetric wells, we expect *on resonance* [18] tunneling process to occur. One can also explain this using the static effective potential $\tilde{V}_{eff}(q)$ that is obtained by eliminating G in $V_{eff}(q,G)$ via. From **Eq.13**:

$$\tilde{V}_{eff}(q)=V(q)+\frac{\hbar}{8G(q)}+\frac{\hbar G(q)}{2}\left[A-\frac{B}{2}q^2+\frac{C}{24}q^4\right]+$$
$$\frac{\hbar^2G^2(q)}{8}\left[-B+\frac{C}{2}q^2\right]+\frac{\hbar^3G^3(q)}{48}C \tag{14}$$

where G(q) is the solution of the following equation:

$$G^4+\frac{4}{\hbar C}\left(-B+\frac{C}{2}q^2\right)G^3+\frac{8}{\hbar^2 C}\left(A-\frac{B}{2}q^2+\frac{C}{24}q^4\right)G^2-\frac{2}{\hbar^2 C}=0 \tag{15}$$

Figure 3, we realise that $\tilde{V}_{eff}(q)$ (Static GEP) changes from $V(q)$ and $V_{eff}(q,G)$ as if there were a phase transition.

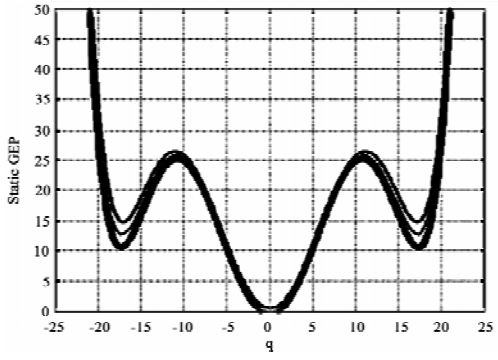


Figure 3. The static effective potential.

3.2. Fix Points and Instabilities

Our potential shows the classical contribution plus the higher order quantum corrections. Note that the number of valleys is in increase as compared to the one in [35]; so does the number of fixed points by two. This increase was earlier predicted by those authors because of the higher order term corrections considered ($\hbar^3 G^3$) in the **GEP**. In the fix points' theory, a hetero structure bears additive properties whose importance is over the aim of this paper. Nevertheless, we focus on the variety of its usage while considering it. Our system offers various possibilities and useful aspects therewith into its rich dynamics (see stationary conditions $\dot{p} = \dot{q} = \dot{G} = \dot{\Pi} = 0$). Concavities observed in the picture are good diagnostics for chaos [30-32]. Really, since there are various combinations of initial conditions corresponding to the same total energy, numerically unobservable chaos may exist at all energies. Nevertheless, it did not happen so; that is why it was difficult to find energy ranges where chaos occurred. As G increases in the **GEP,** three minima appear in the half plane, corresponding to the well minima of the original problem; two saddle points also emerge, interestingly. In addition, the (q, p) system, as driven by the (G,Π) system, is like a non-linearly driven Duffing oscillator [17,39] with back reaction. With these ingredients, added to the highly non-linearity, it is not surprising that our semi quantum equations exhibit chaos. On the other hand, the instabilities are important and their study is based on the motion equations **Eq.10** in the second order form to show the similarity to coupled oscillators:

$$\begin{cases} \ddot{q} + \omega_{eff}^{2}\, q = 0 \\ \ddot{G} + \Omega_{eff}^{2}\, G = 0 \end{cases} \tag{16}$$

with

$$\omega_{eff}^{2} = \left(A - \frac{\hbar G}{2}B + \frac{\hbar^2 G^2}{8}C\right) - \frac{1}{6}\left(B - \frac{\hbar G}{2}C\right)q^2 + \frac{C}{120}q^4 \tag{17}$$

and

$$\Omega_{eff}^{2} = \left(2A - Bq^2 + \frac{C}{12}q^4\right) + \hbar\left(-B + \frac{C}{2}q^2\right)G + \frac{\hbar^2 C}{4}G^2 \tag{18}$$

Two coupled parametric oscillators then describe the dynamic of the system. Both oscillators can become unstable due to exponentially growing modes for a finite range of initial values for q and G. Taking the both effective mass squared terms to be negative, gives the conditions

$$\frac{1}{3}q^2 + \frac{2B}{C} - \frac{1}{2}\left(\frac{8}{45}q^4 + \frac{16}{C^2}B^2 - \frac{32}{C}A\right)^{1/2} < G < \frac{1}{3}q^2 + \frac{2B}{C} + \frac{1}{2}\left(\frac{8}{45}q^4 + \frac{16}{C^2}B^2 - \frac{32}{C}A\right)^{1/2} \tag{19}$$

with the range

$$\left[90\left(\frac{2A}{C} - \frac{B^2}{C^2}\right)\right]^{1/4} \le q \le -\left[90\left(\frac{2A}{C} - \frac{B^2}{C^2}\right)\right]^{1/4} \tag{20}$$

for **Eq.17**; and

$$-q^2 + \frac{2B}{C} - \frac{1}{2}\left(\frac{8}{3}q^4 + \frac{16B^2}{C^2} - \frac{32A}{C}\right)^{1/2} < G < -q^2 + \frac{2B}{C} + \frac{1}{2}\left(\frac{8}{3}q^4 + \frac{16B^2}{C^2} - \frac{32A}{C}\right)^{1/2} \tag{21}$$

with the range

$$\left[6\left(\frac{2A}{C} - \frac{B^2}{C^2}\right)\right]^{1/4} \le q \le -\left[6\left(\frac{2A}{C} - \frac{B^2}{C^2}\right)\right]^{1/4} \tag{22}$$

for **Eq.18**.

The both ranges of q indicate the relationship among the parameters A, B and C to be satisfied for a given potential in order to present oscillations with possible instabilities:

$$\frac{2A}{C} - \frac{B^2}{C^2} \ge 0 \quad \text{that is} \quad A \ge B^2/2C \tag{23}$$

For G to remain positive, the previous conditions give

$$\begin{cases} 0 < G < -\frac{1}{3}q^2 + \frac{2B}{C} + \frac{1}{2}\left(\frac{8}{45}q^4 + \frac{16}{C^2}B^2 - \frac{32A}{C}\right)^{1/2} \\ 0 < G < -q^2 + \frac{2B}{C} + \frac{1}{2}\left(\frac{8}{45}q^4 + \frac{16B^2}{C^2} - \frac{32A}{C}\right)^{1/2} \end{cases} \tag{24}$$

Hence, the instability zone added to the q range obtained from **Eq.17** gives

$$0 < G < -\frac{1}{3}q^2 + \frac{2B}{C} + \frac{1}{2}\left(\frac{8}{45}q^4 + \frac{16B^2}{C^2} - \frac{32A}{C}\right)^{1/2} \tag{25}$$

Note that our parameters satisfy the condition **Eq.23** and the q range for instability is $q \ge 10.9$ or $q \le -10.9$, corresponding to domains containing the two extreme classical potential wells. We can conclude that the equilibrium positions q = 17.2 and $q = -17.2$ of our potential are unstable ones. Nevertheless, note that condition **Eq.25** provides the criterion not only for a better choice of initial G, but also for sensitivity to initial conditions.

4. NUMERICAL STUDIES

In this section, we present the results obtained by direct numerical integrations of our semi quantum equations of motion. Numerical integration is necessary for us in order to confirm the estimates of the theoretical predictions and/or to obtain other results in domains where the **GEP**, and the analytical study cannot be successful. In fact, it is the principal mean that allows knowing about the exact behavior of the solutions for non-trivial nonlinear ordinary differential (semi quantum) equations **Eq.16**.

Unfortunately, unlike an analytical relation from which one can discuss the appearance of chaos for different initial conditions, the numerical integration has the draw back of requiring a discrete variation of the control parameter of the system. Consequently, numerically studying such a semi quantum system for several values of its control parameter, which may vary within intervals of relatively long length, will demand a cumbersome quantity of plots. Thus, despite the fact that we focus great

attention only on the influence of the control parameter, the numerical description involves quite a large number of plots, where lie the key results of this paper. The fourth-order Runge - Kutta algorithm is the scheme we have used. The time step is fixed at Δt = 1e − 3. We always fixed $p(t=0)=\Pi(t=0)=0$ initially and chose G, together with various values for q to represent initial configurations corresponding to varying total energy. Generally, we did not explore the full set of all initial conditions at a given energy. This would require much more extensive numerical calculations which is beyond the scope of the present paper.

We have used five standard indicators including Bifurcation diagrams, phase portraits, frequency spectrum, Poincare sections displays and computation of maximal Lyapunov exponent to characterize the long time dynamics of our model under slight perturbation of initial conditions. These indicators complement each other in the following way: The bifurcation diagram indicates a range where values can be found to obtain regular or chaotic behavior; phase portraits are basically used to appreciate the shape of trajectories in the phase space on which the system evolves in time. They may be sufficient to state whether the dynamic is regular or not.

Nevertheless, they are not practical when the phase space is of dimension greater than two. Moreover, we cannot easily distinguish roughly between chaotic states and some quasi-periodic ones using only phase portraits. Poincare surfaces of section are useful to determine in particular the periodicity of the systems evolution. Strange attractors correspond to surfaces of section made up of an infinite number of points that occupy a bounded domain of the cross section without forming a smooth closed curve. They may be chaotic or not. Thus, we need the maximal Lyapunov exponent to know the nature of the strange attractor, in this case. The frequency spectrum is also useful here to determine, in particular, the value of the frequency in the case of regular periodic motion. Regular spacing shows regular motion.

For a system of first order equations of the form $\dot{X} = F(X,t)$, where $X = (X_1, X_2, ...X_n)^T$, the Lyapunov exponents are defined as the asymptotic values of the eigen values of the solution of the matrix differential equation $\dot{B} = D_F(X(t)) \bullet B$, assuming the initial conditions $B(0) = I_{n\times n}$, where $I_{n\times n}$ is the $n \times n$ square identity matrix. $D_F(X(t))$ represents the Jacobean matrix of the function F evaluated at the solution $X(t)$. The involved Eigenvalues can be computed using the code stated by Wolf *et al.* [37,40].

A chaotic state is the one for which at least one exponent is positive. The bifurcation diagram (see **Figure 4**) shows here the increase of chaotic behavior of the system when moving from classical ($\hbar$ =0) to quantum ($\hbar$ =1). We present here phase portraits and corresponding Poincare sections in the phase space planes of the two conjugate pairs of variable $\{q, p\}$ and $\{G, \Pi\}$ for some values of the total energy closer to the minimum of the classical potential energy. **Figure 5** shows some regularity in the motion of the centroid at low energies.

The first row presents the phase space at E = 4, G = 0.4 together with the corresponding frequency spectrum, which is regularly spaced; indicating regularity. Indeed, the value of the Lyapunov exponent is negative -0.0004. The remaining rows indicate the aspects of the Poincare sections in the both phase spaces for energies 0.9; 4 and 5, respectively. The motion still looks regular and pre-

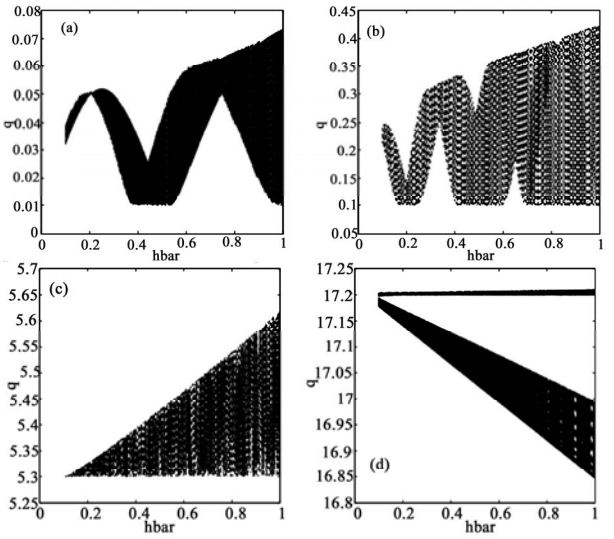


Figure 4. Bifurcation diagrams as ħ moves from 0 to 1, for (a): q = 0.01, G = 0.1, (b): q = 0.01, G = 0.1, (c): q = 5.3, G = 0.09, (d): q = 17.2, G = 0.09.

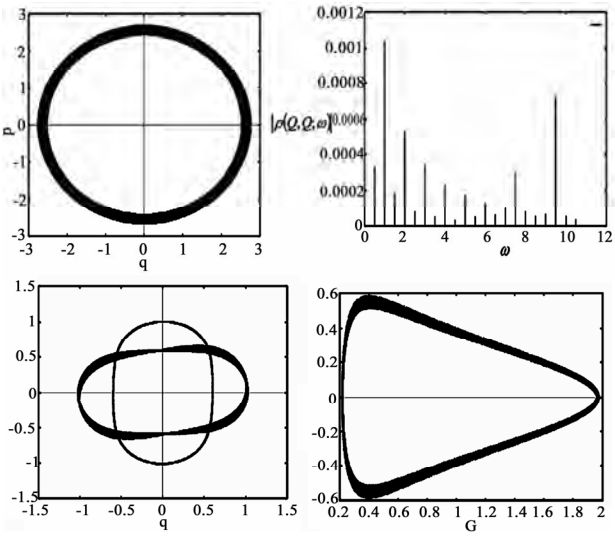


Figure 5. Phase portrait and frequency spectrum in the first line for E = 4, G = 0.4; Poincaré sections in the remaining lines and from top to bottom for E = 0.9, G = 1.2; E = 4, G=1.2 and E = 5, G = 4.

sents a double periodicity with negative values of Lyapunov exponent. This situation is somewhat evident and the analytical studies predicted it. Once the energy keeps increasing, around the secondary minimum $E = 12$, $G = 2.2$, the motion starts to be irregular as shown in **Figure 6** with multi-periodicity. The number of fixed points (five) is clearly observable at this particular state. Regular spacing starts vanishing in the frequency spectrum; Poincare sections present separate closed curves; the Lyapunov exponent remains negative with oscillations over the positive values. This is a sort of phase transition. The significance of this type of motion lies on the definition of KAM tori. At $E = 17$, $G = 10.8$, there is chaos (see **Figure 7**) since the frequency spectrum has irregular spacing and the Lyapunov exponent is positive (not shown).

An attractor seems to appear in its limit cycle around the secondary minimum of the potential energy. However, just at $E = 17.49$, $G = 0.39$, regular motion appears once more (see **Figure 8**) with negative Lyapunov exponent and a four−periodic motion in Poincare sections. Around $E = 18$ and $E = 19$, strange attractor seems to appear, once more, as plotted in **Figure 9**. We find it strange and chaotic since the Lyapunov exponent shifts from positive value (+ 0.224 at $E = 19$) to negative value (−0.0005 at $E = 19.5$). Regular motion is then observed at E=19.5. The particle located in the right potential well evolves regularly, and, chaotically sprays out from the right to the left well and remains there in highly chaotic

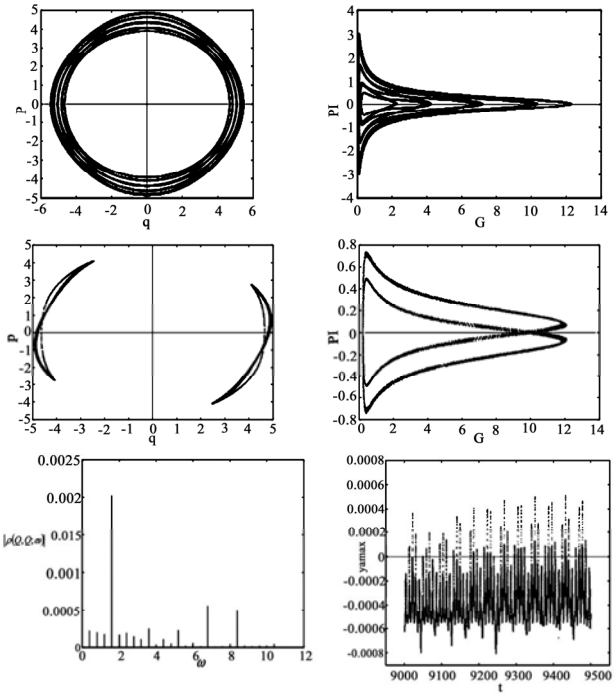


Figure 6. Phase portrait and Poincaré sections in the two first lines and at last, frequency spectrum with Lyapunov exponent for E = 12, G = 2.2.

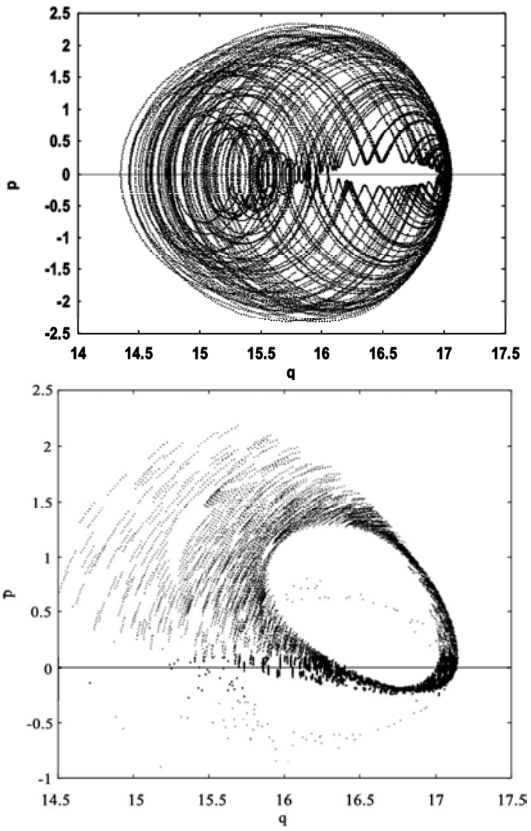


Figure 7. Phase portrait and Poincaré sections for E = 17, G = 10.8.

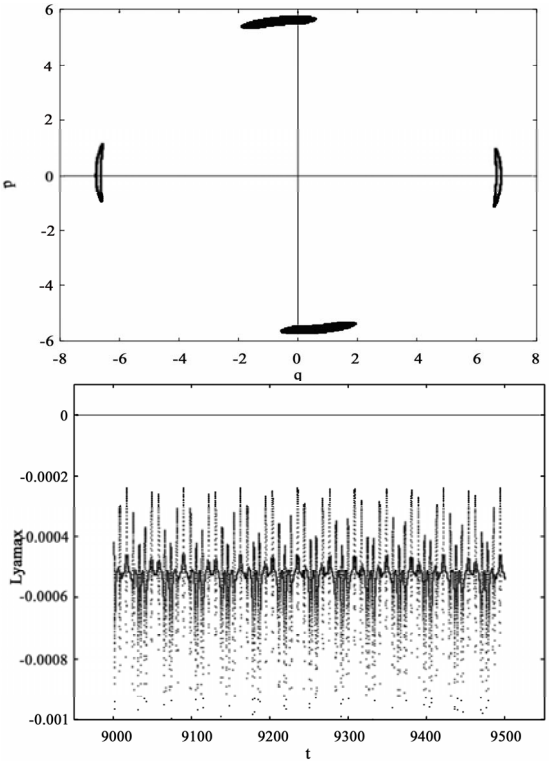


Figure 8. Poincaré sections and Lyapunov exponent for E = 17.49, G = 0.39.

evolution as indicated in **Figure 10**, with Lyapunov exponent equal to + 0.2295. Multiple quantum tunneling arises from $E = 20$ to $E = 23$. Between $E = 25$ and $E = 37$, chaos dominates with complicated trajectories and positive Lyapunov exponents + 0.264 and + 0.297 (not shown). At very high energy, chaos persists (see **Figure 11**) in its different forms of limit cycle, KAM tori or attractors. It is surprising here because, analytically, regular behavior at high energy was awaited. The energy dependent regular and chaotic behaviors hence alternate and make the dynamics very rich.

5. CONCLUSIONS

In this paper, we have analyzed the dynamics of a semi quantum system that has the interesting feature of possessing three potential wells. We focused our attention, in particular, on the non-linearity of the basic set of the

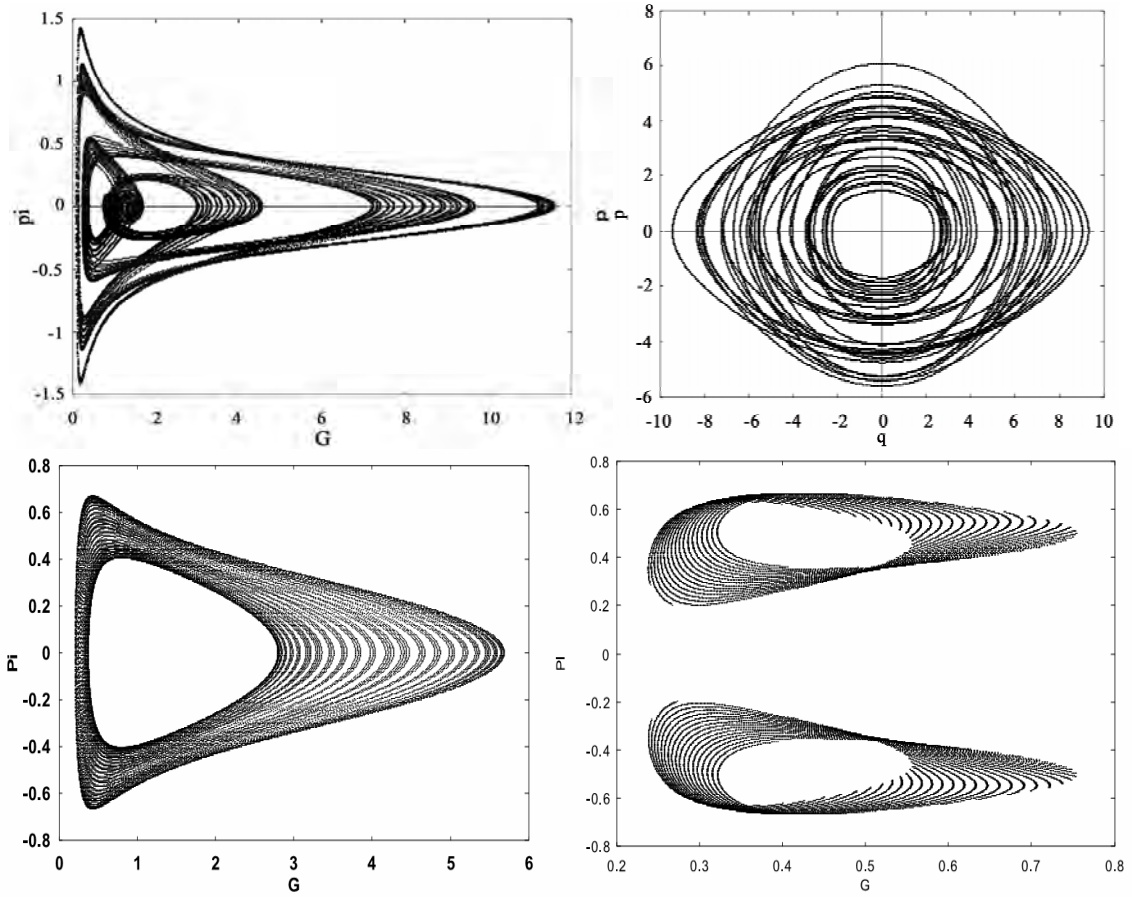


Figure 9. Phase portrait for E = 18, G = 0.09; E = 19, G = 4 and E = 19.5, G = 0.2, respectively and at last, Poincaré section for E = 19.5, G = 0.2.

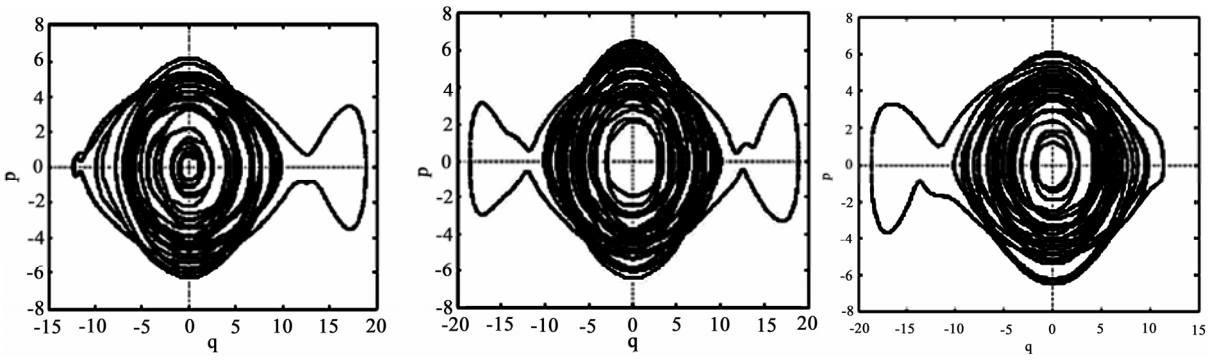


Figure 10. Phase portrait for E = 20, G = 6.3; E = 22, G = 0.2 and E = 23, G = 0.39, respectively.

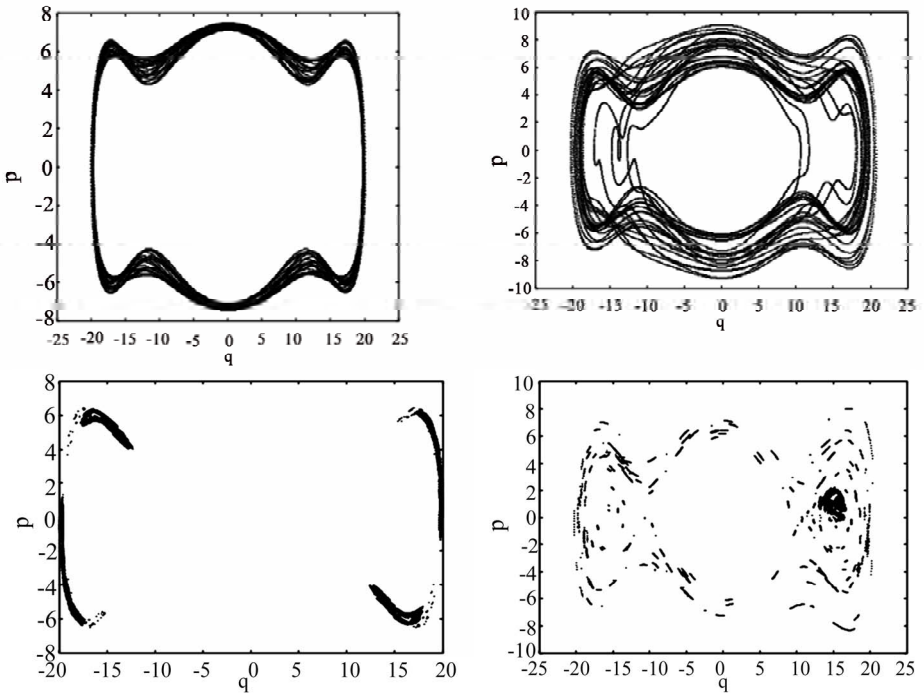


Figure 11. Phase portrait and Poincaré section for E = 38.5, G = 1.2 and E = 50.5, G = 0.2, respectively.

semi quantum equations of motion derived by a Time Dependent Variational Approximation using a general Gaussian as the trial wave function. Proceeding first analytically, we especially used the GEP in both its static and dynamical forms to catch the various instabilities areas and fixed points, tunneling roll over, regions where the system could behave unexpectedly and the non-regular behavior that we characterized as chaotic for the sake of the nonlinear aspect of the equations of motion. Under the control parameter, we turned to numerical investigations to verify and complement these analytical assumptions. With the use of indicators such as phase portraits, Lyapunov Exponents and Poincare sections, the results of numerical analysis on which we definitely rely can be summarized as follows. At energies closer to the minimum, periodic evolution starts. As energy increases, multiperiodic behavior arises leading to chaotic one through multiple tunneling and roll over as earlier predicted by the GEP. Regular and irregular motions hence alternate. Finally and remarkably, at very high energy, the system supposed to behave regularly because the whole structure would be regarded as a unique potential well does not. It rather exhibits chaotic motion. There are not rooms of regular motion at that high energy. This may be related to its complex structure.

The bifurcation diagram earlier predicted that as we move from classical to quantum, chaos increases relevance. The dynamics is found to be very rich because exhibiting various interesting behaviors, which may help to better understand complex phenomena occurring at sub atomic and mesoscopic levels.

However, its complexity makes us ask a question whether the energy is the only parameter to control the system; particularly, at very high energy and in the regions where there occur quantum tunnel effects and chaotic motion. It will be interesting to look for additive parameter on which the system may depend to behave so or the novel phenomena that this chaos may hide. Also, may these results be a bit precise from t he one obtained by the authors of [17] in Gaussian wave packets.

REFERENCES

[1] Cooper, F., Dawson, J.F., Meredith, D. and Shepard, H. (1994) *Physical Review Letters*, **72**, 1337; Lichtenberg, A.J. and Lieberman, M.A. (1983) Regular and stochastic motion. Springer, New York.

[2] Ringot, J., Lecocq, Y., Garreau, J.C. and Szriftgiser, P. (1999) Generation of phase-coherent laser beams for Raman spectroscopy and cooling by direct current modulation of a diode laser.*European Physical Journal D*, **7**, 285; Ringot, J., Szriftgiser, P., Garreau, J.C., and Delande, D. (2000) *Physical Review Letters*, **85,** 2741.

[3] Tabor, M. (1989) Chaos and integrability in non linear dynamics. John Wiley and Sons, New York.

[4] Mackey, R.S. and Meiss, J.D. (1987) Strong variation of global-transport properties in chaotic ensembles. *Hamiltonian dynamical systems*, Adam Hilger, Bristel.

[5] Berry, M.V. (1989) Quantum chaology, not quantum chaos.*Physica Scripta*, **40**, 335.

[6] Ozorio De Almeida, A. (1988) Hamiltonian systems: Chaos and quantization. Cambridge University Press, Cambridge.

[7] Gutzwiller, M.C. (1990) Chaos in classical and quantum mechanics. Springer Verlag, New York.
[8] Reichl, L.E. (1992) The transition to chaos in conservative classical systems: Quantum manifestations. Springer Verlag, New York.
[9] Sengupta, S. and Chattaraj, P.K. (1996) *Physics Letters A*, **215**, 119.
[10] Kluger, Y., Eisenberg, J.M., Svetisky, B., Cooper, F. and Mottola, E. (1991) *Physical Review Letters*, **67**, 2427.
[11] Pradhan, T. and Khare, A.V. (1973) *American Journal of Physics*, **41,** 59.
[12] Jona Lasinio, G., Presilla, C. and Capasso, F. (1992). *Physical Review Letters*, **68**, 2269.
[13] Leo, K., Shah, J., Gobel, E.O., Damen, T.C., Schmitt S., Ring, Schafer, W. and Kholer, K. (1991) *Physical Review Letters*, **66**, 201.
[14] Elze, H.T. (1995) Quantum decoherence, entropy and thermalization in strong interactions at high energy. *Nuclear Physics B*, **436**, 213; *Nuclear Physics B*, **39**, 169.
[15] Jackiw, R. and Kerman, A. (1978) *Physics Letters A*, **71**, 158.
[16] Heller, E.J. (1975) Calculations and mathematical techniques in atomic and molecular physics. *Journal of Chemical Physics*, **62,** 1544; Heller, E.J. and Sundberg, R.L. (1985) Chaotic behavior in quantum systems. Plenum Press, New York, 255.
[17] Pattanayak, A.K. and Schieve, W.C. (1994) *Physical Review Letters*, **72**, 2855; (1992) *Physical Review Letters*, **46,** 1821, Proceedings from workshop in honor of Sundarshan., E.G.G. Gleeson, A.M., Ed., (world scientific, Singapore, in press).
[18] Cooper, F., Pi, S.Y. and Stancioff, P.N. (1986) *Physical Review D*, **34**, 3831.
[19] Kovner, A. and Roseinstein, B. (1983) *Physical Review D*, **39**, 2332.
[20] Littlejohn, R.G. (1988) *Physical Review Letters*, **61**, 2159.
[21] Tannoudji, C.C., Diu, B. and Laloe, F. (1988) Quantum mechanics, Ed., Masson.
[22] Perez, J.P. and Saint Crieq Chery, N. (1986) Relativity and Quantisation, University Paul Sabatier, Toulouse, Ed., Masson.
[23] Gaudaire, M. (1969) Propriete de la matiere: Onde et Matiere, Colorado Dunod University, Orsay.
[24] Ott, E. (1997) Chaos in dynamical systems. Cambridge University Press, Cambridge.
[25] Carlson, L. and Schieve, W.C. (1989) *Physical Review A*, **40**, 5896.
[26] Stevenson, P. (1984) *Physical Review D*, **30**, 1712; (1985) *Physical Review D*, **32**, 1389.
[27] Toda, M. (1974) Instability of trajectories of the lattice with cubic nonlinearity. *Physics Letters A*, **48**, 335.
[28] Brumer, P. and Duff, J.N. (1976) *Journal of Chemical Physics*, **65**, 3566.
[29] Capasso, F. and Datta, S. (1990) Bandgap and interface engineering for advanced electronic and photonic devices.*Physics Today*, **43(2)**, 74.
[30] Kramer, B. (1991) Quantum coherence in mesoscopic systems. Plenum, New York.
[31] Presilla, C., Jona – Lasinio, G.. and Capasso, F. (1991) Nonlinear feedback oscillations in resonant tunneling through double barriers.*Physical Review B*, **43**, 5200
[32] Tchoukuegno, R. and Woafo, P. (2002) *Physical D* **167**, 86; Tchoukuegno, R., Nana, B.R. and Woafo, P. (2002) *Physical Review A*, **304**, 362; (2003) *International Journal of Non-Linear Mechanics*, **38,** 531.
[33] Jing, Z.J., Yang, Z.Y. and Jiang, T. (2006) Bifurcation and chaos in discrete-time predator–prey system.*Chaos, Solitons and Fractals*, **27**, 722.
[34] Sun, Z.K., Xu, W. and Yang, X.L. (2006) *Chaos, Solitons and Fractals*, **27**, 778.
[35] Blum, T.C. and Elze, H.T. (1996) Semiquantum chaos in the double well.*Physical Review E*, **53**, 3123.
[36] Zaslavsky, G.M. (1985) Chaos in dynamic systems. Harwood Academic, Chur, Switzerland.
[37] Wolf, A., Swift, J. B., Swinney, H.L. and Vastano, J.A. (1985) Determining Lyapunov exponents from a time series.*Physical Review D*, **16**, 285.
[38] Melnikov, V.K. (1963) On the stability of a center for time-periodic perturbations. *Transactions of the Moscow Mathematical Society*, **12**, 1.
[39] Gucken, H. J. and Holmes, P. (1983) Nonlinear oscillations, dynamical systems and bifurcations of vector fields. Springer Verlag, Berlin.
[40] Yamgoue, S.B. and Kofane, T.C. (2002) The subharmonic Melnikov theory for damped and driven oscillators revisited .*International Journal of Bifurcation and Chaos*, **8**, 1915; (2003) *Chaos, Solitons & Fractals*, **17**, 155.
[41] Churchill, R.C., Pecelli, G. and Rod, L. (1975) J. Diff. Eqs. **17**, 329; (1977) J. Diff. Eqs. **24**, 329; Churchill, R.C. and Rod, D.L. (1976) Ibid. **21**, 39; (1976) Ibid. **21**, 66; (1980) Ibid. **37**, 23.

Vol.2, No.3, 155-159 (2010)
doi:10.4236/ns.2010.23025

Natural Science

Seismo-Microplasticity phenomenon in the rocks*

Eduard Innokentevich Mashinskii

Institute of Petroleum Geology and Geophysics, Siberian Branch of the RAS, Novosibirsk, Russia; MashinskiiEI@ipgg.nsc.ru

Received 26 November 2009; revised 8 January 2010; accepted 30 January 2010.

ABSTRACT

The seismic records of borehole-to-borehole measurements on frequency of 200 Hz in the microstrain range have been analysed. Microplasticity manifestations caused by seismic wave are detected on seismic records. It is the ladder-like stepwise change in amplitude course in some parts of the seismic trace. The step duration (time plateau) presents the amplitude-dependent time delay that shifts the arrival time and protracts pulse front. The microplastic process occurs owing to the anomalous realignment of the internal stresses on the microstructural defects in "elastic" domain. Result is the useful contribution for improvement of the theory of wave attenuation in the rocks. It can also be used in solving the applied problems in material science, seismic prospecting, diagnostics etc.

Keywords: Anelasticity; Deformation And Time Delay; Anelastic Seismic Parameters; Amplitude Dependence Of Wave Velocity And Attenuation

1. INTRODUCTION

Perfection of the wave attenuation mechanisms is up to now one of main tasks in Earth's sciences. For understanding attenuation mechanisms, the new knowledge is necessary about the rock anelasticity. Viscoelastic model of standard linear solid well describes the dispersion and relaxation but insufficiently correctly explains, for example, the amplitude-dependent effects. So, there are some contradictions relative to character of amplitude dependences of wave velocity and attenuation. There are data about decrease of wave velocity and increase in attenuation with increasing amplitude [1-4]. However, there are also facts when increase in the strain amplitude leads on the contrary to the increase in wave velocity and decrease in attenuation [5-9].

Detection of the quasi-static rock microplasticity has strengthened supposition about some unknown factor in charge of the amplitude dependence of seismic parameters [10-12]. It is necessary to mark that in Earth's sciences the microplastic anelasticity is not yet the generally accepted fact as against solid-state physics. In contrast to the viscoelasticity, microplasticity appears only when stress reaches critical value. The microplastic strain can increase and decrease, appear and vanish during increasing stress. The amplitude-dependent effects in rocks and quartz crystals were interpreted as indirect attributes of microplasticity that is possible even on the small strain amplitudes [13,14]. Assumption about microplasticity processes during seismic wave propagation was also made in works [15-16].

The direct testimony of seismic microplasticity was received in borehole-to-borehole measurements during the amplitude effect study [17]. This paper describes microplasticity manifestations detected on the seismic traces in result of the detailed analysis of the field materials.

2. EXPERIMENTS AND DATA ANALYSIS

The propagation of seismic pulse with diverse amplitudes in the area between two boreholes was studied. The experiments were performed in the Bystrovka research area. The measuring instruments were mounted in two boreholes 110 mm in diameter and 12 m depth spaced 7 m. A source was located in one of the boreholes and a receiver, in the other. The source in borehole 1 and receiver in borehole 2 were successively installed at depth of 2, 6, and 10 m. Detailed description is in the work [17]. The upper part of the section is comparatively homogeneous and is composed of loams as far as a depth of a few tens of meters. Rock is partially water-saturated as far as 8.5 m with compressional wave velocity Vp = 240 – 300 m/sec. After the depth – 9 m the wave travels in completely water-saturated loams with Vp = 1500 m/sec.

The measurements were made in accordance with the following procedure. The source and the receiver were successively located at depths of 2, 6, and 10 m in the diverse combinations. Such source-receiver configuration enables to study the pulse propagation in the different direction. Basic seismic records were made on the location of source–receiver in the lateral (horizontal)

*This work was performed with the support of the Russian Fund of Fundamental Researches, grant N 05-09-00405.

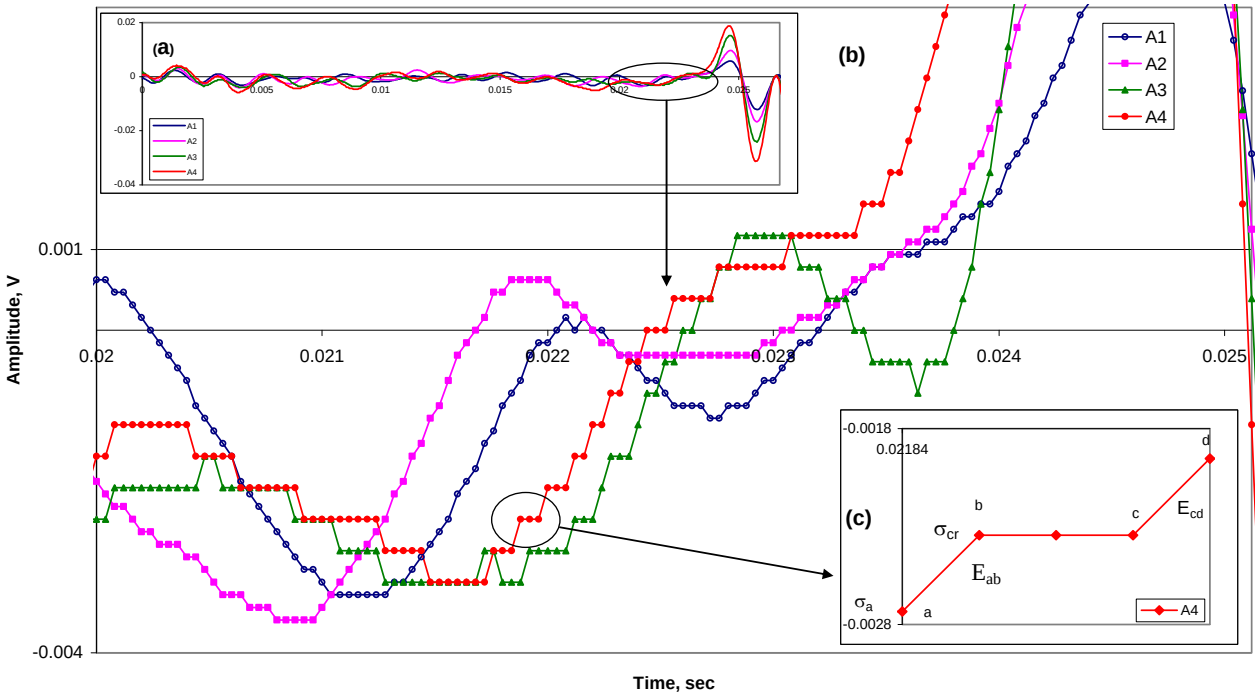


Figure 1. Seismic traces recorded in the 6 m–6 m "source-receiver" location on four amplitude values.

direction: 2–2, 6–6, and 10–10 m. On the short directions, the seismic pulse propagated in the partially saturated rock (6–6 m) and completely saturated rock (10–10 m).

The radiator of seismic signals consists of a set of piezoelectric disks. The signal radiates through a liquid spacer and hermetic elastic jacket contacting the borehole wall. Predominant frequency of P- pulse is about 200 Hz. The pressure receiver has the sensor of the piezoelectric type (PDS-21) and, therefore, records compression extension waves. The receiver contacts the borehole wall via an elastic spacer with a liquid. There is a preamplifier with the amplification coefficients K = 100. The signals were recorded in the digital form high-resolution during of time (Bordo-B-421 system), and were processed on a computer. The digitization time is 8 microseconds and 40 microseconds, the amplitude range is approximately (4 – 50) × 10^{-8}. The discrete amplitude change was fulfiled in a closed cycle, from the minimum to the maximum value and back (A_{min}… → A_{max}… → A_{min}, 4 values upward and 3 values downward).

Some parts of seismic trace have the form of ladder with the horizontal steps or plateau. The steps are evidence of interruption in the stress course and the presence of the time delay and deformation delay. The presumable reason of such effect is microplasticity caused by seismic wave. The signs of microplasticity were detected on many seismic records. The typical fragments from seismic traces with seismic microplasticity manifestations (SMM) are presented on **Figures 1** and **2**. SMM take place for signals with low and high intensity, i.e. small and great deformation rate.

A SMM example with small strain rate is shown in **Figure 1**. Here are presented four seismic traces recorded on four amplitude values in the 6–6 m "source-receiver" location. Change in amplitude value during defined time interval determines the amplitude steepness $R_{A/t}$. The amplitude value A_i is determined by number **n** of the amplitude quantization steps (A_{qu}): A_i = **n** × A_{qu} during one time quantization step, t_{qu}. The amplitude steepness is calculated as $R_{A/t}$ = **n** × A_{qu}/ t_{qu}, i.e. it is determined by way of number **n**. Thus, steepness is **n** = 2 for amplitudes A_1, A_2 and **n** = 5 for amplitudes A_3, A_4. The step length determines the time delay duration. The time delay in this case is from one t_{qu} to more than ten of t_{qu}. On the recordswith the small amplitude steepness, t_{qu} is equal 40 microseconds.

A SMM example of amplitude change with high rate (**n** = 50 and **n** = 100) is presented in **Figure 2.** This record concerns the case when both source and receiver were located at the depth of 10 m. Here the traces are presented for the upward and downward amplitudes. In principle, there is close coincidence in the form of the repeated traces although there are some nuances (see inset *b*). The delay duration in this case is tens of microseconds (t_{qu} = 8 microseconds).

The time delay changes the arrival time (see inset *c* in **Figure 2**) and can also influence on the wave front duration. In the same time, duration of the time delay depends

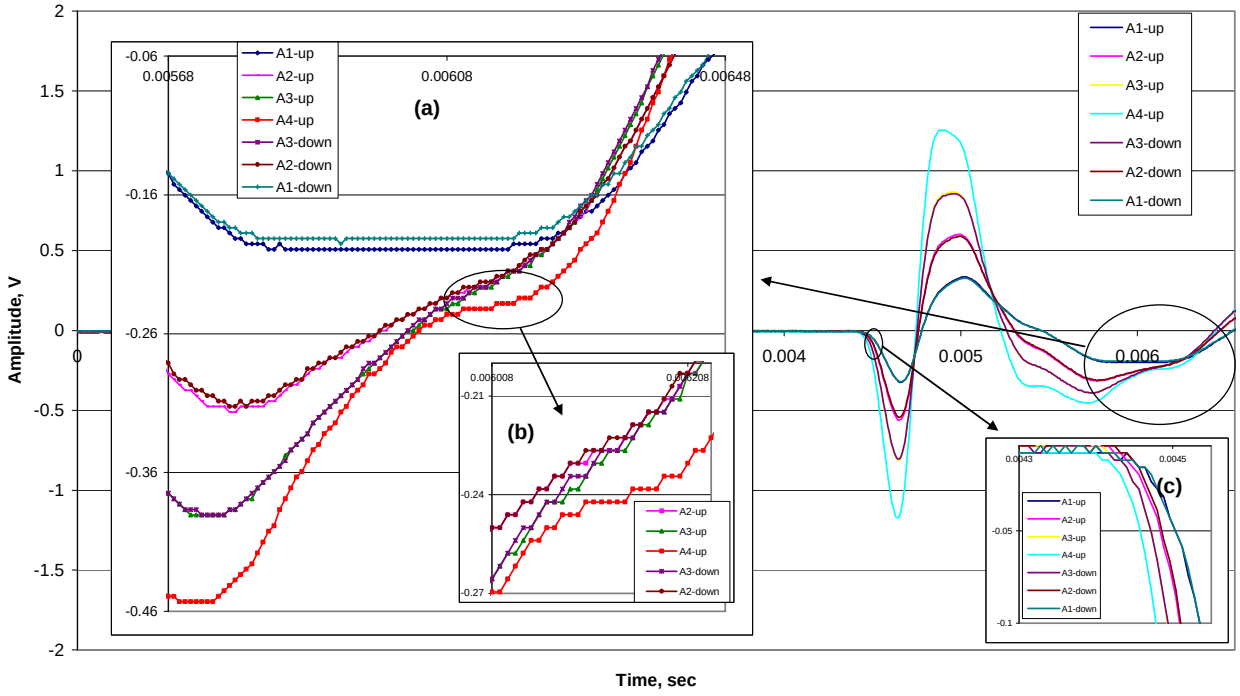


Figure 2. Seismic traces recorded on 4 upward and downward amplitudes in the location of "source– receiver" on 10 m.

on strain amplitude. Therefore, the amplitude variations can lead to the pulse parameters change. The experiments show that the increase in strain amplitude causes the displacement of an arrival time towards smaller time. The protracting of the wave fronts caused by time delay occurs in complex way. In the present time, could say that the intensity increase in wave shows the greater expressiveness of SMM. For the study of these effects, the conducting of the special experiments is needed.

3. DISCUSSIONS

Data about dynamic microplasticity in solid states and quasi-static microplasticity in the rocks confirm possibility of SMM conditioned by seismic wave. The strains level in static test and in seismic wave in the moderate amplitude range is approx selfsame ($\sim 10^{-6} - 10^{-5}$). The difference in the strain rate is not obstacle for the microplasticity process as a physical mechanism in both cases is, seemingly, the same. The point is that microplasticity is the frequency-independent (time-independent) process at least on seismic frequencies. Numerous data testify about prevalence of microplasticity effects in metallic materials, alloys, ceramics, thin-film materials and other solid states in the acoustical and low-frequency range [18,19]. Avowed feature of medium with microplasticity is the dependence of wave attenuation on strain amplitude. As regards rock microplasticity, here there are the obvious gaps in one's knowledge. Inasmuch as the direct manifestations of rock microplasticity have been established only during quasi-static stress in lab conditions (measurements on the samples) [10], any study of rock microplasticity under dynamic force is the considerable advancement in these investigations. Especially importantly, when it concerns the dynamic research in natural conditions.

In itself the physical experiment *in situ* for the purpose of microplasticity detection in the rocks triggered by seismic wave is original. In this experiment, natural medium in which propagation of seismic pulse occurs is test subject. It is unlimited medium in contrast to laboratory samples of small size. This medium consists of the dry and water-saturated rocks. It determines the experiment specificity. Microplasticity detection became possible also thanks to the using of the high-resolution seismic record in the time domain (microseconds) in which there is no need for the usual record of seismic signals. Besides, new information and results originality have been obtained thanks to the sounding by the different-intensity pulses upward and downward. Microplasticity amount depends on the energy level applied. One can surmise that with increasing wave intensity occurs the switching on the sources of microplasticity with the multi-level hierarchy.

SMM can be explained in the following way. Inasmuch as the piezoelectric receiver registers the change in dynamic stress (strain) with time, the flat steps on the trace mean the brief interruption in the course of

stress. Such stress stop supposes similar stop on the stress-strain curve that can occur owing to the microplasticity process. Total strain $\varepsilon = \varepsilon_e + \varepsilon_\mu + \varepsilon_{v\text{-}e}$ in the small-strain range consists mainly of elastic and microplastic component ($\varepsilon_e + \varepsilon_\mu$), as viscoelastic component $\varepsilon_{v\text{-}e}$ is comparatively small [10,11]. During loading (or unloading), both components have own contribution to the total strain. The stop in the course of stress occurs when the stress reaches some critical value σ_{cr} (for example, in the point **b** see inset (**b**) in **Figure 1**). At that moment, the redistribution of contribution between components occurs thus that contribution of the microplastic component become predominant. In the extreme case, the strain increment occurs only thanks to microplasticity as the increment due to the elastic deformation does not occur at all [10]. Therefore, the yield process is also possible. The regrouping in components occurs thus that effective modulus E_{ab} and accordingly the stress and strain remain invariable during interval **b** – **c**. It is possible owing to the distinctive feature of rock microplasticity that can both increase and decrease or even vanish during stress. In the point **c** after delay, the stress recommences own course with the same modulus. There is the hierarchical set of the diverse critical stresses that switch on the microplasticity sources when amplitude increases (decreases). Thus, seismic wave switches on the process of the anomalous redistribution of stresses and strains and respectively adequate to it process of the structural realignment in the rock. It takes a definite time that leads to deformation delay. In order to exclude the instrumental factor in appraisal of SMM, the amplitude error of analog-digital converter is checked. Testing of this device shows that the own error in plateau duration does not exceed 2 microseconds in the broad range of the amplitude steepness.

Similar deformation stops caused by microplasticity were detected in C_{60} single crystals [20]. It is shown that magnetic field manipulations lead to a change in the strain rate, the decrease in the rate being accompanied by a brief interruption of deformation. The deformation delay (incubation period) was observed also in the high-temperature superconductors, possessing by microplasticity [21]. Delay appearance is bound with presence of opposite internal stresses for decrease of which time is required. The wave attenuation mechanism in the rock with microplasticity (referring to the experiments [10]) is theoretically substantiated also in work [22]. As regards the mechanism of wave attenuation, the combination of known mechanisms is possible, for instance, as the hybrid relaxation-hysteresis mechanism [12]. It must not be ruled out also the acoustoplastic effect in the rocks as in metals and alloys [23]. The instance of stress jumps marked on seismic record evokes the great caution in interpretation of this effect, and therefore for the time being, we do not examine this question.

4. CONCLUSIONS

The result of this work is the new knowledge about nature of propagation of the mechanical oscillations in the Earth. The unnoticed heretofore presence in seismic record of microplasticity manifestations was detected thanks to the high-resolution signals measurements. The quasi-static microplasticity of the rocks, dynamic microplasticity of many solid states denoted the prospect of our search. The dynamic microplasticity in the rocks definitely is bound with critical amplitudes but its dependence on strain rate is not yet established. One may surmise data about rock microplasticity will enlarge the comprehension of some known effects that had not formerly of the satisfactory physical explanation. We see it now in the delay effect of the arrival time and the pulse shortening-widening.

As stated above, the nature of microplasticity in the rocks can coincide with known mechanisms in the solid states (for instance, dislocation microplasticity) or to be quite other (quasi-microplasticity). In the last case in spite of the difference in mechanism, the microplasticity manifestations can be the same as in usual solid states. This question requires subsequent clarification. Microplasticity affects the little-known anelastic processes of the small-amplitude wave propagation. The new knowledge about nonlinear-anelastic processes during wave propagation will help in discovery of new diagnostic indications permissive to increase the efficiency of seismic method for search of oil-gas deposits.

5. ACKNOWLEDGMENT

The author thanks G.V. Egorov for the help in the experimental work.

REFERENCES

[1] Mavko, G.M. (1979) Friction attenuation: An inherent amplitude dependence. *Journal of Geophysical Research*, **84(9)**, 4769-4775.

[2] Winkler, K.W., Nur, A. and Gladwin, M. (1998) Friction and seismic attenuation in rock. *Nature*, **274**, 528-531, 1979.

[3] Tutuncu, A.N., Podio, A.L. and M.M. (1994) Sharma An experimental investigation of factors influencing compressional- and shear-wave velocities and attenuations in tight gas sandstones. *Geophysics*, **59(1)**, 77-86.

[4] Ostrovsky, L.A. and Johnson, P.A. (2001) Dynamic nonlinear elasticity in geomaterials. La Rivista del Nuovo Cimento, **24(4)**, 7.

[5] Johnston, D.H. and Toksoz, M.N. (1980) Thermal cracking and amplitude dependent attenuation. *Journal of*

Geophysical Research, **85**, 937-942.

[6] Mashinskii, E.I., Koksharov, V.Z. and Nefedkin, Y.A. (1999) Amplitude-dependent effects in the range of small seismic strains. *Geologiya i Geofizika*, **40**, 611-618.

[7] Mashinskii, E.I. (2004) The variants of the strain-amplitude dependence of elastic wave velocities in the rocks under pressure. *Journal of Geophysics and Engineering*, **1**, 295-306.

[8] Mashinskii, E.I. (2007a) Effect of strain amplitude on the relaxation spectra of attenuation in the dry and saturated sandstone under pressure. *Journal of Geophysics and Engineering*, **4**, 194-203.

[9] Zaitsev, V.Y., Nazarov, V.E. and Talanov, V.I. (1999) Experimental study of the self-action of seismoacoustic waves. *Acoustic Physics*, **45(6)**, 720-726.

[10] Mashinsky, E.I. (1994) Quasi-micro-plasticity processes and nonlinear seismicity. *Physics of the Solid Earth*, **30**, 97-102.

[11] Mashinskii, E.I. (2005b) Non-linear stress-strain relation in sedimentary rocks and its effect on seismic wave velocity. *Geophysica*, **41(1,2)**, 3-17.

[12] Mashinskii, E.I. (2006) Nonlinear amplitude-frequency characteristics of attenuation in rock under pressure. *Journal of Geophysics and Engineering*, **3**, 291-306.

[13] Mashinskii, E.I. (2005a) Experimental study of the amplitude effect on wave velocity and attenuation in consolidated rocks under confining pressure. *Journal of Geophysics and Engineering*, **2**, 199-212.

[14] Mashinskii, E.I. (2008) Amplitude-frequency dependencies of wave attenuation in single-crystal quartz: Experimental study. *Journal of Geophysical Research*, **113**, B11304.

[15] Xu, H., Day, S.M. and Minster, J.-B.H. Model for nonlinear wave propagation derived from rock hysteresis measurements. *Journal of Geophysical Research*, **103(B12)**, 29,915-29,929.

[16] Braccini, S., *et al.* (2000) The maraging-steel blades of the Virgo super attenuator. *Measurement Science and Technology*, **11,** 467-476.

[17] Mashinskii, E.I. (2007) Amplitude-dependent effects of Longitudinal Seismic Wave Propagation in the Interhole Space. Izvestya, *Physics of the Solid Earth*, **43(8),** 683-690.

[18] Nishino, Y. and Asano, S. (1996) Amplitude-dependent internal friction and microplasticity in thin-film materials. *Journal de Physique IV*, **6**, 783-786.

[19] Golovin, I.S., Sinning, H.-R., Goken, J. and Riehemann, W. (2004) Fatigue-related damping in some cellular metallic materials. *Materials Science and Engineering*, **A370**, 537-541.

[20] Smirnov, B.I., Shpeizman, V.V., Peschanskaya, N.N. and Nikolaev, R.K. (2002) Effect of magnetic field on microplastic strain rate for C_{60} single crystals. *Physics of the Solid State*, **44(10)**, 2009-2012.

[21] Davies, P.W. and Wilshire, B. (1971) On internal stress measurement and the mechanism of high temperature creep. *Scripta Metallurgica*, **5(6)**, 475-478.

[22] Yarushina, V.M. and Podladchikov, Y.Y. (2008) Micriscale yielding as mechanism for low-frequency imtrinsic seismic wave attenuation. *70th EAGE Conference & Exhibition*, Rome, Italy, 9-12 June.

[23] Sapozhnikov, K.V., Vetrov, V.V., Pulnev, S.A. and Kustov, S.B. (1996) Acousto-pseudoelastic effect and internal friction during stress-induced martensitic transformations in Cu-Al-Ni single crystals. *Scripta Materialia*, **34(10)**, 1543.

Vol.2, No.3, 160-174 (2010)
doi:10.4236/ns.2010.23026

Natural Science

Surface rupture and hazard characteristics of the Wenchuan Ms 8.0 earthquake, Sichuan, China *

Rong-Jun Zhou[1], Yong Li[2], Liang Yan[2], Jian-Cheng Lei[1], Yi Zhang[2], Yu-Lin He[1], Long-Shen Chen[3], Xiao-Gang Li[1], Shi-Yuan Wang[1], You-Qing Ye[1], Yu-Fa Liu[1], Chuan-Chuan Kang [1], Tian-Yong Ge [1], Qiang He[1], Wei Huang[1]

[1]Institute of Engineering Seismology of Earthquake Administration of Sichuan Province, Chengdu, China
[2]State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Chengdu University of Technology, Chengdu, China; yanliang1003@163.com
[3]The University of Hong Kong, Department of Geology, Hong Kong, China

Received 23 December 2009; revised 14 January 2010; accepted 25 January 2010.

ABSTRACT

Longmen Shan is located the special joint between Tibetan Plateau inland in the west and Yangtze craton in the east. Consisting of a series of parallel imbricated thrust, it develops, from the west to the east, the Maoxian-Wenchuan, Beichuan-Yingxiu and Pengxian-Guanxian faults. The Wenchuan earthquake is a thrust with strike-slip type, and thre surface ruptures are located on the Beichuan-Yingxiu fault zone and Pengxian-Guanxian fault zone. The surface rupture on the Beichuan-Yingxiu fault shows the thrust and dextral slip characteristic. The maximum vertical displacement of the surface rupture is about 10.3 m and the maximum right-lateral displacement is about 5.85m. Though the vertical displacements and the horizontal displacements in the different segments have certain differences, as a whole, the ratio of the vertical displacement and the horizontal displacement is close to 1:1. The surface rupture on the Pengxian-Guanxian fault shows thrust and dextral characteristic. The rates of vertical displacements and the horizontal displacements ones on the most other segments are between 1:3 and 1:2. So the Beichuan- Yingxiu fault is a dextral-slip and thrust fault and the average vertical displacement is equal to the average horizontal displacement, while the Pengxian- Guan xian fault is thrust fault with a little dextral-slip component. The total intensity area above Ⅵ degree of the Wenchuan earthquake is about 333000 km^2. The high earthquake intensity line stretches to N40-50°E along Longmen Shan tectonic belt. The rate of the long axis and the minor is between 8:1 and 10:1. Three Ⅺ intensity regions are isolated in distribution. It presents a multipoint instantaneous characteristic of the rupture.

Keywords: Wenchuan earthquake; Longmen Shan; Surface rupture; Hazard; Sichuan; China

1. INTRODUCTION

According to the record of China earthquake networks, on 12th May, 2008, a huge earthquake Ms8.0 struck Yingxiu town in Wenchuan county, Sichuan province (31°00′N, 103°24′E). The earthquake took place at 14:28:04(Beijing time) and the depth of its hypocenter is about 14 kilometers. This earthquake with high magnitude and shallow hypocenter made an astonished attack on Longmen Shan area, which is located in the west of Sichuan basin, resulting in tremendous casualties and property losses. The earthquake occurred on Longmen Shan tectonic belt at the eastern margin of the Tibet Plateau. Since the late Cenozoic, the thrust-dextral movement of Longmen Shan tectonic belt led to Longmen Shan Mt.'s fast uplifting in opposition to Sichuan basin. Within the range of 50 kilometers along Longmen Shan tectonic belt, topographical fall exceeds 5 kilometers, which produces mountains and canyons region [1-5]. The special terrain and geology condition increased the damage of this earthquake.

As soon as the earthquake occurred, the writer went to Earthquake-stricken area to carry out an on-the-spot investigation for more than twenty days, accompanying the earthquake emergency job team of China Earthquake Administration. Some replenish investigations were carried out later. Mountain fall, the earth splits, changes of mountains and rivers caused by the earthquake made Longmen Shan full of distress everywhere, which was

* This research was supported by China National Natural Science Foundation grant 40841010, 40972083 and China National Science and Technology supporting Plan Foundation grant 2006BAC13B02-07, 2006BAC13B01-604.

elegant and pretty in the past. The enormous casualties and property losses that the earthquake brought about are furthermore grieved. Being greatly shocked and regretful, we illuminate the first step survey results of surface rupture and hazard characteristic of the earthquake and our thoughts about it in this paper, providing a reference for our fraternity. Especially, this paper was devoted to those who had been died or disappeared in this destructive earthquake.

2. THE SURFACE RUPTURE OF THE WENCHUAN EARTHQUAKE

The Longmen Shan tectonic belt stretches through the central of Sichuan province at N40-50°E orientation and forms the boundary of the Songpan-Garzê orogenic belt and the Yangtze Platform[1,6-8]. In late Triassic, the closure of Neo-Tethys Ocean resulted in the Longmen Shan tectonic thrusting from northwest to southeast. It controlled the formation and development of the Cenozonic foreland basin in the southwest of the Sichuan basin. There is a suit of mauve late Triassic - Cenozonic clastic rock of river and lake facies. Since late Cenozonic, with the closure of Neo-Tethys Ocean and fast-uplift of Qinghai-Tibet Plateau, neotectonic activities of Longmen Shan tectonic belt have aroused many researchers' interest. Avouac and Tapponnier [9] believed that Longmen Shan tectonic belt shortens sustainably at the NW-SE orientation, with a rate of 20mm/yr since late Cenozoic. But through field works, Burchfiel [1] found that there was lack of geological evidence for a large-scale shortening of Longmen Shan tectonic belt and the west of Sichuan basin. Data measured by the Global Positioning System(GPS) cross Longmen Shan tectonic belt shows that the shortening rate of it is only 4.0 ± 2.0mm/yr, in opposition to that of Sichuan basin [10]. However, England and Molnar [11] pointed out that convergence between India plate and Eurasia plate was reflected by a simple dextral-shear motion in the east margin of the Qinghai-Tibet Plateau. Furthermore, Li Yong [3] believed that uplift of the Longmen Shan was produced under the co-effect of the dextral motion and erode load-off of the Longmen Shan tectonic belt.

The Longmen Shan tectonic belt is composed of Maoxian-Wenchuan fault (the north-eastern segment is called Pingwu-Qingchuan fault), Beichuan-Yingxiu fault and Pengxian-Guanxian fault. There is a blind fault in the front of the Longmen Shan Mountain (**Figure 1**). All of these faults are late Pleistocence-Holocene active faults. There are divergent opinions on the late Quaternary activities of the Longmen Shan tectonic belt among different scholars. Zhou [12] believe that the Longmen Shan tectonic belt shows a thrust and dextral slip since late Quaternary. The slip rate of single fault is about 1 mm/yr, and the vertical component of displacement is equivalent to the horizontal. However, Densmore [5] believe that the average vertical slip rates of Beichuan-Yingxiu fault and Pengxian-Guanxian fault are less than 1 mm/yr. The horizontal slip rates of the faults are between 1mm/yr and 10mm/yr. And the faults show a dextral-slip characteristic. The Wenchuan Ms8.0 earthquake, on 12th May, 2008, produced coseismic displacements along Beichuan-Yingxiu fault and Pengxian-Guanxian fault coinstantaneously. After the earthquake, a lot of geological experts [13-23] carried out detailed investigations about the surface rupture, caused by the Wenchuan earthquake, in the Longmen Shan area. Therefore, it provides a good study and a validated example for the late Quaternary activity of the Longmen Shan tectonic belt.

2.1. The Surface Rupture on the Beichuan-Yingxiu Fault

Beichuan-Yingxiu fault is also called Beichuan fault [5] and is the major causative fault of the Wenchuan Ms8.0 earthquake on 12th May, 2008. The surface rupture on this fault started near by Yingxiu town in Wenchuan county and extended to the north-east. It is about 220 kilometers long, and crosses Hongkou town，Longmen Shan town, Donglin temple, the north of Hongbai town, Qingping town, Chaping town, Leigu town, Beichuan town, Chenjiaba village, Fenghuang village of Guixi town and Pingtong town, break down near Shikanzi village in the east of Nanba town, Pingwu county.

1) The surface rupture of the Wenchuan Earthquake in Yingxiu town

The epicenter of Wenchuan Ms8.0 earthquake is close to Yingxiu town. The surface rupture of the earthquake crosses the north of Yingxiu town in N60-70°E orientation. According to the different section features of the surface ruptures, it can be inferred that inclination of the fault is about 40-50°. In the north-west of Yingxiu town, the vertical displacement of the national highway (G213) is about 2 meters, and the dextral displacement is about 1.7 meters. A small waterfall was produced on a river bed (**Figure 2**). Zhou [12] described that the vertical displacement of Minjiang river's terrace IV caused by Beichuan-Yingxiu fault was about 40 meters. And the bending fault produced a normal fault graben on the hanging wall of the fault. This earthquake produced two surface rupture zones on the fault scarp (close to its top) with a horizontal distance of 6 meters. The vertical displacement is about 0.75 meters and the horizontal displacement is 1.1 meters at the low elevation place. At the place with higher elevation, the vertical displacement and the horizontal one are 2.4 meters, 1.9 meters, respectively. And the total vertical displacement is 3.15 meters; the total horizontal displacement is 3.0 meters (**Figure 3**). At the normal fault graben, it produced the new normal fault scarp and tensile fractures. It is consistent with the previous cognition. This phenomenon indicates that the fault scarp on terrace IV is the accumula

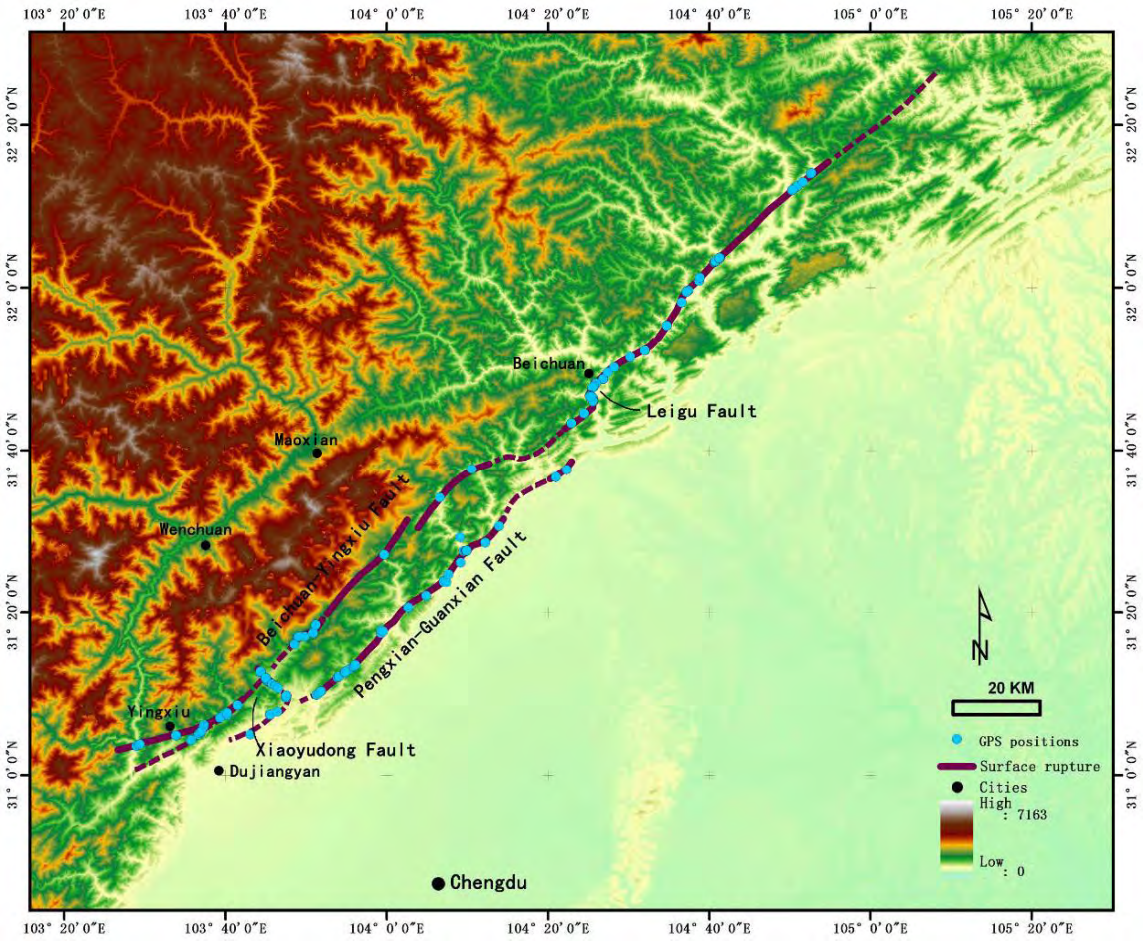


Figure 1. Distribution map for the surface rupture of the Wenchuan earthquake in Longmen Shan.

Figure 2. Sketch for the surface rupture of the Wenchuan earthquake in Yingxiu county.

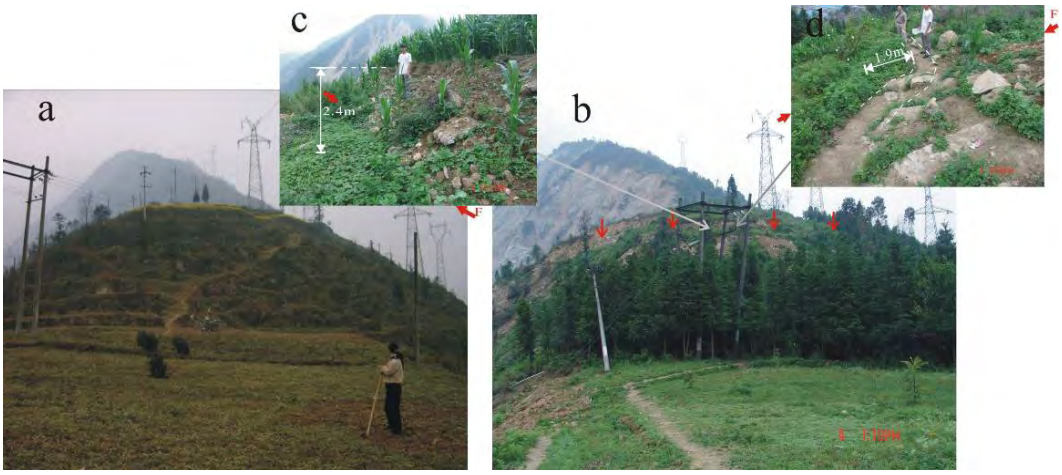


Figure 3. Fault scarp on the terrace IV of Minjiang river in the north of Yingxiu town. The photograph (a) was shot on May, 2002. The photograph (b) was shot on July, 2008. The photograph c shows the vertical component of the surface rupture. The photograph d shows the dextral horizontal displacement of the path.

Figure 4. The surface rupture of the Wenchuan earthquake in Hongkou region, Dujiangyan city. (a) the displacements of the road at Shenxigou village; (b) the fault scarps at Bajiao temple.

tive displacement produced by several movement events of the fault since late Quaternary and it has been undergone a-long-term denudated and back-off process.

2) The surface rupture of the Wenchuan Earthquake in Hongkou town

The surface rupture extends at N50-60°E orientation along the north-west bank of Baisha River near Hongkou town in Dujiangyan city. It is about 16 kilometers long. The maximum vertical displacement and the maximum horizontal one are also about 5 meters. At Shenxigou village, the surface rupture made a path displace with a vertical component of 2.75 meters and a horizontal component of 4.8 meters. Near Bajiaomiao paper mill with a distance of 8 kilometers to the spot, the road, the toft, tarrence I, the wash-land and the latter-day river bed are all displaced 3.5 meters. The fault developed in the Triassic coal stratum. There are two different phases scarps on the fault plane, and the side angle of the horizontal one is 43°. The scarp produced by this earthquake is vertical approximately (**Figure 4**).

3) The surface rupture of the Wenchuan Earthquake in Donglin temple.

Figure 5. Dislocation of road and split of pine tree in Donglin temple, Longmen Shan town, Pengzhou county.

At Donglin temple, 6 kilometers away to Longmen Shan town at NE orientation(Baishui river), the surface rupture zone extends at N45°E orientation, intersects with a pathway with a small angle. The vertical displacement of the pathway is 2.2 meters. A gully was displaced vertically and produced drop water. There is a row of pine trees along one side of the road. According to the local residents' description, the road and pine trees are straight before the earthquake occurred. The road and pine trees along the surface rupture zone produced right-lateral coseismic distortion. The horizontal displacement is 3.6 ± 1.0 m. The instantaneous offset of the fault resulted in the pine trees inclined obviously, which are on the fault scarp. One of the pine trees split and fell down because the effect of gravitation overlaid (**Figure 5**).

4) The surface rupture of the Wenchuan Earthquake in Leigu town, Beichuan county

Near Leigu town of Beichuan county, the Beichuan-Yingxiu fault divides into four branch faults arrange in left-step en echelon. The result in the II proluvial fan and streams vertical and dextral offset. The Gaitou hill in the left-step area uplifted and Qianjiang river course was changed [3,5,12]. The surface rupture of this earthquake at Pingshang village in the north of Leigu town stretches to N40°E, and made a list of gullies, pathways, proluvial fan, road, water pipes, terrain I and river bed produce vertical and right-lateral displacements (**Figure 6**). The surface rupture shows that the north-west wall of the fault uplifted and the south-east wall declined correspondingly. Specially, Zhou [12] found a fault scarp with the height of 22.5m and a gully with the right-lateral displacement of 1.6 m before on the II level pluvial fan [12] (**Figure 7**). At the same place, the surface rupture outcrops along the preexisted fault scarp. The maximum vertical displacement is 4.5 m. The displacements of gully and a pathway are 1.6 m and the right-lateral dislocations of them are 1.43 m (**Figure 8**). At Zhaojiagou village with a distance of 500 m to the spot, the vertical displacement of terrace I is about 6.5m and the right-lateral one is 5.85m. But the landslip maybe impacted on the precision of the displacements measurement here.

The Liulin village in the west of Leigu town is near the left-step area of the fault. There, the show of the surface is more complicated. From north to south, the trend of the surface rupture strikes from NE-trending to SN-trending and NWW-trending (**Figure 9**).The NS-trending surface rupture shows the vertical and left-lateral offset respectively. The vertical displacement and the left-lateral offsets of a water slot are 1.9 m，1.3 m respectively. The NWW-trending surface rupture shows the vertical and right-lateral offset. The foundation of a house is cut off by two cracks. The vertical displacements of two are 1.5 m，0.5 m respectively. And the right-lateral offsets of them are 1.3, 0.3 m respectively (**Figure 10**).

5) The surface rupture of the Wenchuan Earthquake in Maoba village, Beichuan county

Figure 6. Distribution map for the surface ruptures of the Wenchuan Earthquake in Pingshang village, Leigu town. (a) the unworked airscape; (b) the worked airscape; the black lines figure the distribution of the surface rupture, the black-sawtooth lines shows the hanging wall of the rupture; V: horizontal offset; H: vertical offset.

Maoba village is about 300 m away to the northeast of Beichuan county. The surface rupture of Beichuan-Yingxiu fault is N40-50°E trending, and its vertical displacements 4.7 m-10.3 m, which turns out to be a great scarp with a characteristic of antithetic fault (**Figure 11(a)**). The rupture incised the plat in the front of a building with a vertical offset of 10.3 ± 0.2 m, and a horizontal offset of 3.8 m (**Figure 11(b)**). About 200 m away to the northeast of this point, the rupture cut a hummock.

Figure 7. Displacements of the surface ruptures on the terrace I and road in Pingshang village, Leigu town.

As effects of the right lateral component, there is a triangular facet of fault in the edge of the hummock, and the right-lateral offset of a field is 5.3 m (**Figure 11(c)**).

6) The surface rupture of the Wenchuan Earthquake in Chenjiaba village

Chenjiaba village is about 15 kilometers away to the northeast of Beichuan county. Here, the surface rupture is about 200m long. The other segments of the surface rupture are covered by landslides. The rupture is N35°E trending and incised the river bed, floodplain, terrace I and the pathway. The vertical displacements of them are 2.0 m, 1.7 m respectively. The right-lateral offset of them are 1.6 m, 1.35 m respectively (**Figure 12(a)**).

7) The surface rupture of the Wenchuan Earthquake in Fenghuang village, Guixi town

Fenghuang village is about 2 kilometers away to the northwest of Guixi town. The vertical displacement of a

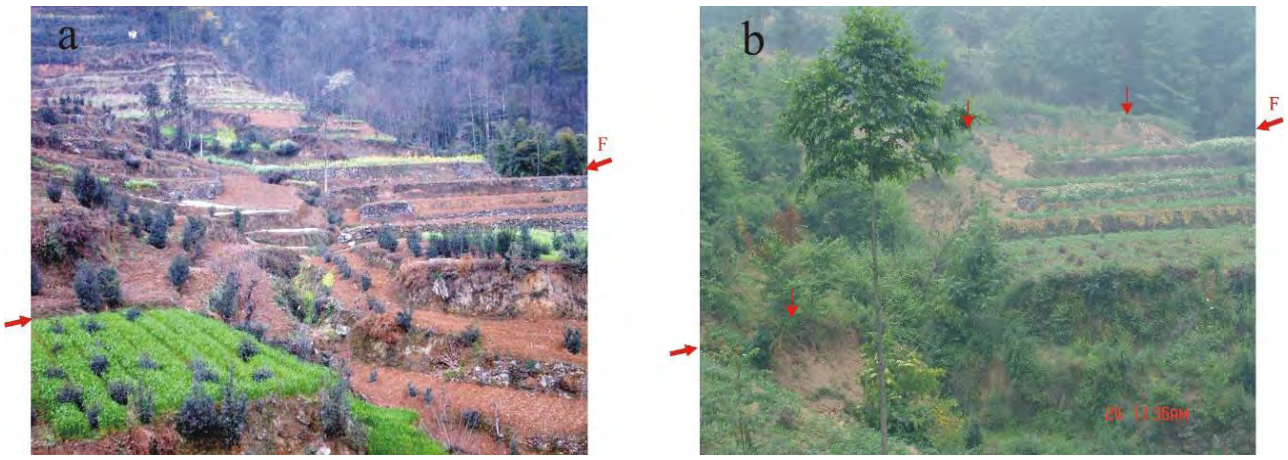


Figure 8. Displacements of the surface ruptures on the terrace II and gully in Pingshang village, Leigu town. (a) The photograph was shot on April, 2002; (b) The photograph was shot on July, 2008; the arrowhead shows the location of surface rupture, the northern-western side of the rupture obviously uplifted relative to the photograph a.

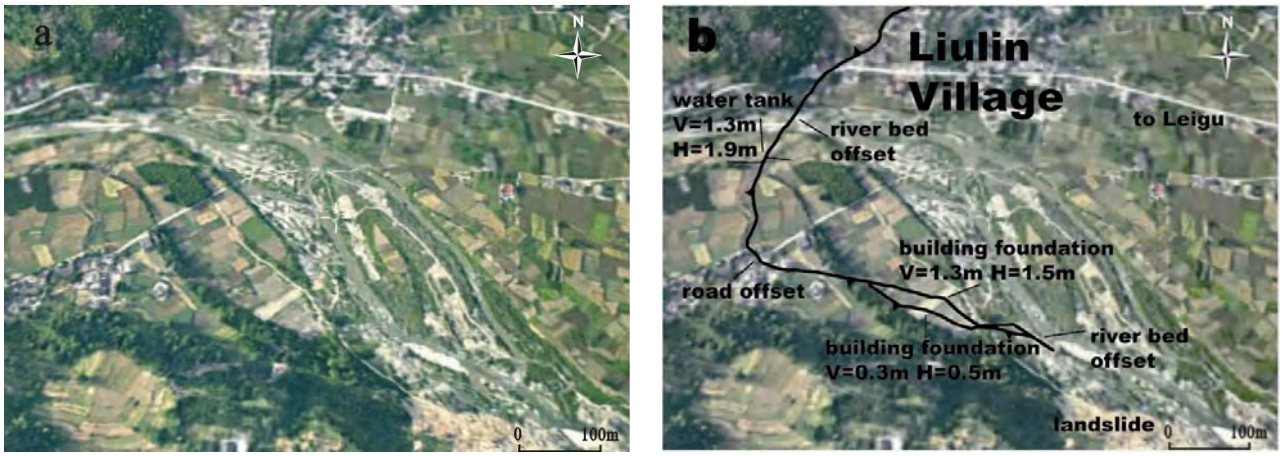


Figure 9. Distribution map for the surface rupture of the Wenchuan Earthquake in Liulin village, Leigu town. (a) the unworked airscape; (b) the worked airscape; the black lines figure the distribution of the surface rupture, the black-sawtooth lines shows the hanging wall of the rupture; V: vertical offset; H: horizontal offset.

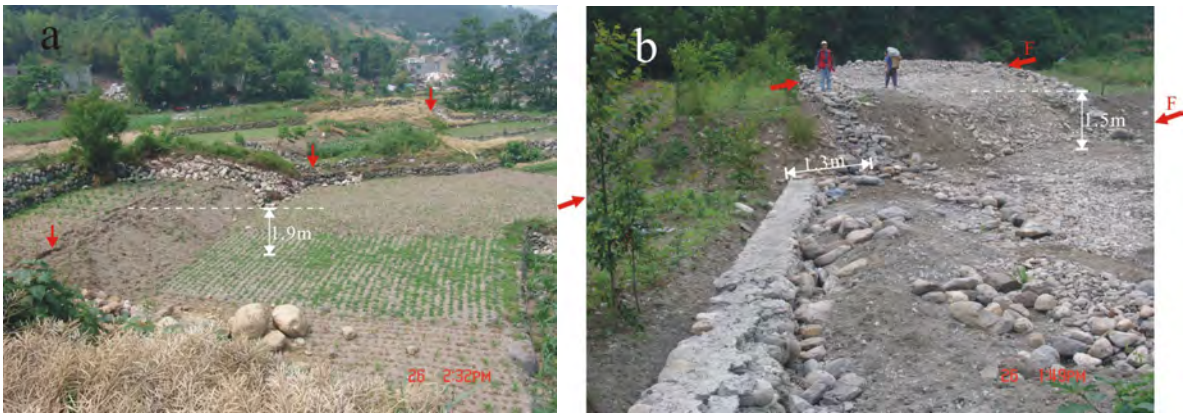


Figure 10. Displacements of the surface ruptures in the west of Leigu town. (a) the vertical displacements of the terrace I and the water slot, the arrowhead shows the locate of the water slot; (b) the offset on the foundation of the house.

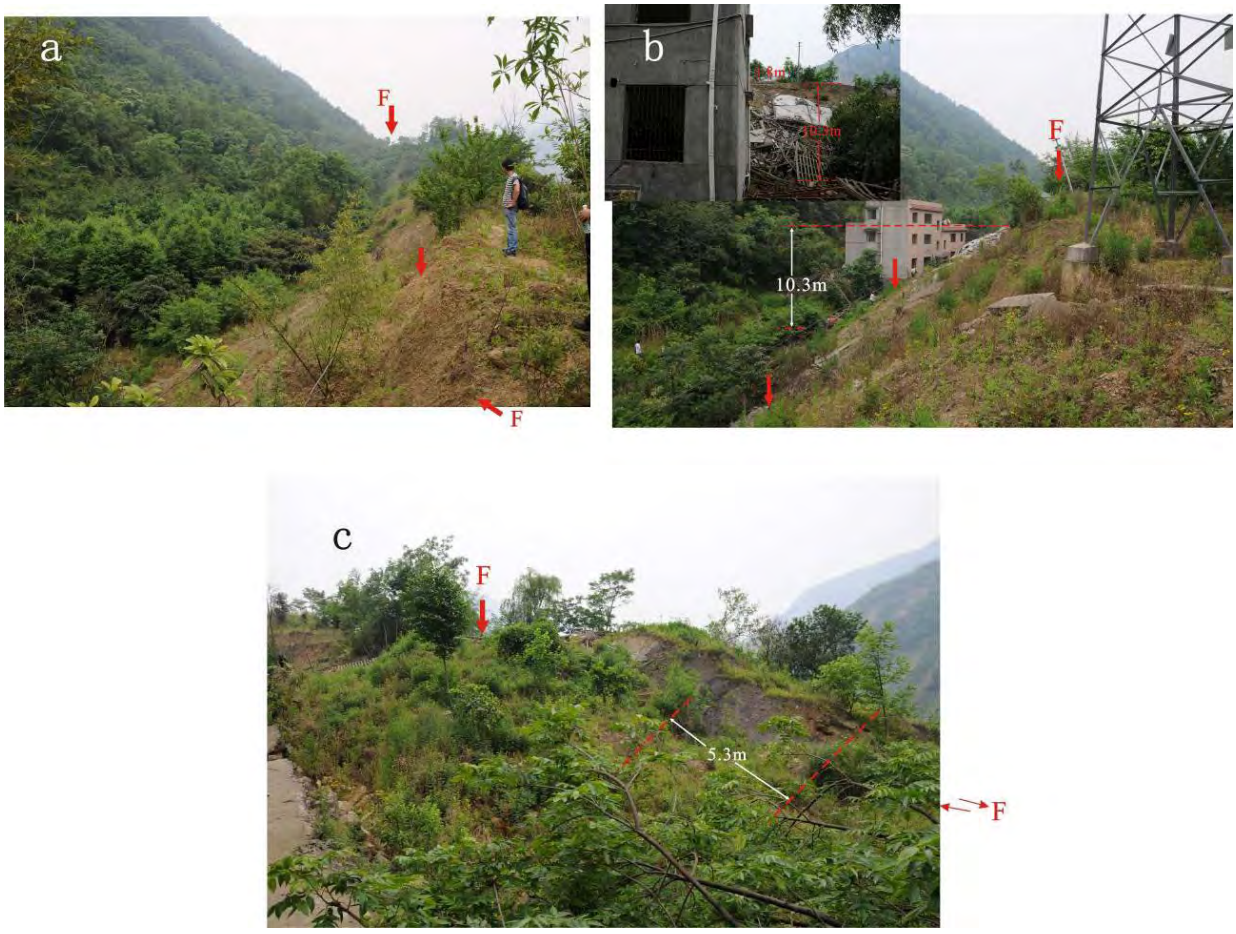


Figure 11. The surface rupture of the Wenchuan Earthquake in Maoba village, Beichuan county. (a) the vertical displacements in Maoba village; (b) the vertical and horizontal displacements of the plat in the front of a building; (c) the right-lateral offset of the field.

pathway in the village is about 2.5 m and the right-lateral offset of it is 2.4 m. The terrace of the river, floodplain and riverbed were incised by the rupture in the south-west of the spot. The fractured zone of bedrock outcrops near the bank of river. The vertical displacement of a pathway on the floodplain is about 2.1 m and the right-lateral of it is about 1.78 m (**Figure 12(b)**).

8) The surface rupture of the Wenchuan Earthquake in Pingtong town

The surface rupture near the Pingtong town is about

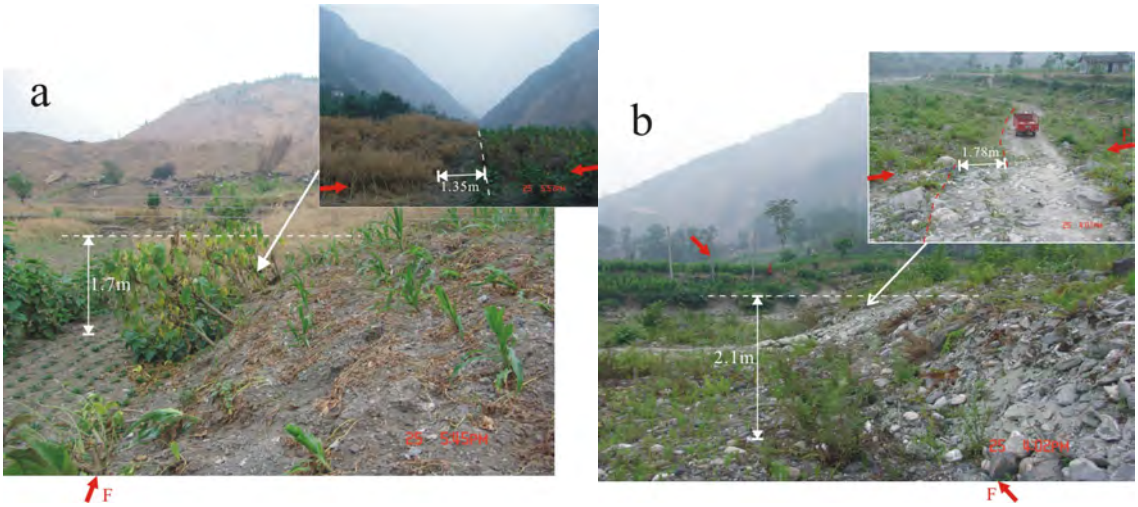


Figure 12. The surface ruptures of the Wenchuan Earthquake in Chenjiaba town and Fenghuang village, Guixi town. (a). the displacement of the terrace Ⅰ in Chenjiaba village; (b). the displacement of a pathway and the floodplain in Fenghuang village, Guixi town.

Figure 13. Waterfall was created by the surface rupture of the Wenchuan Earthquake in Pingtong town.

1km long and is N50-60°E trending. The road, river terrace and the river bed are offset by the rupture (**Figure 13**). The vertical displacement of the road is about 1.9 m.

All above facts indicated that the surface rupture on the Beichuan-Yingxiu fault showed the thrust, dextral slip characteristic of the fault. The maximum vertical displacement of the surface rupture is about 6.5 m and the maximum right-lateral displacement is about 5.85 m. Though the vertical displacements and the horizontal ones of the different segments have certain differences, as a whole, the ratio of the vertical displacement and the horizontal is close to 1:1. This is consistent with the previous studies [12].

2.2. The Surface Rupture on the Pengxian-Guanxian Fault

The Pengxian-Guanxian fault is also called Pengguan fault [5]. The Wenchuan Ms8.0 earthquake occurred on 12th May produced the coseismic surface rupture on this fault. The surface rupture begins near Xiangê town, Dujiangyan city. It sketches to the northeast, passes through Cifeng town in Pengzhou county，Bailu town，Jinhua town in Mianzhu city, Hanwang town, and breaks off near Sangzao town in Anxian county. The total length of it is about 50 kilometers.

1) The surface rupture of the Wenchuan Earthquake in Cifeng town

The southwest end of the surface rupture shows that the fault thrust from the northwest to the southeast and produced a reverse fault scarp. The rupture is N10-50°E trending and shows a broad wave-like on the surface, with the length of 200-300 m. The heights of the fault scarps are between 0.24 m to 0.7 m. The displacements of six ridges in the field are between 0.24 m to 0.5m (**Figure 14(a)**). Here, the vertical displacements are slightly larger than the horizontal ones.

2) The surface rupture of the Wenchuan Earthquake in Bailu town

The surface rupture passes through between two teaching buildings of the central school in Bailu town, Pengzhou city. The height of the fault scarp is about 2 m. According to measuring the length of superposition of the mattess, the shortening amount is 1.65 m (**Figure 14(b)**). The ruptures primarily present the thrust characteristic of the fault and the crustal shortening. It only shows a little horizontal displacement. Several gaps of the mattess are cut off and the maximum displacement is 0.42 m. Northeast about 300 m to this spot, the river bed was cut off and produced a fault scarp with a height of 1.8 m (**Figure 15(a)**). The right-lateral offset of the bank of the river is about 0.8 m (**Figure 15(b)**).

3) The surface rupture of the Wenchuan Earthquake in Jinhua town.

The surface rupture is N65-70°E trending in Jinhua

Figure 14. The surface rupture of the Wenchuan Earthquake in Cifeng town and Bailu central school. (a) the fault scarp and the offset of ridges in a field; (b) the fault scarp on the playground in the Bailu central school.

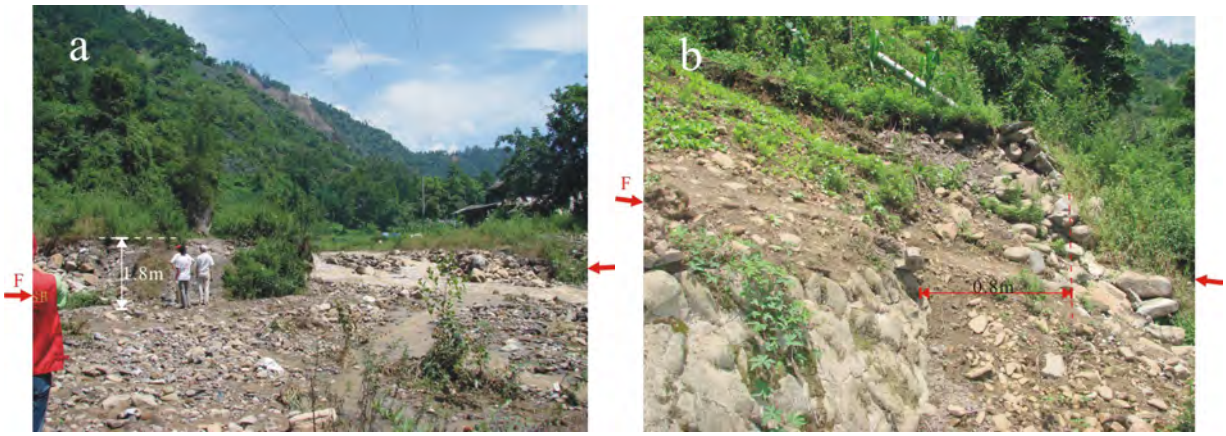


Figure 15. The surface rupture of the Wenchuan Earthquake on the riverbed and bank of Bailu river. (a) the fault scarp and waterfall on the riverbed was created by the Wenchuan Earthquake; (b) the right-lateral offset of the river bank.

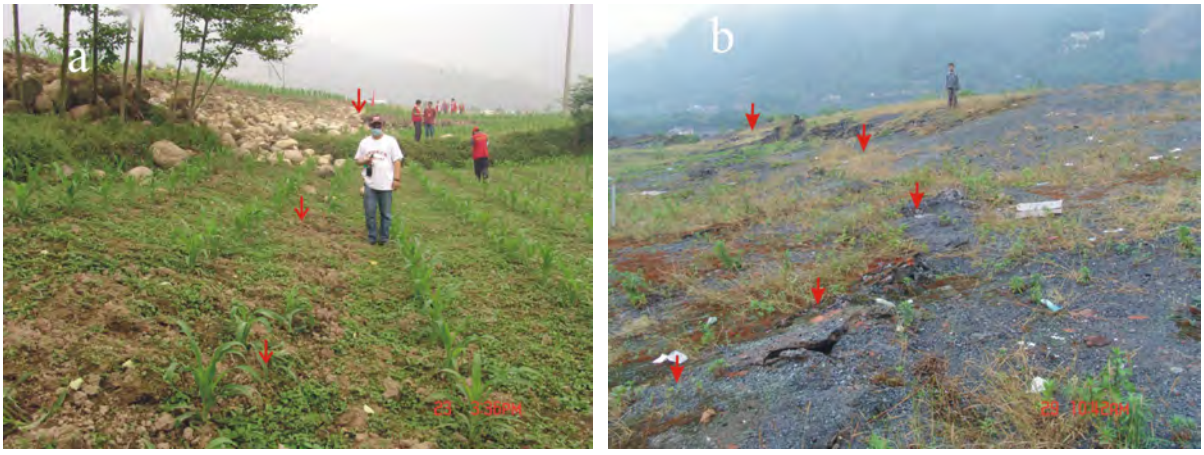


Figure 16. The surface rupture of the Wenchuan Earthquake in Jinhua and Hanwang town, Mianzhu county. (a) Scraps of the Penxian-Guanxian fault in Jinhua town; (b) Scraps of the Penxian-Guanxian fault in Hanwang town, the arrowheads show the location of the surface rupture

town, Mianzhu city (**Figure 16(a)**). The vertical displacement of the riverbed and terrace I is 1.5-2.0 m. The right-lateral dislocation of the lateral edge of the terrace I is 0.7 m. The scarps on the surface of the terrace III are lower relatively and the heights of them are between 0.4 m and 0.7 m. The dextral displacements of a pathway and

the road are 0.18 m，0.5 m respectively.

4) The surface rupture of the Wenchuan Earthquake in Hanwang town

The surface rupture is N48°E trending in the north of Hanwang town, Mianzhu city (**Figure 16(b)**). The right-lateral dislocation of the road in the east bank of the Mianyuan River is 0.3-0.4 m. The vertical displacement of a road in the west bank of the river is about 1.0 m and the dextral dislocation is 0.4-0.6 m. The crustal shortening is 1.3 m.

In a word, it believes that the surface rupture on the Pengxian-Guanxian fault shows thrust and dextral characteristic of the fault. The vertical displacements and the horizontal ones on some segments are approximately equal. But the rates of them on the most other segments are between 1:3 and 1:2. The maximum horizontal shortening is 1.65 m. All of these show that the main characteristic of the fault rupture is thrust and shortening.

According to the surface rupture characteristic of the Wenchuan Ms8.0 earthquake on the Beichuan-Yingxiu fault and Pengxian-Guanxian fault, a seismogenic structures model of this earthquake can be made. The model shows that the inclination of the Beichuan-Yingxiu fault becomes flat gradually and disappears in the horizontal detachment layer about 20 km deep under the surface at last. The inclination of the Pengxian-Guanxian fault becomes flat too from the surface to the deep crust and converges with the Beichuan-Yingxiu fault at last. The collision of India-Asia Plate resulted in Sichuan-Qinghai block slipped to the NEE orientation. The activity of the Longmen Shan tectonic belt transformed brittle and thrust-dextral motion, accumulating tectonic stress. The hypocenter of the Wenchuan Ms8.0 earthquake is on the section of the Beichuan-Yingxiu fault, where the section of the fault up warped from the horizontal detachment layer. It simultaneously made the Pengxian-Guanxian fault produce coseismic dislocations (**Figure 17**).

3. HAZARD OF THE WENCHUAN Ms8.0 EARTHQUAKE

The primary hazard characteristics of the Wenchuan earthquake Ms8.0 are described as follows. 1) The seismic wave produced strong ground motion, destroying the structures. While the earthquake occurred, the writers were working in the seventh floor office in the urban district of Chengdu city. I could felt the strong vibration of the structure and the vibration lasted about two minutes. During this period, the vibration became weak obviously for 3-5 seconds. According to the seismological inversion for the spatiotemporal process of the earthquake's rupture, the first strong vibration that I felt was produced by the initial rupture of the fault near Yingxiu town. The rupture of the fault near Beichuan town resulted in the second strong vibration. 2) The earthquake produced a 220 km long surface rupture on the Beichuan-Yingxiu fault and a 50 km long one on the Pengxian-Guanxian fault. The surface ruptures cut off the structures and all of the structures close to the surface rupture zones almost broke down. The surface rupture zone pass through the central of the Pingtong town and all of the buildings in the rupture zone collapsed. The other structures with 30-50 m distance of the zone did not collapsed, only were severely or moderately destroyed.

3)The strong ground motion and the dislocation of the surface rupture resulted in the secondary geological hazards, such as landslip, collapse, and so on. The hazards pulled down or severely destroyed the structures. The earthquake occurred in the Longmen Shan area. Especially, the Beichuan-Yingxiu fault is the boundary fault between the higher modern geomorphology and the lower. The difference of the topographic height is very big. Thereby under the interaction of the strong ground motion and the dislocation of the surface rupture, the earthquake produced a large-area and large-scale landslide, collapse *et al*. in the high earthquake intensity regions of the Longmen Shan area. On the one hand, the landslides and collapse directly destroyed or buried large numbers of houses. For example, half of the houses were directly destroyed by the landslides. On the other hand, they blocked up the rivers and produced tens of sag lakes. Many croplands and houses were submerged. It formed a large potential threat of the flood disaster to the downstream region.

According to the investigations of the Earthquake Emergency Job Team of China Earthquake Administration, Seismological Bureaus of Gansu province, Shanxi province and Chongqing, it also made reference to the studies of InSAR and records of seismographs and rechecked the earthquake intensity of some major severely damaged regions. Basing on these studies, the isoseismal line of the Wenchuan Ms8.0 earthquake was drew (**Figure 18**). The isoseismal line only shows the general degree of the damages of this earthquake. And we did not considered the effect of the aggravated damages caused by the surface rupture.

1) XI intensity region: There are three XI intensity regions, occurring in Yingxiu-Hongkou, Yuejia Shan-Gaochuan and Xuanping-Beichuan area respectively. They distribute in the Beichuan-Yingxiu fault zone and its hanging wall. The total area of them is about 680 km^2. The houses almost broke down in this area. Especially, the houses in the Yingxiu town and Beichuan county were completely destroyed. Individually, the un-collapsed houses were also severely destroyed. The scale of the surface is large and its continuity is generally well. The displacements of many segments are more than 3 meters. The earthquake resulted in multiple-site collapses and landslides. These second dary geological hazards walled up the rivers and produced the sag

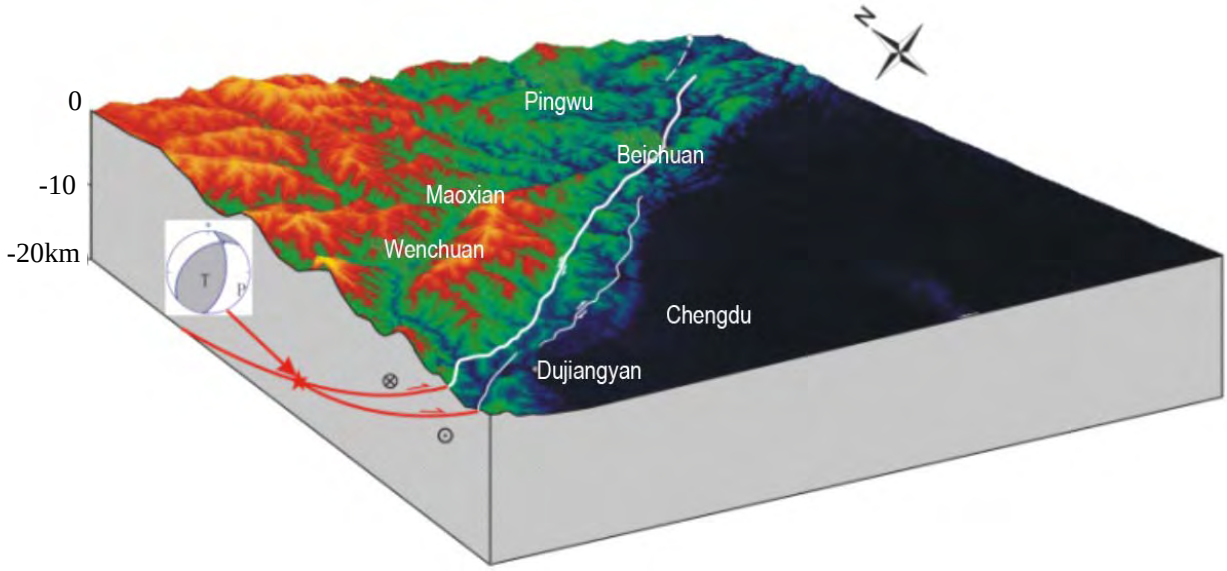


Figure 17. The seismogenic structures model of the Wenchuan Ms8.0 earthquake.

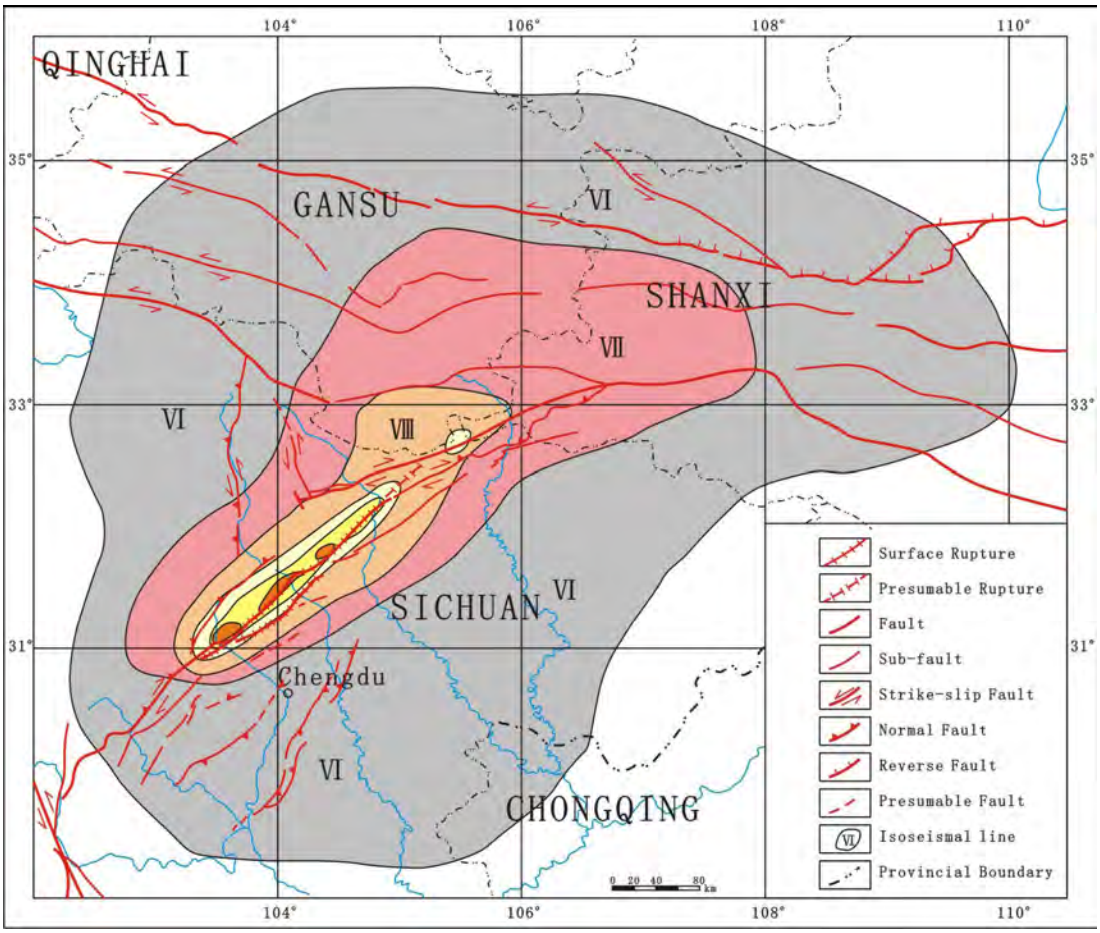


Figure 18. Isoseismal line of the Wenchuan Ms8.0 earthquake on 12th May, 2008.

lakes, such as Qingping sag lake and Tangjia Shan sag lake in Beichuan county, and so on.

2) Ⅹ intensity region: The southwest part of the region begins in the southwest of the Yingxiu town. It

sketches to the northeast and ends at Shikanzi village, Nanba town. The region includes the surface rupture extension of the Beichuan-Yingxiu fault. It distributes as a narrow area with a N50°E-trending long axis. The total area of the region is about 2520 km^2. The houses in this region were generally destroyed and most of them were broken down. There are many huge landslides in this area.

3) Ⅸ intensity region: The region arises from Sanjiang town and Shuimo town, north-eastward to Shuiguan town, Fangshi town, south-eastward to Xicheng town, Hanwang town and Xiaoba town. A Ⅸ intensity abnormal region occurred in the Ⅷ intensity region near the juncture area of Sichuan province, Gansu province and Shaanxi province. This area of the Ⅷ intensity region sums up to 4000 km^2. Most of the houses in there were severely destroyed and many of them collapsed. The collapses and landslides can be seen everywhere.

4) Ⅷ intensity region: This region arises from Wenjing river, Heba village; north-eastward to Yangpingguan village; north-westward to Weimen village in Maoxian town, Pingwu, Wenxian town; south-eastward to Juyuan village, Mianzhu, Yanmen village in Jiange town. A Ⅷ intensity abnormal region occurred in the Hanyuan county. The total area of the Ⅷ intensity region is 17800 km^2. Most of the houses in this region were severely or moderately destroyed, a little collapsed. The collapses and landslides are common too, but the scale of them is relatively small.

5) Ⅶ intensity region: This region arises from Dawei, north-eastward to Taibai, Foping village, north-westward to Shuangliusuo village in Heishui town, Zhenping town in Songpan county, Jiuzhaigou, Lixian county; south-eastward to Deyang, Mianyang, Guangyuan city. The total area of this region is 68000 km^2. Generally, the houses in this region are slightly destroyed. Many of them were moderately destroyed and a little were seriously destroyed. Individually, some adobe houses collapsed. There are some collapses and landslides in the region, but the scale of them is small.

6) Ⅵ intensity region: This region arises from Xiaojin, Tianquan village, north-eastward to Sanyuan, Shanyang village; north-westward to Hongyuan, Lintao, Jingning, Pingliang city; south-eastward to Jianwei town, Zigong, Hechuan county, Pingchang county. The total area of this region is 240000 km^2. Most of the houses were slightly destroyed and a little were moderately destroyed. Individually, some adobe houses were severely destroyed. The collapses and landslides are rare in distribution in this area. Most of them are rolling stones or part of the slopes collapsing.

In a word, the total intensity area above Ⅵ degree of the Wenchuan earthquake is about 333000 km^2. The characteristics of the isoseismal line are described as follows.

1) The high earthquake intensity line stretches to N40-50°E along Longmen Shan tectonic belt. The rate of the long axis and the minor is between 8:1 and 10:1. Three Ⅺ intensity regions are isolated in distribution. It presents a multipoint instantaneous characteristic of the rupture.

2) According to the seismological inversion, the initial rupture of the fault is near Yingxiu town and cracked unilaterally north-eastward along Longmen Shan belt. The isoseismal line decayed rapidly south-westward and slowly norh-eastward. It is consistent with the unilateral rupture process.

3) The isoseismal line of Ⅵ intensity region decayed slowly to Sichuan basin, Hongyuan and Nuoergai county. The reason may be that Sichuan basin is in the rigid Yangtze Platform and seismic wave propagated difficultly. And the soft oil layer in the Hongyuan, Nuoergai area increased the hazard.

4) Ⅵ, Ⅶ and Ⅷ intensity regions stretch to east-west in the north of Sichuan province, Guansu province and the south of Shaanxi province. The intensity of the earthquake may be under the control of the regional major faults.

4. DISCUSSIONS

1) It is generally believed that the Longmen Shan tectonic belt is a deep fault and the boundary fault between the Songpan-Garzê orogenic belt and the Yangtze Carton [1,6,8]. But there are different opinions on the late Quaternary activities of the Longmen Shan tectonic belt among different scholars. Tang [24] believed that some segments of the Longmen Shan tectonic belt were Holocene active faults, while the other segments were early Quaternary faults. Zhou [12] and Densmore [5] believed that most segments of the belt were Holocene active faults and Beichuan-Yingxiu fault was the major active fault. The acquaintanceship change among the scholars has been undergone for 15 years. The deeply incised topography, flourish vegetations, farmland reclamation and low slip rate of the fault make the field study and mapping more difficult and uncertain. Only a few clear active tectonic landforms were found on the Longmen Shan tectonic belt with length of 300 km. Even so, most surface rupture of the Wenchuan earthquake outcropped along the locations of active tectonic landform of the Beichuan-Yingxiu fault found by the scholars before. The surface rupture did not outcrop along the linear structure discovered before near Xujiagou village, Longmen Shan town, but outcropped in the northwest of the linear structure with distance of 0.5-1.0 km. Only a few seismic fractures cut into a sag-ponding discovered before. It presents the dextral slip characteristic of the secondary fault, distributing at an angle 5-10°with the major fault. This may be relative to the strain partition of the complicated fault combination. The outcrop of the surface rupture on the Pengxian-Guanxian fault is more complicated. On the 1:200,000 regional geological map, there are several

northwest-trending faults with different scales and lengths. It is difficult to confirm which is the major fault of the Pengxian-Guanxian fault. The surface rupture stretches through Bailu town with about 7 km distance southeastward the fault, not outcropping along the Holocene active fault in the Daiming temple in the Tongji town, Pengzhou county. It is very difficult to confirm the activity of the fault in Bailu town before the occurrence of earthquake. Therefore, the decision should be made cautiously for the activity of the faults in such region. A few phenomena of the geology and landform are important evidences to confirm the activity of the whole fault.

2) We have known the activity of several major faults of the Longmen Shan tectonic belt. Especially, Beichuan-Yingxiu fault is a Holocene active fault. Only 2-3 trenches in previous field work revealed the paleoearthquakes [3-5]. According to the historical data, only 3-4 earthquake events above Ms6.0 are recorded and the maximum is the Wenchuan Ms6.5 earthquake in 1657. These make us believe that the maximum potential magnitude of the earthquakes on the fault is about Ms7.0. Furthermore, the topographic responses of the active faults almost concentrate on the middle-south segment of the Longmen Shan tectonic belt. The historical strong earthquakes and modern small-moderate earthquakes are dense in a belt along this segment. The seismicities on the middle-north segment are weak. The gradient zone of the crustal thickness along the Longmen Shan tectonic belt changes from NE to NNE near Beichuan county and stretches along the east boundary of the Minshan block. The topography and geomorphology relatively changed there. Therefore, the boundary of the subsection of the Longmen Shan tectonic belt is near Beichuan county [25]. The surface rupture of the earthquake broke through the limit of the fault subsection and stretched north-eastward to 30-40 km. It teminated at Shikanzi village, Nanba town. The earthquake not only produced about 220 km long surface rupture on the Beichuan-Yingxiu fault, but also about 50 km long one on the Pengxian-Guanxian fault. Thus, we should fully consider the scale of active faults on determination of the potential earthquake resource according to sections of active faults. We can't confirm the upper-limit magnitude of the potential earthquake resource through increasing the magnitude of the maximum historical earthquake. Furthermore, it should be fully demonstrated that the boundary of fault subsections could control the propagation of strong earthquakes' rupture permanently. And it should be recognized that several faults in the reverse or reverse-slip tectonic circumstance would be likely to crack coinstantaneously in a strong earthquake. This should be considered for determination of the potential earthquake resource.

3) The unilateral multipoint instantaneous rupture of the earthquake increased the hazards in the north of Sichuan province, Gansu province and the south of Shanxi province. The Ⅵ intensity region attenuated slowly to the Sichuan basin, Hongyuan and Nuoergai county. It presents a special attenuation characteristic of strong earthquake. According to 'National Standard of the People's Republic of China. Seismic ground motion parameter zonation map of China' (GB18306-2001) [26], the area where the surface rupture outcrop are all the partitions of ground motion parameter 0.1gal. It is incompatible with the article of 'National Standard of the People's Republic of China. Code for Seismic Design of Buildings' (GB50011-2001) [27]. In this article, it is considered that the causative fault could be neglectable in the earthquake intensity region under Ⅷ degree. This is because the interval recurrence of the great earthquakes above Ms8.0 is beyond a thousand years or several thousand years. And the peak acceleration of the ground motion could not reach Ⅷ degree in 500yr interval recurrence. Moreover according to the investigation of this earthquake, the structures in the surface rupture zone almost collapsed or severely destroyed. The other structures with 30-50m distance of the zone did not collapse (such as the structures in Hongkou, Jinhua, Pingtong town *et al*.). There is obvious difference in the criterion (GB50011—2001) [27] that the preventing belt should be 200m to the causative fault for the third class buildings in the Ⅷ intensity region. So, it is suggested to cancel the criterion that didn't consider preventing the causative fault under Ⅷ intensity region. We should mainly consider the Holocene active faults. The preventing distance to the causative fault could be modified basing on this earthquake and making reference to the overseas great earthquakes' datum.

4) According to the surface rupture characteristics of the earthquake, the Beichuan-Yingxiu fault is a dextral-slip and thrust fault and the vertical displacement on it is equal to the horizontal. The Pengxian-Guanxian fault is thrust fault with a little dextral-slip component. The slip rates of the major faults of the Longmen Shan tectonic belt are not accurate enough. And the paleoearthquakes is not enough for the determination of the upper-limit magnitude. According to the records of 'Compilation of Historical Data of Earthquakes in China' (Edited by Xie [28]), there were eleven historical earthquakes during May 934A.D. to April-May 953 A.D.. Among them, two earthquake caused serious destroy. One occurred during Nov.16 to Dec.15, 942 A.D.. The other occurred during Nov.7 to Dec.6, 951 A.D.. Considering the bad quality of dwellers' houses at that time, the damage caused by the two historical earthquakes showed that the intensities reached Ⅵ—Ⅶ degree. It is equivalent to the effect of Chengdu city caused by the Wenchuan Ms 8.0 earthquake. Densmore [5] excavated a trench in Xiling town with distance about 70-80km to Chengdu city. The ^{14}C ages constrain the timing of the last paleoearthquake event to 860 ± 40 – 930 ± 40 yr B.P.. The A.D. year is 1090 ± 40 – 1020 ± 40 yr. It is

nearly equivalent to the historical earthquake as mentioned above. Does it imply that a great earthquake with equivalent magnitude as the Wenchuan Ms8.0 earthquake occurred a thousand years ago? The Wenchuan Ms8.0 earthquake provides an opportunity to solve these questions. There is very important scientific and practical meaning for the seismic risk assessment of the Longmen Shan tectonic belt and its adjacent areas and the earthquake disaster prevention of Chengdu Plain with a large population density.

5. ACKNOWLEDGMENT

We are very grateful to the people who provided us supports and help. Finally, we thank the people of the Longmen Shan region for their unending curiosity, hospitality, and generosity.

REFERENCES

[1] Burchfiel, B.C., Chen, Z., Liu, Y. and Royden, L.H. (1995) Tectonics of the longmen shan and adjacent regions. *International Geology Review*, **37**, 661-735.

[2] Kirby, E., Whipple, K.X., Tang, W. and Chen, Z. (2003) Distribution of active rock uplift along the eastern margin of the tibetan plateau:Inferences from bedrock channel longitudinal profiles. *Rnal geophysical research*, **108(B4)**, 2217.

[3] Li, Y., Zhou, R.J., Alexander, L., Densmore, L.A., *et al.* (2006) Continental dynamic processes and geological response in the eastern margin of the Qinghai-Tibet plateau [A]. Geological Publishing House, Beijing, 1-134 (in Chinese).

[4] Zhou, R.J., Li, Y., Alexander, L., Densmore, A.L., Michael, A., Ellis, M.A., He, Y.L., Li, Y.Z. and Li, X.G. (2007) Active tectonics of the longmen shan region on the eastern margin of the Tibetan plateau [J]. *ACTA Geologica Sinica*, **81(4)**, 593-604.

[5] Densmore, A.L., Ellis, M.A., Li, Y., Zhou, R.J., Hancock, G.S. and Richardson, N.J. (2007) Active tectonics of the Beichuan and Pengguan faults at the eastern margin of the Tibetan Plateau. *Tectonics*, TC4005.

[6] Xu, Z.Q., Hou, L.W., and Wang, Z.X. (1992) The orogenic process of Songpan-Ganzi orogenic belt in China. Geology Press, Beijing.

[7] Chen, S.F., Deng, Q.D., Zhao, X.L., Wilson, C.J.L., Dirks, P., Luo, Z.L. and Liu, S.Y. (1994) Deformational characteristics evoloutionary history, and deformation mechanism of the middle Longmenshan thrust-nappes and related tectonics. *Seismology and Geology*, **16(4)**, 404-421.

[8] Li, Y., Zeng, Y.F. and Yi, H.S. (1995) Sedimentary record and tectonic evolution of Longmen Shan foreland basin. Chengdu University of Science and Technology Press [A], Chengdu, 1-9 (in Chinese).

[9] Avouac, J.P. and Tapponnier, P. (1993) Kinematic model of active deformation in Central-Asia. *Geophysical Research Letters*, **20**, 895-898.

[10] Zhang, P.Z., Sheng, Z.K., Wang, M., Gan, W.J., Brögmann, R., Monlar, P., Wang, Q., Niu, Z.J., Sun, J.Z., Wu, J.C., Sun, H.R. and You, X.Z. (2004) Continuous deformation of the Tibetan Plateau from global positioning system data. *Geology*, **32**, 809-812.

[11] England, P.C. and Molnar, P. (1990) Right-lateral shear and rotation as the explanation for strike-slip faulting in eastern Tibet. *Nature*, **344**, 140-142.

[12] Zhou, R.J., Li, Y., Densmore, A.L., Ellis, M.A., He, Y,L., Wang, F.L. and Li, X.G. (2006) Active tectonics of the eastern margin of the Tibet plateau. *Journal of Mineral Petrol*, **26(2)**, 40-51.

[13] Li, Y., Huang, R.Q., Zhou, R.J., Densmore, A.L, Yan, L., Zhang, Y., Nicholas, R., Dong, S.L., Michael, A., Ellis, He, Y.L., Chen, H, Qiao, B.C. and Ma, B.L. (2009) The surface rupture of the Xiaoyudong fault in the Wenchuan Earthquake and its geological significance, Longmen Shan, Sichuan, China. *Quaternary Sciences*, **29(3)**, 506-516.

[14] Xu, X.W., Wen, X.Z., Ye, J.Q., Ma, B.Q., Chen, J., Zhou, R.J., He, H.L., Tian, Q.J., He, Y.L., Wang, Z.C., Sun, Z.M., Feng, X.J., Yu, G.H., Chen, L.C., Chen, G.H., Yu, S.N., Ran, Y.K., Li, X.G., Li, C.X. and An, Y.F. (2008) The Ms8.0 Wenchuan Earthquake surface ruptures and its seismogenic structure. *Seismology and Geology*, **30(3)**, 597-629.

[15] Liu, J., Zhang, Z.H., Wen, Li., Sun, J., Xing, X.C., Hu, G.Y., Xu, Q., Paul, T.E., Zeng, L.S., Ding, L. and Liu, Y.L. (2008) The Ms8.0 Wenchuan Earthquake. Co-seisimic surface rupture and its tectonic implication- an out-of -sequence thrusting event with partitioned on multiple faults. *Acta Geologica Sinica*, **82(12)**, 1707-1722.

[16] Chen, G.H., Xu, X.W., Zheng, R.Z., Yu, G.H., Li, F., Li, C.X., Wen, X.Z., He, Y.L., Ye, Y.Q., Chen, X.C. and Wang, Z.C. (2008) Quantitative analysis of the co-seismic surface rupture of the 2008 Ms8.0 Wenchuan Earthquake, Sichuan China along the Beichuan-Yingxiu fault. *Seismology and Geology*, **30(3)**, 723-738.

[17] Li, Y., Zhou, R.J., Dong, Densmore, A.L., Yan, L., Zhang, Y., Richardson, N., Dong, S.L., Ellis, M.A., He, Y.L., Chen, H., Qiao, B.C. and Ma, B.L. (2008) Surface rupture, thrusting and strike-slipping in Wenchuan earthquake of Sichuan, China. *Journal of Chengdu University of Technology*, **35(4)**, 404-413.

[18] Li, Y., Zhou. R.J., Densmore, A.L., Yan, L., Zhang, Y., Richardson, N., Dong, S.L., Ellis, M.A., He, Y.L. Chen, H., Qiao, B.C. and Ma, B.L. (2008) Surface Rupture and Deformation of the Yingxiu Beichuan Fault by the Wenchuan Earthquake. *Acta Geologica Sinica*, **82(12)**, 1688-1706.

[19] Li, Y., Zhou, R.J., Densmore, A.L., Yan, L., Zhang, Y.,. Richardson, N., Dong, S.L., Ellis, M.A. He, Y.L., Chen, H., Qiao, B.C., and Ma, B.L. (2009) Geological background of Longmen Shan seismic belt and surface ruptures in Wenchuan Earthquake. *Journal of Engineering Geology*, **17(1)**, 3-18.

[20] Li, Y., Zhou, R.J., Densmore, A.L. Yan, L., Zhang, Y., Richardson, N., Dong, S.L., Ellis, M.A., He, Y.L., Chen, H., Qiao, B.C. and Ma, B.L. (2009) Active tectonics and surface rupture of the Pengxian-Guanxian Fault, Longmenshan, Sichuan, China. *Quaternary Sciences*, **29(3)**, 405-417.

[21] Li, H.B., Fu, X.F., Woerd, J.V.D., Si, J.L.,Wang, Z.X., Hou, L.W., Qiu, Z.L., Li, N., Wu, F.Y. and Xu, Z.Q. (2008) Co-seisimic surface rupture and dextral-slip oblique thrusting of Ms8.0 Wenchuan Earthquake. *Acta Geologica Sinica*, **82(12)**, 1623-1643.

[22] Fu, B.H., Shi, P.L., Wang, P., Li, Q., Kong, P. and Deng, G.D. (2009) Geometry and kinematics of the 2008 Wenchuan earthquake surface ruptures around the Qushan Town of Beichuan County, Sichuan: Implications for mitigation of seismic and geologic disasters. *Chinese Journal of Geophysics*, **52(2)**, 485-495.

[23] Ran, Y.K., Chen, L.C., Chen, G.H., Yin, J.H., Chen, J., Gong, H.L., Shi, X. and Li, C.X. (2008) Primary analyses of in-situ recurrence of large earthquake along seismogenic fault of the Ms8.0 Wenchuan Earthquake. *Seismology and Geology*, **30(3)**, 630-634.

[24] Tang, R.C. and Han, W.B., Ed. (1993) Active Faults and Earthquakes in Sichuan Province. Seismological Press, Beijing, 368.

[25] Chen, G.G., Ji, F.J., Zhou, R.J., Xu, J., Zhou, B.G., Li, X.G. and Ye, Y.Q. (2007) Primary research of activity segmentation of Longmenshan fault zone since later quaternary. *Seismology and Geology*, **29(3)**, 657-673.

[26] State Bureau of Quality and Technical Supervision. (2001) National Standard of the People's Republic of China. Seismic grouond motion parameter zonation map of China [A].1-2 (in Chinese).

[27] Ministry of Housing and Urban-Rural Development of the People's Republic of China, General Administration of Quality Supervision, Inspection and Quarantine of the People's Repubilic of China. (2001) National Standard of the People's Republic of China. Code for Seismic Design of Buildings [A]. China Architecture & Building Press, Beijing, 17 (in Chinese).

[28] Xie, Y.S. and Cai, M.B. (1983) Compilation of historical data of earthquakes in China (the first volume) [A]. Science Press, Beijing, 105-107 (in Chinese).

Vol.2, No.3, 175-183 (2010)
doi:10.4236/ns.2010.23027

Natural Science

The antioxidant activity and hypolipidemic activity of the total flavonoids from the fruit of Rosa laevigata Michx

Yue-Tao Liu, Bi-Nan Lu, Li-Na Xu, Lian-Hong Yin, Xiao-Na Wang, Jin-Yong Peng*, Ke-Xin Liu

College of Pharmacy, Dalian Medical University, Dalian, China; jinyongpeng2005@163.com

Received 21 December 2009; revised 8 January 2010; accepted 30 January 2010.

ABSTRACT

In the present work, the antioxidant activity *in vitro* and hypolipidemic activity of the total flavonoids (TFs) from the *Rosa laevigata* Michx fruit was evaluated, and the antioxidant effect *in vivo* was also discussed. The TFs exhibited a high scavenging effect on 2, 2-diphenyl-1-picrylhydrazyl (·DPPH) with IC_{50} values 0.01 mg/mL, and a stong reduce power in the test. Hyperlipemic mice were intragastric administrated with TFs (25, 50 mg/kg/day) for 4 weeks, and fenofibrate was used as the positive reference substance. After the experiment, the levels of TC (total cholesterol), TG (triglyceride), LDL- C (low density lipoprotein-cholesterol) of the mice administrated with high-dose of TFs were markedly declined by 45.02%, 33.86% and 73.68%, respectively, while HDL-C (high density lipoprotein-cholesterol) was significantly increased compared with model group. To investigate the hepatoprotective effect, histopathological assay, ALT (alanine aminotransferase), AST (aspartate aminotransefrase) and ALP (alkaline phosphatase) were also studied, and the results showed that TFs exhibited a favorable effect on liver protection, of which the levels of ALT, AST and ALP were elevated by 55.85%, 29.15% and 25.68%, respectively. Furthermore, the TFs could significantly decrease the MDA (malondialdehyde) level and improve the levels of CAT (Catalas), SOD (superoxide dismutase), GSH (reduced glutathione), and GPX (glutathione peroxidase) compared with hyperlipemia mice. Our results suggested that TFs has a high antioxidant activity and hypolipidemic activity, which can be used as a potential medicine for cardiovascular diseases.

Keywords: *Rosa laevigata* Michx; Total Flavonoids; Antioxidant Activity; Hypolipidemic Activity

1. INTRODUCTION

Many studies have proved that reactive oxygen species (ROS) and free radicals play a vital role in maintaining human health. When the balance between the generating and scavenging of ROS and free radicals *in vivo* is destroyed, an oxidative stress would happen, which might lead to extensive oxidative damage to cellular biomolecules, such as DNA, proteins and lipids. Many chronic-diseases, such as hyperlipemia, hyperpiesia and cancer, have proved to be associated with the existence of oxidative stress.

Hyperlipemia is considered as a risk factor involved in the development of cardiovascular disease [1]. High lipid levels can harden the arteries or speed up the process of atherosclerosis. Nowadays, there are numerous hypolipidemic drugs for clinical use, however, the pharmacologists and chemists have been perplexing by the characteristic profiles of toxic side effects including numerous harmful syndromes [2,3], which can increase the risk of heart disease, stroke, and other vascular diseases. Thus, an investigation of hypolipidemic agents with negligible side effect seems important. In recent researches, there has been a growing interest in the ingredients of natural plants, vegetables and cereals, not only for their radical-scavenging activities, but also due to their neglectable physical side effects. There are numerous herbal medicines exerting good hypolipidemic actions with few side effects in Asian countries. Plant materials have long been used as traditional medicines for the treatment of a wide variety of ailments and diseases, such as the polysaccharides from *Lycium barbarum*, Saponins in *Ipomoea batatas* tubers, total flavonoids of *Litsea coreana* leaf, the flavonoid-riched extract from *Eugenia jambolana* seeds. Vast numbers of plants have been shown to lower plasma lipid levels [4-7].

R. laevigata Michx is a famous medicinal plant in China, and its fruit is widely used as the invigorator, paregoric and astringent. Now chemical and pharmacological researches have demonstrated that this medicinal plant can cure hyperpiesia, chronic cough, and dermatogic disease, as well as inhibit experimental arterial sclerosis [8]. Flavonoids are primarily considered as the pig-

ments responsible for the autumnal burst of hues and many shades of yellow, orange and red in flowers and food. The flavonoids have long been applied to possess anti-inflammatory, antioxidant, antiallergic, hepatoprotective, antithrombotic, antiviral and anticarcinogenic activities [9-11]. Up to now, the report about the antioxidant activity and hypolipidemic activity of the flavonids from this medicinal plant fruit was not discovered.

The major of the present work is to study the antioxdant activity and hypolipidemic activity of the total flavonoids (TFs) from the fruis of *Rosa laevigata* Michx on hyperlipemic mouse. In the test, we measured the diversity on biotical parameters on hyperlipidemic mice and the antioxidant effect *in vivo*. This is the first cover about it as far as we known.

2. MATERIALS AND METHODS

2.1. Plant Material

The dried fruit of *R. laevigata* Michx was purchased from Yunnan Qiancaoyuan Pharmaceutical Company Co. LTD (Yunnan, China), and authenticated by Dr. Yunpeng Diao (School of Pharmacy, Dalian Medical University, Dalian, China). Voucher specimen was deposited in College of Pharmacy, Dalian Medical University (Dalian, China).

2.2. Chemicals

$NaNO_2$, Al $(NO_3)_3$, NaOH, $FeCl_3$, potassium ferricyanide, cholesterol , sodium cholate (biochemical reagent), trichloroacetic acid (TCA) and vitamin C (VitC) were purchased from Shenlian Chemical Company (Shenyang, China). Rutin was purchased from TCM institute of Chinese Materia Medica (Nanjing, China). 2, 2-diphenyl-1-picrylhydrazyl (·DPPH) was purchased from Sigma-Aldrich (St. Louis, USA). Commercial kits used for determination of TC (total cholesterol), TG (triglyceride), HDL-C (high density lipoprotein-cholesterol), ALT (alanine aminotransferase), ALP (alkaline phosphatase), AST (aspartate aminotranse- frase), MDA (malondialdehyde), CAT (Catalase), SOD (superoxide dismutase), GSH (reduced glutathione), and GPX (glutathione peroxidase) were all purchased from Jiancheng Institute of Biotechnology (Nanjing, China). Fenofibrate was purchased from Laboratoires Fournier S.A. (France). All the chemicals were of analytical grade.

2.3. High-Fat Diet and Animals

High-fat diet was made according to the method depicted by Experimental methodology of pharmacology [12], containing normal pulverized food (97.5%), cholesterol (2%) and sodium cholate (0.2%). The cake was cut into pieces and dried at room temperature for 3 days before feeding to mouse.

Male Kunming mice were obtained from the Experimental Animal Center of Dalian Medical University (Dalian, China). The animals, weighted 20 ± 2g, were group-housed and kept in a regulated environment at 25 ± 1°C and 60 ± 5% relatively humidity under 12 h light/12 h dark conditions. The animals had free access to water and normal or high-fat diet.

2.4. Preparation of TFs from the Medicinal Plant

The fruit was ground into powder. Samples of 1.0 kg were weighted and mixed with 60% aqueous ethanol (solvent: sample= 8:1, v/w) for extraction. The process was refluxed in haven for two times and 2 h for each. The extracted solutions were filtered and evaporated to 1000 mL under reduced pressure at 60°C. Then, 100 mL of the residue was added into a glass column (4.0 cm×60 cm, contained 200.0 g D101 macroporous resin purchased from the Chemical Plant of Nankai University, Tianjin, China). The column was first washed by water (600 mL) to remove the un-desired compounds, and then the resin was eluted by 40% aqueous ethanol (800 mL) to elute the targets. The 40% aqueous ethanol elution was collected and evaporated under reduced pressure at 60°C to dryness, and 5.44 g powder was produced, which was stored in a refrigerator for subsequent experiments.

2.5. Determination the Content of Tfs the Medicinal Plant

The content of TFs was determined with a colorimetric method described by China Pharmacopeia [13].

2.6. ·DPPH Radical Scavenging Activity Assay

Different concentrations (0-1 mg/mL) of the TFs solutions were produced by dissolving the samples in deionised water. Each sample solution (2 mL) was mixed with 2 mL of ethanolic solution containing 2×10^{-4} mM ·DPPH. The mixture was shaken vigorously and left to stand for 30 min in dark place, and then the absorbance was measured at 517 nm against a blank [14]. VitC was used as the positive control.

2.7. Reducing Power Assay

The reducing power was determined according to the literature with some modifications [15]. VitC was used as the reference compound.

2.8. Experimental Procedure

After 1 week of acclimatization to the home cage, the mice were randomly divided into five groups with each group containing 10 mice. Group I (controlled group): animals were fed the normal laboratory diet daily for 4 weeks; Group II (model group): mice were fed the

high-fat diet daily for 4 weeks; Group III (positive group): mice were fed the high-fat diet plus fenofibrate (50 mg/kg/day, i.g.) daily for 4 weeks; Group IV: mice were fed the high-fat diet plus TFs (25 mg/kg/day, i.g.) daily for 4 weeks; Group V: mice were fed the high-fat diet plus TFs (50 mg/kg/day, i.g.) daily for 4 weeks. During the experimental procedure, the treated animals were given sufficient normal or high-fat diet.

2.9. Preparation of Biotical Samples and Protein Assay

After the experiments, the mice were killed and then the blood was collected, and the liver was quickly removed. The collected blood was placed 30-40 min for clot formation, and then the serum was separated by centrifugation at 3000 rpm for 15 min. Each liver tissue was immediately rinsed with saline, blotted on filter paper, weighed and finally stored at -70˚C pending biochemical analyses. One part of liver was cut and put into a flasket containing 10% buffered formalin solution for the following histopathology analysis. The other part of the liver tissues was sampled quickly and washed with chilled normal saline water. 10% (w/v) of tissue homogenate was prepared in cold normal saline, and centrifuged at 3000 rpm at 4˚C for 15 min, and the supernatants were preserved for the next step assay. Lowry method [16] was employed to measure the protein concentration in the homogenate with bovine serum albumin as the standard.

2.10. Histopathological Assay

Liver tissues were fixed with 10% neutral formalin and embedded in paraplast. Tissue sections (5 μm) were cut and stained by hematoxylin and eosin.

2.11. Measurement of Biochemical Parameters and Hepatic Enzymes in Serum

The levels of TC, TG and HDL-C in serum were determined using enzymatic kits according to the manufacture's instructions. The LDL-C was estimated by the method of Friedwald *et al*. [17]. ALT, AST and ALP were all assayed using the corresponding commercial kits.

2.12. Measurement of Hepatic Lipid Peroxidation and Antioxidant Enzymes

The assays of the levels of MDA, CAT, SOD, GSH and GPX were measured following the kits' instruction.

2.13. Statistical Analysis

All values were expressed as mean ± S.D. The significance of differences between the means of the treated and un-treated groups have been compared by one-way analysis of variance (ANOVA), followed by Student's *t*-test and *p*-values less than 0.05 were considered significant.

3. RESULTS

3.1. Tfs Purification Protocol

Before pharmacological investigation, the TFs from the fruit of R. laevigata Michx were required to prepare. In the present study, a kind of macroporous resin (MR) named D101, which has been widely used to purify flavonoids from medicinal plants as previously reported [18,19], was selected to accomplish the work in this study.

In MR column chromatography, 100 mL residue of the extraction (1 g plant material/mL) was added into a glass column (4.0 × 60 cm, containing 200 g D101 MR). Water (600 mL) and different concentrations of aqueous ethanol (10%, 20%, 30%, 40%, 50%, and 70%, each 600 mL) were used to elute the column in serials, and the elution solutions were collected individually to produce 42 different fractions (100 mL for each fraction) (**Figure 1(a)**). All the fractions were evaporated to dryness under reduced pressure at 50℃, and then the contents of the

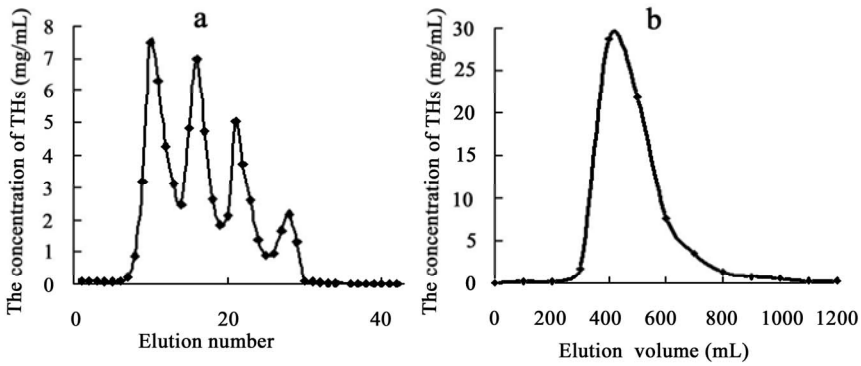


Figure 1. (a) The elution curve of the TFs on D101 MR column using water and different concentrations of ethanol as the elution solvents; (b) The elution curve of the TFs on D101 MR column using 40% aqueous ethanol as the elution solvent.

TFs were determined. And then the recovery of every concentrations of alcohol was calculated through summing every fraction in one concentration. It was obvious that the flavonoids mainly existed in 10%, 20%, 30% and 40% aqueous ethanol elution solutions. There was nearly nothing in water, 50%, and 75% aqueous ethanol solutions. Thus, water was first used to elute the water soluble chemicals, which were the undesired components and not supposed to be retained on D101 MR. Then, 40% aqueous ethanol was selected to elute the targets.

Secondly, the required amount of water to remove undesired constituents and 40% aqueous ethanol to collect the TFs were detrmined. When water was applied to the column, each 100 mL was collected individually, and then the contents of the TFs in each fraction were detected. The results showed no flavonoid was eluted out untill water volume increasing to 600 mL. However, with more water was used, part of flavonoids appeared in water. Therefore, 600 mL was set as the water quantity to clean the column in our research. Then, 40% aqueous ethanol was used to elute the flavonoids, and each 100 mL of the ethanol solution was collected individually and evaporated to dryness. The contents of the TFs were determined. As shown in **Figure 1(b)**, it was apparently us that when 800 mL of 40% aqueous ethanol was used to elute the column, almost all the TFs was eluted out. Thus, it is reasonable to choose 800 mL as the elution volume.

After optimization, the TFs purification was carried out in triplicate as the mentioned protocol. The production rate, purity and recovery of the TFs were obtained, of which the production rate and the recovery were calculated as the following formula.

Production rate $(\%) = \frac{W_1}{W_2} \times 100\%$, where W_1 represents the amount of the crude extract; W_2 represents the amount of plant material used for MR column chromatography.

Recovery $(\%) = (\frac{P \times W}{C \times V}) \times 100\%$, where P is the purity of the TFs in crude extract; W represents the amount of the crude extract; C is the TFs concentration in residue before MR column chromatography, and V is the volume of the residue added into the column for purification. In the study, the production rate of the crude extract was 5.44 ± 0.07%, of which the content of the TFs reach to 78.45 ± 2.41%.

3.2. ·DPPH Radical Scavenging Activity

DPPH, a stable free radical, has widely been used as a substance to evaluate the antioxidant activity of various samples. The method is based on the reduction of the absorbance of ·DPPH solution at 517 nm in the presence of proton-donating substance, due to the formation of the diamagnetic molecule by accepting an en electron or hydrogen radical [20]. The TFs exhibted good scavenging activities to scavenge the stable radical ·DPPH to yellow-colored diphenyl picrylhydrazine in a dose dependent. As shown in **Figure 2(a)**, the TFs showed strong scavenging activity on ·DPPH with no difference compared with VitC, and the estimated IC_{50} value was 0.01 mg/mL. TFs can clear almost 90% of the radical at 0.15 mg/mL. From our test, the TFs can significantly clear ·DPPH in vitro with a dose-depended manner from 0 to 0.15 mg/mL compatible with the positive drug.

3.3. Reducing Power Assay

The reducing power of a compound may serve as a significant indicator of its potential antioxidant activity. The absorbance at 700 nm was used to demonstrate the reducing power. Increased absorbance of the reaction mixture indicates increased reducing power of the sample. As shown in **Figure 2(b)**, the reducing power of the TFs increased slightly in a dose-depended manner when

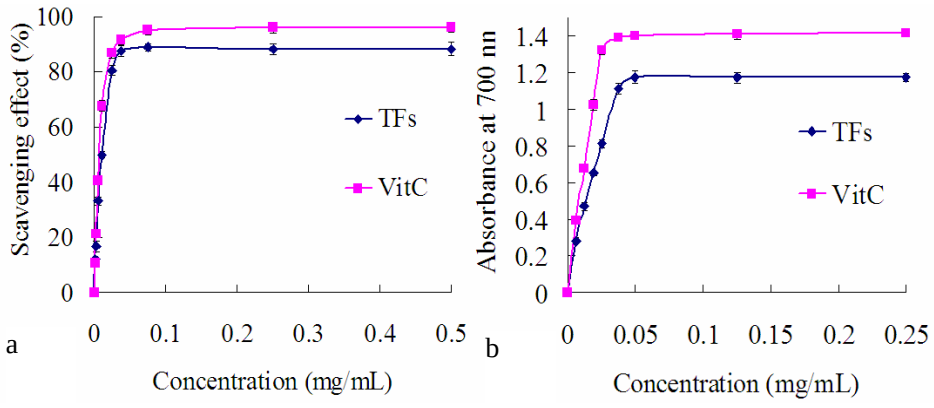


Figure 2. (a) Scavenging effect of the TFs and VitC at different concentrations on DPPH. (b) Reducing power of the TFs and VitC at different concentrations. Data was expressed as means ± S.D. (n = 5).

then it reached a plateau. The reducing power of the TFs was 0.812 at 0.025 mg/mL and 1.17 at 0.125 mg/mL. However, VitC only showed slightly higher activity, with the reducing power of 1.32 at 0.025 mg/mL and 1.45 at 0.125 mg/mL.

3.4. Histopathological Assay

As shown in **Figure 3**, histopathology of the liver in normal mouse **(a)** was that the central vein was surrounded by hepatic cord of cells, while the liver of high-fat diet treated mouse **(b)** showed patches of liver cell necrosis with macrovesicular and microvesicular steatosis, and massive fatty changes. Whereas, the livers of the animals trested by fenofibrate **(c)** and TFs (E, 50 mg/kg/day) showed the absence of vesicular steatosis to display good hepatoprotective actions. The low-dose of TFs **(d)**, 25 mg/kg/day) treated groups showed less vesicular steatosis.

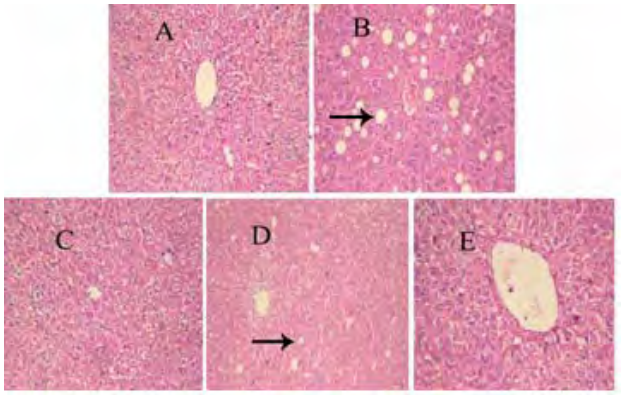


Figure 3. Protective effect of TFs on liver injury induced by high-fat diet in mice. (a) Group I; (b) Group II; (c) Group III; (d) Group IV; (e) Group V. high-fat diet treatment induced vesicular steatosis (black arrows). Hematoxylin and eosin staining; Original magnification, ×100.

Table 1. Effect of TFs on serum lipid profile in experimental animals.

Group	TC (mg/dL)	TG (mg/dL)	HDL-C (mg/dL)	LDL-C (mg/dL)
I	72.63 ± 7.22	53.86 ± 9.19	48.44 ± 5.49	13.40 ± 4.27
II	143.33 ± 9.71[b]	100.1 ± 17.3[b]	47.03 ± 4.84	83.55 ± 14.23[b]
III	87.12 ± 7.31[d]	64.0 ± 8.41[d]	63.09 ± 5.49[c]	16.18 ± 7.06[d]
IV	98.64 ± 15.74[c]	86.18 ± 4.15[c]	64.11 ± 13.49[c]	27.40 ± 16.53[d]
V	78.8 ± 9.80[d]	66.18 ± 4.75[d]	63.84 ± 13.41[c]	18.31 ± 3.40[d]

Values are given as mean ± S.D. (n=10). [a] $p<0.05$ compared with Group I . [b] $p < 0.01$ compared with Group I. [c] $p < 0.05$ compared with Group II. [d] $p < 0.01$ compared with Group II.

Table 2. Effect of TFs on serum liver enzymes in experimental animals.

Group	ALT (IU/L)	AST (IU/L)	ALP (IU/L)
I	22.70 ± 3.40	27.70 ± 5.40	321.92 ± 130.33
II	102.60 ± 24.50[b]	76.50 ± 17.40[b]	495.38 ± 126.02[a]
III	45.30 ± 18.0[d]	71.80 ± 14.86	336.77 ± 71.15[c]
IV	91.00 ± 27.20[c]	60.29 ± 7.90[c]	382.28 ± 83.69[c]
V	45.8 ± 2.20[d]	54.20 ± 17.5[c]	368.16 ± 128.08[c]

Values are given as mean ± S.D. (n=10). [a] $p<0.05$ compared with Group I . [b] $p < 0.01$ compared with Group I. [c] $p < 0.05$ compared with Group II. [d] $p < 0.01$ compared with Group II.

3.5. Effects of TFs on Serum Lipid Profiles

As seen in Table 1, the serum levels of TC, TG, and LDL-C were significantly increased in animals fed with high-fat diet compared with those in Group I. The levels of TC, TG and LDL-C of those animals given fenofibrate or TFs were significantly decreased compared with hyperlipemic animals. High-dose of TFs (50 mg/kg/day) was able to markedly eliminate TC, TG, and LDL-C by 45.02%, 33.86% and 73.68%, respectively, with no difference compared with the positive group. Although there was no statistically significant difference in the serum HDL-C levels between normal and hyperlipemia mice, an increasing tendency was exhibited in the groups administrated with TFs and fenofibrate.

3.6. Effect of TFs on Hepatic Enzymes

The levels of ALT, AST and ALP in serum were performed to evaluate liver function. As can be observed in Table 2, animals fed with high-fat diet exhibited a significantly elevation in serum ALT, AST and ALP compared with normal group, which suggested that the liver was markedly damaged. Compared with hyperlipemic mice, the activities of ALT, AST and ALP were reduced to 55.85%, 29.15% and 25.68%, respectively, of the animals treated by TFs at the dosage of 50 mg/kg/day, which was no difference compared with fenofibate-treated group. The results implied that the TFs executed a protective effect against liver damage induced by high-fat diet.

3.7. Effect of TFs on Hepatic Lipid Peroxidation

Figure 4 showed that the MDA levels of the liver in all experimental groups. The hepatic MDA level of model group was increased significantly compared with normal group ($p < 0.01$). The decline extents of TFs-treated groups (25 and 50 mg/kg/day) were 31.09% ($p < 0.05$) and 53.65% ($p < 0.01$) compared with hyperlipemic mouse. Meanwhile, there was no difference between fenofibrate- and TFs (50 mg/kg/day)-treated groups.

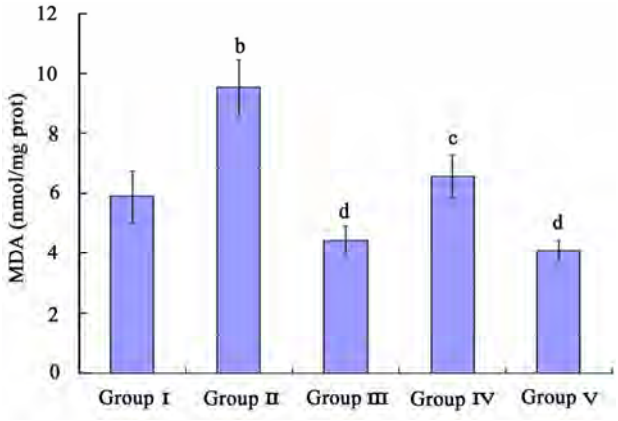


Figure 4. Effect of TFs on hepatic level of MDA in mouse fed with high-fat diet. Values are expressed as mean ± S.D. (n = 10). [b] $p < 0.01$ compared with Group I. [c] $p < 0.05$ compared with Group II. [d] $p < 0.01$ compared with Group II.

3.8. Effect of TFs on Hepatic Antioxidant Enzymes

The variations of CAT, SOD, GSH and GPX in liver among the experimental groups were shown in **Figure 5.** Contrasting with normal animals, the levels of CAT, SOD, GSH and GPX of the liver tissues were significantly decreased after feeding high-fat diet for 4 weeks ($p < 0.05$). Fenofibrate and TFs both could amend the antioxidant status *in vivo*. The improvement capacity of CAT and GPX was TFs (50 mg/kg/day) > fenofibate (50 mg/kg/day) > TFs (25 mg/kg/day). Furthermore, the improvement capacities of SOD and GSH were fenofibate (50 mg/kg/day) > TFs (50 mg/kg/day) > TFs (25 mg/kg/day).

4. DISCUSSIONS

Hyperlipemia is the largest endocrine disease in the world, involving metabolic disorders of carbohydrate, fat and protein. Therefore, it is necessary to search for new drugs that can be used to amendment this metabolic disorder without any side effect. Oxidative stress is currently suggested as a mechanism underlying hyperlipemia, which is one of the major risk factors for coronary artery diseases [21]. In our study, the TFs with high flavonoids showed a good antioxidant activity *in vitro*, which encouraged us to check the antioxidant activity in vivo in hyperlipemic mice.

MDA, one of the lipid peroxidation products, could make the deomosine between the cellulose molecular relax or inhibit protein synthesis. The hepatic MDA level significantly improved in hyperlipemia animals. In the present experiment, the MDA level was significantly decreased in mouse fed with TFs. These findings are in accordance with those of other investigators [22].

Furthermore, Biological antioxidants are natural compounds which can prevent the un-controlled formation of free radicals and activated oxygen species, or inhibit their reaction with biological structures. These compounds include antioxidative enzymes exist in all oxygen-metabolizing cells, such as superoxide dismutase (SOD), catalase (CAT), glutathione peroxidase (GPX) and non-enzymatic antioxidants, such as glutathione (GSH), vitamin C and vitamin E [23]. The main role of the CAT, mainly existing in some cells, is to catalytic the decomposition of hydrogen peroxide [24]. It could prevent hydrogen peroxide to form hydroxyl radical, which is the most harmful radical *in vivo* [25]. SOD mutates the superoxide radicals to form molecular oxygen and H_2O_2. GPX, one of the most important hepatic detoxification

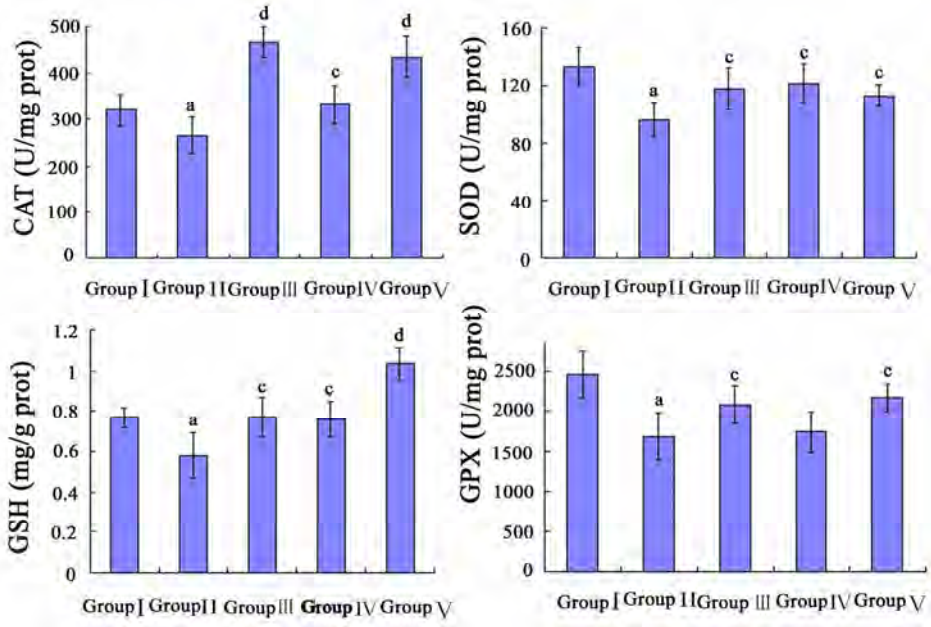


Figure 5. Effect of TFs on hepatic levels of CAT, SOD, GSH and GPX in mouse fed with high-fat diet. Values are expressed as mean ± S.D. (n = 10). [a] $p < 0.05$ compared with Group I . [b] $p < 0.01$ compared with Group I. [c] $p < 0.05$ compared with Group II. [d] $p < 0.01$ compared with Group II.

elements [26], almost presents in the whole biological tissues, especially the liver and RBC. It can scavenge lipid peroxidation induced by hydroxyl radical, postpone the aging of cells and clear the nucleic acid peroxide. Moreover, GPX could remove hydrogen peroxide generating in the tissues. Meanwhile, GSH is the most important biomolecule against chemically induced toxicity and can participate in the elimination of reactive intermediates by reducing hydroperoxides in the presence of GPX [27,28]. GSH also functions as a free radical scavenger and in the repair of free radical-induced biological damages [28]. The decrease in the GSH level represents increased utilization due to oxidative stress [29].

The efficiency of this defense system is apparently weakened in hyperlipemia condition resulting in ineffective scavenging of free radicals and lipid peroxidantion products, which can improve the levels of tissue lipid profiles *in vivo*. The inhibition of CAT, SOD and GPX activities are found to be involved in many degenerative diseases. Likewise, in our study, high-fat diet caused a markedly decrease in SOD, GPX, CAT activities and nonenzymatic antioxidant (GSH) levels in test organs in mice. These adaptive decreases in antioxidant enzyme activities protect the organs against lipid peroxidation mouse [30]. TFs significantly increased hepatic CAT, SOD, GSH and GPX levels. Our data also showed that TFs significantly reduced the hepatic MDA concentration in a dose-dependent manner. These results suggested that, the antioxidative effect of TFs might reduce oxidative stress, resulting in a lower lipid peroxidation in liver. These results indicated that the antioxidative effect of TFs in the liver may only be observed in the presence of severe oxidative stress, suggesting that TFs may act as a chemopreventive agent with inhibition activity on oxidative-induced cell damage.

Elevated blood triglyceride and cholesterol, especially low-density-lipoprotein cholesterol (LDL-C), is a major risk factor in development of cardiovascular disease [31]. It is well-known that cholesterol is a fatty substance which is important to the membrane of cells in the animal body. TC is a measure of the total amount of all cholesterol in blood at a given time and is the sum of HDL-C, LDL-C. TG composed of three fatty acids and glycerol. Liking cholesterol, they circulate in blood, but are stored in body fat and used when body needs extra energy. HDL-C removes excess cholesterol from arteries and moves it to liver for further processing or to be eliminated from the body. The higher serum HDL-C is the better. Therefore, the HDL-C is called ‘‘good’’ cholesterol. It also plays a key role in the protection against oxidative damage of membrane. LDL-C contributes to buildup of fat deposits in the arteries (atherosclerosis), which can cause decreased blood flow and head attack. So it is always called ‘‘bad’’ cholesterol, and a less levels are desirable. The value of AI indicates the deposition of foam cells or plaque or fatty infiltration or lipids in heart, coronaries, aorta, liver and kidney. In this study, TC, TG and LDL-C levels significantly increased in the hyperlipemic animals fed a high-fat diet for 4 weeks, but with no significantly decreased in HDL-C level (Table 1). All these results showed that the mouse hyperlipemic model was established successfully by feeding a high-fat

diet for 4 weeks. After the experiment, the animals administration of fenofibrate and TFs were all exhibited a decrease tendency ($p < 0.05$). As to the data of HDL-C in the experiment, there was also a remarkable increase of the serum HDL-C levels in the mouse fed a high-fat diet for 4 weeks. The increase of 'good' cholesterol does not mean high-fat diet is favorable to serum lipid profile levels, because the increase of 'good' probably does not surpass that of 'bad', which has negative effects on lipid-lowering. In summary, the administration of TFs could decrease the serum levels of TC, TG and LDL-C.

The liver is capable of removing cholesterol from the blood circulation as well as manufacturing cholesterol and secreting cholesterol into the blood circulation and liver damages are generally induced in the condition of hyperlipemia as dramatic increase of serum ALT and AST levels. Furthermore, ALP, one of hepatic enzymes, also reflects the damage induced by high-fat diet. In our test, the damage to liver induced by high-fat diet was also investigated, which included hepatic enzymes and histopathological examination. These results indicated that TFs was characterized by an ameliorating effect on fatty liver.

5. CONCLUSIONS

The present study demonstrated TFs from *R. laevigata* Michx has favorable potency to develop a hypolipidemic and hepatoprotective activities, of which the levels may be mediated, in part, by enhancing the system of antioxidant defense. It can be used as a potential medicine for cardiovascular diseases. The toxicity, clinical application, and chemical constituents of TFs are not clear and need further investigation.

6. ACKONWLEDGEMENT

This research was supported by funds of Liaoning BaiQianWan Talents Program, Liaoning, China.

REFERENCES

[1] Frishman, W.H. (1998) Biologic Markers as Predictors of Cardiovascular Disease. *The American Journal of Medicine*, **104(6A)**, 18S-27S.

[2] Williams, G. and Pickup, J. Ed., (1991) New drugs in the management of diabetes mellitus. Textbook of Diabetes II. Blackwell, Oxford, 977-993.

[3] Kameswara Rao, B., Kesavulu, M.M., Giri, R. and Apparao, Ch. (1999) Antidiabetic and hypolipidemic effects of *Momordica cymbalaria* Hook. Fruit powder in alloxan diabetic rats. *Journal of Ethnopharmacology*, **67(1)**, 103-109.

[4] Sharma, B., Balomajumder, C. and Roy, P. (2008) Hypoglycemic and hypolipidemic effects of flavonoid rich extract from *Eugenia jambolana* seeds on streptozotocin induced diabetic rats. *Food Chemisty Toxicology*, **46(7)**, 2376-2383.

[5] Dini, I., Tenore, G.C. and Dini, A. (2009) Saponins in Ipomoea batatas tubers: Isolation, characterization, quantification and antioxidant properties. *Food Chemistry*, **113(2)**, 411-419.

[6] Ma, M., Liu, G.H., Yu, Z.H., Chen, G. and Zhang, X. (2009) Effect of the *Lycium barbarum* polysaccharides administration on blood lipid metabolism and oxidative stress of mice fed high-fat diet in vivo. *Food Chemistry*, **113(4)**, 872-877.

[7] Wang, J.Q., Li, J., Zou, Y.H., Cheng, W.M., Lu, C., Zhang, L., Ge, J.F., Huang, C., Jin, Y., Lv, X.W., Hu, C.M. and Liu, L.P. (2009) Preventive effects of total flavonoids of *Litsea coreana* leve on hepatic steatosis in rats fed with high fat diet. *Journal of Ethnopharmacology*, **121(1)**, 54-60.

[8] Zhang, T.Y., Nie, L.W., Wu, B.J., Yang, Y., Zhao, S.S., and Jin, T. (2004) Hypolipedemic activity of the polysaccharose from *Rosa laevigata* Michx fruit. *Chinese Journal of Public Health*, **20(7)**, 829-830.

[9] Welton, A.F., Hurley, J. and Will, P. (1988) Flavonoids and arachidonic acid metabolism. *Progress in Clinical and Biological Research*, **280**, 301-312.

[10] Sharma, B., Balomajumder, C. and Roy, P. (2008) Hypoglycemic and hypolipidemic effects of flavonoid rich extract from *Eugenia jambolana* seeds on streptozotocin induced diabetic rats. *Food Chemistry and Toxicology*, **46(7)**, 2376-2383.

[11] Diouf, P.N., Stevanovic, T. and Cloutier, A. (2009) Study on chemical composition, antioxidant and anti-inflammatory activities of hot water extract from *Picea mariana* bark and its proanthocyanidin-rich fractions. *Food Chemistry*, **113(4)**, 897-902.

[12] Xu, S.Y., Bian, R.L. and Chen, X. Ed., (2002) Experimental methodology of pharmacology. People' Medical Publishment House, Beijing, 1202.

[13] China Pharmacopoeia Committee. (2005) Pharmacopoeia of the People's Republic of China, the first division of 2005 edition Ed., 291-292, China Chemical Industry Press. Beijing.

[14] Hsu, B., Coupar, I.M., and Ng, K. (2006) Antioxidant activity of hot water extract from the fruit of the Doum palm, *Hyphaene thebaic*a. *Food Chemistry*, **98(2)**, 317-328.

[15] Tsai, S.Y., Huang, S.J. and Mau, J.L. (2006) Antioxidant properties of hot water extracts from *Agrocybe cylindracea*. *Food Chemistry*, **98(4)**, 670-677.

[16] Lowry, O.H., Rosebrough, N.J. and Far, A.L. (1951) Protein measurement with the Folin phenol reagent. *Journal of Biological Chemistry*, **193(1)**, 265-275.

[17] Friedwald, W.T., Levy, R.J. and Fredricken, D.S. (1972) Estimation of HDL-C in the plasma without the use of preparative ultracentrifuge. *Clinical Chemistry*, **18**, 449.

[18] Fu, B., Liu, J., Li, H., Li, L., Lee, F.S.C. and Wang, X. (2005) The application of macroporous resins in the separation of licorice flavonoids and glycyrrhizic acid. *Journal of Chromatography A*, **1089(1-2)**, 18-24.

[19] Peng, J.Y, Yang, G.J., Fan, G.R. and Wu, Y.T. (2005) Preparative isolation and separation of a novel and two

known flavonoids from *Patrinia villosa* Juss by high-speed counter-current chromatography. *Journal of Chromatography A*, **1092(2)**, 235-240.

[20] Soares, J.R., Dins, T.C, Cunha, A.P. and Ameida, L.M. (1997) Antioxidant activity of some extracts of *Thymus zygis*. *Free Radical Research*, **26**, 469-478.

[21] Fruchart, J.C. and Duriez, P. (1998) High-density lipoproteins and coronary heart disease future prospects in gene therapy. *Biochimie*, **80(2)**, 167-172.

[22] Mahfouz, M.M. and Kummerow, F.A. (2000) Cholesterol-rich diets have different effect on lipid peroxidation, cholesterol oxides, and antioxidant enzymes in rats and rabbits. *The Journal of Nutritional Biochemistry*, **11(5)**, 293-302.

[23] Fridovich, I. (1999) Fundamental aspects of reactive oxygen species, or what's the matter with oxygen? *The New York Academy of Sciences*, **893(1)**, 13-18.

[24] Yao, D.C., Shi, W.B., Gou, Y.L., Zhou, X.R., Tak, Y.A. and Zhou, Y.K. (2005) Fatty acidmediated intracellular iron translocation: A synergistic mechanism of oxidative injury. *Free Radical Biology and Medicine*, **39**, 1385-1398.

[25] Chance, B., Sies, H. and Boveris, A. (1979) Hydroperoxide metabolism in mammalian organs. *Physiological Reviews*, **59(3)**, 527-605.

[26] Shaw, S., Rubin, K. and Lieber, C.S. (1986) Depressed hepatic glutathione and increased diene conjugates in alcoholic liver disease: Evidence of lipid peroxidation. *Digestive Diseases and Sciences*, **28**, 585-589.

[27] Nicotera, P. and Orrenius, S. (1986) Role of thiols in protection against biological reactive intermediates. *Advances in Experimental Medicine and Biology*, **197**, 41-51.

[28] Lin, T. and Yang, M.S. (2007) Benzo [a] pyrene-induced elevation of GSH level protects against oxidative stress and enhances xenobiotic detoxification in human HepG2 cells. *Toxicology*, **235(1-2)**, 1-10.

[29] Anuradha, C.V. and Selvam, R. (1993) Effect of oral methionine on tissue lipid peroxidation and antioxidants in alloxan-induced diabetic rats. *The Journal of Nutritional Biochemistry*, **4(4)**, 212-217.

[30] Padmavathi, R., Senthilnathan, P., Chodon, D. and Sakthisekaran, D. (2006) Therapeutic effect of paclitaxel and propolis on lipid peroxidation and antioxidant system in 7, 12 dimethyl benz (a) anthracene-induced breast cancer in female Sprague Dawley rats. *Life Science*, **78(24)**, 2820-2825.

[31] Anderson, J.W. and Tietyen-Clark, J. (1986) Dietary fiber: Hyperlipidemia, hypertension, and coronary heart disease. *The American Journal of Gastroenterology*, **81(10)**, 907-919.

Vol.2, No.3, 184-189 (2010)
doi:10.4236/ns.2010.23028

Natural Science

A novel analytic potential function applied to neutral diatomic molecules and charged Ions

Chang-Feng Yu, Chang-Jun Zhu, Chong-Hui Zhang, Li-Xun Song, Qiu-Pin Wang

Department of physics, School of Science, Xi'an Polytechnic University, Xi'an, China; yuh55@126.com

Received 24 November 2009; revised 8 January 2010; accepted 30 January 2010.

ABSTRACT

In this paper, a new method on constructing analytical potential energy functions is presented, and from this a analytical potential energy function applied to both neutral diatomic molecules and charged diatomic molecular ions is obtained. This potential energy function includes three dimensionless undetermined parameters which can be determined uniquely by solving linear equations with the experimental spectroscopic parameters of molecules. The solutions of the dimensionless undetermined parameters are real numbers rather than complex numbers, this ensures that the analytical potential energy function has extensive universality. Finally, the potential energy function is examined with four kinds of diatomic molecules or ions—homonuclear neutral diatomic molecule $H_2(X^1\sum_g^+)$, $K_2(B^1\prod_u)$ and $Li_2(B^1\prod_u)$, homonuclear charged diatomic molecular ion $He_2^+(X^2\sum_u^+)$, $N_2^+(X^2\sum_g^+)$ and $O_2^+(X^2\prod_g)$, heternuclear neutral diatomic Molecule $AlBr(A^1\prod)$, $PuO(X^1\sum_g^+)$ and $NaLi(X^1\sum_g^+)$, heternuclear charged diatomic Molecular ion $BC^-(X^3\prod)$, $MgH^+(X^1\sum^+)$ and $HCl^+(X^2\prod_i)$, as a consequence, good results are obtained.

Keywords: Diatomic Molecules And Ions; Potential Energy Function; Force Constants; Spectroscopic Parameters; Phase Factor

1. INTRODUCTION

Analytical potential energy functions are of great significance in the study of material science, molecular spectrum, reaction dynamics of atoms and molecules, vibrational and rotational energy-level structures of molecules, interactions between laser and matter, photoionization *etc*. [1-3] Due to the importance and extensive applications of the potential energy function, the corresponding research works have been carried on all along [4-6]. So far, the representative analytical potential energy function proposed have Morse potential [7], Rydberg potential [8], Murrell-Sorbie potential (M-S) [9] and Huxley-Murrell-Sorbie potential (HMS) [10] *etc*.

Recently, Sun Weiguo *et al* have proposed an energy consistent method (ECM) and constructed a new physically well behaved analytical potential function of a diatomic system called ECM potential [11]. These potential functions above have merits and defects respectively, they are valid in describing the behaviors of some individual or classificatory diatoms and molecules. But none of them can describe both neutral diatomic molecules and charged diatomic molecular ions and describe precisely the behaviors of potential energy function over the whole range of internuclear distance. Seen from expressional forms, most of these potential energy functions adopt the forms of ploynomial and exponential. In this paper, a cosine function with a phase factor is used as basic potential energy function and, through renormalization to the phase factor, a universal potential energy function applied to four kinds of diatomic molecules or ions — homonuclear neutral diatomic molecules, homonuclear charged diatomic molecular ions, heternuclear neutral diatomic molecules and heternuclear charged diatomic molecular ions is given. Finally, the potential energy function is examined with twelve different kinds of diatomic molecules and ions *etc*., as a consequence, good results are obtained.

2. FUNDAMENTAL SUPPOSITIONS, AND DERIVATION OF A UNIVERSAL ANALYTIC POTENTIAL FUNCTION

Suppose that the potential function of diatomic molecular satisfies the following relation

$$V(r) = A\cos\psi(r) + B \tag{1}$$

where $\psi(r) = \varphi + \arccos(\rho / r)$ (2)

where, A, B are undetermined constants, $\psi(r)$ is a

phase factor related to ϕ, ρ and the internuclear distance r, here ϕ is equivalent phase difference between two interacting atoms, ρ is equilibrium internuclear distance. Substituting **Eq.2** into **Eq.1**, yields

$$V(r) = A\cos\left(\varphi + \arccos(\rho / r)\right) + B$$

$$= A\left((\rho / r)\cos\varphi - \sqrt{1-\rho^2/r^2}\sin\varphi\right) + B \quad (3)$$

Eq.3 is a basic analytical potential energy function**.** In order to obtain the universal analytical potential function of diatomic molecules and ions, renormalization should be needed for the term $\sqrt{1-\rho^2/r^2}$ in **Eq.3**, so as to ensure that the derivatives of each order of the **Eq.3** are continuous and finite at equilibrium distance $r = \rho$. Thus we can expand the term into binomial series

$$\sqrt{1-\rho^2/r^2} = -\sum_{i=0}^{\infty}\frac{(2i)!}{4^i(i!)^2(2i-1)}(\frac{\rho}{r})^{2i} \quad (4)$$

Here, **Eq.4** is a infinite series, it need to be truncateed into finite terms and its following infinite terms should be absorbed into three undetermined coefficients a, b, c, so from **Eq.4**, we have

$$\sqrt{1-\rho^2/r^2} \approx -\left\{\sum_{i=0}^{n}\frac{(2i)!}{4^i(i!)^2(2i-1)}(\frac{\rho}{r})^{2i} + a(\frac{\rho}{r})^{2n+2}\right\}$$

$$= -\left\{\sum_{i=0}^{n}H(i)(\frac{\rho}{r})^{2i} + a(\frac{\rho}{r})^{2n+2} + b(\frac{\rho}{r})^{2n+4} + c(\frac{\rho}{r})^{2n+6}\right\} \quad (5)$$

$$\text{where} \quad H(i) = \frac{(2i)!}{4^i(i!)^2(2i-1)} \quad (6)$$

Generally, the potential energy function satisfies asymptotic condition $\lim_{r\to\infty}V(r) = 0$, so from **Eq.3**

$$\text{we have} \quad B = A\sin\phi \quad (7)$$

Substituting **Eq.5** and **Eq.7** into **Eq.3**, and notice $H(0) = -1$, yields

$$V(r) = \frac{A\rho}{r}\cos\phi + A\sin\phi\left\{\sum_{i=1}^{n}H(i)(\frac{\rho}{r})^{2i} + a(\frac{\rho}{r})^{2n+2} + b(\frac{\rho}{r})^{2n+4} + c(\frac{\rho}{r})^{2n+6}\right\} \quad (8)$$

In **Eq.8**, the undetermined constant A can be determined according to the properties of potential energy function. At the equilibrium distance $r = \rho$, the potential value is equal to the negative value of dissociation energy D_e, *i.e.* $V(\rho) = -D_e$, and the first derivatives of $V(r)$ with respect to r is zero.

So from **Eq.8**, we obtain

$$V(\rho) = A\cos\phi + A\sin\phi\left\{\sum_{i=1}^{n}H(i) + a + b + c\right\} = D_e \quad (9)$$

$$\left(\frac{\mathrm{d}V}{\mathrm{d}r}\right)_{r=\rho} = -\frac{A}{\rho}\cos\phi - \frac{A}{\rho}\sin\phi\{\sum_{i=1}^{n}H(i)(2i) + a(2n+2) + b(2n+4) + c(2n+6)\} = 0 \quad (10)$$

From **Eq.9** and **Eq.10**, the solutions of A and $\cos\phi$ can be given as follows

$$A = \frac{D_e}{\sin\phi}\left(\sum_{i=1}^{n}H(i)(2i-1) + a(2n+1) + b(2n+3) + c(2n+5)\right)^{-1} \quad (11)$$

$$\cos\phi = -\sin\phi\{\sum_{i=1}^{n}H(i)(2i) + a(2n+2) + b(2n+4) + c(2n+6)\} \quad (12)$$

Substituting **Eq.11** and **Eq.12** into **Eq.8**, yields

$$V(r) = D_e\left\{\sum_{i=1}^{n}H(i)(\rho/r)^{2i} + a(\rho/r)^{2n+2} + b(\rho/r)^{2n+4} + c(\frac{\rho}{r})^{2n+6} - \left[\sum_{i=1}^{n}H(i)(2i) + a(2n+2) + b(2n+4) + c(2n+6)\right]\frac{\rho}{r}\right\}\left(\sum_{i=1}^{n}H(i)(2i-1) + a(2n+1) + b(2n+3) + c(2n+5)\right)^{-1} \quad (13)$$

Eq.13 is the universal analytical potential energy function that is required. The undetermined parameters a, b, c can be determined with the experimental spectroscopic parameters (ω_e, $\omega_e\chi_e$, α_e B_e) of molecules or fitting method using singlepoint potential energy scanning. When $n = 1, 2, 3$, from **Eq.13**,we have

$$V(r) = \left(\frac{2D_e}{1+6a+10b+14c}\right)\left\{\frac{1}{2}\left(\frac{\rho}{r}\right)^2 + a\left(\frac{\rho}{r}\right)^4 + b\left(\frac{\rho}{r}\right)^6 + c\left(\frac{\rho}{r}\right)^8 - (1+4a+6b+8c)\frac{\rho}{r}\right\}, \quad (n=1) \quad (14)$$

$$V(r) = \left(\frac{8D_e}{7+40a+56b+72c}\right)\left\{\frac{1}{2}\left(\frac{\rho}{r}\right)^2 + \frac{1}{8}\left(\frac{\rho}{r}\right)^4 + a\left(\frac{\rho}{r}\right)^6 + b\left(\frac{\rho}{r}\right)^8 + c\left(\frac{\rho}{r}\right)^{10} - \left(\frac{3}{2}+6a+8b+10c\right)\frac{\rho}{r}\right\} \quad (n=2) \quad (15)$$

$$V(r)=\left(\frac{16D_e}{19+112a+144b+176c}\right)\left\{\frac{1}{2}\left(\frac{\rho}{r}\right)^2+\frac{1}{8}\left(\frac{\rho}{r}\right)^4+\frac{1}{16}\left(\frac{\rho}{r}\right)^6+a\left(\frac{\rho}{r}\right)^8+b\left(\frac{\rho}{r}\right)^{10}+c\left(\frac{\rho}{r}\right)^{12}-(15/8+8a+10b+12c)\frac{\rho}{r}\right\} \quad (n=3) \tag{16}$$

3. USING EXPERIMENTAL SPECTROSCOPIC PARAMETERS TO DETERMINE *a,b,c*

The undetermined parameters a, b, c can be determined with the experimental spectroscopic parameters (ω_e, $\omega_e\chi_e$, α_e B_e) of diatomic molecules or ions. The principle of this method is, according to the relationship between undetermined parameters and force constants, to obtain a, b, c by solving linear equations. From **Eq.13**, the general expression of force constants at the equilibrium internuclear distance can be given as follows

$$f_m=\left(\frac{\mathrm{d}^mV}{\mathrm{d}r^m}\right)_{r=\rho}=\frac{(-1)^mV_0}{\rho^m}\left\{\sum_{i=1}^{n}[H(i)\prod_{j=0}^{m-1}(2i+j)]+a\frac{(2n+1+m)!}{(2n+1)!}+b\frac{(2n+3+m)!}{(2n+3)!}+c\frac{(2n+5+m)!}{(2n+5)!}-[\sum_{i=1}^{n}H(i)(2i)+a(2n+2)+b(2n+4)+c(2n+6)]\ m\ !\right\} \quad (m=2,\ 3,\ 4) \tag{17}$$

where

$$V_0=\frac{D_e}{\sum_{i=1}^{n}H(i)(2i-1)+a(2n+1)+b(2n+3)+c(2n+5)} \tag{18}$$

From **Eq.17** and **Eq.18**, when $n=1,\ 2,\ 3$, the following linear equations can be obtained

$$\begin{cases}\dfrac{1+12a+30b+56c}{1+6a+10b+14c}=X=\dfrac{f_2\rho^2}{2D_e}\\[2ex]\dfrac{2+32a+100b+224c}{1+6a+10b+14c}=Y=-\dfrac{f_3\rho^3}{6D_e}\\[2ex]\dfrac{3+62a+240b+644c}{1+6a+10b+14c}=Z=\dfrac{f_4\rho^4}{24D_e}\end{cases} \tag{19}$$

$(n=1,\ \text{and}\ 62X-25Y+4Z-32\neq 0)$

$$\begin{cases}\dfrac{10+120a+224b+360c}{7+40a+56b+72c}=X=\dfrac{f_2\rho^2}{2D_e}\\[2ex]\dfrac{24+400a+896b+1680c}{7+40a+56b+72c}=Y=-\dfrac{f_3\rho^3}{6D_e}\\[2ex]\dfrac{43+960a+2576b+5640c}{7+40a+56b+72c}=Z=\dfrac{f_4\rho^4}{24D_e}\end{cases} \tag{20}$$

$(n=2,\ \text{and}\ 98X-31Y+4Z-80\neq 0)$

$$\begin{cases}\dfrac{35+448a+720b+1056c}{19+112a+144b+176c}=X=\dfrac{f_2\rho^2}{2D_e}\\[2ex]\dfrac{98+1792a+3360b+5632c}{19+112a+144b+176c}=Y=-\dfrac{f_3\rho^3}{6D_e}\\[2ex]\dfrac{206+5152a+11280b+21648c}{19+112a+144b+176c}=Z=\dfrac{f_4\rho^4}{24D_e}\end{cases} \tag{21}$$

$(n=3,\ \text{and}\ 142X-37Y+4Z-160\neq 0)$

In **Eqs.19-21**, the relationships between force constants and spectroscopic parameters are as follows

$$f_2=4\pi^2\mu\omega_e^2c^2 \tag{22}$$

$$f_3=-\frac{3f_2}{\rho}(1+\frac{\alpha_e\omega_e}{6B_e^2}) \tag{23}$$

$$f_4=\frac{f_2}{\rho^2}\left[15(1+\frac{\alpha_e\omega_e}{6B_e^2})-\frac{8\omega_e\chi_e}{B}\right] \tag{24}$$

The **Eqs.19-21** above are all linear equations, when the conditions of $62X-25Y+4Z-32\neq 0$, $98X-31Y+4Z-80\neq 0$ and $142X-37Y+4Z-160\neq 0$ are satisfied with respect to **Eqs.19-21**, they have unique real number solutions for the undetermined parameters a, b, c. Calculations show that the conditions above are always tenable in general. This ensures that the analytical potential function **Eq.13** has extensive universality, which can describe any of diatomic molecules and ions especially the behaviors of molecules near equilibrium internuclear distance. So far, the most extensively used analytical potential energy function is Murrel-Sorbie (M-S) potential. The undetermined parameters in Murrel-sorbie potential which are determined by experimental spectroscopic parameters have no unique solutions and contain complex number solutions. Thus, the M-S potential is extremely limited in applications to some diatomic molecules and ions. [12]

4. APPLIED EXAMPLES OF THE UNIVERSAL ANALYTICAL POTENTIAL ENERGY FUNCTION

For examining potential energy function **Eq.13**, fifty kinds of neutral diatomic molecules and charged diatomic

Table 1. Experimental spectroscopic parameters of diatomic molecules and ions.

states	ω_e / cm^{-1}	$\omega_e\chi_e$ / cm^{-1}	B_e / cm^{-1}	α_e / cm^{-1}	ρ / nm	D_e / eV	Refs.
$H_2(X^1\Sigma_g^+)$	4401.21	121.34	60.809	3.062	0.0741	4.747	[13]
$K_2(B^1\Pi_u)$	75.00	0.3876	0.04824	0.000235	0.4235	0.514	[14]
$Li_2(B^1\Pi_u)$	270.7	2.9530	0.5770	0.0083	0.2936	0.3700	[15]
$He_2^+(X^2\Sigma_u^+)$	1698.52	35.30	7.211	0.2240	0.1080	2.475	[16]
$N_2^+(X^2\Sigma_g^+)$	2207.20	16.1360	1.9320	0.0200	0.1116	6.341	[14]
$O_2^+(X^2\Pi_g)$	1905.30	16.304	1.6905	0.0189	0.1117	6.7792	[17]
$AlBr(A^1\Pi)$	297.2	6.400	0.1555	0.00216	0.2322	2.400	[14]
$PuO(X^1\Sigma_g^+)$	822.28	2.500	0.3365	0.00146	0.1830	7.3372	[18]
$NaLi(X^1\Sigma_g^+)$	256.80	1.610	0.3960	0.0036	0.2810	0.8570	[13]
$BC^-(X^3\Pi)$	1301.4	9.820	1.418	0.0155	0.1445	5.588	[19]
$MgH^+(X^1\Sigma^+)$	1226.60	16.300	3.321	0.0640	0.16530	2.100	[14]
$HCl^+(X^2\Pi_i)$	2675.4	53.50	9.9463	0.3183	0.13152	4.480	[14]

Table 2. Potential parameters and force constants of diatomic molecules and ions.

states	D_e / eV	ρ / nm	n.	a	b	c	f_2 /10^{2}aJ·nm^{-2}	f_3 /10^{3}aJ·nm^{-3}	f_4 /10^{4}aJ·nm^{-4}
$H_2(X^1\Sigma_g^+)$	4.747	0.0741	1	–0.4615	0.2008	–0.0367	5.752	–37.43	238.7
$K_2(B^1\Pi_u)$	0.514	0.4235	3	–0.64605	0.47875	–0.10628	0.0646	–0.1035	0.045
$Li_2(B^1\Pi_u)$	0.370	0.2936	3	–0.69431	0.57753	–0.14667	0.1463	–0.3177	0.4546
$He_2^+(X^2\Sigma_u^+)$	2.475	0.1080	3	–0.8073	0.6766	–0.1729	3.401	–20.97	101.23
$N_2^+(X^2\Sigma_g^+)$	6.341	0.1116	3	–0.5889	0.44335	–0.10575	20.11	–160.61	1059.1
$O_2^+(X^2\Pi_g)$	6.7792	0.1117	3	–0.55136	0.37972	–0.0829	17.09	–142.33	918.04
$AlBr(A^1\Pi)$	2.400	0.2322	1	–0.4874	0.2759	–0.0577	1.049	–7.357	21.837
$PuO(X^1\Sigma_g^+)$	7.3372	0.183	3	–0.64234	0.49584	–0.11975	5.959	–27.03	98.61
$NaLi(X^1\Sigma_g^+)$	0.857	0.281	3	–0.78369	0.67093	–0.17435	0.2095	–0.4434	0.7012
$BC^-(X^3\Pi)$	5.588	0.1445	3	–0.6900	0.5404	–0.1317	5.677	–31.49	140.50
$MgH^+(X^1\Sigma^+)$	2.10	0.1653	3	–0.74505	0.62333	–0.15926	1.6468	–6.5343	19.547
$HCl^+(X^2\Pi_i)$	4.480	0.13152	3	–0.80085	0.66178	–0.16814	4.3433	–25.61	138.1

Table 3. Potential parameters of Murrel-Sorbie potential of diatomic molecules and ions.

states	a_1 / nm^{-1}	a_2 / nm^{-2}	a_3 / nm^{-3}	ρ / nm	D_e / eV
$H_2(X^1\Sigma_g^+)$	39.601	405.91	3577.1	0.0741	4.747
$K_2(B^1\Pi_u)$	1.227	–38.457	161.65	0.4235	0.514
$Li_2(B^1\Pi_u)$	28.79	291.03	1317.4	0.2936	0.370
$He_2^+(X^2\Sigma_u^+)$	32.363	94.792	584.08	0.108	2.475
$N_2^+(X^2\Sigma_g^+)$	70.966	1528.3	15675	0.1116	6.341
$O_2^+(X^2\Pi_g)$	1.376	68.968	1085.1	0.1117	6.7792
$AlBr(A^1\Pi)$	9.1044	–95.068	2072.2	0.2322	2.400
$PuO(X^1\Sigma_g^+)$	30.377	207.88	804.25	0.183	7.3372
$NaLi(X^1\Sigma_g^+)$	19.865	121.01	329.24	0.281	0.857
$BC^-(X^3\Pi)$	27.880	71.540	634.30	0.1445	5.588
$MgH^+(X^1\Sigma^+)$	33.953	331.66	1450.8	0.1653	2.10
$HCl^+(X^2\Pi_i)$	29.618	150.74	1133.1	0.13152	4.48

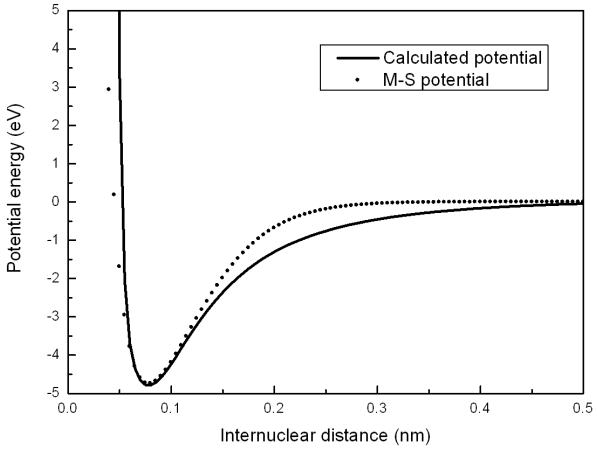


Figure 1. Potential curve of $H_2 - X^1\Sigma_g^+$.

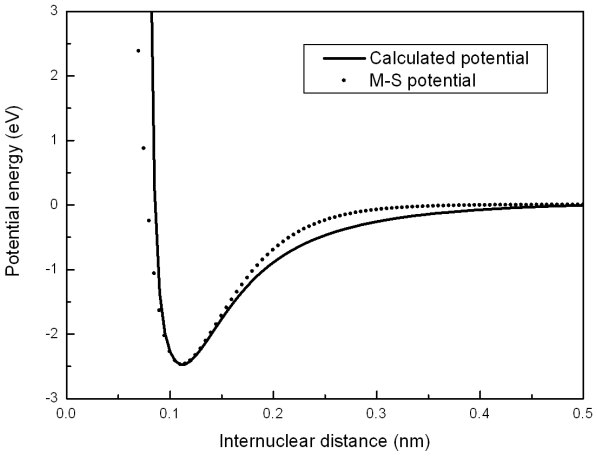


Figure 2. Potential curve of $He_2^+(X^2\Sigma_u^+)$

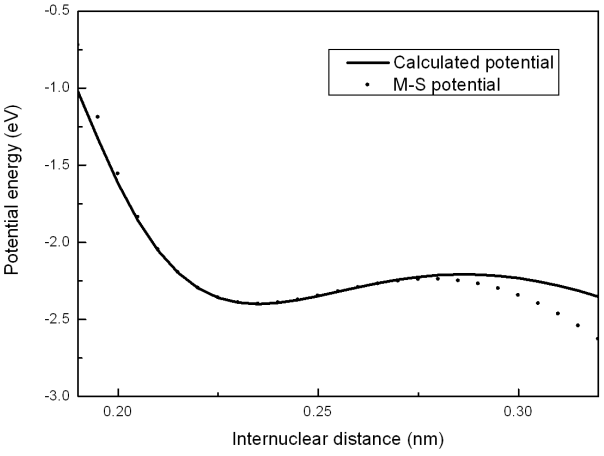


Figure 3. Potential curve of $AlBr - A^1\prod$.

molecular ions have ever been investigated and good results are obtained. Calculations show that two common potential energy curves, *i.e.* steadystate and metastable state

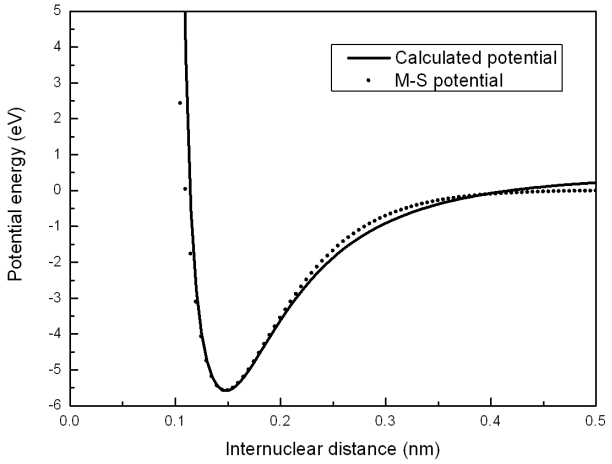


Figure 4. Potential curve of $BC^-(X^3\prod)$.

can be given by using the potential energy function determined with experimental spectroscopic parameters. The experimental spectroscopic parameters of $H_2(X^1\Sigma_g^+)$, $He_2^+(X^2\Sigma_u^+)$, $AlBr(A^1\prod)$ and $BC^-(X^3\prod)$ *etc*. are listed in **Table 1**. According to **Eqs.22-24**, the corresponding force constants can be obtained by using the experimental spectroscopic parameters above, and substituting these force constants into **Eq.19** or **Eq.21**, then the undetermined parameters a, b, c can be calculated by solving the linear equations. The calculation values are listed in **Table 2**. The potential energy curves (to be calculated and plotted by using **Eq.14** and **Eq.16** directly with Origin 7.0 software) plotted by **Eq.14** and **Eq.16** of $H_2(X^1\Sigma_g^+)$, $He_2^+(X^2\Sigma_u^+)$, $AlBr(A^1\prod)$ and $BC^-(X^3\prod)$ are illustrated in **Figures 1-4**. As comparison, in the Figs., the dot lines are the potential curves which are plotted by using the most extensively used Murrel-Sorbie Potential. The M-S potential expression is as follows

$$V(r) = -D_e\left[1 + a_1(r-\rho) + a_2(r-\rho)^2 + a_3(r-\rho)^3\right]\exp\left[-a_1(r-\rho)\right] \quad (25)$$

The relationships between undetermined parameters of M-S potential and force constants are as follows

$$D_e(a_1^2 - 2a_2) = f_2 \quad (26)$$

$$2D_e(3a_1a_2 - 3a_3 - a_1^3) = f_3 \quad (27)$$

$$D_e(3a_1^4 - 12a_1^2a_2 + 24a_1a_3) = f_4 \quad (28)$$

5. CONCLUSIONS

In this paper, we first introduce the phase concept to the studies of analytical potential energy functions and get

good results. This shows that the method of constructing analytical potential energy function by means of phase is effective and reliable. Compared with other potential energy functions, the potential energy function given in this paper has two merits: 1) The undetermined parameter equations determined by experimental spectroscopic parameters are linear equations. Because these linear equations have unique real number solutions, so this potential energy function has a extensive universality; 2) This potential energy function can describe four different kinds of diatomic molecules or ions—homonuclear neutral diatomic molecules, homonuclear charged diatomic molecular ions, heternuclear neutral diatomic molecules and heternuclear charged diatomic molecular ions; In addition, This potential energy function can also describe accurately the behaviors of potential curves over a fairly wide range of internuclear distance.

Potential energy functions of diatomic molecules are the basis to the studies of multi-atomic molecules, ions and clusters, which have extremely significances and applied values in the study of material science, molecular spectrum chemical reaction *etc*. Chemical reaction, molecular collision and many other problems need precise analytical potential energy functions. Thus, the studies of analytical potential energy function will still be important subject in atomic and molecular physics.

REFERENCES

[1] Yiannopoulou, A., Jeung, G.-H., Su, J.P., *et al*. (1999) Potential-energy curves for highly excited electronic states in diatomic molecules related to the atomic orbital undulations [J]. *Physical Review A*, **59(2)**, 1178-1186

[2] Liu, G.Y., Sun, W.G. and Feng, H. (2004) Studies on the analytical potential energy function of diatomic molecular Ion XY^+ using variational method [J]. *Science in China (Series G)*, **47(2)**, 154-164.

[3] Maniero, M. and Acioli, P.H. Full configuration interaction pseudopotential determination of the ground-state potential energy curves of Li_2 and LiH [J]. *International Journal of Quantum Chemistry*, **103(5)**, 711-717.

[4] Xie, R.H. and Gong, J.B. (2005) A Simple Three- parameter model potential for diatomic systems: From weakly and strong to metastable molecular ions [J]. *Physical Review Letters*, **95**, 263202.

[5] Yu, C.f., Yan, K. and Liu, D.Z. (2006) A universal analytic potential-energy function based on a phase factor [J]. *Acta Metallurgica Sinca (English Letters)*, **19(6)**, 455-468.

[6] Esteves, C.S., de Oliveira, H.C.B., Ribeiro, L., *et al*. (2006) Modeling diatomic potential energy curves through the generalized exponential function [J]. *Chemical Physics Letters*, **427(1)**, 10-13.

[7] Morse, P.M. (1929) Diatomic molecules according to the wave mechanics. Ⅱ. Vibrational levels [J]. *Physical Review Letters*, **34**: 57-64.

[8] Rydberg, R. (1931) Graphische Darstellung einiger Bandenspe-ktroskopischer Ergebnisse [J]. *Z Physics*, **73**: 376-385.

[9] Murrel, J.N. and Sorbie, K.S. (1974) New analytic from for the potential energy curves of stable diatomic state [J]. *Journal of the Chemical Society, Faraday Trans, II*, **70**, 1552-1557.

[10] Huxley, P. and Murrel, J.N. (1983) Ground-state diatomic potential [J]. *Journal of the Chemical Society, Faraday Trans Ⅱ*, **79**, 323-328.

[11] Sun, W.G. and Feng, H. (1999) An energy-consistent method for potential energy curves of diatomic molecules [J]. *Journal of Physics B: Atomic, Molecular and Optical Physics*, **32(21)**, 5109-5121.

[12] Zhu, Z.H. and Yu, H.G. (1997) Molecular structure and molecular potential energy function [M]. Science Press, Beijing, 38-123.

[13] Geng, Z.D., Fan, X.W. and Zhang, Y.S. (2006) Structure and potential energy function of the ground state of *XY* (H, Li, Na).[J] *Acta Physica Sinica*, **55(05)**, 2175-2179.

[14] Herzberg, G. (1983) Molecular spectra and molecular structure (i. spectra of diatomic molecules) [M]. Science Press, Beijing, 397-454

[15] Yu, B.H., Shi, D.H., Sun, J.F., Zhu, Z.L., Liu, Y.F. and Yang, X.D. (2007) Ab initio Calculation on Accurate Analytic Potential Energy Functions and Harmonic Frequencies of $c^3\sum_g^+$ and $B^1\prod_u$ States of Dimer 7Li2 [J]. *Chinese Physics*, **16(8)**, 2371-2377.

[16] Haberland, H., Issendorff, B.V., Frochtenicht, R., *et al.* (1995) Absorption Spectroscopy and Photodissociation Dynamics of Small Helium Cluster Ions [J] *Journal of Chemical Physics*, **102(22)**, 8773-8779.

[17] Sarvpreet, K. and Mahajan, C.G. (1999) Wei hua's four-parameter potential comments and computation of molecular constants_ α_e and $\omega_e\chi_e$ [J]. *Pramana journal of physics*, **52(4)**, 409-420.

[18] Gao, T., Wang, H.Y., Yi, Y.-G., Tan, M.-L., Zhu, Z.-H., Sun, Y., Wang, X.-L. and Fu, Y.-B. (1999) Ab intio calculation of the potential energy function and thermodynamic functions for ground state $X^5\Sigma^-$ of PuO [J]. *Acta. Physica Sinica*, **48(12)**, 2222-2227.

[19] Tzeli, D. and Mavridis, A. (2001) Continuing our study on the electronic structure of the carbides BC and AlC [J]. *Journal of Physical Chemistry*, **A105**, 1175-1184.

Vol.2, No.3, 190-196 (2010)
doi:10.4236/ns.2010.23029

Natural Science

Solid polymeric electrolyte of poly(ethylene)oxide-50% epoxidized natural rubber-lithium triflate (PEO-ENR50-$LiCF_3SO_3$)

Siti Aminah bt. Mohd Noor[1], Azizan Ahmad[1], Mohd. Yusri bin Abd. Rahman[2]*, Ibrahim Abu Talib Talib[3]

[1]School of Chemical Science and Food Technology, Faculty of Science and Technology, National University of Malaysia, Selangor, Malaysia; sahabat1711@yahoo.com, azizanukm@yahoo.com
[2]College of Engineering, Universiti Tenaga Nasional, Selangor, Malaysia; yusri@uniten.edu.my
[3]School of Applied Physics, Faculty of Science and Technology, National University of Malaysia, Selangor, Malaysia; ibatal@ukm.my

Received 19 November 2009; revised 28 December 2009; accepted 25 January 2010.

ABSTRACT

A solid polymer electrolyte (SPE) films consisting of polyethylene oxide (PEO), 50% epoxidized natural rubber (ENR50) and $LiCF_3SO_3$ with various compositions of PEO-ENR50 and various weight percentage of $LiCF_3SO_3$ were prepared by solution casting technique. The polymer electrolyte films were characterized using DSC, XRD and AC impedance spectroscopy. The SPE with the PEO-ENR50 composition of 70-30 shows the highest conductivity of 4.2 × 10^{-5} Scm^{-1} at the 15 wt.% of $LiCF_3SO_3$ compared with the other composition of PEO/ENR50. This composition was then chosen to investigate the effect of $LiCF_3SO_3$ on the thermal property, structure and conductivity of the electrolyte. The highest room temperature conductivity of 1.4 × 10^{-4} Scm^{-1} was achieved at 20 wt.% of $LiCF_3SO_3$. The conductivity result is supported by the DSC and XRD analysis which showed the semi- crystalline nature of PEO turning to amorphous state due to the increase in $LiCF_3SO_3$ content.

Keywords: A. Polymers; C. Differential Scanning Calorimetry (DSC); C. X-Ray Diffraction; D. Electrical Conductivity; D. Electrical Properties

1. INTRODUCTION

Main application of polymer electrolytes lies in electrochemical device such as display, sensor, electrochemical window, super-capacitor, rechargeable battery and photoelectrochemical cell. However, the main attention of many solid state researchers is the development of secondary lithium batteries [1-4]. Ionic conducting polymer electrolytes were first suggested by Fenton *et al.* in 1973 [5], who reported that PEO-salt complexes are capable of exhibiting ionic conductivity and subsequently substantial activities were directed towards the development of a wide variety of solid polymer electrolytes using different combinations of polymer and salt. The most reported and extensively studied polymer host for solid polymer electrolyte is PEO [6-12]. This is due to a high electrochemical stability of PEO in comparison with other polyethers, copolymers or PEO branched polymers [9]. Besides, PEO has a good solubility for many salts such as LiTFSI, Li_2SO_4, NH_4ClO_4 and $LiCF_3SO_3$. Even though PEO has a good electrochemical stability, PEO exists in semi-crystalline state at room temperature and does not have an excellent contact with the electrodes.

Solid polymer electrolytes (SPE) for lithium batteries have many advantages over its counterpart liquid electrolyte, such as mechanical stability and processing flexibility. However, the conductivity at room temperature is usually too low to be utilized in ionic device. There are various techniques that can be employed to enhance the conductivity of SPE at room temperature. The recent technique developed for solid polymer electrolytes is the dispersion of nanosize inorganic ceramic filler particles such as Al_2O_3, SiO_2, TiO_2, SnO_2, ZnO and ZrO in the electrolyte systems [6-10]. Propylene carbonate (PC), ethylene carbonate (EC) and dibutyl phthalate (DBP) are also used as plasticizers for conventional polymer electrolyte systems [11-13]. An alternative to ceramic fillers or plasticizers is to employ a complexing polymer such as modified natural rubber. This is due to their distinctive characteristics such as low glass transition temperature, soft elastomeric characteristic at room temperature, good elasticity and adhesion that makes them a suitable candidate as a filler for polymeric electrolyte systems [1,11,12]. **Figure 1** shows the structure of ENR50 dimmer.

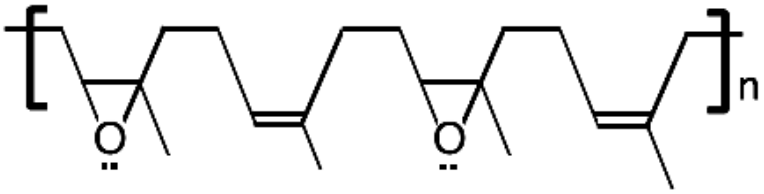


Figure 1. Structure of ENR50 dimmer.

This work describes the effect of ENR50 content on the conductivity of PEO-ENR50-$LiCF_3SO_3$ electrolyte at the fixed concentration of the $LiCF_3SO_3$ salt. The PEO-ENR50 composition that gave the highest room temperature conductivity was chosen to study the effect of the salt content on the conductivity of the electrolyte. The blended polymer of PEO-ENR50 as a polymer host was doped with $LiCF_3SO_3$ using the solution casting technique [14] to obtain a solid polymeric electrolyte film. This work also reports the effect of $LiCF_3SO_3$ content on thermal property and structure of the electrolyte fixed at PEO-ENR50 composition of 70/30.

2. EXPERIMENTAL

2.1. Sample Preparation

The SPE films were prepared by means of the solution-cast technique using a single solvent of THF. The host materials, poly(ethylene oxide) (PEO) (M_W = 600,000) was obtained from Sigma Aldrich, while 50% epoxidized natural rubber (ENR50) (M_W = 592,487) was supplied by Malaysian Rubber Board. The required amounts of PEO and ENR50 were dissolved in tetrahydrofuran (THF) that was obtained from SYSTERM ChemAR. Lithium triflate ($LiCF_3SO_3$), obtained from Fluka was used as the doping salt. All the materials were used without further purification. Subsequently, the desired amounts of PEO-ENR50 and various weight percentage of $LiCF_3SO_3$ were dissolved separately in THF, mixed together and stirred efficiently for 24 hours in order to achieve a homogeneous mixture. The solutions obtained was cast on a teflon mould and was allowed to evaporate completely at room temperature to produce an electrolyte film. Then, the film was dried in a vacuum oven at 60°C for 24 hours to remove the residual solvent. This procedure provided a mechanically stable, free standing and flexible films with the thickness ranging from 150 to 250 μm. The preparation of SPE was carried out in an atmosphere environment at room temperature [12,14].

2.2. Sample Characterizations

The AC impedance measurement was carried out at room temperature using a high frequency response analyzer (HFRA Solartron 1256, Schlumberger) in the frequency range of 100 Hz-1 MHz with 30 mV amplitude. The electrolyte was sandwiched between the stainless steel ion-blocking electrodes with a surface contact area of 2.0 cm^2. The bulk resistance, (R_b) of the electrolyte was determined from the equivalent circuit analysis assisted with Z-View software. The conductivity, (σ) have been calculated from the equation $\sigma = (1/R_b)(t/A)$, where t is the film thickness and A is the active area of the electrode. The thermal property measurement was conducted using a differential scanning calorimeter model 8822^e Mettler Toledo from -60°C and + 150°C at a scanning rate of 10°C min^{-1} under nitrogen atmosphere. Previously, a pure indium and tin were used for temperature and enthalpy calibrations for DSC measurements. Approximately, 1 to 3 mg of the electrolyte film specimen was used for each DSC measurement. The X-ray diffraction technique was performed on the electrolyte films at room temperature using a Siemens model D5000 (Cu-Kα; α = 1.5418 Å). The diffraction angle, 2θ was 10° to 45° using a step size of 0.025° to determine the crystallinity of the electrolyte samples. The Scherrer length was then automatically calculated by evaluation (EVA) software.

3. RESULTS AND DISCUSSION

3.1. Effect of ENR50 Content on Conductivity of PEO- ENR50-$LiCF_3SO_3$ Electrolyte

The freestanding electrolyte films of PEO-ENR50-$LiCF_3SO_3$ have been produced by using solution casting technique. **Table 1** shows the conductivity the electrolyte with various compositions of PEO-ENR50 at the fixed concentration of $LiCF_3SO_3$ (15 wt.%). The percentage by weight of ENR50 was chosen up to 30% since after this level, the electrolyte became a gel electrolyte and was not a free standing film. It made the electrolyte film

Table 1. Room temperature conductivity for various composition of PEO-ENR50 at 15 wt.% $LiCF_3SO_3$.

Composition	Ionic conductivity (Scm^{-1})
100-0	2.1×10^{-6}
90-10	4.1×10^{-6}
80-20	1.0×10^{-5}
70-30	4.2×10^{-5}

was difficult to peel off from the mould. From **Table 1**, the conductivity of the electrolyte without ENR50 was 2.1×10^{-6} Scm^{-1}. The conductivity was found to increase with the increasing percentage of ENR50 in the polymer host. The highest room temperature conductivity of 4.2×10^{-5} Scm^{-1} was achieved at 70-30 composition of PEO- ENR50. This observation shows that the ENR50 content affected the conductivity of the electrolyte. It shows that the elastomeric characteristic of ENR50 reduced the glassy nature of PEO. This condition will enhance the segmental motion of the polymer host. The effect of segmental flexibility is responsible for the formation of free volume which enhances the ionic conductivity. The strong adhesive property can give an efficient contact between the electrodes in electrochemical contacts [12]. This behavior can be seen in the electrolyte system which is due to a good elasticity and adhesion property of ENR50. Razali *et al.* [12] reported that ENR50-LiTf with 100% wt. of EC/PC gave ionic conductivity of 4.0×10^{-5} Scm^{-1} at room temperature. This conductivity value was almost similar with our electrolyte system for 70-30 composition without any plasticizer. It shows that with the absence of a costly plasticizer, the conductivity of the electrolyte has been improved by blending PEO with ENR50. In addition, this electrolyte system produced a low cost and an environmental friendly electrolyte. Besides, the use of PC and EC can corrode the lithium metal electrode in electrochemical cell [16]. Glasse *et al.* [11] reported the incorporation of PEO with ENR50 can reduce the sticky property of ENR50 and made the film could be peeled off easily from the mould.

3.2. Effect of $LiCF_3SO_3$ Salt on Thermal Property, Structure and Conductivity of PEO-ENR50-$LiCF_3SO_3$ Electrolyte

The effect of $LiCF_3SO_3$ salt on the above properties of a PEO-ENR50-$LiCF_3SO_3$ electrolyte was investigated by choosing the electrolyte with the 70-30 composition of PEO-ENR50. This is because this composition has shown the best performance in term of ionic conductivity as compared with the other compositions of PEO-ENR50. The electrolytes were prepared by varying the percentage by weight of $LiCF_3SO_3$ from 5% to 25% at 5% interval. In order to investigate the effect of the salt content on the thermal property of the electrolyte such as melting point, DSC measurement was performed on the prepared electrolyte samples. **Figure 2** shows the DSC thermographs for all samples. The sharp endothermic peak observed at 68°C, corresponds to the crystalline melting temperature (T_m) of the pure PEO [15]. The endothermic peak for pure PEO showed the transition from 68°C to 60°C by addition of 5 wt.% of the salt. The T_m value decreases dramatically to 49°C with addition of 20 wt.% of the salt. This observation shows the reduction in T_m value by addition of the salt. The endothermic curves also indicate a reduction of PEO crystallinity. The relative percentage of crystallinity (λ) of PEO has been calculated by using the relation, $\lambda = (\Delta H_m/\Delta H_m^o) \times 100\%$, where, ΔH_m is the melting enthalpy estimated experimentally and ΔH_m^o used as referenced is the melting enthalpy for 100% crystalline PEO (213.7 Jg^{-1}) [16]. The calculated values of λ are summarized in **Table 2.** The crystallinity degree of the electrolyte decreases with the wt.% of the salt which causes an increase in the amorphous phase. The polymeric chain in the amorphous phase is more flexible, which results in the enhancement of segmental motion of the polymer [17]. The T_m and λ values obtained from this work for polymer electrolyte based on PEO closely agree with the values reported in literature [8,18]. The reduction of T_m and λ suggest that the interaction between the polymer host backbone and $LiCF_3SO_3$ affects the dynamic main chain of the polymer. This will promote the amorphous phase which is ex-

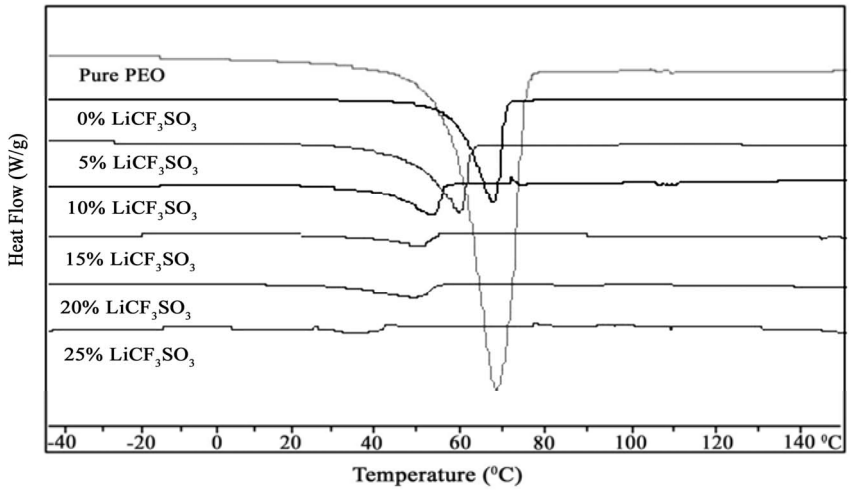


Figure 2. DSC thermograms of PEO,PEO-ENR50 (70-30) with different wt.% $LiCF_3SO_3$.

Table 2. O/Li mole ratio, crystallinity, (λ) and melting temperature, (T_m) of PEO-ENR50-$LiCF_3SO_3$ electrolyte.

$LiCF_3SO_3$ (wt%) to PEO/ENR50 (70/30)	O/Li	T_m of PEO (^oC)	Relative percentage of crystalline phase (λ) (%)
Pure PEO	-	68	78.5
Pure ENR	-	-	
0%	0/1	67	49.9
5%	16/1	60	35.9
10%	8/1	53	31.6
15%	5/1	50	24.5
20%	4/1	49	19.1
25%	3/1	-	-

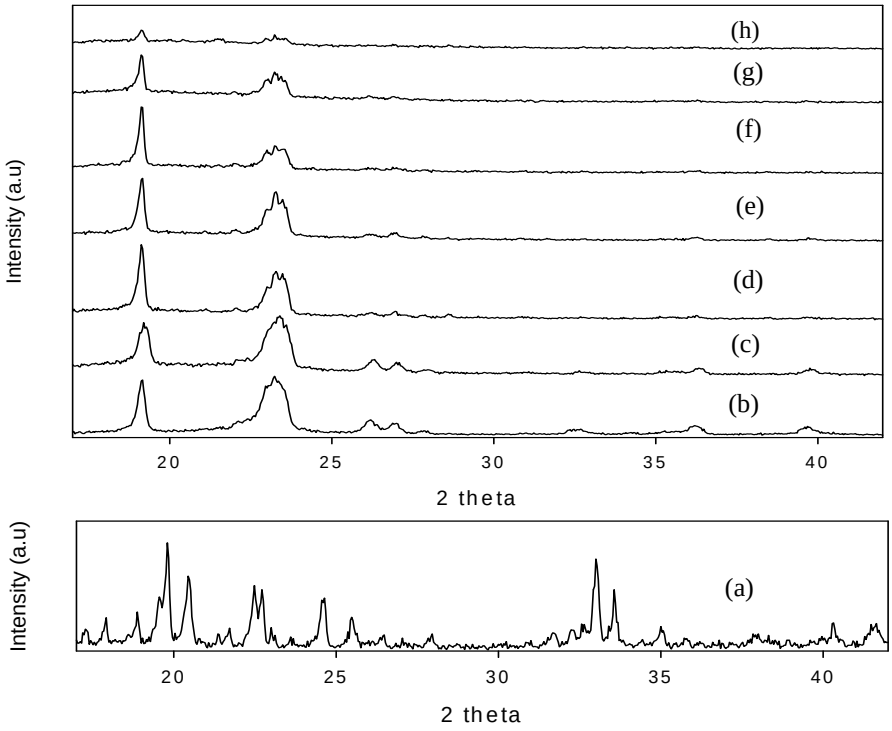


Figure 3. XRD patterns of PEO-ENR50 (70/30) with $LiCF_3SO_3$ with different weight ratios (a) Pure salt (b) Pure PEO (c) 0% (d) 5% (e) 10% (f) 15% (g) 20% (h) 25%.

pected to favour ion transport, thus enhancing the electrolyte conductivity [2]. DSC investigation confirmed that PEO crystallinity is reduced with the addition of ENR50 and the salt which has a plasticizing effect on the polymeric chain.

The structure of the electrolyte samples was investigated by XRD analysis. The XRD pattern of the electrolyte samples with various concentrations of $LiCF_3SO_3$, and that for pure PEO and pure $LiCF_3SO_3$ salt are shown in **Figure 3**. The indexing of the two PEO peaks are reported to be (120) at 19.5^o and (032) + (112) at 23.5^o of the monoclinic system [19]. The characteristic diffraction peaks of crystalline PEO are apparent between 2θ = 15^o – 40^o [3,20]. The two prominent peaks for PEO at 2θ = 19.2^o and 23.2^o are present in all patterns, indicating the presence of pure polymer. Nevertheless, the patterns show that the intensities of the prominent peaks decrease and become broader with the increasing wt.% of $LiCF_3SO_3$. The broadening of sharp peaks of PEO could be due to the disruption of the PEO crystalline structure by $LiCF_3SO_3$ and indicates that the complexation has taken place in the amorphous phase [10]. These results confirm both the decrease in degree of crystallinity and the lamellae size of PEO with the presence of plasticizing $LiCF_3SO_3$ [15]. The $LiCF_3SO_3$ peaks are not seen in any of the polymer-salt complexes and this signifies that $LiCF_3SO_3$ solvates very well in PEO-ENR50 matrix, resulting in the absence of pure salt phase in the complexes. The main XRD peaks of PEO are characterized by significant changes in the full width at half maximum (FWHM) with the increasing salt concentration. The peak broadening estimated in terms of FWHM has been used to evaluate the Scherrer length (*l)*

of main PEO peak which was 22.90, 18.20, 18.12, 18.07, 15.84, 12.51 and 8.14 nm for the electrolytes shown in **Figure 3(b)-3(h)**, respectively. The length was automatically calculated using evaluation (EVA) software with the formula (l) = [0.9λ]/[B$\cos\theta_B$] where λ is the wavelength of the X-rays used and B is FWHM in radians [21]. The l of main PEO peak was 22.9 nm. This value was decreased to 18.20 nm when PEO was blended with ENR50 and become 8.14 nm at 20 wt.% of $LiCF_3SO_3$. The Scherrer length characterizes the crystallite size of PEO and it gives a picture of changes in the PEO crystallinity as a polymer host upon the salt addition. These results suggested that the degree of crystallinity of PEO decreases with the $LiCF_3SO_3$ content.

The ionic conductivity measurement was carried out with the aim to observe the effect of $LiCF_3SO_3$ addition on the ionic conductivity of the electrolyte. The AC spectra for two of the PEO-ENR50-$LiCF_3S0_3$ electrolyte films are shown in **Figure 4** in form of Cole-Cole plot. The high frequency semicircle gives the information about the bulk properties of the electrolyte, such as bulk resistance (R_b) and bulk capacitance (C_b) which arises from the migration of lithium ions and the dielectric polarization of the electrolyte film, respectively. In the low frequency response region, the appearance of a non vertical spike is attributed to the additional capacitance and resistance, arising from dielectric relaxation and ion trapping in PEO-ENR50-$LiCF_3SO_3$ electrolyte [13]. At room temperature, the conductivity of the electrolyte without $LiCF_3SO_3$ is 1.5 × 10^{-9} Scm^{-1}. The ionic conductivity increases rapidly by two orders of magnitude with addition of 5 wt.% of $LiCF_3SO_3$ as Li^+ charge carriers were added to the system. It was observed that the conductivity increases gradually as the wt.% of $LiCF_3SO_3$ increases to 20 wt.% of $LiCF_3SO_3$ and began to decrease at 25 wt.% of $LiCF3SO_3$. The maxi

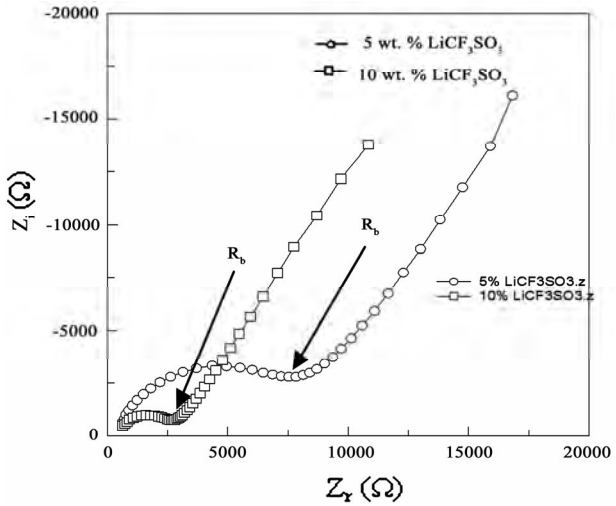


Figure 4. Cole-cole plot for two PEO-ENR50 (70/30)-$LiCF_3SO_3$ electrolytes.

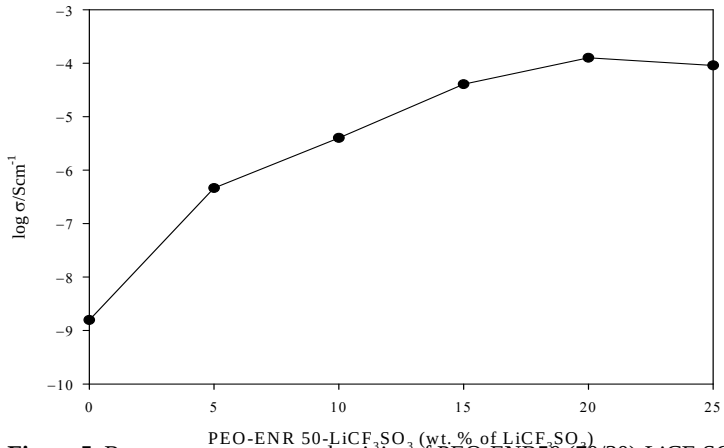


Figure 5. Room temperature conductivity of PEO-ENR50 (70/30)-$LiCF_3SO_3$ electrolyte a function of $LiCF_3SO_3$ concentration.

mum conductivity of 1.4 × 10-4 Scm^{-1} was achieved at 20 wt.% of LiCF3SO_3 as presented in **Figure 5**. An additional factor that favoured the salt dissociation was that the salt with a large monovalent anions such as $LiCF_3SO_3$ has a low lattice energy and as a result, it will easily dissolve in polyether [15]. The conductivity is reduced after the maximum value due to an ever-increasing number of transient crosslink in the system, thus reducing the chain mobility. Besides, the formation of immobile aggregated species will contribute to the fall in conductivity.

The conductivity increases with the lithium salt concentration due to the increase in charge carrier density and also due to the plasticizing affects of the $LiCF_3SO_3$ which decreases PEO crystallinity, thus increasing the pathway for the polymer segmental motion [21]. Ramesh *et al.* [14] reported that the highest conductivity achieved for PEO-$LiCF_3SO_3$ electrolyte was 1.1 × 10^{-6} Scm^{-1} at room temperature. However, the result is slightly lower than the ionic conductivity obtained from this work. In conclusion, the electrolyte based on PEO blended with modified natural rubber possess higher ionic conductivity than conventional PEO-based electrolytes at ambient temperature [11,12]. This result is due to the molecular structure of ENR50. In term of molecular structure of the polymer, ENR50 has active oxygen in the epoxy group attached to the main chain as shown in **Figure 1**. It is assumed that the ENR50 oxygen atom takes a role similar to the ether group in the PEO polymer structure and provide co-ordination sites for Li^+ ion conduction [1,11,12]. Glasse *et al.* [11] reported a gel electrolyte of PEO-ENR-LiTf-EC/PC and found the highest conductivity around 10^{-4} Scm^{-1} at the fixed salt concentration. This value was also in the range of the electrolyte with 20 wt.% of $LiCF_3SO_3$. It shows that the conductivity was influenced by the salt concentration. However, the excessive of salt will increase the transient crosslink in the electrolyte, thus reducing its chain mobility. This observation also shows without using the EC/PC, high room temperature conductivity can be achieved for solid electrolyte system by increasing the wt.% of salt. The increase in ionic conductivity with the wt.% of $LiCF_3SO_3$ clearly suggests that the major contribution to the conductivity enhancement is associated with structural modification. This can be observed by the increase in amorphous phase with Li salt content as observed from the DSC and XRD results. According to Chandra & Chandra [20], there is a correlation between Scherrer length and conductivity. As the crystallite size decreases, the ionic conductivity increases. Hence, it can be proven that the conductivity of our electrolyte is greatly influenced by the wt.% of $LiCF_3SO_3$ through a consequential change in the crystallite size of the host matrix.

4. CONCLUSIONS

A solid polymeric electrolyte of PEO-ENR50-$LiCF_3SO_3$ has been successfully prepared by solution casting technique. The effect of ENR50 content on the conductivity of the electrolyte was found to increase with the weight percentage of ENR50. The highest room temperature conductivity of 4.2 × 10^{-5} Scm^{-1} was achieved at 70-30 composition of PEO-ENR50 host and at 15 wt.% $LiCF_3SO_3$ salt. The effect of the salt content on the thermal property, structure and conductivity of the electrolyte was investigated for the electrolyte with 70-30 composition of PEO-ENR50. The maximum room temperature conductivity was 1.4 × 10^{-4} Scm^{-1} corresponding to 20 wt.% of $LiCF_3SO_3$. This result was supported by the lowest relative percentage of crystalline phase and the decrease in crystalline melting temperature of the electrolyte. Also, XRD analysis showed that the semi-crystalline of the electrolyte became more amorphous with the increase in $LiCF_3SO_3$ content.

5. ACKNOWLEDGEMENT

The authors would like to thank UKM and MOSTI for the provision of grant 03-01-02-SF0423.

REFERENCES

[1] Latif, F., Madzlan, A., Nasir, K., Abd, M.M.A. and Muhd, Z.Y. (2006) The role and impact of rubber in poly(methyl metacrylate)/lithium triflate. *Journal of Power Sources*, **159**, 1401-1404.

[2] Dissanayake, M.A.K.L., Bandara, L.R.A.K., Karaliyadda, L.H., Jayathilaka, P.A.R.D. and Bokalawala, R.S.P. (2006) Thermal and electrical properties for SPE $PEO_9Mg(ClO_4)_2$ incoprating nano porous Al_2O_3 filler. *Solid State Ionics*, **177**, 343-346.

[3] Reddy, V.S.C., Wu, G.P., Zhao, C.X., Jin, W., Zhu, Q.Y., Chen, W. and Sun-il, M. (2007) Mesoporous silica (MCM-41) effect on PEO+$LiAsF_6$ solid polymer electrolyte. *Current Applied Physics*, **7**, 655-661.

[4] Scrosati, B., Croce, F., and Panero, S. (2001) Progress in lithium polymer battery R&D. *Journal of Power Sources*, **100**, 93-100.

[5] Fenton, D.E., Parker, J.M. and Wright, P.V. (1973) Complexes of alkali metal ions with poly (ethylene oxide). *Polymer*, **14**, 589.

[6] Kumar, B., Rodrigues, S.J. and Koka, S. (2002) The crystalline to amorphous transition in PEO-based composite electrolytes: Role of lithium salts. *Electrochimica Acta*, **47**, 4125-4131.

[7] Natesan, B., Karan, N.K., Rivera, M.B., Aliev, F.M. and Katiyar, R.S. (2006) Segmental relaxation and ion transport in polymer electrolyte films by dielectric spectroscopy. *Journal of Non-Crystalline Solids*, **352**, 5205-5209.

[8] Fan, L., Dang, Z., Wei, G., Ce-Wen, N. and Li, M. (2003) Effect of nanosized ZnO on the electrical properties of $(PEO)_{16}LiClO_4$ electrolytes. *Material Science and Engineering B*, **99**, 340-343.

[9] Mohapatra, S.R., Thakur, A.K. and Choudhary, R.N.P., (2008) Studies on PEO-based sodium ion conducting

composite polymer films. *Ionics*, **14**, 255-262.

[10] Xiong, H.M., Zhao, K.K., Zhao, X., Wang, Y.W. and Chen, J.S. (2003) Elucidating the conductivity enhancement effect of nano sized SnO_2 fillers in hybrid polymer electrolyte PEO-SnO_2-$LiClO_4$. *Solid State Ionics*, **159**, 89-95.

[11] Glasse, M.D., Idris, R., Latham, R.J., Linford, R.G. and Schlindwein, W.S. (2002) Polymer electrolytes based on modified natural rubber. *Solid State Ionics*, **147**, 289-294.

[12] Idris, R., Glasse, M.D., Latham, R.J., Linford, R.G. and Schlindwein, W.S. (2001) Polymer electrolytes based on modified natural rubber for use in rechargeable lithium batteries. *Journal of Power Sources*, **94**, 206-211.

[13] Benedict, T.J. Banumathi, S. Veluchamy, A. Gangadharan, R. Zulfihar, A.A. and Rajendran, S. (1998) Characterization of plastisized solid polymer electrolyte by XRD and AC Impedance methods. *Journal of Power Sources*, **75**, 171-174.

[14] Ramesh, S., Tai, F.Y. and Chia, J.S. (2008) Conductivity and FTIR studies on PEO-LiX [X: $CF_3SO_3^-$, SO_4^{2-}] polymer electrolytes. *Spectrocimica Acta*: *Part A*, **69**, 670-675.

[15] Pitawala, H.M.J.C., Dissanayake, M.A.K.L. and Seneviratne, V.A. (2007) Combined effect of Al_2O_3 nano-fillers and EC plasticizer on ionic conductivity enhancement in solid polymer electrolyte $(PEO)_9$ LiTf. *Solid State Ionics*, **178**, 885-888.

[16] Ali, A.M.M., Subban, R.H.Y., Bahron, H., Winnie, T., Latif, F. and Yahya, M.Z. (2008) Grafted natural rubber based polymer electrolytes: ATR-FTIR and conductivity studies. *Ionics*, **14**, 491-500.

[17] Lu, G., Li, Z.F., Li, S.D. and Xie, J. (2001) Blends of natural rubber latex and methyl methacrylate-grafted rubber latex. *Journal of Polymer Science*, **85**, 1736-1741.

[18] Mihaylova, M.D., Krestev, V.P., Kresteva, M.N., Amzill, A. and Berlinova, I.V. (2001) Amphiphilic graft copolymers with poly(oxy ethylene) side chains: Supermolecular structure in solid state I. WAXS studies, *European Polymer Journal*, **37**, 233-239.

[19] Chu, P.P., Reddy, M.J. and Kao, H.M. (2003) Novel composite polymer electrolyte comprising mesoporous structured SiO_2 and PEO/Li. *Solid State Ionics*, **156**, 141-153.

[20] Chandra, A. and Chandra, S. (1994) Effect of alumina dispersal on the conductivity and crystallite size of polymer electrolyte. *Proceeding of the 4th Asian Conference on Solid State Ionics*, Kuala Lumpur,Malaysia, 2-6 August, 355-359.

[21] Ahn, J.H., Wang, G.X., Liu, H.K. and Duo, S.X. (2003) Nanoparticle-dispersed PEO polymer electrolytes for Li batteries. *Journal of Power Sources*, **119**, 422-426.

Vol.2, No.3, 197-200 (2010)
doi:10.4236/ns.2010.23030

Natural Science

Effect of anneal temperature on electrical and optical properties of SnS:Ag thin films

Hong-Jie Jia[1], Shu-Ying Cheng[1,3], Xin-Kun Wu[1], Yong-Li Yang[2]

[1]College of Physics and Information Engineering, and Institute of Micro-Nano Devices and Solar Cells, Fuzhou University, Fuzhou, China
[2]Ministry of Education Key Laboratory of Analysis and Detection Technology for Food Safety, and Department of Chemistry, Fuzhou University, Fuzhou, China
[3]Corresponding author at: College of Physics and Information Engineering, Fuzhou University, Fuzhou, China; sycheng@fzu.edu.cn

Received 14 September 2009; revised 13 January 2010; accepted 25 January 2010.

ABSTRACT

SnS and Ag films were deposited on glass substrates by vacuum thermal evaporation technique successively, and then the films were annealed at different temperatures (0-300℃) in N_2 atmosphere for 2h in order to obtain silver-doped SnS (SnS:Ag) films. The phases of SnS:Ag films were analyzed by X-ray diffraction (XRD) system, which indicated that the films were polycrystalline SnS with orthogonal structure, and the crystallites in the films were exclusively oriented along the（111）direction. With the increase of the annealing temperature, the carrier concentration and mobility of the films first rose and then dropped, whereas their resistivity and direct band gap E_g showed the contrary trend. At the annealing temperature of 260℃, the SnS:Ag films had the best properties: the direct bandgap was 1.3 eV, the carrier concentration was up to 1.132 × 10^{17} cm^{-3}, and the resistivity was about 3.1 Ωcm.

Keywords: Sns:Ag Films; Thermal Evaporation; Annealing; Electrical And Optical Properties

1. INTRODUCTION

Emphasis on cost-competitive photovoltaic cells has been involved in the study of low-cost, non-toxic materials. SnS could be of interest for photovoltaic cells, since its optical energy gap of 1.3 eV [1,2] is close to the optimum energy gap 1.5 eV of solar cells, and it has a high absorption coefficient (> 10^4 cm^{-1}) and a high conversion efficiency of about 25% [3,4]. In addition, the constituent elements Sn and S are non-toxic and abundant in nature.

In spite of the above advantages, the electrical properties of SnS thin films still need to be improved in order to make good SnS thin film solar cells. W. Albers *et al*. [5] investigated Sb- and Ag-doped SnS single crystals and observed n-type conductivity with carrier concentration of ~10^{19} cm^{-3} in Sb-doped SnS crystals and p-type conductivity with ~10^{18} cm^{-3} concentration in Ag-doped SnS crystals. Devika *et al.* [6] investigated Ag-doped SnS films grown by thermal evaporation technique and observed that the resistivity of the SnS layers reached a minimum value of 6.98 Ω cm at 15 atom% of Ag. However, some properties of SnS:Ag thin films are unclear and waiting for further study. In particular, how the annealing temperatures affect the microstructure and physical properties of SnS:Ag thin films prepared by vacuum thermal evaporation technique has not been investigated. Therefore, in this paper we investigate the influence of annealing temperatures on the films in order to improve optical and electrical properties of SnS:Ag thin films.

2. EXPERIMENTAL

SnS and Ag films were deposited on glass substrates by vacuum thermal evaporation technique successively. The thermal evaporation system is DMDE-450 deposition equipment (made in China). The SnS powder with 99.5% purity and the Ag grains with 99.9% purity were used as source materials and were loaded onto a ceramic crucible and a molybdenum boat, respectively. The chamber was evacuated down to 5.6 × 10^{-3} Pa. The source-to-substrate distance was about 10 cm. The as-prepared films were annealed at different temperatures of 210, 260 and 300℃ in N_2 atmosphere for 2h respectively so that Ag-doped SnS films can be obtained.

The structure of the films was characterized by a Philips X'Pert-MPD X-ray diffraction (XRD) system with a Cu K radiation source. The surface roughness was analyzed by a CSPM5000 Scanning Probe Microscopy (AFM), and the thickness of the films was measured by

a Veeco Dektak 6M stylus profiler. The transmission and reflectance spectra were carried out with a Varian Cary 500 UV−VIS−NIR spectrophotometer in the range 400 – 1600 nm. Based on Van der Pauw method, the electrical properties were determined by a HMS-3000 Hall measurement system.

3. RESULTS AND DISCUSSION

3.1. Structural Analysis

Figure 1 (a)-(d) show the XRD patterns of the samples as-prepared and annealed at different temperatures (samples S1-S4 corresponding to unannealing, 210, 260 and 300℃), respectively. It can be seen from **Figure 1 (a)-(c)** that all the diffraction peaks are attributed to SnS (JCPDS39-354) phase with orthorhombic structure. It indicates that the films are polycrystalline SnS with a strong {111} preferred orientation. With the increase of the annealing temperature, the main diffraction peaks of the samples become stronger. Sample S3 exhibits the best crystallization. But the crystallization of sample S4 becomes weaker and there is an obvious SnO_2 peak at 2θ = 33.17°. Maybe a partial of SnS was oxidized into tin dioxide (SnO_2) at the higher annealing temperature because of a low vacuum annealing condition. Therefore, when the annealing temperature is higher, there is probably a SnO_2 phase in the SnS:Ag films.

3.2. Optical Properties

Figure 2 shows absorption coefficient vs photon energy ($h\nu$) curves of the SnS:Ag thin films at different annealing temperatures. As a whole, with the increasing of photon energy, the absorption coefficient increases rapidly and then almost stabilizes at $h\nu > 2.5$ eV. The maximum absorption coefficients of all the films are greater than $1.3 \times 10^5\ cm^{-1}$. With the increase of the annealing temperature, the absorption coefficient increases correspondingly. Because the improved crystallization and uniformity, and the reduced defect density in the films will make the scattered light loss decrease, thus leading to the increase of the absorption coefficient. However, when the annealing temperature is equal to or greater than 300℃, the absorption coefficient decreases probably due to the presence of SnO_2 in the films.

Figure 3 shows a curve of $(\alpha h\nu)^2$ vs. $h\nu$ for sample S3 and the estimated *E*g value of 1.30 eV (here only show sample S3 for simplicity). The *Eg* values of all the four samples are shown in **Table 1**. With the increase of the annealing temperature, the E_g first drops and then rises. Because, with the increasing of the annealing temperature, Ag atoms are easier to be diffused and doped in the polycrystalline SnS films, and the doped-Ag can drop the band gap [6]. But the *Eg* value of sample S4 is larger than that of sample S3, this is perhaps due to the presence of SnO_2 (3.4 ~ 4.6eV) [7] in sample S4.

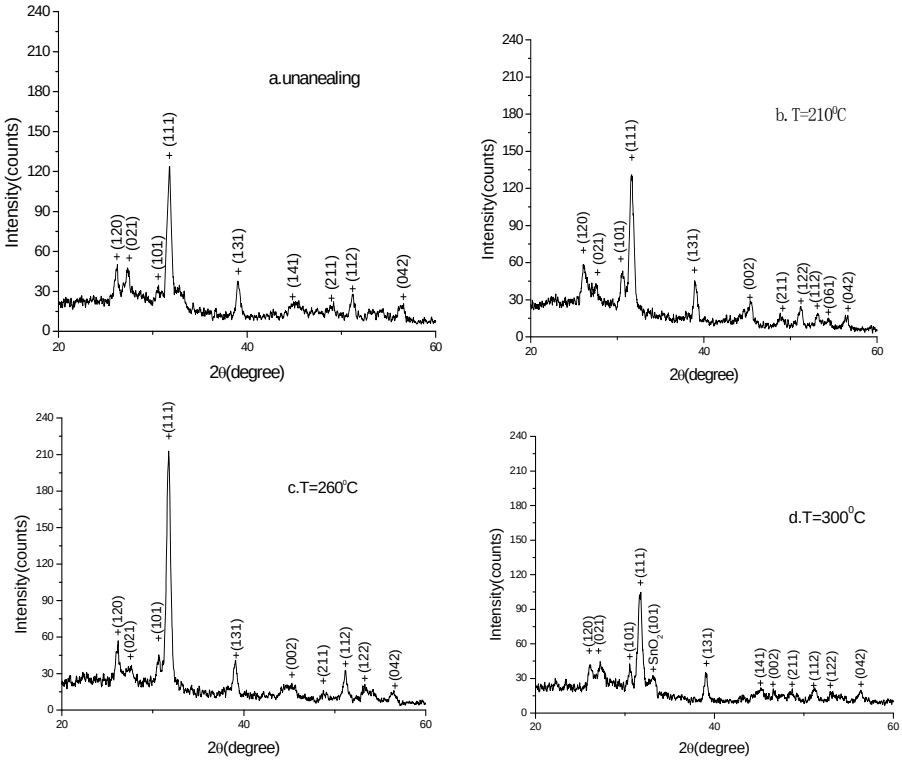


Figure 1. XRD patterns of SnS:Ag thin films at different annealing temperatures.

Table 1. The bandgap and absorption edge of SnS:Ag thin films at different annealing temperatures.

Samples	Annealing Temperature/℃	Band gap Eg /eV	Absorption Edge /nm
S1	unannealed	1.34	925.4
S2	210	1.32	939.4
S3	260	1.30	953.8
S4	300	1.52	815.8

Table 2. Hall measurement results for SnS:Ag thin films at different annealing temperatures.

Samples	temperature/℃	Bulk concentration/cm^{-3}	Mobility/ ($cm^2.v^{-1}.s^{-1}$)	Resistivity/ (Ωcm)	Average Hall Coefficient/(m^2c^{-1})
S1	unannealed	7.242×10^{14}	14.3	601.1	8.631×10^{3}
S2	210	3.212×10^{16}	16.3	11.9	1.945×10^{2}
S3	260	1.132×10^{17}	17.8	3.1	5.522×10^{1}
S4	300	2.601×10^{16}	15.7	15.3	2.402×10^{2}

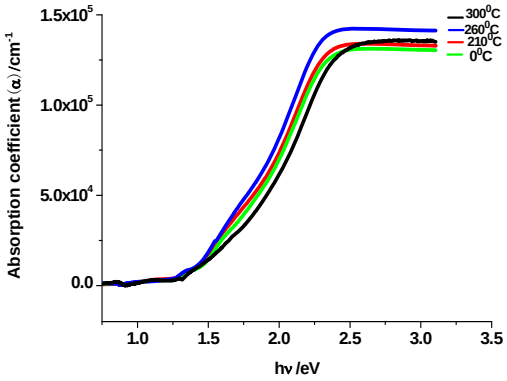


Figure 2. α vs. $h\nu$ curves of SnS:Ag thin films at different annealing temperatures.

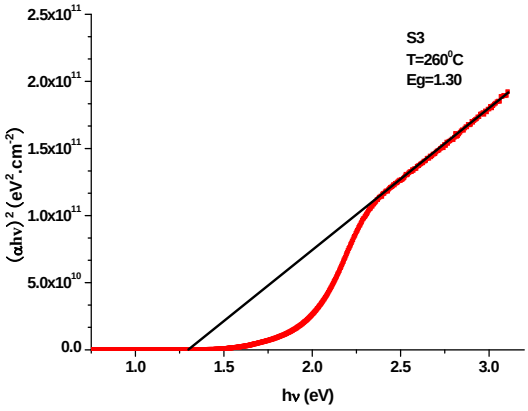


Figure 3. $(\alpha h\nu)^2$ vs. $h\nu$ curve of the SnS:Ag thin film annealed at 260℃.

3.3. Electrical Properties

At room temperature, semiconducting properties of the films were measured by a HMS-3000 Hall measurement system. The results are listed in **Table 2.** Compared with the unannealed sample, the semiconducting properties of the annealed samples have been improved. With the increase of the annealing temperature, the carrier concentration and mobility increase whereas the resistivity decreases except that of films annealed at 300℃. When the annealing temperature equals to 260 ℃, the carrier concentration and mobility reach the maximum values of 1.132×10^{17} cm^{-3} and 17.8 $cm^2.v^{-1}.s^{-1}$, respectively. But, when the annealing temperature is up to 300℃, the carrier concentration decreases to 2.601×10^{16} cm^{-3} with a mobility of 15.7 $cm^2.v^{-1}.s^{-1}$. The phenomenon can be explained by the fact that, with the increase of the annealing temperature, better crystallization and greater grain size in the films lead to the decrease of defects density and crystal-boundary, therefore the resistivity decreases. However, higher temperature urges crystal lattice vibration stronger and results in some crystal lattice defects. These defects become dispersion centers, causing the increase of resistivity of the films. In addition, the presence of SnO_2 can also cause the resistivity of the films increasing. Notably, the average Hall coefficients of all the samples are positive, which proves that the SnS:Ag films are of p-type conduction.

4. CONCLUSIONS

The SnS:Ag thin films were deposited on glass substrates using thermal evaporation technique and post-annealing, and effect of annealing on the films was investigated. The above results indicate that appropriate annealing temperature can increase the grain size of the films, improve the uniformity and crystallization of the films, decrease the resistivities of the films, and increase the absorption coefficients of the films. However, if the annealing temperature is higher than 300℃, crystallization of the films become weaker and the films can be oxidized due to low vacuum, thereby electrical and optical properties of the films become poor. At an annealing temperature of 260℃, the SnS:Ag films have the best properties: the direct bandgap is 1.3eV, the carrier concentration is up to 1.132×10^{17} cm^{-3}, and the resistivity is about 3.1 Ωcm.

5. ACKNOWLEDGMENT

The project-sponsored by SRF for ROCS, SEM (LXKQ0801) is gratefully acknowledged. The authors also wish to express their gratitude to

funding from Fujian Provincial Department of Science & Technology and Department of Education, China (2008I0019, 2009J01285, JB09008, JB09010).

REFERENCES

[1] El-Nahass, M.M., Zeyada, H.M., Aziz, M.S., *et al.* (2002) Optical properties of thermally evaporated SnS thin films [J]. *Optical Materials*, **20**, 159-170.

[2] Devika, M., Koteeswara Reddy, N., Sreekantha Reddy, D., *et al.* (2008) Synthesis and characterization of nanocrystalline SnS films grown by thermal evaporation technique [J]. *Journal of the Electrochemical Society*, **155**, 130-135.

[3] Takeuchi, K., Ichimura, K.M., Arai, M.E., *et al.* (2003) SnS thin films fabricated by pulsed and normal electro chemical deposition [J]. *Solar Energy Material and Solar Cells*, **75**, 427-432.

[4] Muller R. S. (1991) Microsensor [M]. The Institute of Electrical and Electronics Engineers Inc, New York.

[5] Albers, W., Hass, C., Vink, H.J., *et al.* (1961) Investigations on SnS [J]. *Applied Physics*, **32(10)**, 2220-2225.

[6] Devika, M., Koteeswara, R.N., Ramesh, K., *et al.* (2006) Low resistive micrometer-thick SnS:Ag films for optoelectronic applications [J]. *Journal of the Electrochemical Society*, **153**,727-733.

[7] Yakuphanoglu, F. (2009) Electrical conductivity, Seebeck coefficient and optical properties of SnO_2 film deposited on ITO by dip coating [J]. *Journal of Alloys and Compounds*, **470**, 55-59.

Vol.2, No.3, 201-204 (2010)
doi:10.4236/ns.2010.23031

Natural Science

Gaussian beam with non-spiral optical vortex

Xiu-Min Gao[1, 2], Song Hu[1], Jin-Song Li[3], Han-Ming Guo[2], Jian Wang[1], Song-Lin Zhuang[2]

[1]Electronics and Information College, Hangzhou Dianzi University, Hangzhou, China; xiumin_gao@yahoo.com.cn
[2]Optics and Electronics College, University of shanghai for Science and Technology, Shanghai, China
[3]Optics and Electronics College, China Jiliang University, Hangzhou, China

Received 24 November 2009; revised 28 December 2009; accepted 25 January 2010.

ABSTRACT

Optical vortex has attracted much interest recently due to its novel properties and applications. In this paper, the focusing properties of Gaussian beam containing one non-spiral optical vortex are investigated by means of scalar diffraction theory. Simulation results show that topological charge of non-spiral optical vortex affects optical intensity distribution in focal region considerably, and non-spiral focal pattern may also occur. Multiple intensity peaks may appear companying with center main focal spot under condition of higher topological charge. In addition, the number of weak intensity peak outside of the center main intensity peak is related to the value of topological charge.

Keywords: Focusing Properties; Gaussian Beam; Optical Vortex; Scalar Diffraction Theory

1. INTRODUCTION

Singular optics studying optical vortex has grown rapidly recently because optical vortex has some interesting properties [1-5] and promising applications [6-8]. And common optical vortex refers the phase singularity in light wavefront. In practice, optical vortex contains optical orbital angular momentum, which can be used to construct highly versatile optical tweezers [9]. In the focal reign of focusing beam that contains optical vortex, the optical momentum can be transferred to micro-particle, namely, light energy can be transformed directly to dynamic energy of particle, which has been used to construct experimentally microscopic optical rotator. In recent years, there also has been much interest in optical vortex propagation [10-16]. From a fundamental perspective they enable the study of phase singularities in a dynamical context [11,12], where twists, loops and knots in the path of the vortices have been shown to appear [13-15].

Spiral optical vortex has attracted much and has been investigated intensively and extensively, however, non-spiral optical vortex induces little attention. In fact, non-spiral optical vortex may own more interesting properties due to more flexible vortex pattern. The present paper is aimed at studying focusing properties of Gaussian beam containing one non-spiral optical vortex are investigated by scalar diffraction theory. The principle of the focusing this non-spiral vortex Gaussian beam is given in Section 2. Section 3 shows the simulation results and discussions. The conclusions are summarized in Section 4.

2. PRINCIPLE OF THE FOCUSING GAUSSIAN BEAM WITH ONE ASYMMET4IC OPTICAL VORTEX

In the focusing system we investigated, the Gaussian beam contains one non-spiral optical vortex, and convergences through a lens. According to scalar diffractive theory, the relative amplitude distribution in focal region of the focusing optical system is [17],

$$U(r,\psi,z)=\frac{i}{\lambda}\int_0^{\alpha}\int_0^{2\pi}E(\theta,\phi)\exp(i\varphi_j)\exp\left[-ikr\sin(\theta)\cdot\cos(\phi-\psi)-ikz\cos(\theta)\right]\cdot\sin(\theta)\,d\theta d\phi \quad (1)$$

where $E(\theta,\varphi)$ indicates the electric the amplitude of electric field of Gaussian beam that contains one non-spiral optical vortex, and can be written as,

$$E(\theta,\phi)=A_0\cdot\exp\left[-\frac{\sin^2(\theta)}{NA^2\cdot w^2}\right]\cdot\exp\left[i\pi\sin(m\phi)\right] \quad (2)$$

where A_0 is a constant, where $w=w_0/r_p$, w_0 is waist width (defined as radius) of the incident Gaussian beam.

Parameter r_p is the outer radius of the beam, and m is the topological charge of the non-spiral optical vortex. The phase is the sine function of azimuthal angle, which is not spiral distribution and can be changed by topological charge. The optical intensity distribution in focal region can be calculated quantitatively by means of substituting **Eq.2** into **Eq.1**. It should be note that

the intensity distribution in focal region is proportional to the modulus square of **Eq.1**, and in this paper the modulus square of **Eq.1** is calculated numerically.

3. RESULTS AND DISCUSSIONS

Without losing generality and validity, the intensity is normalized by optical intensity maximum, and it was proposed that $NA = 0.6$, $w = 1$. Firstly, the topological charge of the optical vortex is chosen as 1. **Figure 1(a)** illustrates the phase wavefront distribution under condition of m = 1 in polar coordinates. The radial coordinate indicates the phase value. It can be seen that the optical vortex is non-spiral and turns on off-axis distribution.

The **Figure 1(b)** shows that optical intensity in focal region of this kind of vortex Gaussian beam. It should be noted that Vr and Ψ indicate radial and azimuthal coordinates, respectively. Ψ ranges from 0 to 2π. And the distance unit in radial direction is k^{-1}, here k is the wave number of the focusing beam. We can see that the focal spot is asymmetric remarkably, and does not locate at optical axis. The cause of this focal pattern is the asymmetry characteristics of the optical vortex that was embedded in the incident Gaussian beam. There the symmetry property of the optical vortex affects focal pattern.

Topological charge is important parameter optical vortex, here, the effect of different topological charge on focal pattern is investigated. **Figure 2(a)** illustrates the phase distribution for m = 2. It can be seen that there are four phase maximums and locate around optical axis. Topological charge m affects phase wavefront distribution considerably. **Figure 2(b)** show us the intensity distribution in focal region under condition of m = 2.

From this figure, we can see that there is one center main intensity spot on optical axis. Four optical intensity peaks come into being outside of center main intensity peak. And the center main intensity peaks also extends outside to form four intensity branches. Therefore, there are four local intensity minimums appear between center main peak and outside four subordinate intensity peaks. This phenomenon is very interesting and can be used to restrict those micro particles whose refractive index is lower than that of the surrounding medium, and this kind of condition is common in life science. In addition, the whole focal pattern may also be employed to construct novel optical trap. By comparing **Figure 2** with **Figure 1**, the optical intensity distribution in focal region of the Gaussian beam containing one non-spiral optical vortex can be altered remarkably by the topological charge.

In order to get insight into the effect of topological change more deeply, focusing properties for other dif-

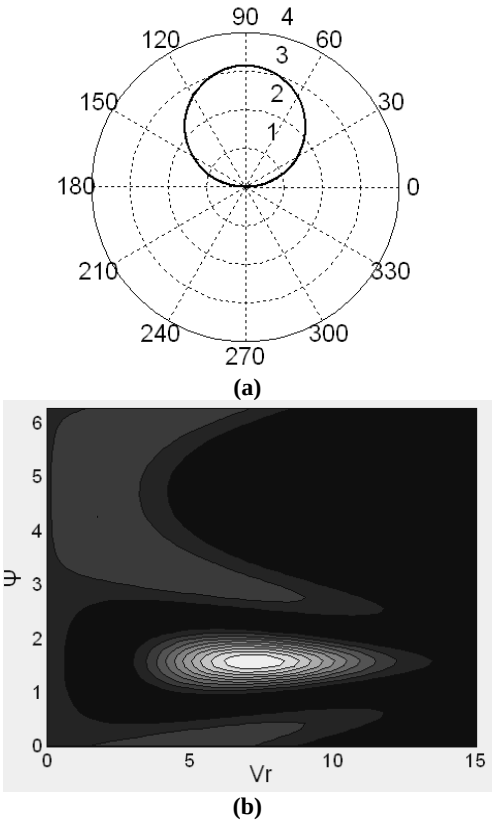


Figure 1. The (a) phase distribution for m = 1 and the (b) corresponding focal intensity distribution.

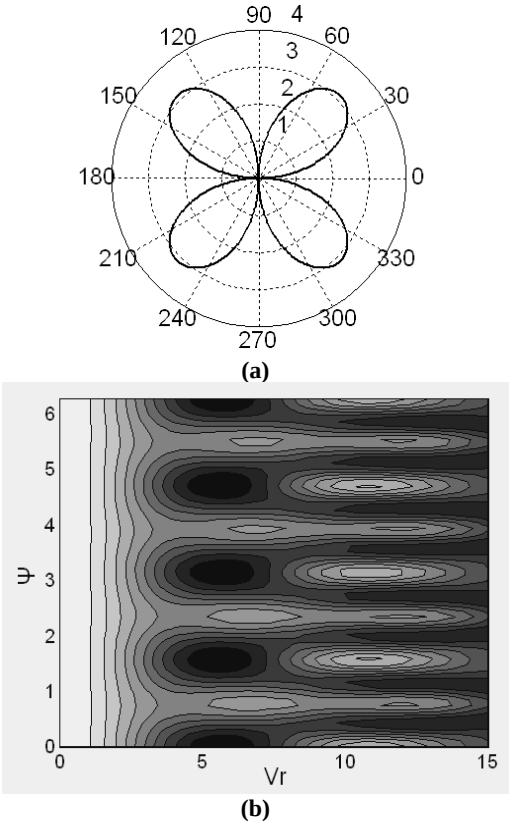


Figure 2. The (a) phase distribution for m = 2 and the (b) corresponding focal intensity distribution.

ferent m are also investigated. **Figure 3** shows the phase distribution for m = 3 and the corresponding focal intensity distribution. There are three phase maximums and turns symmetric distribution around optical axis. From **Figure 3(b)**, it can be seen that for this case, the focal pattern contains one center main intensity peak with three relative weak intensity peaks outside. And there are no local intensity minimums between center peak and outside three peaks. Number of subordinate intensity peaks is related to the number of topological charge m. Simulation shows that number of subordinate intensity peaks equals m when m is odd integral number.

We also studied effect of some other value of topological charge on focal pattern. **Figure 4** illustrates phase distribution for m = 4 and the corresponding focal intensity distribution. It can be seen that there are eight phase maximums, and also eight weak intensity peaks around center main intensity peak in focal region, as shown in **Figure 4(b).** The number of subordinate intensity peaks is two times of the parameter m when m is even integral number.

From all above focal pattern evolution on increasing topological charge, we can seen that topological charge of non-spiral optical vortex affects optical intensity distribution in focal region very obviously, and non-spiral focal pattern may also occur. Multiple intensity peaks may appear companying with center main focal spot under condition of higher topological charge. In addition, the number of weak intensity peak outside of the center main intensity peak is related to the value of topological charge. When topological charge is odd number, the number of weak intensity peak equals the value of topological charge, while, the number of weak intensity peak is two times the value of topological charge if topological charge is even number. In optical trapping system, it is usually deemed that the forces exerted on the particles in light field include two kinds of forces, one is the gradient force, which is proportional to the intensity gradient; the other is the scattering force, which is proportional to the optical intensity [18]. Therefore, the tunable

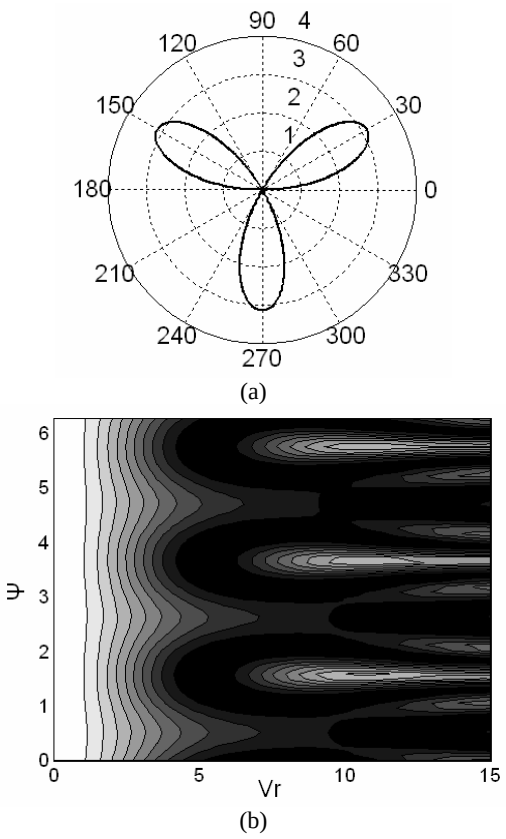


Figure 3. The (a) phase distribution for m=3 and the (b) corresponding focal intensity distribution.

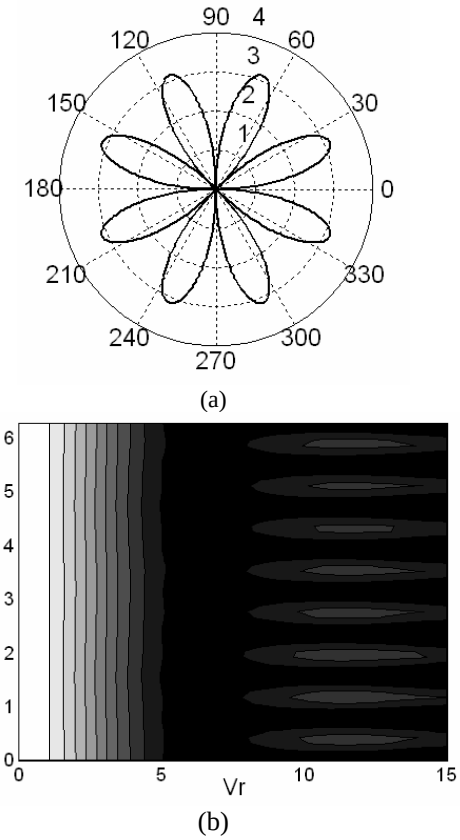


Figure 4. The (a) phase distribution for m=4 and the (b) corresponding focal intensity distribution.

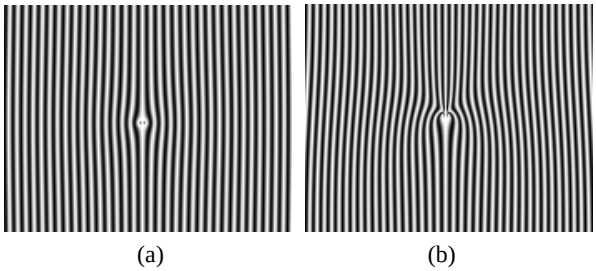


Figure 5. Holograms for generation this non-spiral optical vortex for (a) m = 1 and (b) m = 2, respectively.

focal intensity distribution predicts that the focusing properties of this kind of beam can be employed to construct controllable trap.

We are now going on investigating the focusing properties of this kind of vortex beam. Next step, the attention will be focused on the experimental research and extends study into vector optical domain. **Figure 5** illustrates holograms for generating this non-spiral optical vortex, which may be employed in our future experiment. These holograms can be obtained conveniently by calculating numerically optical interference diagram.

4. CONCLUSIONS

The focusing properties of Gaussian beam containing one non-spiral optical vortex are investigated by scalar diffraction theory in this paper. Calculation results show that topological charge of non-spiral optical vortex affects optical intensity distribution in focal region remarkably, and asymmetric focal pattern may also occur for lower topological charge. Multiple intensity peaks may come into being companying with center main focal spot under condition of higher topological charge. In addition, the number of weak intensity peak equals the value of topological charge under condition of odd number topological charge, while, the number of weak intensity peak is twice times the number of topological charge for even number topological charge.

5. ACKNOWLEDGMENT

This work was supported by National Basic Research Program of China (2005CB724304), National Natural Science Foundation of China (60708002, 60878024, 60778022, 60807007), China Postdoctoral Science Foundation (20080430086), Shanghai Postdoctoral Science Foundation of China (08R214141), and Shanghai Leading Academic Discipline Project (S30502).

REFERENCES

[1] Thomas, V.D. and Taco, D.V. (2009) Evolution of singularities in a partially coherent vortex beam. *Optical Society of America B*, **26**, 741-744.

[2] Coullet, P., Gil, L. and Rocca, F. (1989) Optical vortices. *Optics Communications*, **73**, 403-408.

[3] Palacios, D.M., Maleev, I.D., Marathay, A.S. and Swartzlander, G.A. (2004) Spatial correlation singularity of a vortex field. *Physical Review Letters*, **92**, 143905-4.

[4] Rozas, D.Z., Sacks, S. and Swartzlander, G.A. (1997) Experimental observation of fluidlike motion of optical vortices. *Physical Review Letters*, **79**, 3399-3402.

[5] Filippus, S.R. (2004) Distribution of angular momentum and vortex morphology in optical beams. *Optics Communications*, **242**, 45-55.

[6] Lee, W.M., Yuan, X.C. and Cheong, W.C. (2004) Optical vortex beam shaping by use of highly efficient irregular spiral phase plates for optical micromanipulation. *Optics Letters*, **29**, 1796-1798.

[7] Ladavac, K. and Grier, D.G. (2004) Microoptomechanical pumps assembled and driven by holographic optical vortex arrays. *Optics Express*, **12**, 1144-1149.

[8] Cojoc, D., Garbin, V., Ferrari, E., Businaro, L., Romamato, F. and Fabrizio, E.D. (2005) Laser trapping and micro-manipulation using optical vortices. *Microelectronic Engineering*, 77-78, 125-131.

[9] Vladlen, G., Shvedov, A.S., Desyatnikov, A.V., Krolikowski, R.W. and Kivshar, Y.S. (2009) Optical guiding of absorbing nanoclusters in air. *Optics Express*, **17**, 5743-5757.

[10] Soskin, M.S. and Vasnetsov, M.V. Singular Optics, in Progress in Optics 42, Wolf, E. Ed. (Elsevier 2001).

[11] Indebetow, G. (1993) Optical vortices and their propagation. *Journal of Modern Optics*, **40**, 73-87.

[12] Rozas, D., law, C.T. and Swartzlander, G.A. (1997) Propagation dynamics of optical vortices. *Journal of the Optical Society of America B*, **14**, 3054-3065.

[13] Berry, M.V. and Dennis, M.R. (2001) Knotted and linked phase singularities in monochromatic waves. *Proceedings of the Royal Society A*, **457**, 2251-2263.

[14] Berry, M.V. and Dennis, M.R. (2001) Knotting and unknotting of phase singularitites: Helmholtz waves, paraxialwaves and waves in 2 + 1 spacetime. *Journal of Physical A*, **34**, 8877-8888.

[15] Leach, J., Dennis, M., Courtial, R.J. and Padgett, M.J., (2005) Vortex knots in light. *New Journal of Physics*, **7**, 1-11.

[16] Baumann, S.M., Kalb, D.M., MacMillan, L.H. and Galvez, E.J. (2009) Propagation dynamics of optical vortices due to Gouy phase. *Optics Express*, **17**, 9818-9827.

[17] Gu, M. (2000) Advanced Optical Imaging Theory (Springer, Heidelberg)

[18] Visscher, K. and Brakenhoff, G.J. (1992) Theoretical study of optically induced forces on spherical particles in a single beam trap I: Rayleigh scatterers. *Optik*, **89** 174-180.

Vol.2, No.3, 205-211 (2010)
doi:10.4236/ns.2010.23032

Natural Science

The nature of neuronal words and language

Morris Henry Baslow

Nathan S. Kline Institute for Psychiatric Research, Center for Neurochemistry, New York, USA; Baslow@nki.rfmh.org

Received 5 December 2009; revised 13 January 2010; accepted 30 January 2010.

ABSTRACT

Individual neurons in the brain possess natural language in the form of energy-dependent action potentials or spikes (S) operating on a millisecond timescale that, along with pauses (P) between spikes, constitute a two letter (S, P) "alphabet" that is used to generate meaningful frequency-encoded neuronal "words". These words are then used to transmit information to other neurons in the form of phrases consisting of two or more words that are contained within longer pause-delineated structured declarative sentences. In this article, the nature of neuronal words and language are described, and examples provided that illustrate the way in which neuronal language is used by the brain to interact with and interpret both its internal and external environments. It is hoped that a fuller understanding of the language used by neurons to communicate may lead to development of novel treatments for a number of human neuropathies.

Keywords: Biosemiotics; Brain; Cognition; Language; Learning; Memory; Neurons

1. INTRODUCTION

Semiotics is broadly defined as a study of signs, meanings and their significance in communication, and biosemiotics as a study that is based on the general assertion that all living organisms including cells are semiotic systems. While information processing in the brain is highly complex, each neuron uses a simple code mechanism for transmitting information. This is in the form of electrophysiological action potentials or spikes (S) of about a 1 millisecond (ms) duration that, along with pauses (P) between spikes constitute a two letter "alphabet" that generates meaningful frequency-encoded signals or neuronal S/P "words" [1]. The term "word" for neuronal activity in spikes/s has previously been used to describe the electrophysiological activity of the visual interneuron of the blowfly where each action potential (AP) was considered a "short" word and trains of AP's considered "longer" words [2]. In this article a neuronal word is defined as a single AP together with the total pause before the next AP; a phrase is defined as a group of two or more words, and a sentence as a temporal grouping of such words containing a subject, and a predicate that expresses what is stated about the subject [1]. All neuron codes, whether continuous trains of spikes that are characteristic of information transmitted at low frequencies, or bursts of spikes that are characteristic of information transmitted at high frequencies, are made up of interactions between spikes and pauses. There are two theories regarding how these spikes carry encoded information [3]. One is the "spike rate code" which suggests that information is carried in the average rate at which the neuron fires, and that the timing of each spike is random. The other is the "spike timing code" that suggests that specific information is carried not only in the average spike rate at which a neuron fires, but also in the precise timings between each spike. In this article, it is proposed that neurons transmit meaningful information not only in their spike rate and spike pause timing, but also in the spike duration or length (S_L) and its relationship to the pause length (P_L) between spikes. It is this relationship that constitutes a neuronal "S/P word", and adds an additional dimension to the potential richness of communication between neurons.

2. DISCUSSIONS

2.1 Anatomy of a Neuronal Word

The communication method used by the brain is generated at the cellular level where individual neurons transmit information to one another by generation a series of wave-like depolarizations. This depolarization's start at the plasma membrane of their cell bodies (soma) and pass along the plasma membranes of axon hillocks, axons and pre-synaptic dendrites. Each wave-like spike consists of two components, first, a rapid depolarization, involving the efflux of K^+ and influx of Na^+ at the plasma membrane down their respective intracellular/extra cellular gradients, a process that requires little or

no additional energy. This is followed by a re-polarization process that occurs within an absolute refractory period during which time the intracellular negative membrane K^+/Na^+ potential is restored using energy derived from intracellular stores of adenosine triphosphate (ATP) via a plasma membrane surface enzyme, Na^+/K^+ ATPase, in order to transport these ions against their respective gradients [4]. In single afferents of human mechanoreceptive sensory nerve fibers of the skin of the hand, spike times-to-peak were 0.12-0.45 ms with a mean of 0.21 ms [5]. While the depolarization process is pre-energized and trigger-ready, re-polarization is a slower enzyme-dependent process that is subject to rate regulation by all factors that can affect enzyme activity. These include enzyme, substrate, and end-product concentrations, focal ionic makeup, pH, temperature as well as the presence of competitive and non-competitive inhibitors. All of these factors can affect the rate of re-polarization and the length of the absolute refractory period which in this study varied from 0.31-0.75 ms with a mean of 0.50 ms indicating that the energy consuming re-polarization period was 2.38 times longer than the depolarization period such that the length of the re-polarization period dominated the total spike period. In this study, the average length of a spike including the depolarization and longer re-polarization periods in the sensory nerve fibers was 0.71 ms with a range from 0.43-1.20 ms. The relationship between the depolarization and re-polarization periods of a single spike is illustrated in **Figure 1**. Since the depolarization is pre-primed and trigger ready, and the re-polarization period requires energy to reestablish ionic gradients, re-polarization is therefore subject to cellular control. Thus, the S_L can be varied to incorporate information that is transmitted in the form of a S_L/P_L ratio over and above information that transmitted by Hz and/or specific peak-to-peak inter-spike intervals. Data from [5].

During the spiking process which includes the absolute refractory period, and until the ionic gradients are reestablished, a second spike cannot be generated. In addition, following the absolute refractory period there is a relative refractory period where only a strong stimulus can generate a second AP, and following this period there is a period of longer pause with no electrophysiological activity. A spike including its absolute refractory period may vary, but an average spike-absolute refractive period length is about 1 ms, and is followed by a relative refractory period of about 2 ms in duration [6]. This in turn is followed by a pause of varying lengths before the next spike is generated. Thus, there are two phases of signal generation, each consisting of two subcomponents. The total spike length (S_L) is made up of the sum of the depolarization and absolute refractory (re-polarization) periods, and the total pause length (P_L) is made up of the relative refractory period and the pause until the next spike is initiated. The total pause length can vary from 0 to approximately 10,000 ms under normal physiological conditions. Based on these neuronal spike and pause components, a single neuronal S/P word has been defined as the sum of its total spike length and its total pause length in ms, and different words are characterized by differences in the ratio of its S_L to its P_L [1].

The spike-generation process that results in S/P words is also is metabolically costly, calculated at 2.2×10^9 ATP molecules per spike [7], which requires that ATP supplies be constantly replenished or the timing of the spike-refractory periods will be altered, and meaningful frequency-encoded information lost [1]. An illustration of the components of a neuronal word as described is graphically presented in **Figure 2.**

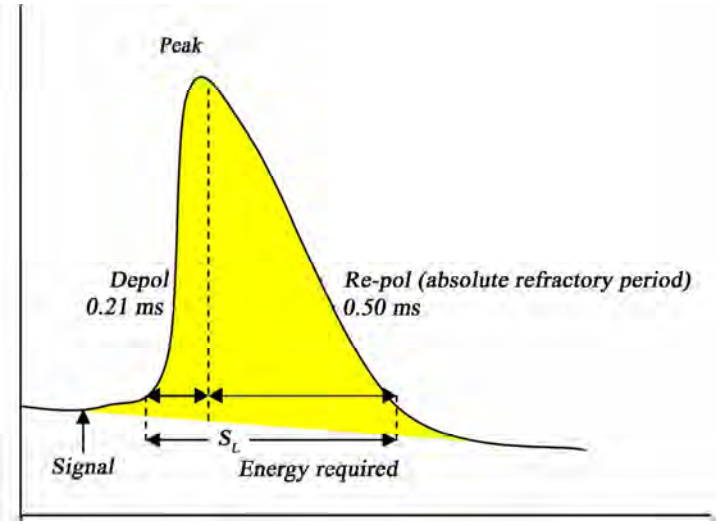


Figure 1. Cartoon showing the relative depolarization and repolarization times of a neuronal spike.

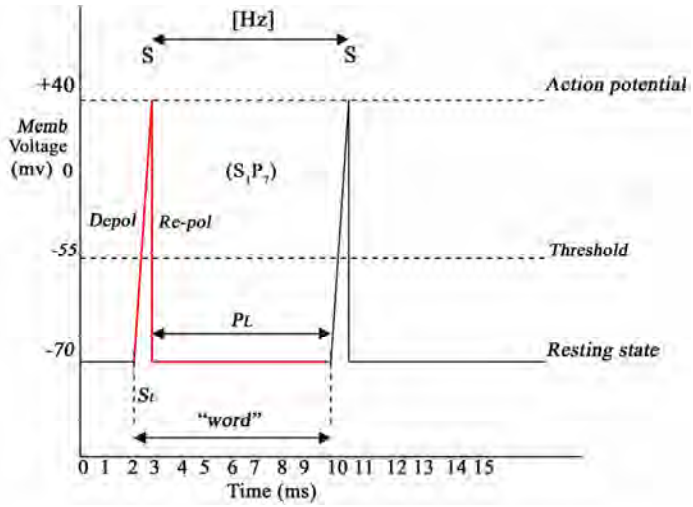


Figure 2. Cartoon showing the components of the neuronal word, S_1P_7.

In this illustration, each neuronal word (S_1P_7) consists of a depolarizing spike (0.5 ms), an absolute refractory period when the plasma membrane is being re-polarized using ATP (0.5 ms) for a total spike length (S_L) of 1.0 ms; a relative refractory period where the neuron can be activated but only by a strong stimulus (2.0 ms); and a period of inactivity (5.0 ms). The spike frequency is (1000 ms/8 ms) or 125 Hertz (Hz). Unless the pause is terminated by a second spike within a physiologically relevant period, the word is incomplete and without any quantitative meaning. Thus, a meaningful phrase must consist of at least two words, and a sentence of at least one phrase. The word S_1P_7 presented in this cartoon reflects the S/P message generated upon exposure of the isolated salt receptor of the blowfly to a 3 molar NaCl solution [8].

2.2. Conversion of Frequency in Hz to S/P Words

As shown in **Figure 2**, information contained in Hz format differs from that in S/P word format in that in Hz, the signal is a measure of the time between two spikes that is initiated at the peak of the first spike and is constituted by the re-polarization period of that spike, its relative refractory period, the pause until initiation of a second spike and the depolarization period of a second spike. In essence, the signal recorded in Hz splits a spike into its two components such that a spike in Hz is a composite formed from elements of two AP's. On the other hand, an S/P word as defined includes the depolarization and re-polarization components of a single spike, along with the relative refractory period and the pause until the next spike is initiated. However, the two formats are clearly related, and while Hz is a commonly used method of recording and describing neuronal activity, a reasonable estimation of the length of a neuronal word can also be derived from Hz. Thus, from Hz, the total time in ms of an average S+P word can be determined: Total time per word: $(S_L + P_L)_{ms} = 1000/Hz$.

At a frequency of 100 Hz, the mean S+P word length would be 10 ms. However, the total value in ms for an S + P word does not provide sufficient information to construct a specific neuronal S/P word. For this, one requires in addition either the total spike length (S_L) or total pause length (P_L). Thus, for a S_L of 0.5 at 100 Hz, the S/P word would be $S_{0.5}/P_{9.5}$, for an S_L of 1.0, the S/P word would be $S_{1.0}/P_{9.0}$, and for a S_L of 1.5, the S/P word would be $S_{1.5}/P_{8.5}$, representing three different S/P words out of a great number of possible S/P words that are differentiated by the relative lengths of their spikes and pauses. It is important to note that all of these S/P words are different and may transmit different meanings, but that all are being communicated at the same frequency of 100 Hz. Clearly, a measure of frequency alone for neuronal communications cannot be physiologically accurate. When afferent neuronal messages are converted from electrophysiological to chemical and back into electrophysiological messages at synaptic interfaces for transmission to efferent neurons, it is highly likely that differences in the S_L/P_L ratios are important elements in modulation of the messages at these interfaces. Since much of the published neuronal response data is presented in the form of Hz, and while less accurate than actual S/P words, in the absence of a measured S_L, and by using an average S_L of 1.0, one can still derive useful information based on Hz about how neurons communicate with one another concerning the nature of their environment.

2.3. Differences in Spike Rate, Spike Timing and Neuronal Word Codes in Information Transfer

The spike rate code hypothesis proposes that information is processed in the form of the average time between

spikes, and the spike timing code hypothesis proposes that information can also be contained in the differences in pauses between individual spikes, an analysis that can be used to identify harmonics and periodic oscillations in spike trains. Both of these hypotheses treat a spike as a point source and both are valid interpretations of recorded data. However, these hypotheses are mirror images of one another in that when viewed at the level of only two spikes, as presented in **Figure 1**, it is clear that the time between individual spikes treated as point sources, and the pause time between spikes will always be the same. On the other hand, the neuronal word code hypothesis as detailed above takes into consideration that individual spikes have a significant time dimension and shows that even at the level of only two spikes, there can be an enormous number of specific S/P words that can be generated and used to transmit different kinds of information. This is because each neuronal word is formed by individual components of the AP process associated with spike generation, and differences in each component can alter the ratio between the total spike length and the total pause length. For the three examples provided above, $S_{0.5}/P_{9.5}$, $S_{1.0}/P_{9.0}$ and $S_{1.5}/P_{8.5}$, their S/P ratios are 0.052, 0.111 and 0.176 respectively, and all are capable of being transmitted at the same frequency (Hz) where both the inter-spike time and pause time are the same. While all three hypotheses reflect frequency-based neuronal codes, the neuronal word code with its spike time dimension is the most physiologically realistic since it is the only one that identifies a mechanism whereby the AP can interact with the rate of depolarization-induced release of neurotransmitters from presynaptic vesicles. Thus, only the neuronal word code appears to have the potential for an enhanced richness in information transfer, especially at the low frequencies and short periods dictated by ongoing bioenergetic and temporal physiological constraints. Moreover, measurement of the ratios that comprise S/P words may provide a new investigational tool, not available using the other "codes", with which to assess differences in neuronal function between normal and pathological conditions.

2.4. Nature of Neuronal Language

There are three components to neuronal language. First, neurons are "wired" in the sense that each neuron in the corpus occupies a specific place on a brain "map" of a given organism. Second, by virtue of its phenotype, each neuron sends specific kinds of messages for interpretation within the CNS network. For environmental sensing neurons, these may identify the qualitative nature of the messages being communicated such as where in the brain or corpus it originates, and whether the message is a general call for stored information, or whether it is from a specific region associated with some specific factor such as sound, light or availability of "glucose" (Glc). Lastly, the messages sent by neurons must also concern the quantitative nature of sensed information such as the importance of message itself, or the specific levels of light, sound, pain, pressure, stretch, or substance concentrations sensed. It is this last component, the quantitative nature of information within the brain, where specific frequency-encoded neuronal words appear to play an important role.

As an example, based on a recorded neuronal frequency of 100 Hz, a neuronal spike with a total spike length of 1 ms and a total pause length of 9 ms would constitute the neuronal S/P word (S_1P_9) of 10 ms duration. When this word is repeated, it becomes a phrase. In the blowfly, this phrase corresponds to immersion of its isolated salt receptor in 2.0 M NaCl [8]. Based on using a value for total S_L of 1 ms, the specific words in "fly-speak" for immersion in additional NaCl concentrations of 0.5 and 1.0 M are S_1P_{18} (52 Hz) and S_1P_{12} (80 Hz) respectively. In addition, if the S_L is known along with the total S + P time in ms derived from Hz, the total "pause" time in ms can be derived using the following general relationship: $P = [1000\ ms - (Hz \times S_L)]/Hz$.

It has also been proposed, that the language of neurons can be written in any of a number of frequency-encoded formats including Hz, S/P and musical notations and can be translated into any of the oral, written or symbolic human languages as well [1]. Importantly, using artificial electronically generated frequencies that mimic the frequencies of recorded natural neuronal words, it has been demonstrated that it is possible not only to record these neuronal words, but also to communicate directly with the central nervous system (CNS) of rats. Where this has been done, the manufactured words have been shown to elicit the same behavioral responses in rats that would have been generated by exposure to a specific environmental stimulus [9].

2.5. The Dictionary of Neuronal Words and their Representational Nature

The dictionary of neuronal S/P words in an ms timeframe is relatively short, being limited at its upper level by the absolute and relative neuronal refractory periods and at its lower level by physiological requirements such as maintenance of the human heart rate at 68 beats/min (1.13 Hz). With the average spike including an absolute refractory period of 1 ms duration followed by a relative refractory period of about 2 ms, the highest possible frequency would be S_1P_2 or 333 Hz. If the system were driven by a strong enough input eliminating the relative refractory period altogether, the highest possible frequency might be S_1P_0 or 1,000 Hz.

Since the brain's primary form of communication, both internally and with the outside world, is in the form of these electrophysiological messages, it follows that neuronal words must also be highly representational in nature. For example, audible sound in many animals may range up to 20,000 Hz and visible light is in the

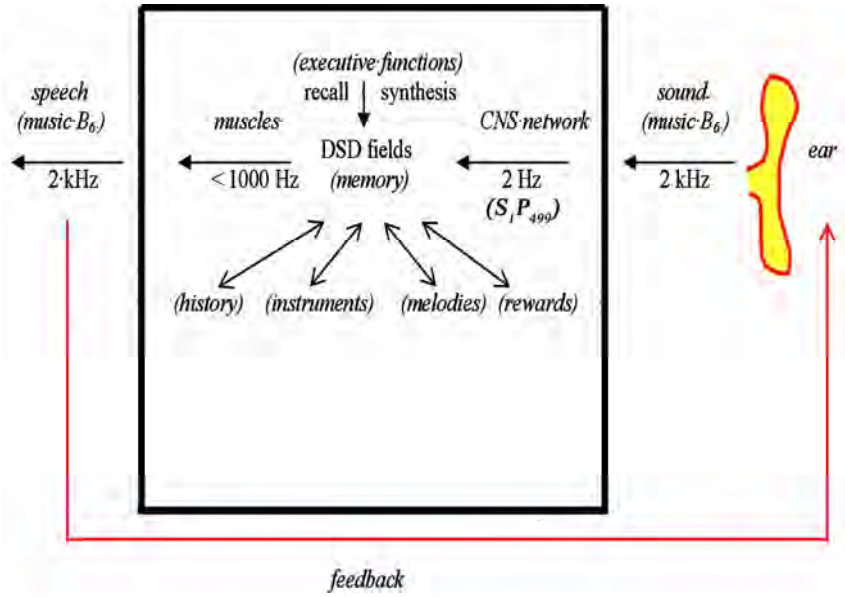


Figure 3. A "dark room" paradigm of brain function.

Table 1. List of several neuronal S/P words and their meanings.

Organism	Sensor	Word	Meaning	Reference
Blowfly				
sensillum	salt	S_1P_{499}	NaCl 0.10 M	[8]
		S_1P_{45}	0.20	
		S_1P_{27}	0.25	
		S_1P_{18}	0.50	
		S_1P_{12}	1.00	
		S_1P_9	2.00	
		S_1P_7	3.00	
Ground beetle				
Sensillum	salt	S_1P_{1000}	NaCl 0.001 M	[11]
		S_1P_{71}	0.01	
		S_1P_{29}	0.10	
		S_1P_{23}	1.00	
Hamster				
taste cells	acid	S_1P_{286}	pH 2.5	[12]
		S_1P_{312}	3.0	
		S_1P_{1999}	4.3	
		S_1P_{4000}	5.0	
Cat brain				
hypothalamus	glucose			
glucosensing	steady-state	S_1P_{10000}	normal Glc level	[13]
	Glc i.v. 30 min	S_1P_{2000}	elevated Glc level	
Rat				
taste cells	quinine	S_1P_{21}	quinine 0.01 mM	[9]
Rat brain				
hippocampal	sound	S_1P_{499}	frequency 2 kHz	[10]
neuron		S_1P_{142}	10 kHz	
Guinea pig brain				
GABAergic	H_2O_2	S_1P_{53}	H_2O_2 0.0 mM	[14]
neurons		S_1P_{40}	1.5	

range of 10^{14} Hz. For the brain to comprehend its external environment, these frequencies must first be converted into representations of those wavelengths at less than 1000 Hz, with the same meanings as the original signals, which can then be transmitted by neurons within the CNS network. These representations can also be "memorized" in the form of structural and biochemical changes in the dendrite-synapse-dendrite (DSD) fields

between neurons, and later recalled to "imagine" the nature of the original wavelengths even without external input. For example, one can imagine "in the minds eye" the color "red" with the eyes closed, and with eyes open can also compare a newly observed representation of red with that which was previously seen and stored in memory. This phenomenon is illustrated in **Figure 3** where a specific sound tone of 2 kHz impinging on the ear is converted into a representational neuronal word at 2 Hz for CNS network processing.

In this illustration, the translation by a sense organ of the specific frequency-encoded environmental information for the approximate musical tone B_6 (1975.53 Hz) is illustrated. The sound sense organ first converts the tone into a brain frequency-encoded S/P word representation for B_6 at 2 Hz (S_1P_{499}) that is then compared within the CNS network to auditory area DSD memory constructs. During this comparison related historical documents in other brain areas may also be probed for relevant information. Upon an executive command, appropriate muscles can also be activated to reproduce the 1975.53 Hz frequency, exporting it to another organism and also sending it back to the sense organ where it can be compared to the original tone. Importantly, the S/P word for the tone B_6 or any other musical tone can also be synthesized *de novo* by the brain from memory engrams without external input and can be "heard" within the brain as recalled sounds or melodies. All of these functions occur in a ms/sub-ms timeframe and apply to any other sensed parameter, whether derived internally or external to the brain. The specific auditory data presented in **Figure 3** is adapted from rat experiments [10].

2.6. Examples of Neuronal Words and their Meanings

In **Table 1** descriptions of some S/P words and their meanings are presented. In each of these cases, the S/P words have been derived from frequency data by using an average S_L including an absolute refractory period of 1.0 ms, and rounded to the nearest integer. However, it is also apparent that neurons can operate in a sub-millisecond timeframe [10], which is in keeping with the microsecond timeframe dictated by physiological demands and communication networking requirements of the brain.

3. CONCLUSIONS

In this paper it is proposed that neurons possess natural language in the form of specific neuronal words, phrases and sentences, and evidence is provided that when specific neuronal words are known, it is possible to mimic them and generate electronic words to communicate intelligently with the CNS. While most investigative recordings of neuronal activity are in the form of Hz, the differences between Hz and word formats have been described, and a case has been made that the word format is the only format that can be physiologically associated with the release of specific quanta of neurotransmitters at synapses and is therefore more representative of the potential richness of information that can be transmitted by neurons. In addition, based on physiological limitations associated with the bioenergetic nature of neuronal AP's, it has been shown that much of the information about the external environment that is transmitted to the brain must first be reduced to a frequency-encoded representation of that information that can be communicated at < 1000 Hz. These representations can be acted upon immediately, memorized when appropriate, and the memorized information recalled as needed in the form of such representations within the brain. Of note, in humans and other animals these internal representations can also be converted back into their original electromagnetic spectrum frequencies by externally re-creating the original environmental source such as by reproducing a particular memorized sound frequency. Finally, it is hoped that this paper will encourage the continued analysis of the nature of neuronal words and language, and that such studies will lead to additional development of novel electronic methods for communicating directly and intelligently with the CNS as has already been attempted in treatment of several human neuropathies [15]. With such specific frequency-encoded words as tools, it may be possible to alter or induce new behaviors, and thus lead to successful treatment stratagems for a variety of human psychological and neurological conditions.

REFERENCES

[1] Baslow, M.H. (2009) The languages of neurons; An analysis of coding mechanisms by which neurons communicate, learn and store information. *Entropy*, **11**, 782-797.

[2] Borst, A. (2003) Noise, not stimulus entropy, determines neuronal information rate. *Journal of Computational Neuroscence*, **14**, 23-31.

[3] Clifford, C.W. and Ibbotson, M.R. (2009) Response variability and information transfer in directional neurons of the mammalian horizontal optokinetic system. *Visual Neuroscience*, **17**, 207-215.

[4] Baslow, M.H. and Guilfoyle, D.N. (2007) Using proton magnetic resonance imaging and spectroscopy to understand brain activation. *Brain and Language*, **102**, 153-164.

[5] Brink, E.E. and Mackel, R.G. (1993) Time course of action potentials recorded from single human afferents. *Brain*, **116**, 415-432.

[6] Berry II, M.J. and Meister M. (1998) Refractoriness and neural precision. *Journal of Neuroscience*, **18**, 2200-2211.

[7] Lennie, P. (2003) The cost of cortical computation. *Current Biology*, **13**, 493-497.

[8] Gillary, H.L. (1966) Stimulation of the salt receptor of the blowfly. II. Temperature. *Journal of General Physiology*, **50** 351-357.

[9] Di Lorenzo, P.M., Leshchinskiy, S., Moroney, D.N. and Ozdoba, J.M. (2009) Making time count: Functional evidence for temporal coding of taste sensation. *Behavioral Neuroscience*, **123**, 14-25.

[10] Takahashi, S. and Sakurai, Y. (2009) Sub-millisecond firing synchrony of closely neighboring pyramidal neurons in hippocampal CA1 of rats during delayed nonmatching to sample task. *Frontiers in Neural Circuits*, **3(9)**

[11] Merivee, E., Renou, M., Mand, M., Luik, A., Heidemaa, M. and Ploomi, A. (2004) Electrophysiological responses to salts from antennal chaetoid taste sensilla of the ground beetle *Pterostichus aethiops. Journal of Insect Physiology*, **50**, 1001-1013.

[12] Gilbertson, T.A., Avenet, P., Kinnamon, S.C. and Roper, S.D. (1992) Proton currents through amiloride-sensitive Na channels in hamster taste cells. *Journal of General Physiology*, **100**, 803-824.

[13] Anand, B.K., Chhina, G.S., Sharma, K.N., Dua, S. and Singh, B. (1964) Activity of single neurons in the hypothamic feeding centers: effect of glucose. *American Journal of Physiology*, **207**, 1146-1154.

[14] Lee, C.R. and Rice, M.E. (2008) Hydrogen peroxide increases the excitability of substantia nigra pars reticulata GABAergic neurons. *Society for Neuroscience Meeting*, November 16, Program 179.2, Poster # QQ37.

[15] George, M.S. and Aston-Jones, G. Noninvasive (2010) techniques for probing neurocircuitry and treating illness: Vagus nerve stimulation (VNS), transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (tDCS). *Neuropsychopharmacology*, **35**, 301-316.

Vol.2, No.3, 212-219 (2010)
doi:10.4236/ns.2010.23033

Natural Science

What does the “arrow of time” stand for?

Etienne Klein

Laboratoire de recherche sur les sciences de la matiere (larsim)centre d’etudes de saclay 91191 gif sur yvette cedex France;
etienne.klein@cea.fr

Received 24 November 2009; revised 28 December 2009; accepted 28 January 2010.

ABSTRACT

One hundred and thirty years after the work of Ludwig Boltzmann on the interpretation of the irreversibility of physical phenomena, and one century after Einstein's formulation of Special Relativity, we are still not sure what we mean when we talk of “time” or “arrow of time”. We shall try to show that one source of this difficulty is our tendency to confuse, at least verbally, time and becoming, *i.e.* the course of time and the arrow of time, two concepts that the formalisms of modern physics are careful to distinguish. The course of time is represented by a time line that leads us to define time as the producer of duration. It is customary to place on this time line a small arrow that, ironically, must not be confused with the “arrow of time”. This small arrow is only there to indicate that the course of time is oriented, has a well-defined direction, even if this direction is arbitrary. The arrow of time, on the other hand, indicates the possibility for physical systems to experience, over the course of time, changes or transformations that prevent them from returning to their initial state forever. Contrary to what the expression “arrow of time” suggests, it is therefore not a property of time itself but a property of certain physical phenomena whose dynamic is irreversible. By its very definition, the arrow of time presupposes the existence of a well-established course of time within which – in addition – certain phenomena have their own temporal orientation. We think that it is worthwhile to emphasize the difference between several issues traditionally subsumed under the label “the problem of the direction of time”. If the expressions “course of time”, “direction of time” and “arrow of time” were better defined, systematically distinguished from one another and always used in their strictest sense, the debate about time, irreversibility and becoming in physics would become clearer.

Keywords: Time, Time’S Arrow; Temporal Asymmetry; Principle of Causality; Irreversibility

1. INTRODUCTION

Each one of us can make the following observation: we often talk about time as if it corresponded only to a “becoming”, to the stream of changes affecting a thing, a person, an institution, a physical system. Certainly, change is truly the phenomenon that best suggests the idea of time, and one can easily understand why: we never encounter a specific and directly perceptible reality that would be the time. We only see around us *changing* things, things *becoming* others, and it is therefore through the concrete effect of change that the course of time first appears to us. But to conclude from this, as our natural language does, that *time* and *becoming* are the same is a step that is too easily made: without proceeding to further investigation, time is mostly referred to as if it resembled what it holds. It is said to stop or disappear when nothing seems to be happening as if all its dynamics depended not upon time itself but upon its contents.

Such short-cuts serve to answer, before it is even formulated, the question of the relationship between time and events, as well as the question of its dynamics: is time an abstract structure into which events are inserted, that is to say a reality in itself preceding all possible events, or is it composed of the stream of events itself? In other words, is it legitimate to differentiate time from the concrete succession of events? If the answer is yes, then what makes time go forward? If the answer is no, do events determine its flow?

Those questions are neither purely academic nor purely theoretical. They deserve to be thoroughly examined since concealing them creates a confusion between time and becoming, which leads to a kind of intellectual vagueness through repetition: concepts get lost, clouded and confused with one another. Moreover, this confusion constantly exposes our conception of temporality to metaphors that ultimately discourage us from thinking about time for what it is. By willing too much to harmonize the representation of time with the experience of becoming, common thinking in fact engenders confusion

between time and irreversible physical phenomena. At least for the sake of hypothesis, and without going back to a Newtonian conception of time, why not conceive of time as a more fundamental entity than temporal phenomena? A kind of skeletal being whose flesh would be made of events? The discreet background of any concrete becoming? Those are the questions we would like to discuss by examining the kind of answers physics provide…

For this purpose, we have studied the theoretical constructions specific to conventional physics (classical physics, quantum mechanics, special or general relativity). As we will see, this study shows that through its operational formalisms, physics distinguishes time from becoming. It even distinguishes them completely; on one side, there is *the course of time*, a primitive entity, on the other side there is the *arrow of time*, which is not a property of time but of the majority of phenomena taking place in time, specifically irreversible temporal phenomena.

The course of time establishes an asymmetry between past and future: if two events are not simultaneous, then one of them is earlier than the other one. It also establishes a difference of position (but not of nature) between past and future moments: on a timeline, tomorrow isn't set where yesterday is – a certain amount of time separates them definitely. *So defined, the course of time expresses the irreversibility of time itself.* As for the arrow of time, it represents the fact that some physical systems evolve in an irreversible way throughout time: they won't go back to their previous states. *So defined, the arrow of time expresses the irreversibility of phenomena within the course of time*. It is the concrete manifestation of becoming.

Today, physics has become so spectacularly effective that it is possible to imagine that the distinction it makes between time and becoming could be transferred to philosophy, which often aggregates the two notions. Especially because physics not only distinguishes them but almost opposes them. In a way, it considers the course of time as that which never becomes, in the sense that it never changes its way of renewing the present moment or, to put it differently, its way of being time. Therefore, almost ignoring the meaning of words and going against how we normally think of time when we don't really think about it, physics conceives of the existence, within the course of time itself, of a principle that remains and never changes: within passing time, there is something that doesn't pass, something that time doesn't affect.

2. TIME AND TEMPORAL PHENOMENA: DON'T JUDGE BY APPEARANCES

Because of a transference process between concepts, we don't distinguish between time and what happens within it: we automatically mistake the contents for its container. Because we notice that cyclical phenomena exist around us, we pretend that the course of time is cyclical. Or, if our schedules get crammed, if we manufactured objects at a frantic pace, we declare that it is time itself that's speeding up. Our notion of time is always of a time marked by events, by the phenomena it contains, a full time excluding any notion of emptiness or abstraction.

It is of course possible to pretend that those are only figures of speech, frequently used expressions that don't hinder our understanding of time in any way. But this would be to ignore that this shift, this simplification of language doesn't exclusively happen in daily life. Some philosophers have repeated it: as soon as they learned that scientists have discovered a new category of phenomena, they declare that time itself has therefore been modified even that it has just been the object of an ontological revolution.

On the subject of chaos theory, which shows that a system defined by deterministic equations doesn't necessarily have a predictable evolution, the French philosopher Michel Serres wrote: "Time does not always flow according to a line nor according to a plan but, rather, according to an extraordinarily complex mixture, as though it reflected stopping points, ruptures, deep wells, chimneys of thunderous acceleration, rending, gaps – all sown at random, at least in a visible disorder. Thus, the development of history truly resembles what chaos theory describes."

What does this mean? That the existence of chaotic phenomena necessarily implies that the course of time itself is chaotic? But chaos theory, as revolutionary as it first appears, can be explained through classical Newtonian mechanics. It even belongs to it completely. Neither the status nor the representation of time have been modified by this theory or even questioned by it: time in chaos theory is the same as Newtonian time. The discovery of a new typology of phenomena doesn't necessarily require a new conception of time in order to be characterized.

Similar remarks could be made about a book from 1979 that has had a lasting influence: *The New Alliance* by Ilya Prigogine and Isabelle Stengers, which is often presented as the manifesto for "*temps retrouvé*" (the rediscovery of time) supposedly because its authors conclude that: "Today's physics no longer denies time. It acknowledges the irreversible time of evolutions toward equilibrium, the rhythmical time of structures whose impulse feeds on the world that runs through them, the zigzagging time of evolutions through instabilities and amplifications of fluctuations, and even the microscopic time that reflects the indeterminacy of macroscopic physical evolutions.[2]" But has physics ever "denied" time? And what is "rhythmical time"? What about "zig-

[1] *Conversations on Science, Culture and Time*, Michel Serres with Bruno Latour, University of Michigan Press, 1995, p. 57.
[2] Ilya Prigogine, Isabelle Stengers, *La Nouvelle Alliance*, Paris, Gallimard, 1979, p. 275.

zagging time"? And what kind of time would not be "irreversible"?

To postulate that time is equivalent to what it contains, or even that it is generated by phenomena is to concede explicitly that a multiplicity of time exists: there would then be the same number of times than the number of temporalities. In this sense, time is labelled with several adjectives that specify the type of phenomena associated with it: there is a psychological time, a geological time, a astrophysical time, a subjective time, one or several historical times, even a "rhythmical time" or a "zigzagging time", because those phenomena inhabit time in a way that varies according to each… Such posture makes believe that multiple kinds of time can coexist. But could there seriously be a time for rocks and one for atoms, a time for stars and one for galaxies?

Ilya Prigogine found it useful to add another element to the list: the "entropic time", which is supposed to reflect the irreversible evolution of a system by quantifying the scope of changes that have affected it. This "time" is proportional to the variation of a system's entropy during a given irreversible process.[3] What does this mean?

As we know, the entropy of a system is a quantity that characterizes the system's ability to experience spontaneous changes: the bigger the entropy, the smaller the system's ability to transform itself. The second law of thermodynamics states that a closed system's entropy can only increase throughout time. This means that by changing, the system necessarily loses some of its ability to change even more. A closed system spontaneously tends toward a state of maximum entropy, in which any spontaneous transformation will become impossible. Let's take an example: the total entropy of a sugar cube and of a cup of coffee being inferior to the entropy of a cup of coffee with sugar, a sugar cube dropped into a cup of coffee has no choice but to be dissolved into the coffee. This phenomenon is irreversible: the sugar cube dissolved in the coffee cup will never return to its original square shape, nor will it recover its whiteness, and the coffee won't ever get its bitterness back. As such, the second law of thermodynamics is in accordance with the fact that physical phenomena seem to go in a truly specific direction.

Another way of defining entropy is to say that it measures the energy quality available within a system. The energy is conserved and can therefore only be transformed. But during this transformation, its quality lessens and it becomes less and less usable. Good quality energy is organized energy with little entropy, like the entropy of a waterfall whose overall falling movement can activate a water turbine. At the bottom of the fall, water molecules have lost the vertical organization they had during the fall due to the action of gravity. Their energy is now of lower quality: it can't be used so easily.

That being said, let's go back to Ilya Prigogine who, by using the entropic variation, links the course of physical time to processes that happen within it. In fact, he creates a kind of time that resembles the phenomena it contains and clearly depends on the system. Somehow, he "mixes up" the length of time during which the system changes and the intensity of these changes: if the system doesn't change, its entropic time doesn't increase; if it changes, its entropic time shifts from physical time and this shift depends on the production of entropy. The specificity of entropic time is that it stops as soon as the system's entropy becomes constant while physical time keeps on passing. Entropic time is similar to Bergsonian time: it only passes when it creates novelty. This time is only valid for the changes it brings and stops as soon as the system can't evolve anymore[4]. Besides, the course of entropic time has a clearly defined motor: it is "pulled" or "pushed" by the irreversibility of temporal processes that occur within it.

But Prigogine is far from having created a new time since his definition of entropic time explicitly relies on physical time. Entropy variation being the manifestation of any irreversible change, it can certainly be used to characterize a system's inner dynamics through its relationship to the duration of that system's change. But is it therefore judicious to measure a system's irreversible evolution using a variable called "time" if this variable has actually been set to proceed outside of…time? "Entropic time" appears as a misleading expression in the sense that it implies that entropy could be generated by time even though it is a quantity linked to *what is happening* throughout time.

This kind of confusion between time and temporal phenomena seems implicit, evident, but is it even relevant or justified? Can such assimilation of time with temporal developments be found among the formalisms of physics?

3. TIME AND BECOMING: PHYSICS SEES DOUBLE

Physics formalisms are mostly composed of equations, which condense fundamental relationships, reflect essential properties, and reveal things more profound than our human discourse can express however subtle it might be. But equations don't speak, at least in the common sense of the verb. So what can we say about time and becoming that the equations would say if they

[3]If dS is the variation of entropy for a system and P(t) the production of entropy per unit of physical time, then of course: $dS = P(t)dt$. The idea on which the concept of entropic time is based consists in building a "time" t* for the variation of entropy during the time interval dt to be proportionnal to dt*: $dS = P(t)dt = a.dt^*$, where a is a constant. We therefore have: $\int dt^* = (1/a)\int P(t)dt$.

[4]« Time is invention or it is nothing at all », Henri Bergson, *Creative Evolution*, Courier Dover Publication,1998, p. 341.

could speak?

Here is the short answer to this question: modern physics has been constructed through formalisms that make a distinction between time and becoming, or more precisely between the *course of time* and the *arrow of time*. Since its birth, that is since the apparition of Newtonian mechanics, physics has conceived a time that doesn't need something to happen somewhere to pass. Physics' operational effectiveness, its experimental successes have become so impressive that it seems right to believe that the distinction it establishes between time and becoming represents a "negative philosophical discovery" of the highest importance since it alters the terms in which the philosophical question of becoming is stated.

We are now going to explain the comment above. What does *course of time* and *arrow of time* mean?

The course of time only reflects the passing of time. In a way, it is time itself. It is represented by a line, a timeline on which a little arrow is usually drawn, *an arrow that is not the arrow known in physics as the arrow of time*. It is there to indicate that time has a single direction, and that time travel is indeed impossible; our presence on the timeline at a specific moment conditions our subsequent positions on that line. It is also impossible to come back, nor to go through the same moment twice.

We have to notice that this depiction of time as a line is fundamentally incomplete because it omits indicating how this line is built. Since the present does not bring another present by itself, there has to be something, an "engine" of time, to do this "work". This little engine is responsible of the course of time, in the sense that it continuously renews the present: without it, the newness of each instant could not arise. Where does this engine come from? Is it a property of time itself or a property of the arrangement of things in time? Is it linked to a global property of the Universe or to our consciousness? The answers to these questions have still to be elucidated, so that we have to consider that the mystery of time exists less in the line through which we represent it than in the hidden dynamic that builds this line.

The course of time has a structure, which by itself guarantees that each event is necessarily definitive. As soon as the event occurs, this fact can't be erased by anything. If the imprint that the event might have left can be erased, this erasure will change the close or distant future, but not the past. More generally, any present action can only have consequences in the future, never in the past. The course of time thus maintains a constant possibility of distinguishing the past from the future.

As for the arrow of time, what is it if not the little arrow placed on a timeline? Contrary to what the expression might suggest, it is not related to time itself but to what happens within it. It is not an attribute of time, but a potential property of physical phenomena; most of what exists at our scale is transformed irreversibly throughout time and can't return to its original state. The dynamics of those physical phenomena is then marked with an arrow, wrongly called the "arrow of time" (since this arrow is linked to the dynamics, not to the time itself).

The problem of the arrow of time is often summarized by the following question: why do we remember the past and not the future? The answer usually given is that the only way of distinguishing between past and future is by means of the second law of thermodynamics. But in fact, the question asked doesn't concern the arrow of time since the invocation of the course of time is enough to answer it: if we do not remember the future, it is because we have not yet been present in… the future! Asking "Why are we in a different state in the future than in the past?" is quite another question (whose answer can be, this time, the second law of thermodynamics) that has to be distinguished from the first one.

This example of confusion shows that it is worthwhile to emphasize the difference between several issues traditionally subsumed under the label "the problem of the direction of time". The most invoked concepts are the concepts of irreversibility and of time-reversal invariance. *Time-reversal invariance* is a property of physical laws: a law is time-reversal invariant when it is expressed by a differential equation which is invariant under the transformation $t \rightarrow -t$. By contrast, irreversibility is a property of processes: a process is irreversible if it is always observed in the same temporal order, and never in the inverse one. The problem of the arrow of time consists in finding out how irreversible processes can be explained by means of time-reversal invariant laws (as we see, this problem is conceptually different from the explanation of the origin of the course of time, and also from the description of the engine of time).

To better grasp this distinction between the course of time, the arrow of time and the engine of time, we should remember that since Newton, the principle of causality has always constrained *from the outside* the representation of the course of time in physics. This principle is generally summarized by saying that every event has a cause that precedes it, but a point should be clarified here.

Even if it might not seem so, the notion of cause, in the strict sense of the term, gave a lot of trouble to physicists who ended up abandoning it. After playing an essential role in 17th and 18th century physics, its importance declined with the emergence of probabilities in statistical physics. In the 20th century, quantum physics definitely wiped it out. Indeed, quantum physics uses probabilities in a way that precludes referring to a cause, in the strict sense of the term, when talking about quantum processes. Therefore, the notion of cause slowly disappeared from scientific theories in favour of the notion of physical laws, or was absorbed by the dynamics of systems.

The fact remains that the principle of causality, even freed of the notion of cause, has enabled the development of modern physics' theories, and deeply structures those currently being elaborated, such as superstring theory. It sets an absolute and compulsory order between several types of phenomena, even if none can be presented as the cause of another, and thus imposes a direction to time. In short, the principle is reduced to a classification method for events that fall under its influence, a rule that organizes them according to a systematically constraining order.

In practice, the different formalisms of physics adapt the principle of causality to themselves by giving it a form that depends on how events and phenomena are represented. Its consequences are always constraining. In Newtonian physics, causality implies that time is linear and non cyclical (which is enough to guarantee that an effect can't influence its cause retroactively). In special relativity, it posits that a particle can't travel faster than the speed of light (which is enough to render travelling to the past impossible). In non-relativistic quantum physics, causality is guaranteed by the structure of Schrödinger's equation[5]. In particle physics, causality made it possible to predict the existence of antimatter, and it is now formally expressed by CPT invariance to which the dynamics of physical phenomena must respond. What does CPT invariance represent? The fact that physical laws ruling our universe are perfectly identical to the rules of a universe in which matter and antimatter would interchange their roles, observed in a mirror, and where time would go backward (CPT invariance imposes that the mass and lifetime of particles must be strictly equal to the mass and lifetime of their antiparticles).

Putting aside their most technical aspects, these declinations and implications of the principle of causality are quite clear, so clear that they tend to hide a fundamental conceptual problem. In fact, the notion of causality can't be thought of, or even defined, outside of events that embody it. How it relates to the course of time then becomes partly ambiguous: if the principle of causality constrains the course of time, this means that the latter is indirectly "contaminated" by causally related phenomena that occur within it. In other words, the principle of causality (partly) aggregates the course of time and temporal phenomena despite the distinction established between them.

But we would like to emphasize one thing: in every physical theory, once the course of time is dependent upon the principle of causality, it then becomes completely irreversible, in the sense that an instant can't occur twice. This irreversibility can never be compensated or erased by the reversibility of any movement or dynamical process; as fast as one can possibly return from Paris after being to Genève, time has irreversibly passed during the trip and one is therefore a bit older. More generally, the absence of the arrow of time doesn't stop the hours from passing.

The course of time possesses a direction that is quite different from the arrow of time. When it exists, the arrow of time appears *in addition*, "filling up" the irreversible course of time with irreversible phenomena. We shall later see that physicists have identified possible explanations for the irreversibility of phenomena. All of them presuppose the existence of a set course of time within which time-oriented phenomena take place.

While time passes, the course of time in itself doesn't change (all successive instants are equal in the sense that they have the same status with respect to physical laws). It doesn't change throughout time its way of being time. Thus it escapes becoming (the variable t, designing the time, doesn't vary according to time). It is the arrow of time that constitutes the true expression of becoming. It manifests itself within the course of time, which it doesn't affect in any way but which it overruns with mostly irreversible phenomena. In some respects, the notion of "the course of time" therefore precedes the notion of becoming.

This is also what the second law of thermodynamics suggests: the assertion that the entropy of isolated systems going through a spontaneous transformation only increases with time implicitly presupposes that the transformation in question follows the direction of a time that leads us from our "past" toward our "future", and not in the opposite direction. It presupposes that the course of time has first been defined.

It is possible to illustrate the distinction between time and becoming, to make it visible. Let's look at the work of Roman Opalka who, every day since 1965, has been painting a series of integers on canvases then photographing himself after each work session. The succession of numbers materializes the irreversible course of time, which happens even if nothing happens; if each number drawn is new (and each moment is completely new), it is always obtained by adding a unit to the preceding number. As for the photographs the artist takes of himself regularly under unchanging conditions (on a white background, with a white shirt, under a white light, with the same facial expression), they show a series of physical changes over time, that is to say the irreversibility of his own becoming. On the one hand, the course of time is represented by the succession of numbers and the accumulation of canvases; on the other, becoming is represented by a series of photographs of the same being changing. This dual representation is enough to prove, if not to demonstrate, that these two kinds of irreversibility, which always appear entangled to the point that they seem whole, can in fact be separated. But since they're so often combined and have simultaneous effects, it is

[5]In quantum physics, the Hamiltonian is the mathematical operator that describes a physical system's evolution throughout time. Schrödinger's equation makes this operator into the infinitesimal generator of time translations. The principle of causality is therefore respected.

hard for us not to confuse them even though they are always just superposed.

4. WHERE DOES THE ARROW OF TIME COME FROM?

Why are some temporal phenomena irreversible or others not? When a phenomena is irreversible, that is to say when an arrow of time appears, what is its origin?

The arrow of time wasn't part of physics fundamental formalisms from the start, neither in classical mechanics, nor in quantum physics and the theory of relativity. Therefore how do we understand it? Where could it have come from?

The question appeared only a century and half ago, when physicists started to ask themselves if physical phenomena could "go in both directions": can a dynamic process capable of changing a system from a state A to a state B make it change from a state B to a state A? This question was born of the conjunction of two apparently contradictory observations:

1) Daily, we can observe around us many physical processes for which corresponding reverse processes have never been observed or are exceptional. Therefore these are, by definition, irreversible phenomena.

2) Yet none of the dynamics laws that govern these processes contain temporal asymmetries, that is to say that they would be the same if the course of time was going in the opposition direction. If they allow a certain process to occur when time goes into one direction, they allow it to happen when it goes into the opposite direction: the initial and final states could be interchanged. Such equations are called "T-invariant equations": if a system can go from state A to state B, it should be able to go from state B to state A (in that case, the system isn't concerned with the arrow of time).

Therefore, why are there some irreversible phenomena? Why is there an arrow of time, that is to say an asymmetry in the dynamics of certain phenomena that we observe, even though the equations of physics have no room for it?

In view of what we have stated above, these questions can't be answered by explaining "the direction of time", by setting out the reasons why it flows in one direction rather than another, or even less by explaining why we don't remember the future. The issue is solely related to the *asymmetry of physical processes* within time and not to the asymmetry of time itself. It is an asymmetry of the "contents" of time, not an asymmetry of the container itself.

To try to solve this riddle, physicists advance four categories of argument that can delimit the origins of the arrow of time, and they also study their possible inter-relations. We will present them briefly:

- *The second law of thermodynamics*, or the increase of the entropy of isolated systems. In Boltzmann's interpretation, which underlies this principle, there is no arrow of time at the microscopic level, but on a macroscopic level, one can get the impression that one exists.
- *The process of measurement in quantum physics*, which has been the subject of intense debate for eighty years. Generally, it is understood as a temporally asymmetrical process.
- *The violation of CP symmetry during certain phenomena governed by the weak interaction:* some unstable particles, for example neutral kaons, don't behave exactly like their anti-particles. More specifically, they don't disintegrate into other particles at the same pace than their antiparticles. This means that they disintegrate according to a temporally asymmetrical law. The fundamental reason for this temporal asymmetry, which remains hard to interpret, is not completely understood. It raises the question of the existence of an "arrow of time" at the microscopic level ;
- *The expansion of the universe*, which would make it impossible for any system to return to its initial state because the universe itself is evolving. This can appear contradictory since the equations of general relativity are temporally symmetrical, but in reality their cosmological solutions, which are supposed to govern the evolution of the universe, are not. The universe they describe is either expanding, or contracting, as represented by the existence of an arrow of cosmic time related to the conditions at the limits of the universe. Some theorists, including Stephen Hawking and Roger Penrose, think that this arrow of time could be the arrow mastering all the others, but not all physicists share this position.

We've gone far enough to be able to make two remarks.

The first one is that the attempts to explain the arrow of time resort to arguments that all differ from the restrictions imposed on the course of time by the principle of causality. (We mentioned them earlier: linear time, the impossibility of going beyond the speed of light, the existence of anti-matter, CPT invariance). *In conclusion, the course of time is accounted for in ways that never coincide with ways in which the arrow of time is justified*. This indicates – or even demonstrates – that the course of time and the arrow of time are two distinct things in contemporary physics; *the irreversibility of phenomena doesn't come from the irreversibility of time and vice versa.*

The second remark is that none of the explanations given for the arrow of time is likely to constitute a real theory. They are closer to an *interpretation* of such or such physical theories, but are not *incorporated* into any formalism. There is indeed no operating physical theory that integrates becoming from the start (through the use

of irreversible fundamental equations). Consequently, becoming can only be accounted for in physics through the reading of theories that don't include it among their principles. So interpretations of the arrow of time's origins end up mixing physics and philosophy. Thus, they can be subject to disagreement and are indeed very ardently disputed. These disagreements are not without similarities to the debate between supporters of Parmenides and supporters of Heraclitus. Some physicists think this is only a fake problem: on the pretext that no arrow of time appears in physics' fundamental equations, they believe, like Parmenides, that becoming is only pure appearance and is closely related to how our limited senses make us perceive the world. Others, following the Heraclitian tradition, consider that because actual physics can't explicitly account for becoming, it is either wrong, or incomplete.

These two positions can be defended as long as there is agreement on the meaning of words. And also as long as no one is claiming that physics has negated time just because its formalisms don't include the arrow of time. For if becoming wasn't integrated directly into its principles, physics has always referred to the course of time. One can regret that physics hasn't integrated becoming from the start – or better suggest how physics could make room for becoming in its formalisms -, but it can't be blamed for forgetting to integrate the course of time.

Although, *on paper*, it is possible to change the sign of time in a physical equation, this doesn't imply that the course of time can be physically reversed. Only the direction of phenomena can be physically reversed, not the direction of the course of time.

However let us be clear: it is not excluded that the flow of time and the arrow of time come from the same source, more profound than they both; that they are by-products of underlying phenomena that a "new physics" might reveal. Moreover, some progress has recently been made in that direction, in characterizing causality independently of any concept of time and deriving both time and becoming from an ordering relation on sets of events taken as primitives. It may thus appear that causality cannot be understood as a feature of the world that would exist independently of any phenomenon. Causality would be intrinsically shaped by the phenomena. As the principle of causality underlies our representation of time in physics, this would give some formal foundation to a close connection between time and becoming. But for the time being, it would be wiser to formally distinguish them in order to make the arguments clearer.

This is not always the case. For example, in the cosmological context, it is often told that the problem of the arrow of time owes its origin to the intuitive asymmetry between past and future. According to what we have shown, the expression "course of time" would be there more appropriate. Because of this confusion, many authors argue that the direction past-to-future is related to the direction of the gradient of the entropy function of the universe, forgetting that the definition of the entropy of the universe is a very controversial matter: entropy is a thermodynamic magnitude that is typically associated with subsystems of the universe, and not with the universe as a whole. To evade this difficulty, some authors[6] have tried to define a course of time for the universe only on the basis of the geometrical properties of space-time, independently of any entropic consideration. Unfortunately, they have called this direction past-to-future the "global arrow of time". The expression "global course of time" would have suited better.

5. WHAT MAKES THE TIME FLOW?

The nature of the "engine of time" that makes us feel the flow of time has not been elucidated but a lot of theoretical work is now being devoted to this problem. Different avenues are being explored. In fact, there have been three major theories of time's flow. The first, and most popular among physicists, is that the flow is an illusion, the product of the faulty river metaphor. The second is that it is not an illusion but rather is subjective being deeply ingrained due to the nature of our minds. The third is that it is objective, a feature of the mind-independent reality that is to be found in, say, today scientific laws, or, if it has been missed there, then in future scientific laws.

The first theory, rooted in the theory of relativity, represents space-time as a fixed whole and suggests that the flow of time is a pure illusion: the entire universe just is, with no special meaning attached to the present time. All past and future times are equally present, have the same degree of existence along time, just as different locations coexist along space. According to this view, there is nothing special about the "now". Incidentally, in the special theory of relativity, there is an uncountable infinity of nows, and the standard symmetries assure that none of them can have special significance.

In the second theory, time would only be a psychological feature linked to the very complex structure of our brain; in the space-time region we are observing, we have the feeling that time passes "from the bottom to the top" of space-time, but in reality the space-time is a rigid block without any internal dynamics. We observers would unfold the thread of time ourselves. In other words, we would be the "engine" of time.

The third theory considers that time's apparent flow is real, that it corresponds to a true physical reality. At any moment in time an observer perceives a "now"; future events are not only unknown but objectively

[6]See for example: Mario Castagnino, Olimpia Lombardi, Luis Lara, *The Global Arrow of Time as a Geometrical Property of the Universe*, Foundations of Physics, Vol. 33, N° 6, June 2003; G. Matthews, "Time's arrow and the structure of space-time", *Philosophy of Science* **46**, 82-97 (1979).

non-existent, to be created later as the now advances. Thus physics should grant time's flow a well defined place in its formalisms[7].

It is not our purpose here to discuss these theories in detail or to argue for or against any one of them. We merely wished to stress that the common semantic carelessness when it comes to the expressions "course of time", "direction of time" and "arrow of time" makes the arguments of all parties more confusing than they really should be. If these expressions were better defined, systematically distinguished from one another and always used in their strictest sense, the debate about time, irreversibility and becoming in physics would become clearer.

6. CONCLUSIONS

What can these considerations teach us? That a more carefully chosen vocabulary and a more rigorous conceptualization would give us a chance to show how the different theories formalize the course of time, interpret the arrow of time, and relate time and becoming. It would allow us to better think about the question of time.

Therefore, the principle of causality could benefit from being renamed "antecedence principle" or "principle of chronological protection", as Stephen Hawking proposed it. Similarly, when referring to a physical process, the quite awkward expression "time reversal" could be replaced by the expression "movement reversal", since the intention is not to create a time machine but to reverse the speed of the physical entities concerned. When a phenomenon's dynamics are reversible, the course of time (or the direction of time) is indeed arbitrary, but once it has been chosen, it can't simply be reverted.

Finally, the situation is the same with the course of time as with electrical charges. Saying that the electron carries a negative charge and the proton a positive one results from a convention. To change this convention and declare that an electron's charge is positive and a proton's negative wouldn't change anything to the laws of physics or the universe. Beginning with a conventional choice, it is therefore possible to design physical laws that are unconventional.

To claim that the course of time doesn't exist according to physics under the pretence the laws of physics are time-reversal invariant (so that the direction of time is arbitrary) is equivalent to saying that electrical charges have no reality because physical laws don't change if each charge's sign is reversed.

REFERENCES

[1] Davies, P.C.W. (1974) The physics of time asymmetry. University of California Press, berkeley-losangeles.

[2] Hawking, S. (1994) The no boundary proposal and the arrow of time. In Halliwell, J., Pérez-Mercader, J., Zurek, W., Ed. Physical origins of time asymmetry. Cambridge University Press, Cambridge.

[3] Haddad, W.M., Chellaboina, V. and Nersesov, S.G. (2005) Thermodynamics: A dynamical systems approach. Princeton University Press, Princeton.

[4] Horwich, P. (1987) Asymmetries in time, MIT Press, Cambridge.

[5] Klein, E. (2005) Chronos, How time shapes our universe. Thunder's Mouth Press, New York.

[6] Price, H. (1996) Time's arrow and archimedes' point. Oxford University Press, new tork.

[7] Prigogine, I. (1980) From being to becoming. Freeman, W. H. San Francisco.

[8] Reichenbach, H. (1956) The direction of time. University of California Press, Berkeley.

[9] Sachs, R.G. (1987) The Physics of time reversal. Chicago University Press, Chicago.

[10] Savitt, S., Ed. (1995) Time's arrows today, recent physical and philosophical work on the direction of time. Cambridge University Press, Cambridge.

[11] Zeh, H.D. (1989) The physical basis of the direction of time. Springer Verlag, Berlin.

[7]See for example "*Becoming as a bridge between quantum physics and relativity*", A. C. Elitzur, S. Dolev, in *Endophysics, Time, Quantum and the Subjective*, 589-606, R. Buccheri *et al* (eds), World Scientific Publishing Co., 2005.

Vol.2, No.3, 220-227 (2010)
doi:10.4236/ns.2010.23034

Natural Science

Protein phase instability developed in plasma of sick patients: clinical observations and model experiments

Tatiana Yakhno

Department of Radio-Physical Methods in Medicine, Institute of Applied Physics RAS, Nizhny Novgorod, Russia;
tanya-yakhno@rambler.ru

Received 6 November 2009; revised 1 January 2010; accepted 26 January 2010.

ABSTRACT

This article discusses the causes of formation of micron-size protein structures in liquid plasma or serum of the patients with different diseases, which are accompanied by inflammatory reactions. Self-organizing processes in sessile drying drops of natural and model biological liquids are used for study of possible mechanisms of development the protein phase instability in serum. There was shown that violation of optimal ratio between albumin and osmotic active components could lead to loss of albumin aggregative stability and albumin coagulation structures formation. Possible role of these structures in pathogenesis of inflammation is discussed.

Keywords: Protein Aggregation in Serum; Mechanisms of Inflammation

1. INTRODUCTION

In 2003 we reported that micron-size protein structures were observed in liquid samples of serum and plasma of patients with different diseases in acute stage [1]. We could not give a correct explanation of this phenomenon at that time, but it seemed to be very interesting and important for blood physiology as one of possible nonspecific reactions of blood proteins accompanying inflammation. When these samples of plasma or serum were placed in the form of drops on glass for drying, the protein structures usually partly melted, and formed a structured film consisting of agglutinated particles. In serum and plasma of healthy people such structures appeared only at a certain stage of drying, forming a light-diffusing ring around the central part of the dried drops [2]. Thus, we suggested that appropriate conditions for protein structure formation in normal biological fluids were developed at a definite moment during drop drying. We believe that these conditions may be found in model experiments using protein-salt water solutions.

Recently, protein phase transitions are widely studied in the bulk of model solutions by means of static and dynamic light scattering, small-angle neutron and x-ray scattering [3-8]. Phase separation is achieved by adding precipitation agents: inorganic salts [4,5,8], polyelectrolytes and organic solvents [3,4,8], as well as by changing the concentration, temperature, and pH level [4,5,7,9]. Experimental studies show that protein precipitation by salts requires an electrolyte concentration in the range of 1-10 mole [9]. Protein aggregation in cells and biological fluids is also a subject for investigation from the point of view of crowding, because biochemical processes proceed in a medium with high concentration (50-400 mg/l) of macromolecules of different types [10-12]. This may lead to formation of nonfunctional protein aggregates, which upset balance between the active components, corrupting their functions [12] and may even be toxic [13].

The critical protein charge required to induce protein-polyelectrolyte complex formation between bovine serum albumin (BSA) and some synthetic polyelectrolytes was found to vary linearly as a function of square root of ionic strength [14]. Formation of intrapolymer complexes between human serum albumin (HSA) and polyethylene glycol (PEG) showed formation of a water-soluble complex the size of which varied depending on both, ionic strength and molecular weight of PEG but remained unaltered when the mixing ratio of PEG and HSA was varied [15]. Serum glycosaminoglycans, such as heparin, chondroitin and glucuronic acid, are native polyelectrolytes [16], so it is reasonable to expect their complex formation with HSA under certain conditions. It is well known that abnormal protein aggregation is responsible for a variety of serious diseases, including eye cataracts [17-19], sickle-cell anemia [20] and Alzheimer disease [21]. It was shown also that the dynamics of phase transitions in sessile drying drops of biological liquids can be used as a sensitive parameter for medical diagnostics [22-24].

The goal of this work is to attract the attention of medical blood researchers to the phenomenon of protein structure formation in native human biological fluids and

to display factors that may be responsible for this phenomenon using the drop drying model. Here we show the data of model experiments with protein-salt water solutions, and compare them with the data of our clinical investigations carried out in 2000-2003.

2. MATERIALS AND METHODS

We used bovine serum albumin solutions (BSA, 68 kDa, Sigma, USA) in distilled water or in salt-water solutions with different protein/salt ratios. Salts (NaCl, KCl) were labeled "chemically pure" ("Reaktiv, Inc.", Russia). Experiments were also performed with lyophilized preparations of human serum proteins supplied by OOO IMTEK, Russia: 1) Human serum albumin (HSA, Sigma, USA, #A-1653, 67 kDa, pI ≈ 4.7, *E* (280 nm, 1%) = 6.7); 2) immunoglobulin G (IgG, 150 kDa, *E* (280 nm, 1%) = 14); 3) immunoglobulin M (IgM, 900 kDa, pI ≈ 7.2, *E* (280 nm, 1%) = 12); and 4) fibronectin (Fn, 420 kDa, *E* (280 nm, 1%) = 13). Also 10% wt food gelatin solution in distilled water or in physiological salt solution was used. To clarify a contribution of macromolecular serum components to albumin aggregation, model protein mix was prepared (**Table 1**). The mixed sample was compared with pure HSA-salt water solution of the same concentration. All solutions were prepared without buffering, a day prior to experimentation, refrigerated overnight and allowed to come to room temperature before testing. The samples under study were placed, using micropipette, onto clean glasses in the form of drops 3 μl in volume (6-8 drops for each sample), and let for drying in room conditions. Morphological observations were carried out during drying, and 2-3 days after placing on the glasses, using МБС-10 and Люмам-И-3 microscopes with video camera computer setup.

Samples of blood plasma and serum were obtained from 30 clinically healthy donors (the material supplied by Hemotransfusion Station, Nizhny Novgorod); 18 patients with viral hepatitis B and C in acute stage (the material supplied by the Hepatological Center, Nizhny Novgorod); 30 patients with burn disease, and 8 patients with diseases of articulations of inflammatory and degenerative character (supplied by the Federal Burn Treatment Center, Nizhny Novgorod Research Institute of Traumatology and Orthopedics); 40 women after normal or premature (second- and third-trimester) childbirth (supplied by the maternity and child-welfare services of Nizhny Novgorod); one patient with Waldenstrom's macroglobulinemia, and seven patients with paraproteinemic hemoblastosis, with genetically modified IgG (supplied by the Research Institute of Epidemiology and Microbiology, Nizhny Novgorod). The samples were transported at + 4°C, heated up to room temperature, and then analyzed. The test fluids were applied on chemically clean glass slides either as small drops (six to seven 5-ul drops per slide) or as large drops (0.5 ml per slide). The slides were dried in room conditions (T = 18 – 22° C, H = 60 – 70%) for 24-48 h. Thus, drying of the fluids occurred under different thermodynamic conditions depending on their volume and the shape of the drops. Some of the samples (small droplets) were dried under an МБС-10 microscope fitted with a television camera connected to a computer, so that phase transitions in the fluid could be recorded. After drying, the drops were studied in a Люмам-И3 or МБС-10 microscopes under conventional illumination conditions. Biochemical data of blood of the patients under investigation were obtained from the corresponding clinical laboratories, and statistical data were calculated using the Excel program.

3. RESULTS

The dried drops of protein solutions had a disc-like shape with a thick protein ring around the edge and a thinner circular central zone. Salt-free solutions formed homogenous transparent dried spots (**Figure 1(a,b)**), while protein-salt water dried drops had distinct concentric zones: external protein ring, ring of light-diffusing protein structures, ring of clear gel, and central zone of shrinking gel (**Figure 1(c,d)**) due to salt crystallization in it [2]. The above-named protein structures became visible mostly on the second day of drying, after partial evaporation of film water. So, when a drop is watched from the edge to the center in a transmitted light microscope, a homogenous layer of protein film is visualized first, then single micron-size protein aggregates are settled on the protein layer, after that fractal complexes are formed of these separated structures which finally turn into gel (**Figure 1(e,f)**). This principal sequence is the same for globular (BSA, HSA) and non-globular (gelatin) protein-salt water solutions regardless of protein and salt type and con centration. The difference is in morphological features and in the width of these zones, depending on the ionic strength/protein ratio. The protein ring in

Table 1. Content of model protein solutions in 0.9% NaCl water solution.

№	Sample content	Total protein, g/l	HSA, g/l	Fn, g/l	IgG, g/l	IgM, g/l
1	HSA + Fn + IgG + IgM	84.5	70.0	0.3	12	2.2
2	HSA (control)	84.5	84.5	0.0	0.0	0.0

dried drops moves to the edge when the salt/protein ratio in the initial solution increases (**Figure 2**).

Earlier we found that IgG, IgM and Fn exert a considerable effect on both, the morphological and dynamic parameters of fluid structuring in drying drops through a change in the surface tension and viscoelastic properties of the adsorption layer at the fluid–air interface [25,26]. The goal of the current work is to elucidate how these macromolecular serum components influence protein structure formation during drop drying. We replaced part of HSA in physiological salt solution by different serum macromolecules with the same weight proportions (**Table 1**), and saw a different type of protein structures in dried drops in comparison with the control sample (**Figure 3**). They looked like large-scale agglutinated particles following the homogenous protein ring at the drop's edge. We chose the concentrations of the macromolecular components approximately close to their normal concentration in blood [27].

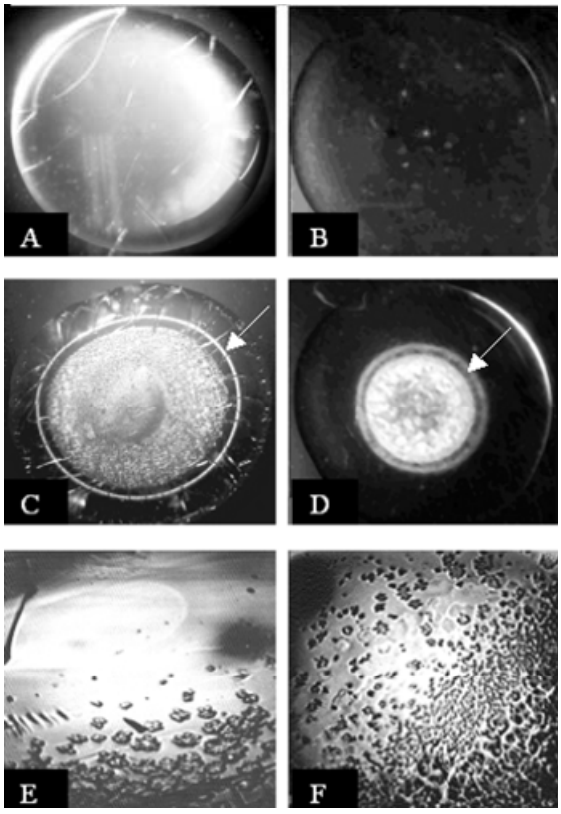


Figure 1. Dried drops of protein water solutions and their fragments. (a) – 7% wt BSA in distilled water; (b) – 10% wt gelatin in distilled water; (c) - – 7% wt BSA in 0.9% wt NaCl solution; (d) - 10% wt gelatin in 0.9% wt NaCl solution; (e) – fragment of a peripheral zone of a dried drop C with protein structures on the homogenous protein layer; (f) – fragment of a peripheral zone of a dried drop d with protein structures on a homogenous protein layer. Arrows (c, d) show light-diffusion rings of protein structures. a-d shown at × 28 original magnification using МБИ-3 microscope. (e) and (f) shown with original magnification x 140 (e) and 70 (f) using Люмам-И3 microscope and video-camera computer setup.

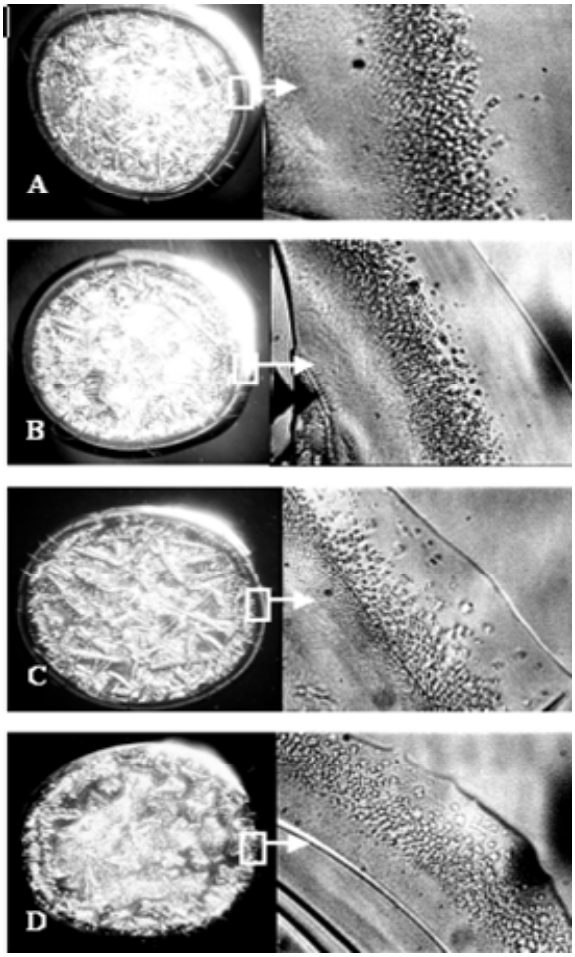


Figure 2. Dried drops of NaCl/BSA water solutions with different NaCl/BSA ratio (right), and their peripheral zones, are labeled with rectangle (left). NaCl/BSA weight concentrations are the following: (a) – 1.3/6.0; (b) – 1.5/5.0; (c) – 1.8/4.3; (d) – 2.2/3.4. Increasing in salt/albumin ratio leads to moving protein structures zone to the drop's edge (left). (a-d) (left) shown at × 28 original magnification using a microscope МБИ-3. (a-d) (right) shown at origin magnification × 70 using a microscope Люмам-И3 and video-camera computer setup.

These observations suggest that protein structures begin to form at a definite stage of drying of protein-salt water solutions, and this phenomenon is inherent in all the samples under study. Ionic strength, protein concentration and serum macromolecule components present in the initial solution influence structure formation. Surprisingly, sometimes we could see micron-size protein structures in liquid samples of plasma or serum of patients with different diseases (**Figure 4(b-d)**). They were never observed in liquid samples of healthy people (**Figure 4(a)**), but were usually formed at a definite stage of drop drying. Peripheral protein zones in dried drops of blood serum of healthy patients had the same order as

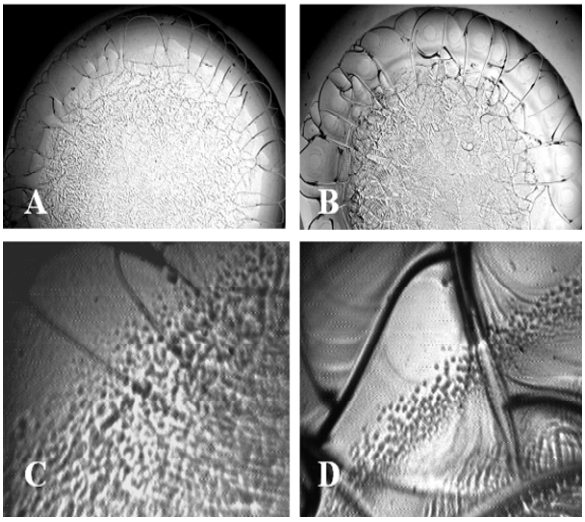


Figure 3. Influence of macromolecular serum components on protein structure formation in drying drops. (a) – a dried drop of mixed protein solution (HSA+Fn+IgG+IgM) in 0.9% wt NaCl water solution with total protein concentration 84.5% wt; (c) – fragment of peripheral zone of the drop a; (b) – a dried drop of 84.5% wt HSA solution in 0.9% wt NaCl water solution (control, see **Table 1**); (d) – fragment of peripheral zone of the drop b. In contrast to the control sample, zone of protein structures in c looks as large-scale agglutinated particles. a and b shown at × 28 original magnification using МБИ-3 microscope. c and d shown at original magnification × 70 using Люмам-И3 microscope and video-camera computer setup.

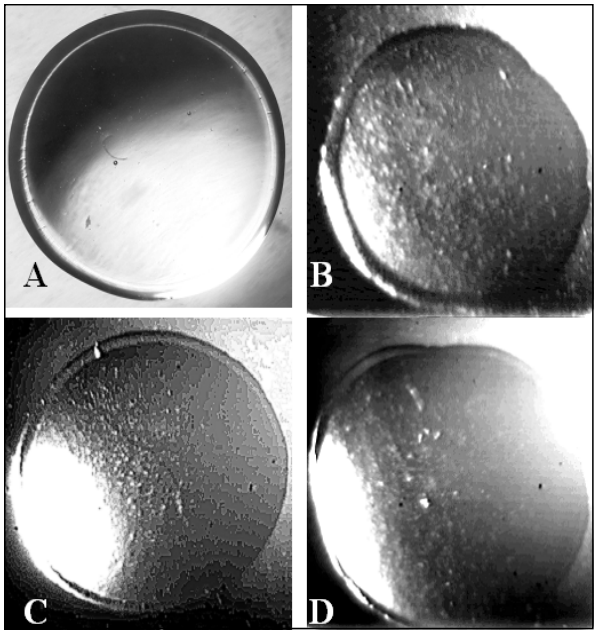


Figure 4. Fluid sessile drops of biological liquids at the very beginning stage of drying. (a) – serum of practically healthy donor; (b) – serum of patient with chronic hepatitis b and chronic hepatitis c; (c) – plasma of patient with burn disease; (d) – serum of patient with coxarthrosis. Shown at original magnification using a microscope МБИ-3. Drop diameter is 5 mm.

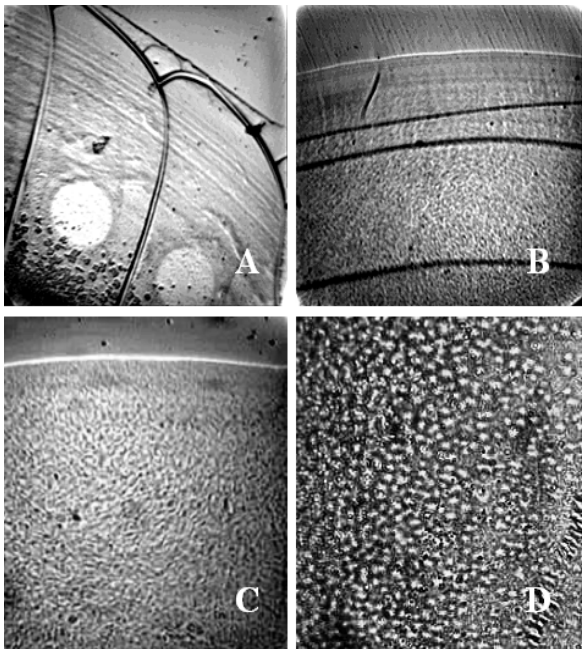


Figure 5. Peripheral zones of dried drops of serum. (a) – practically healthy donor; (b) and (c) – two patients with burn disease; (d) – patient with paraproteinemia (overproduction of genetically modified IgG). The drop a has a homogenous protein film over the edge, then zone of protein structures begins. In the drops b-d zone of homogenous protein is absent. This zone is represented by agglutinated protein particles. Shown at original magnification × 70 using Люмам-И3 microscope and video-camera computer setup.

model protein-salt water solutions: homogenous protein, ring of protein structures, and gel (**Figure 5(a)**). In the serum of patients with burn disease and paraproteinemia, the very first zone
was a structural protein film instead of a homogenous one (**Figure 5(b-d)**). This film consisted of a lot of small agglutinated protein structures. Microscopic investigation of big (0.5 ml) drops showed that irrespective of the disease type, these protein structures didn't differ from each other in principle (**Figure 6**).

Albuminemia is a usual state of acute-phase reactions of different diseases [28]. Thus, average content of total protein and albumin was estimated by the example of burn disease patients and healthy donors. Also, plasma osmolarity was calculated using the empiric equation:

Osmolarity (mosm) = 195.1 + 0.74 Na (mM/l) + 0.25 N of blood urea (mg %) + 0.03 Glucose (mg %).

Figure 7(a-c) shows the results of the calculations. Thus, burn disease was characterized by total proteinemia due to albuminemia and a tendency to excess of plasma osmolarity. It is clear that the ratio of osmotic active components to albumin content in plasma increases in burn disease, and tends to norm during treatment (**Figure 7(d)**). At the same time, the peripheral zone of dried serum drops becomes more homogeneous.

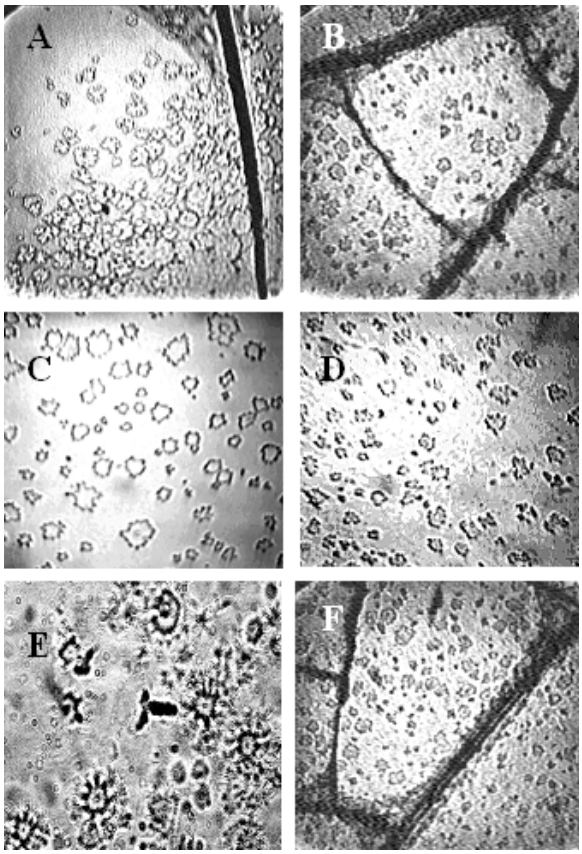


Figure 6. Protein structures in dried samples of serum of patients with different diseases. (a) and (b) – two patients with acute viral hepatitis b; c and d – two patients with burn disease; (e) – a patient with Waldenstrom's macroglobulinemia; (f) – a woman after premature birth. Shown at original magnification × 70 (a-d and f) and 140 (e) using Люмам-И3 microscope and video-camera computer setup.

4. DISCUSSIONS

Desiccated sessile drops of colloidal suspensions are good objects for studying liquid instabilities and non-equilibrium pattern formation [29]. The main principles of deposit ring formation over the circle ("coffee drop deposit"), as well as pattern formation, depending on particle size, concentration, ionic strength, and surfactant presence [30] were revealed in drying drops of model colloidal suspensions of polystyrene microspheres in water. Detailed description can also be found in [31,32]. The point is that a redistribution of protein content along drop radius occurs at the beginning of drop drying: up to 70% of albumin moves to the drop's edge and solidifies there, forming a ring, whereas the central part of the drop is still a liquid [2]. At the same time, salt concentration enhances in the liquid center due to water evaporation, and the salt/protein ratio increases. It can lead firstly to liquid-liquid protein separation in the form of coalescention, and subsequent solidification of neogenic micelles because of outside osmotic pressure increases. AFM investigation of appropriate zone of a dried drop of a protein-salt water solution allowed finding a structure that seemed to be a precursor of a micron-size protein aggregates that were observed via optical microscope. It consisted of rounded substances–subunits with a radius about 50 nm [2]. Then, those quasi-solid albumin particles began to aggregate into micron-size structures forming fractal clusters, which finally transformed into gel. Thus, a gradual increase of the salt/albumin ratio was a driving force for this cascade of albumin structures formation in drying drops of albumin-salt water solutions. In a similar manner, increasing of osmolarity/albumin ratio in the samples of biological liquids was one of the main promoters of structure formation too. Increasing in concentration of macromolecules of different nature also promoted loss of colloid stability. The capacity of large unfolded protein molecules to bind and transform water to a structured state [33,34] decreases the volume of free biologically active water, which plays an important role in prevention of coagulation. On the other hand, surfactants and polyelectrolytes under appropriate conditions can interplay with albumin, forming coacervates.

Cooper, C.L., *et al*. [35] (**Figure 4**) succeeded in viewing by eye in a tube coacervate/dilute equilibrium phase separation and polyelectrolyte-protein complexes formation (BSA, labeled with fluorescein isothiocynate, and poly (diallyldimethylammonium chloride).

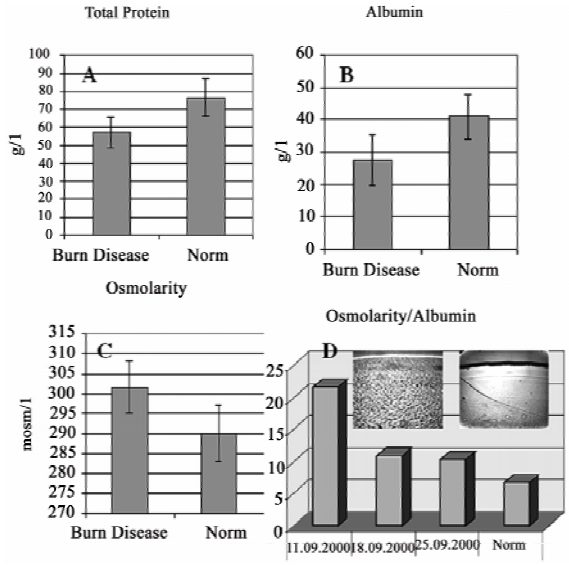


Figure 7. The data of serum protein content and osmolarity in norm compared with patients with burn disease. (a) – Total protein (g/l); (b) – Albumin (g/l); (c) – osmolarity (mosm/l); (d) – the dynamics of osmolarity/albumin ratio during treatment in a patient with burn disease. Microscopic images show the peripheral zones of dried drops at the beginning of treatment (left) and after two weeks of treatment (right). One can see the transformation of a structural protein film to a homogenous film.

Coacervates were prepared at pH 9.5 and I = 0.1M NaCl. It is interesting to compare this data with our observations of the same phenomenon in serum of a patient with Waldenstrom's macroglobulinemia [1], when serum in a tube segregated spontaneously into two transparent liquid fractions differing in light-refracting capacity: lower density at the top and higher density at the bottom. Upon drying, the structures of interest were observed in the bottom fraction only, as well as in the thoroughly mixed whole serum. The transparency of these complexes meant that they were well hydrated, and "invisible". During film water evaporation they lost their water shell, and became visible. Appearance of such structures in liquid samples of biological fluids means that this complex system – serum or plasma – transforms to a non-equilibrium state due to the presence of blood macromolecules, including acute-phase proteins, and osmotic active components which violate mutual balance and lead to loss of colloid stability.

Albumin is the most mysterious multifunctional blood protein. In norm, albumin realizes ligand-binding and transport functions, as well as supports colloid osmotic pressure of plasma and regulates neutrophyl function [36]. It has already been found that albumin specifically binds with low molecular weight molecules that might be important diagnostic and prognostic indicators of diseases [37]. In clinics, albumin has been intensively used in critical conditions, including vascular collaps in seriously ill patients [38]. At the same time, it was shown that highly purified commercial albumin preparations for laboratory use may contain significant amounts of albumin polymers [39-41]. There was an increase in death among patients who were treated with human albumin solution for burns, hypoalbuminaemia and hypotension [42] by up to 6%. Pro and contra for using albumin in clinical practice are widely discussed in the literature [43].

It was shown that albumin polymers [44] and their glycation olygomers [45] promote erythrocyte rouleaux formation in blood, whereas molecular protein prevents erythrocyte aggregation [46]. Armstrong, J.K., *et al.* (2004) [47] have found that dimension of hydrodynamic radius of polymer or macromolecule is the main criterion of its erythrocyte aggregation capacity: if it does not exceed 4 nm, it prevents aggregation, and if it is more than 4 nm, it promotes it. Thus, in accordance with these data, molecular albumin is a promoter of erythrocyte disaggregation, and albumin polymers (or albumin coagulation structures) are promoters of their aggregation. Shacter, E., *et al.* (1993) [48]have reported that albumin polymers presented in commercial samples of albumin can stimulate interleukin-6 and prostaglandin E_2 production by macrophags in vitro and in vivo. The albumin fraction with molecular weight $\geq$ 600 kD had the highest stimulation activity. It means that this polymer contains 10 or more monomers. The authors [48] believe that this is a new mechanism of cytokine activation that triggers a cascade of inflammatory reactions in organism.

This short review demonstrates that protein coagulation structures formed in biological fluids due to loss of their colloid stability can start the same cascade of pathophysiological reactions. In addition, protein coagulation leads to decreasing of blood oncotic pressure and to water penetration into intercellular space which encloses a vicious circle forming edema and worsening microcirculation. Thus, this phenomenon requires further investigation, strong laboratory control and elaboration of methods for its physical-chemical corrections if necessary. The drop drying process is a natural model of "functional loading" on biological liquids by decreasing albumin concentration and increasing ionic strength, which allows one to evaluate the degree of its colloid stability.

REFERENCES

[1] Yakhno, T.A., Sedova, O.A., Sanin, A.G. and Pelyushenko, A.S. (2003) On the existence of regular structures in the liquid human blood serum (plasma) and phase transitions in the course of its drying. *Technical Physics*, **48(4)**, 399-403.

[2] Yakhno, T. (2008) Salt-induced protein phase transitions in drying drops. *Journal of Colloid and Interface Science* ,**318**, 225-230.

[3] Velev, O.D., Kaler, E.W. and Lenhoff, A.M. (1998) Protein interactions in solution characterized by light and neutron scattering: comparison of lysozyme and chymotripsinogen. *Biophysical Journal*, **75**, 2682-2697.

[4] Thomson, J.A., Schurtenberger, P., Thurston, G.M. and Benedek, G.B. (1987) Binary liquid phase separation and critical phenomena in protein/water solution, *Proceedings of the National Academy of Sciences, USA*, **84**, 7079-7083.

[5] Kaibara, K., Okazaki, T., Bohidar, H.B. and Dubin, P.L. (2000) pH-induced coacervation in complexes of bovine serum albumin and cationic polyelectrolytes, *Biomacromolecules*, **1**,100-107.

[6] Seyrek, E., Dubin, P.L., Tribet, C. and Gamble, E.A. (2003) Ionic strength dependence of protein-polyelectrolyte interactions. *Biomacromolecules*, **4**, 273-282.

[7] Wang, Y., Kimura, K., Huang, Q., Jaeger, W. and Dubin, P.L. (1999) Effects of salt on polyelectrolyte-micelle coacervation. *Macromolecules*, **32**, 7128-7134.

[8] Tanaka, S., Yamamoto, M., Ito, K., Hayakawa, R., Ataka, M. (1997) Relation between the phase separation and the crystallization in protein solutions. *Physical Review E*, **56(1)**, R67-R69.

[9] Chang, B.H. and Bae, Y.C. (2003) Salting-out in the aqueous single-protein solution: the effect of shape factor. *Biophysical Chemistry*, **104**, 523-533.

[10] Fulton, A.B. (1982) How crowded is the cytoplasm? *Cell*, **30(2)**, 345–347.

[11] Ellis, R.J. and Minton, A.P. (2006) Protein aggregation in crowded environments. *Journal of Biological Chemistry*, **387**, 485-497.

[12] Zimmerman, S.B. and Trach, S.O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia Coli. *Journal of Molecular Biology*, **222**, 599-620.

[13] Vazquez, A., Beg, Q.K., Demenezes, M.A. and Ernst, J., Bar-Joseph, Z., Barabasi, A.L., Boros, L.G., Oltavi, Z.N. (2008) Impact of the solvent capacity constraint on E. Coli metabolism. *BMC Systems Biology*, **2**, 7-10.

[14] Mattison, K.W., Dubin, P.L. and Brittain, I.J. (1998) Complex formation between bovine serum albumin and strong polyelectrolytes: Effect of polymer charge density. *Journal of Physical Chemistry B*, **102**, 3830-3836.

[15] Azegami, S, Tsuboi, A., Izumi, T., Hirata, M., Dubin, P. L., Wang, B. and Kokufuta, E. (1999) Fotmation of an intrapolymer complex from human serum albumin and poly(ethylene glycol). *Langmuir*, **15**, 940-947.

[16] Tripp, B.C., Magda, J. J. and Andrade, J.D. (1995) Adsorption of globular proteins at the air/water interface, as measured via dynamic surface tension. Concentration dependence, mass-transfer considerations, and adsorption kinetics. *Journal of Colloid and Interface Science*, **173**, 16-27.

[17] Pande A., Pande, J., Asherie, N., Lomakin, A., Ogun, O., King, J.A., Lubsen, N.H., Walton, D. and Benedek, G.B. (2000) Molecular basis of a progressive juvenile-onset hereditary cataract. *Proceedings of the National Academy of Sciences*, **97(5)**, 1993-1998.

[18] Simpanya, M.F., Ansari, R.R., Suh, K.I., Leverenz, V.R. and Giblin, G.J. (2005) Aggregation of lens crystallins in an in vivo hyperbaric oxygen Guinea pig model of nuclear cataract: dynamic light-scattering and HPLC analysis. *Investigative Ophthalmology & Visual Science*, **46**, 4642-4651.

[19] Stradner, A., Foffi, G., Dorsaz, N., Thurston, G. and Schurtenberger, P. (2007) New Insight into cataract formation: Enhanced stability through mutual attraction. *Physical Review Letters*, **99**, 198103.

[20] Green, M.A., Noguchi, C.T., Keydan, A.J., Marwah, S.S. and Stuart, J. (1998) Polymerization of sickle cell hemoglobin at arterial oxygen saturation impairs erythrocyte deformability. *Journal of Clinical Investigation*, **81**, 1669-1674.

[21] Ahn, S.-M., Byun, K., Cho, K., Kim, J. Y., Yoo, J. S., Kim, D., Paek, S. H., Kim, S. U., Simpson, R. J. and Lee, B. (2008) Human microglial cells synthesize albumin in brain. *Plos ONE*, **3(7)**, e2829.

[22] Yakhno, T.A., Yakhno, V.G., Sanin, A.G., Sanina, O.A. and Pelyushenko, A.S. (2003) A Method for Liquid Analysis by means of Phase Transitions during drop drying. *Proceedings of SPIE, Bioengineered and Bioinspired Systems*, 19-21 May, Maspolamas, Gran Canaria, Spain, **5119**, 87-99.

[23] Yakhno, T., Sanin, A., Yakhno, V., Pelyushenko, A., Egorova, N.A., Terentiev, I.G., Smetanina, S.V., Korochkina, O.V. and Yashukova, E.V. (2005) The informative-capacity phenomenon of drying drops. Aptitude test in medical diagnostics. *IEEE Engineering in Medicine and Biology Society*, **24**, 2, 96-104.

[24] Yakhno, T., Sanin, A., Pelyushenko, A., Kazakov, V., Shaposhnikova, O., Chernov, A., Yakhno, V., Vacca, C., Falcone, F. and Johnson, B. (2007) Uncoated quartz resonator as a universal biosensor. *Biosensors and Bioelectronics*, **22**, 9-10, 2127-2131.

[25] Yakhno, T.A., Kazakov, V.V., Sanin, A.G., Shaposhnikova, O.B. and Chernov, A.S. (2007) Dynamics of Phase Transitions in Drying Drops of Human Serum Protein Solutions. *Technical Physics*, **52(4)**, 515–520.

[26] Yakhno, T.A., Kazakov, V.V., Sanin, A.G., Shaposhnikova, O.B. and Chernov, A.S. (2007) Mechanical Properties of Adsorption Layers in Solutions of Human Blood Serum Proteins: A Comparative Assessment. *Technical Physics*, **52(4)**, 510–514.

[27] Heil, W., Koberstein, R. and Zawta, B. (2001) Reference Ranges for Adults and Children: Pre-Analytical Considerations. Boehringer, Mannheim, 1997, Labpress, 176, Moscow.

[28] Ben-Ami, R., Barshtein, G, Mardi, T., Deutch, V., Elkayam, O. and Berliner, S. (2003) A synergistic effect of albumin and fibrinogen on immunoglobulin-induced red blood cell aggregation. *American Journal of Physiology, Heart and Circulatory Physiology*, **285**, H2663-H2669.

[29] Shacter, E., Arzadon, G.K. and Williams, J.A. (1993) Stimulation of interleukin-6 and prostaglandin E2 secretion from peritoneal macrophages by polymers of albumin. *Blood*, **82**, 2853-2864.

[30] Deegan, R.D. (2000) Pattern formation in drying drops. *Physical Review E*, **61(1)**, 475-485.

[31] Popov, Y. (2005) Evaporative deposition patterns: Spatial dimensions of the deposit. *Physical Review Letters*, **71**, 036313.

[32] Yakhno, T. and Yakhno, V. (2009) Structural evolution of drying drops of biological fluids. *Technical Physics*, **54(8)**, 1219-1227.

[33] Ling, G.N. (2000) Life at the cell and bellow-cell level. The hidden history of a fundamental revolution in biology. Pacific Press, 280. Also available: http://biopa radigma.narod.ru/hidden_history/ling_newbook.htm

[34] Pollak, G.H. (2001) Cells, gels and the engines of life. *Ebner & Sons, Seattle*, 301. WA, USA.

[35] Cooper, C.L., Dubin, P.L., Kayitmazer, A.B., Turksen, S. (2005) Polyelectrolyte-protein complexes. *Current Opinion in Colloid and Interface Science*, **10**, 52-78.

[36] Peters, T. (1996) All about albumin: Biochemistry, genetics, and medical applications. San Diego, Academic press.

[37] Petricoin, E.F., Belluco, C., Araujo, R.P., Liotta, L.A. (2006) The blood peptidome: A higher dimension of information content for cancer biomarker discovery. *Nature Reviews Cancer*, **6**, 961-967.

[38] Quinlan, G.J., Martin, G.S., Evans, T.W. 2005. Albumin: Biochemical properties and potential. *Hepathology*, **41**, 1211-1219.

[39] Keyser, J.W. (1995) Standardization of dye-binding methods for estimation of serum albumin. *Clinica Chimica Acta*, **11**, 477-9.

[40] Blaaberg, O. and Hylton, P.P. (1979) Effect of aggregates on albumin standardization. *Scandinavian Journal of Clinical and Laboratory Investigation*, **39**, 751-7.

[41] Bormer, O.P., Amlie, L.M., Paus, E. and Kongsgard, U. (1999) Automated albumin method underestimates pharmaceutical-grade albumin in vivo. *Clinical Chemistry*, **45(7)**, 1082-1084.

[42] Reviewers, C.I.G.A. (1998) Human albumin administration in critically ill patients: Systematic review of randomised controlled trials. *British Medical Journal*, **317**, 235-240.

[43] Pulimood, T.B. and Park, G.R. (2000) Debate: Albumin administration should be avoided in the critically ill. *Critical Care*, **4**, 151-155. Also available: http://ccforum/content/4/3/151

[44] Forsdyke, D.R., Palfree, R.G.E. and Takeda, A. (1982) Formation of erythrocyte rouleaux in preheated normal serum: roles of albumin polymers and lysophosphatidylcholine. *Canadian Journal of Biochemistry*, **60(7)**, 705-711.

[45] Candilores, H., Muller, S., Ziegler, O., Donner, M. and Drouin, P. (1996) Role of albumin glycation on the erythrocyte aggregation: An in vitro study. *Diabetic Medicine*, **13(7)**, 646-650.

[46] Reinke, W., Gaehtgens, P. and Johnson, P.C. (1987) Blood viscosity in small tubes: effect of shear rate, aggregation, and sedimentation. *American Journal of Physiology*, **253**(3), H540-547.

[47] Armstrong, J.K., Wenby, R.B., Meiselman, H.J. and Fisher, T.C. (2004) The hydrodynamic radii of macromolecules and their effect on red blood cell aggregation. *Biophysical Journal*, **87**, 4259-4270.

[48] Deegan, R.D., Bakajin, O., Dupont, T.F., Huber, G., Nagel, S.R., Witten, T.A. (2000) Contact line deposits in an evaporating drop. *Physical Review E*, **62(1)**, 756-776.

Vol.2, No.3, 228-245 (2010)
doi:10.4236/ns.2010.23035

A retrospective view on the history of natural sciences in XX-XXI

Vladislav Sergeyevich Olkhovsky

Institute for Nuclear Research, Kiev, Ukraine; olkhovsky@mail.ru

Received 1 December 2009; revised 8 January 2010; accepted 30 January 2010.

ABSTRACT

The presented paper is dedicated to a new retrospective view on the history of natural sciences in XX-XXI cc, partially including the science philosophy (mainly, the problems of the scientific realism, *i.e.* the correspondence of science to reality) and also a novel scheme for different classes of sciences with different objects and paradigms. There are analyzed the chosen "great" and "grand" problems of physics (including the comprehension of quantum mechanics, with a recently elaborated new chapter, connected with time as a quantum observable and time analysis of quantum processes) and also of natural sciences as a whole. The particular attention is paid to the interpretation questions and slightly to the aspects, inevitably connected with the world- views of the researchers (which do often constitute a part of the interpretation questions).

Keywords: science history; science realism; paradigm; problem of interpretation and comprehension of quantum mechanics; the wave-function collapse; the Einstein-Podolsky-Rosen paradox; time as a quantum observable, canonically conjugated to energy; maximal hermitian time operator; time analysis of quantum processes; relationship between physics and biology; problem of origin of biologic life; reductionism; cosmologic problem; Big Bang; anthropic principle

1. INTRODUCTION

In the science history and in the science philosophy of XX-XXI (especially in the field of the natural sciences and most of all in physics) there has been a lot of interesting things, which had not obtained a sufficiently complete elucidation and analysis yet. Firstly, under the influence of scientific and technological progress a great attention has been paid to the development of such direction in the science philosophy as the scientific realism (*i.e.* the correspondence of the science to the reality), which has successively acquired three forms: the naïve realism, the usual realism and the critical science realism. Secondly, some new important problems of physics (especially the problem of the essentially probabilistic description of the reality of the microscopic world, the problem of the essential influence of the observer on the reality, the collapse of the wave function and the Einstein-Podolsky-Rosen paradox) had been revealed in the development of quantum mechanics; *the continuously complicated explanation of the Universe origin and the expansion after the Big Bang*; and *no* succeeded attempt in explaining the origin of the biological life in terms of physics and other natural sciences, all being with a variety of interpretation versions (often connected with the world-views of the researchers), cause to undertake a new view of the science history. And thirdly, a clear analysis of the variety of known sciences brings to the re-considering of the science classification and a novel scheme for different classes of natural sciences with quite different objects (including not only simple natural phenomena and processes, but also the *human intelligent design* and the *origin* of the Universe and life). Finally, a simple study of the enlargement of mathematics in practically all of sciences did not only indicate that mathematics became the branch of the natural sciences (as to the opinion of some scientists, such as N.N.Bogolyubov and others) but has also in fact induced the solution of the long-standing problem of time in quantum mechanics.

2. THE SCIENTIFIC REALISM AND ITS DIFFERENT KINDS

If the science does correspond the reality? In the science philosophy the term *reality* defines the direction, postulating the existence of the reality, *independent* from the cognitive subject. *The scientific realism* postulates the existence of the objective truth, the aim of the scientific theories is being declared the revelation of the real truth, the moving force of the scientific progress is declared

the approach to the truth (the truth is explained as entirely adequate description of the reality). The scientific theories, if they ate really truthful, do describe in an adequate way the reality and the essences, which are postulated by the well tested theories, are existing really.

R.Boyd [1] selected three types of the scientific realism:

The ontological realism assumes that the reality, which is described by the scientific realism, does not depend, on the whole, from our thinking and from the theoretic assumptions. The ontological realism responds to the questions like "what essences are real?", "if the world, which is independent from the observer, does exist?"

The epistemological realism assumes that the scientific theories are confirmed as true and in fact are often approach to the reality. The epistemological realism responds to the question: "is the knowledge about the world possible?"

The semantic realism assumes that "the theoretical terms" of the scientific theories indicate to the realistic essences, *i.e.* the theories have to be interpreted realistically. The semantic realism responds the question: "if the truth of the objective world is expressed by the scientific language?"

Respectively, the scientific realism on the whole assumes that the scientific theories tend to give the truthful description of the reality which is existing independently ("the truth" signifies here the complete correspondence between the science and the reality). If the scientific theory is really true, then the unobserved essences, which it postulates, are really existing.

A.Bird [2] had formulated the short thesis of the scientific realism. He states that the scientific theories:

a) can be estimated in the terms of the truthfulness or the approach to the truthfulness;

b) their reasonable aim is the truth or the approach to the truth;

c) their success, confirmed by the scientific progress, testimonies their truth;

d) if they are true, then the unobserved essences, which they assume, are really existing;

e) if they are true, they will explain the observed phenomena.

The main argument for the realism is the *conclusion on the best explanation of the reality*: the scientific realism is the only science philosophy that can explain the scientific progress. The scientific realism is exposed to the critics from the antirealism (antirealism, appeared in the second part of the XX, does represent the scientific philosophy, which is opposite to the realism). Antirealists state that to consider the scientific theories to be true is too risky. Some previous scientific theories were false, for example, the theories of heat matter, phlogiston, the ether conception. So, modern theories can be also false. The position of the scientific realism is criticized by antirealists, although it has a lot of supporters.

In the non-uniform current of the scientific realism there are known 3 kinds: naïve, usual and critical.

The naïve realism is the position of the majority of men from the point of view of the common sense. [The common sense is acquired by all normal men during the natural living process, in the overall man communications and in actions with the objects of our usual experience. It is like the assimilation of the natal language with which the common sense is therefore closely connected. In many situations the common sense is used as a matter of fact the primordial universal kind of knowledge]. According to this position, the world is such, which is represented by the modern (however, pre-quantum!) science: those essences, the existence of which are postulated by the well-supported scientific theory, are really existing. And only objects, which are described by the scientific theories, have the authentically real ontological status, and the scientific knowledge as an epistemological base of the science does also represent the realism.

The usual realism is the position of the investigators like [1,2], and also somewhat different positions of a series of other authors (for instance, such as J.Smart, R.Harre, H.Putnam *etc*.).

Later the more "weak" realistic position appeared – the *critical realism*, with the more modest declarations of its supporters. The critical scientific realism had been declared by the not very equal positions of many various authors. The position of the critical realism had been rather clearly formulated by I.Niiniluotto from the Helsinki University [3], which added some more precise definitions to the position of the critical realism. In particular, he recognizes the conceptual pluralism, under the influence of the uncertainty thesis of W.Quine [4]: our appeal to the world does always occur in some linguistic frame. The thesis of the pessimistic induction forces him to accept the possibility of the untrue theories: the knowledge about the reality is not very trustworthy and demands certain corrections, and even the best scientific theories can contain the mistakes, however the successful theories approach to the reality. The thesis about "the human insertion" signifies that the reality is partially (but only partially!) constructed by the humanity. And in the whole the position of Niiniluotto conserves the realistic optimism: the scientific progress can be rationally explained. As before, the best explanation does consist in that the scientific theories approximate to reveal the truth. Concretely I.Niiniluotto had formulated such thesis of the critical realism [3]:

a) At least, the part of the reality is ontologically independent from the human intelligence.

b) The truth is the semantic relation between the language and the reality.

c) The conception of the truth or the falsification can de used to all linguistic derivatives of the scientific activity including the reports on the observations, the laws

and the theories. In particular, the statements on the existence have the truthfulness significations.

d) The attainment of the truth is the main aim of the science. The truth is not being found and recognized in a simple way and even the best scientific theories can be false. Nevertheless, it is possible to approach to the truth and obtain the rational conception on the cognitive process.

e) The best explanation of the practical success of the science consists in the approximate truthfulness or successful approaching to the revelation of the truth about the reality. And therefore the scientific progress can be rationally explained.

Not always it is possible clearly distinct the usual scientific realism and the critical scientific realism because in the scientific thinking a lot of attention is paid to the critical analysis of the cognition methods and the scientific knowledge with utilization of all logic cognition methods with understanding of their limitedness. But with the sufficiently completeness of the utilization of the logic cognition methods, the position of the critical realism is better defended against the standard arguments utilized against the realism (usual and especially naïve). If earlier the main attention in the science philosophy had been paid to the justification of the scientific method, during the last dozens of years mainly the questions of the ontological status of objects, introduced by the scientific theories, are discussed. It must be noted that the problems of the quantum theory, revealed as a result of the long (during many years) *discussion of N.Bohr with A.Einstein*, had seriously undermined the traditional forms of the naïve realism in the science and have strongly influenced not only on physics but also on other kinds of knowledge and on our understanding of the human knowledge at all.

The quantum mechanics, in difference from the classical (non-quantum) physics, revealed that on the microscopic level there is *the un-removable indeterminism*, represented by the uncertainty relations of Heisenberg, by the essential non-locality of the particle-waves (still unmeasured, *i.e.* before measurements) and also by the measurements with the discrete interaction of the microscopic objects and the measurement devices (when, for instance, there are the photons are emitted and absorbed). Then in quantum mechanics there is the problem of the interpretation of the quantum measurements and particularly the wave-function collapse *etc*, when the state of the measured system is formed by the observer [5,6]. All these problems and paradoxes had arisen as a challenge to the philosophy and even now bring to the acute discussions [5,6]. And if the majority of the physicists agree with the Bohr Copenhagen interpretation of the quantum mechanics, a certain part of the physicists still assumes that A.Einstein was righteous in his statement that the quantum theory (in its Copenhagen interpretation) does not directly describe the reality. Still the more acute situation had arisen from other quantum phenomena such as the Einstein-Podolsky-Rosen paradox. But as to opinion to some ideologists, such conclusions are justified only in the frame of the physical description and even in such description many of these problems are open even now, so there is no the final necessity now to extrapolate them into the philosophy and theology with profound "revolutionary" philosophical and theological conclusions.

3. THE DIFFERENT CLASSES OF PARADIGMS IN THE DIFFERENT CLASSES OF NATURAL SCIENCES

If the objects of natural sciences (physics, chemistry, biology, geology, astronomy *etc*) are limited by the only *natural* events and processes, the objects of some other sciences include in their objects also the *artificial* facts (arte-facts) as *creations of the human intelligent design* (there are archeology, medicine, criminalistics, and, moreover, mathematics, cybernetics, informatics, and also such humanitarian sciences as history, economics and political sciences). There are also the particular sciences where the origin and history of the Universe or the origin and history of the biological life (including genetics) are studied and where side by side with the scientific method a metaphysical worldview approach of the investigators does, almost inevitably, also take place: Due to the cardinal separation of the investigators because of the incompatibility of their worldviews as to the problems of the origin, the dilemma of the following choice had appeared: either 1) the self-organization of the matter from the null or a less organized level into the much more organized level by virtue of a certain irrational chance or by virtue of unknown now synergetic processes (or phase transitions), or 2) the origin of the Universe and of the life inside it as a result of *the supreme intelligent design of a certain super-human creative basis (or a Creator)*.

And in these three classes of sciences with the different research objects now there are co-existing for scientific researches three different classes of paradigms (the term "paradigms" had been introduced by T. Kuhn in [7]) which exist inside the sciences of their application: the class of paradigms for research of the laws of functioning of the natural processes; the class of paradigms of introducing (or inserting) the human intelligent design inside the natural processes or in the human activity; and finally the class of paradigms of the research of the mechanisms of the origin of the Universe and the life.

Moreover, the first two classes are now already observed to be sometimes overlapped: for instance, in quantum mechanics it is known that the state of the measured system can be in fact formed by the observer [5], *i.e.* the human intelligent design can actively influence on the currency of the observed natural processes! And in the third class of sciences, which deals with the origin prob-

lems, for a long time there are known acute collisions of different worldviews which are sometimes expanding into the second group, including mathematics, informatics and cybernetics. Besides the confusion of the different classes of paradigms, such discussions between the supporters of the different worldviews have sometimes become more acute because of the complete incompatibility of the researcher's worldviews, taking the especially sharp forms between the evolutionists and creationists. Even A. Einstein in his last-life period had participated, at least partially and philosophically, in these discussions (see, for instance, [8]): "*Considero le dottrine evoluzionistiche di Darwin, Haeckel, Huxley, come tramontate senza speranza*" (in English: "*I consider the evolutionism doctrines of Darwin, Haeckel and Huxley as being outgoing without any hope to revive* ").

As to mathematics, one can note that usually the object of every mathematical discipline or theory is taken as the system of the exactly formulated axioms, and the methodology of mathematics consists in the derivations of the logical conclusions and theorems from the chosen axioms. Previously C.Gauss referred to mathematics as "the Queen of the Sciences" [9]. Later it was observed that certain scientific fields (such as theoretical physics) are mathematics with axioms that are intended to correspond to reality. In any case, mathematics shares much in common with many fields in the physical sciences, notably the exploration of the logical consequences of assumptions. And K. Popper concluded that "*most mathematical theories are, like those of physics and biology, hypothetico-deductive: pure mathematics there- fore turns out to be much closer to the natural sciences whose hypotheses are conjectures, than it seemed even recently*"[10. Moreover, in XX-XXI some mathematicians consider that mathematics has already become practically a branch of natural sciences (theoretical physics) – see, for instance, [11]. And it is in agreement with the known statement of Galileo: "*Il libro della natura è scritto in lingua matematica*" (in English: "*The book of nature is written by the mathematical language*") [12]. Now in the theory of quantum collisions and, in particular, in the theory of the dispersion relations the initial assumptions of these theories do contain, besides the physical principles, also the mathematical principle of certain analytic properties of the *S*-matrix in the complex plane of energies (momenta) [11, the first reference]). Finally, namely mathematics generated the solution of the long-standing problem of time as a quantum observable, canonically conjugated to energy, and self-consistent time analysis of quantum processes.

4. AS TO "GREAT" AND "GRAND" PROBLEMS OF NATURAL SCIENCES

There is an extensive introduction in the large number of open problems in many fields of physics, published by the Russian physicist V.Ginzburg in [13], which is rather interesting to study. Inside this large list of open problems of modern physics (and in a certain degree of modern natural sciences), represented by V.Ginzburg repeatedly in Russian editions, some of them are marked him "great" or "grand" problems. Between namely these problems I would like to separate three of them.

a) *The problem of interpretation and comprehension of quantum mechanics* (*even of the non-relativistic quantum theory*) *remains still topical*.

The majority of critics of quantum mechanics are unsatisfied with the probabilistic nature of its predictions. One can add here also the questions and paradoxes of the theory of quantum measurements theory, especially like the wave-function reduction and the Einstein-Podolsky-mRosen paradox. The appearance of quantum mechanics, and, in particular, the discussion of N.Bohr with A. Einstein (lasting many years), had seriously undermined the traditional forms of the naïve realism in the philosophy of the scientific realism and had strongly influenced (and are continuating to influence) not only on physics but also on other kinds of knowledge in the sense of the dependence of the reality on the observer and, moreover, on our understanding of the human knowledge at all. The problem of the relativistic quantum mechanics and quantum field theory is even much more sharp because of the incompatibility of the main premises of the quantum theory and of the relativity theory.

b) *The relationship between physics and biology and, specifically, the problem of reductionism*.

The main problem, according to V.Ginzburg, is connected with the explanation of *the origin of the biologic life* and *the origin of the human abstract thinking* (but the second one, as to me, is connected not with biology but with the origin of the human spiritual life which is far beyond natural sciences). V.Ginzburg assumes that for a possible explanation of the origin of the biologic life one can naturally imagine a certain jump which is similar to some kind of phase transition (or, may be, certain synergetic process). But there are other points of view too.

c) *The cosmological problem* (in other words, *the problem of the Universe origin*).

According to V.Ginzburg, it is also a grand problem, or strictly speaking, a great complex of cosmic problems many of which is also far from the solution.

5. MORE DETAILED COMMENTS ON THE PROBLEM OF COMPREHENSION OF QUANTUM MECHANICS

Various interpretations of quantum mechanics. Not only philosophers of scientific critical realism, but also up to now a certain part of physicists, beginning from A. Ein-

stein, D.Bohm, Y.Aharonov and some others, did not agree with the Copenhagen interpretation of quantum mechanics and, moreover, had constructed alternative versions of interpretation (see, for instance, [14-22]).

Einstein never accepted quantum mechanics as a "real" and complete theory, struggling to the end of his life for an interpretation that could comply with relativity without complying with the Heisenberg uncertainty principle. As he once said: "*God does not play dice*", skeptically referring to the Copenhagen interpretation of quantum mechanics which says there exists no objective physical reality other than that which is revealed through measurement and observation.

In 1935 Einstein, Podolsky, and Rosen were formulated their thought experiment, which had been called the *EPR paradox* (which is also referred to as the *EPRB paradox* after Bohm, who improved the formulation of the thought experiment). It draws attention to a phenomenon predicted by quantum mechanics known as *quantum entanglement*, in which measurements on spatially separated quantum systems can instantaneously influence one another. As a result, quantum mechanics violates a principle formulated by Einstein, known as the principle of locality or local realism, which states that changes, performed on one physical system, should have no immediate effect on another spatially separated system. The principle of locality seems to be persuasive, because, according to relativity, information can never be transmitted faster than the speed of light, or causality would be violated. Any theory, violating causality, would be deeply unsatisfying. However, a detailed analysis of the EPR scenario shows that quantum mechanics violates locality without violating causality, because no information can be transmitted using quantum entanglement.

Nevertheless, the principle of locality appeals powerfully to physical intuition, and Einstein, Podolsky and Rosen were unwilling to abandon it. They suggested that quantum mechanics is not a complete theory, just an (admittedly successful) statistical approximation to some yet-undiscovered description of nature. Several such descriptions of quantum mechanics, known as "*local hidden variable parameters*", were proposed. These deterministically assign definite values to all the physical quantities at all times, and explicitly preserve the principle of locality.

Of the several objections to the then current interpretation of the quantum mechanics spearheaded by Einstein, the EPR paradox was the subtlest and most successful. The EPR paradox has not been resolved or explained, in a way, which agrees with classical intuition, up to this day. It brought a new clarity and permanent shift in thinking about 'what is reality' and what is a 'state of a physical system'.

The shift was caused by the EPR thought experiment, which has shown how to measure the property of a particle, such as a position, without disturbing it. In today's terminology, we would say that they did the determination by measuring the state of a distant but entangled particle. *Quantum entanglement* is a property of a system of two or more particles (objects) in which the quantum states of the constituting objects are linked together so that one object can no longer be adequately described without full mention of its counterpart - even if the individual objects are spatially separated. According to quantum mechanics, the state of the counterpart particle will instantly change even though we did not disturb it in any local way. It conflicts with our classical intuition with the relativistic principle of locality. Different views on the essence of the quantum entanglement bring to different interpretations of quantum mechanics. The very concept of quantum entanglement also conflicts with our intuition the same way.

However, experiments have shown that entanglement does occur, and in fact quantum entanglement has practical applications in the field of q*uantum cryptography* and *quantum computation*. Earlier quantum entanglement had been utilized in experiments with *quantum teleportation*. *Quantum teleportation* is a technique used to transfer quantum information from one quantum system to another. It does not transport the system itself, nor does it allow communication of information at superluminal (faster than light) speed. Its distinguishing feature is that it can transmit the information present in a quantum superposition, useful for quantum communication and quantum computation. In quantum cryptography, an entangled signal is sent down a communications channel making it impossible to intercept and rebroadcast that signal without leaving a trace. In quantum computation, entangled states allow simultaneous computations to occur in one step.

Entanglement has many applications in quantum information theory [23-31]. Mixed state entanglement can be viewed as a resource for quantum communication. With the aid of entanglement, otherwise impossible tasks may be achieved. Among the best known applications of entanglement there is super-dense coding.

In 1964 J.Bell had shown that many theories, known as hidden variable theories, are either non-local or known as satisfying Bell inequality [16]. Quantum mechanics predicts that this inequality is not satisfied. To make sure, additional experiments were made to confirm that predicted action at distance is indeed instant. Today most physicists agree that local hidden variable theories are untenable and that the principle of locality does not hold. Therefore, the EPR paradox would only be a paradox because our physical intuition does not correspond to physical reality. But even now the topic remains active and some people are still looking for Quantum. quantum-quantumquantummechanics is neither "real" (since measurements do not *state*, but instead *prepare* properties of the system) nor "local" (since the state vector comprises the simultaneous probability amplitudes for all

positions), and the properties of entanglement are some of the many reasons why the Copenhagen Interpretation is no longer considered standard by a large proportion of the scientific community. So, the discussion of N.Bohr with A.Einstein had originated so many interesting fundamental results, experimental applications and other (already second or derived) discussions, which have endless continuation up to now, that it was unique in the history of physics.

And now, let us speak some words on the *many-world interpretation* (MWI) in quantum mechanics (and in quantum cosmology). In this interpretation one assumes the existence of the parallel universes, in every of which the same nature laws and physical constants are acting, but all of them are found in different states. MWI refuses an indeterminate collapse of the wave function which is connected with the measurement in the Copenhagen interpretation. The ideas of MWI had been originated in the phd-thesis of H.Everett bat the term MWI had been proposed by B.S.M. de Witt who had developed that idea, and then the various authors had participated in the further development of that topic [32-41].

In various versions of MWI there two main points: The first one consists in the existence of the wave function for the total Universe, described by the Schroedinger equation, but without any in-determined collapse. The second one consists in that such state of the Universes is the quantum superposition of several (and may be, of the infinite number of) states of the equal parallel universes which are non-interacting among themselves.

According to the modern criteria of the scientific theories, MWI is not experimentally verificable and not falsified, and therefore is not scientific! However, any other interpretation of quantum mechanics, including the Copenhagen one, is also not scientific but philosophical and therefore the usefulness of the quantum-mechanical interpretation is determined by its pragmatism. And, although the analysis of some problems in the MWI brings to the same results as in any other interpretation, but these results are more simple logically, so they had been resulted to some physicists to be more popular in quantum mechanics (and quantum cosmology).

May be, it seems that the majority of the opponents of the MWI reject it because, for them, introducing a very large number of worlds that we do not see is an extreme violation of Ockham's principle: "Entities are not to be multiplied beyond necessity". However, in judging physical theories one could reasonably argue that one should not multiply physical laws beyond necessity either (such a verion of Ockham's Razor has been applied in the past), and in this respect the MWI is the most economical theory. Indeed, it has all the laws of the standard quantum theory, but without the collapse postulate, the most problematic of physical laws.

The reason for adopting the MWI is that it avoids the collapse of the quantum wave. And there is no experimental evidence in favor of collapse and against the MWI. We need not assume that Nature plays dice. The MWI is a deterministic theory for a physical Universe and it explains why a world appears to be in-deterministic for human observers.

The MWI exhibits some kind of non-locality: "world" is a non-local concept, but it avoids action at a distance and, therefore, it is not in conflict with the relativistic quantum mechanics .

The MWI is not the most accepted interpretation of quantum theory among physicists, but it is becoming increasingly popular .

The strongest proponents of the MWI can be found in the communities of quantum cosmology and quantum computing. In quantum cosmology it makes it possible to discuss the whole Universe avoiding the difficulty of the standard interpretation which requires an external observer. In quantum computing, the key issue is the parallel processing performed on the same computer; this is very similar to the basic picture of the MWI. However, the advantage of the MWI is that it allows us to view quantum mechanics as a complete and consistent physical theory which agrees with all experimental results obtained to date. And also, the elegant conception of the de-coherence, proposed in 1970 by Dieter Zeh, explains that the various branches of the single wave function, which describe these worlds, are oscillating in time with the different phases and so as if do not exist each for other [42].

As a whole, the problem of the final interpretation of quantum mechanics and of quantum theory of measurements is far from the total consensus and still remains open for both physicists and philosophers (in the science philosophy).

One can add here that the still inherent incompatibility of the postulates of quantum theory as non-local theory and relativity theory (both special and general) as local theory is the main root of the impossibility to construct the self-consistent relativistic quantum mechanics, quantum field theory and the quantum cosmology even in quasi-linear approximation.

Another long incompleteness of non-relativistic quantum mechanics (even in the Copenhagen interpretation) is connected with the problem of time as a quantum observable, which is, moreover, canonically conjugated to energy. It has been known from the beginning of twentieth of XX (see [43] and later also the discussion of Y.Aharonov and D.Bohm with Fock [44,45]) till the last years, when it has in fact been resolved practically by using the mathematical means.

6. *TIME* IS REALLY A QUANTUM OBSERVABLE, CANONICALLY CONJUGATED TO ENERGY

Introduction to the history of the problem. During almost ninety years (see, for example, [43]) it is known that

time cannot be represented by a self-adjoint operator, with the possible exception of special abstract systems (such as an electrically charged particle in an infinite uniform electric field[1] and a system with the limited from both below and above energy spectrum (to see later)). This fact results to be in contrast with the known circumstance that time, as well as space, in some cases plays the role just of a parameter, while in some other cases *is* a physical observable which *ought* to be represented by an operator. The list of papers devoted to the problem of time in quantum mechanics is extremely large (see, for instance, [46-82], and references therein). The same situation had to be faced also in quantum electrodynamics and, more in general, in relativistic quantum field theory (see, for instance, [53,81,82]).

As to quantum mechanics, the first set of known and cited articles is [54-60]. The second set of papers on time as an observable in quantum physics [61-82] appeared from the end of the eighties and chiefly in the nineties and more recently, stimulated mainly by the need of a self-consistent definition for collision duration and tunnelling time. It is noticeable that many of this second set of papers appeared however to ignore the Naimark theorem from [83], which had previously constituted an important basis for the results in refs. [54-60]. This Naimark theorem states [83] that the non-orthogonal spectral decomposition $E(\lambda)$ of a hermitian operator H is of the Carleman type (which is unique for the maximal hermitian operator), *i.e.* it *can be approximated* by a succession of the self-adjoint operators, the spectral functions of which do weakly converge to the spectral function $E(\lambda)$ of the operator H.

Namely, by exploiting that Naimark theorem, it has been shown in [54-59] (more details having been added in [60,65,66,78,81,82]) that, for systems with continuous energy spectra, time can be introduced as a quantum-mechanical observable, canonically conjugate to energy. More precisely, the time operator resulted to be maximal hermitian, even if not self-adjoint. Then, in [59 (1),66(3),81,82] it was clarified that time can be introduced also for the systems with energy discrete spectra as a quantum-mechanical observable, canonically conjugate to energy, and the time operator resulted to be quasi-self-adjoint (more precisely, it can be chosen as an almost self-adjoint operator with practically almost any degree of the accuracy).

We have also to note that there is known in the literature the so-called *positive-operator-value-measure* (POVM) approach, often used in the second set of papers on time in quantum physics (for instance, in [61-64, 67-77,79,80]. This approach, in general, is well-known in the various approaches to the quantum theory of measurements approximately from the sixties and had been applied in the simplest form for the time-operator problem in the case of the free motion already in [84]. Then, in [61-64, 67-77,79,80] (often with certain simpli- fications and abbreviations) it was affirmed that the generalized decomposition of unity (or POV measures) is reproduced from any self-adjoint extension of the time operator into the space of the extended Hilbert space (usually, with negative values of energy E in the left semi-axis) citing the Naimark's dilation theorem from [85]. As to our approach, it is based on *another* Naimark's theorem (from [83]), cited above, and *without* any extension of the physical Hilbert space of usual wave functions (wave packets) with the subsequent return projection to the previous space of wave functions; and, moreover, it had been published in [54-57,59,60] earlier than [61-64,67-77, 79,80]. Being based on the earlier published remarkable Naimark theorem [83], it is *much more direct, simple and general, and at the same time mathematically not less rigorous* than POVM approach!

From the simple analysis of the articles [57,78,81,82], based on the remarkable Naimark theorem [83], one can see that the appearance of these articles, does demonstrate that the problem of time as an observable in quantum mechanics is factually and practically resolved for the systems with continuous spectra, and the alternative approach presented in the articles [73-77,79,80], based on the another Naimark theorem [85], which does not contradict this conclusion, in fact does partially support it.

Time as a quantum observable in quantum mechanics for systems with continuous spectra. For systems with continuous energy spectra, the following simple operator, canonically conjugate to energy, can be introduced for time:

$$\hat{t} = \begin{cases} t & \text{in the (time) } t\text{-representation,} \quad (1a) \\ -i\hbar \dfrac{\partial}{\partial E} & \text{in the (energy) } E\text{-representation} \quad (1b) \end{cases}$$

which is not self-adjoint, but is *hermitian*, and acts on square-integrable space-time wavepackets in representation (1a), and on their Fourier-transforms in representation (1b), once the point $E=0$ is eliminated (*i.e.*, once one deals only with moving packets, *i.e.*, excludes any *non-moving* back tails, as well as, of course, the zero flux cases)[2]. It has been shown already in [54-57,59,60]. The elimination of the point E=0 is not restrictive since the "rest" states with the zero velocity, the wave-packets with *non-moving* rear tails, and the wave-packets with zero flux are unobservable.

Operator (1b) is defined as acting on the space P of the continuous, differentiable, square-integrable functions $f(E)$ that satisfy the conditions

[1]Namely that fact that time cannot be represented by a self-adjoint operator is known to follow from the semi-boundedness of the continuous energy spectra, which are bounded from below (usually by the value zero). Only for an electrically charged particle in an infinite uniform electric field, and for other very rare special systems, the continuous energy spectrum is not bounded and extends over the whole energy axis from $-\infty$ to ∞.

$$\int_0^\infty |f\ (E)|2\ \mathrm{d}E\ <\ \infty,\ \int_0^\infty \left|\partial f\,(E)/\partial E\right|^2 dE < \infty,$$

$$\int_0^\infty |f\ (E)|2\ E2\ \mathrm{d}E\ <\ \infty \qquad (2)$$

(the notation $< \infty$ in (2) denotes the finite value of the integrals from the left) and the condition

$$f(0)\ =\ 0 \qquad (3)$$

which is a space *P dense* in the Hilbert space of L^2 functions defined (only) over the semi-axis $0 \leq E < \infty$. Obviously, the operator (1a,b) is hermitian, *i.e.* the relation $(f_1,\ \hat{t}\ f_2) = ((\ \hat{t}\ f_1),\ f_2)$ holds, only if all square-integrable functions $f(E)$ in the space on which it is defined vanish for $E=0$. And also the operator $\hat{t}^{\,2}$ is hermitian, *i.e.* the relation $(f_1,\ \hat{t}^{\,2} f_2) = ((\hat{t}\ f_1),\ ((\hat{t}\ f_2)) = (\hat{t}^{\,2} f_1,\ f_2)$ holds under the same conditions.

Operator $\hat{t}$ has no hermitian extension because otherwise one could find at least one function $f_0(E)$ which satisfies the condition $f_0(0) \neq 0$ but that is inconsistent with the propriety of being hermitian. So, according to [86], $\hat{t}$ is a maximal hermitian operator.

Essentially because of these reasons, earlier Pauli (see, for instance, [45]) rejected the use of a time operator: and this had the result of practically stopping studies on this subject for about forty years.

However, as far back as in [87] von Neumann had claimed that considering in quantum mechanics only self-adjoint operators could be too restrictive. To clarify this issue, let us quote an explanatory example set forth by von Neumann himself [87]: Let us consider a particle, free to move in a spatial semi-axis ($0 \leq x < \infty$) bounded by a rigid wall located at $x = 0$. Consequently, the operator for the momentum x- component of the particle, which reads

$$\hat{p}_x = -i\hbar \frac{\partial}{\partial x}\ ,$$

is defined as acting on the space of the continuous, differentiable, square-integrable functions $f\ (x)$ that satisfy the conditions

$$\int_0^\infty |f\,(x)|^2 \mathrm{d}x < \infty,\ \int_0^\infty |\partial f\,(x)/\partial x|^2 \mathrm{d}x < \infty,$$

$$\int_0^\infty |f\,(x)|^2\, x^2\ \mathrm{d}x < \infty$$

(here the notation $< \infty$ denotes the finite value of the integrals from the left) and the condition

$$f\,(0) = 0$$

which is a space *Q dense* in the Hilbert space of L^2 functions defined (only) over the spatial semi-axis $0 \leq x < \infty$. Therefore, operator $\hat{p}_x = -i\hbar \frac{\partial}{\partial x}$ has the same mathematical properties as operator $\hat{t}$ (1a,b) and consequently it is not a self-adjoint operator but it is only a maximal hermitian operator. Nevertheless, it is an observable with an obvious physical meaning. And the same properties has also the radial momentum operator

$$\hat{p}_r = -i\hbar \frac{\partial}{\partial r} + \frac{1}{r}\ \ (0 < r < \infty).$$

By the way, one can easily demonstrate (see, for instance, [47]) that in the case of (hypotetical) quantum-mechanical systems with the continuous energy spectra *bounded from below and from above* ($E_{min} < E < E_{max}$) the time operator (1a,b) becomes a *really self-adjoint operator* and has a discrete time spestrum, with the "the time quantum" $\tau = \hbar / d$, where $d = \left| E_{max} - E_{min} \right|$.

In order to consider *time as an observable in quantum mechanics* and to define the observable mean times and durations, one needs to introduce not only the time operator, but also, in a self-consistent way, the measure (or weight) of averaging over time. In the simple one-dimensional (1D) and one-directional motion such *measure* (weight) can be obtained by the simple quantity:

$$W(x,t)\mathrm{d}t = \frac{j(x,t)\mathrm{d}t}{\int_{-\infty}^{\infty} j(x,t)\mathrm{d}t}, \qquad (4)$$

where the probabilistic interpretation of $j(x,t)$ (namely in *time*) corresponds to the flux probability density of a particle passing through point x at time t (more precisely, passing through x during a unit time interval, centered at t), when travelling in the positive x-direction.. Such a measure had not been postulated, but is just a direct consequence of the well-known probabilistic (*spatial*) interpretation of $\rho\ (x,\ t)$ and of the continuity relation

$$\partial\rho(x,t)/\partial t\ +\ \mathrm{div} j(x,t) = 0 \qquad (5)$$

for particle motion in the field of any hamiltonian in the description of the 1D Schroedinger equation. Quantity $\rho(x,t)$ is the probability of finding a moving particle inside a unit space interval, centered at point x, at time t. The probability density $\rho(x,t)$ and the flux probability-density $j(x,t)$ are related with the wave function $\Psi\ (x,t)$ by the usual definitions $\rho(x,t) = |\Psi(x,t)|^2$ and $j(x,t) = \mathrm{Re}\ [\Psi^*(x,t)\ (\hbar/i\mu)\ \partial\Psi(x,t)/\partial x]$. The measure (4) was firstly investigated in [57,59,60,65,66].

When the flux density $j(x,t)$ changes its sign, the quantity $W(x,t)\mathrm{d}t$ is no longer positive definite and it acquires a physical meaning of a probability density *only* during those partial time-intervals in which the flux density $j(x,t)$ does keep its sign. Therefore, let us introduce the two measures, by separating the positive and the

[2]Such a condition is enough for operator (1a,b) to be a "*maximal hermitian*" (or "*maximal symmetric*") *operator* [53-57,59,60,78,81,82], according to Akhiezer & Glazman's terminology.

negative flux-direction values (*i.e.*, flux signs):

$$W_{\pm}(x,t)\mathrm{d}t = \frac{j_{\pm}(x,t)\mathrm{d}t}{\int_{-\infty}^{\infty} j_{\pm}(x,t)\mathrm{d}t} \qquad (4a)$$

with $j_{\pm}(x,t)=j(x,t)\Theta(\pm j)$ where $\Theta(z)$ is the Heaviside step function. It had been made firstly in [60,65,66]. Actually, one can rewrite the continuity relation (5) for those time intervals, for which $j = j_+$ or $j = j_-$ as follows:

$$\frac{\partial \rho_{>}(x,t)}{\partial t} = -\frac{\partial j_{+}(x,t)}{\partial x} \quad \text{and} \quad \frac{\partial \rho_{<}(x,t)}{\partial t} = -\frac{\partial j_{-}(x,t)}{\partial x} \qquad (6)$$

(the equalities (6) do formally serve also as a definitions of $\rho_{>}(x,t)$ and $\rho_{<}(x,t)$), respectively.

Then, one can eventually define the mean value $< t(x) >$ of the time t at which a particle passes through position x (when travelling in only one positive x-direction) , and $< t_{\pm}(x) >$ of the time t at which a particle passes through position x, when travelling in the positive or negative direction, respectively :

$$<t(x)> = \frac{\int_{-\infty}^{\infty} tj(x,t)dt}{\int_{-\infty}^{\infty} j(x,t)dt} =$$

$$\frac{\int_{0}^{\infty} dE \frac{1}{2}[G^{*}(x,E)\hat{t}vG(x,E) + vG^{*}(x,E)\hat{t}G(x,E)]}{\int_{0}^{\infty} dEv|G(x,E)|^{2}} \qquad (7a)$$

where $G(x,E)$ is the Fourier-transform of the moving one-dimensional (1D) wave packet

$$\Psi(x,t) = \int_{0}^{\infty} G(x,E)\exp(-iEt/\hbar)\,\mathrm{d}E$$

$$= \int_{0}^{\infty} g(E)\varphi(x,E)\exp(-iEt/\hbar)\mathrm{d}E \qquad (8)$$

when going on from the time representation to the energy one,

$$<t_{\pm}(x)> = \frac{\int_{-\infty}^{\infty} tj_{\pm}(x,t)dt}{\int_{-\infty}^{\infty} j_{\pm}(x,t)dt} \quad , \qquad (7b)$$

and also the mean durations of particle 1D transmission from x_i to $x_f > x_i$ and 1D particle reflection from the region (x_i, ∞) into $x_f \leq x_i$:

$$<\tau_T(x_i, x_f)> = <t_+(x_f)> - <t_+(x_i)> \text{ and}$$
$$<\tau_R(x_i, x_f)> = <t_-(x_f)> - <t_+(x_i)> , \qquad (7c)$$

respectively. Of course, it is possible to pass in **Eq.7b** also to integrals $\int_{0}^{\infty} dE\ldots$, similarly to (7a) and (8) by using the unique Fourier (Laplace) - transformations and the energy expansion of $j_{\pm}(x,t)=j(x,t)\Theta(\pm j)$, but it is evident that they result to be rather bulky. The generalization for the three-dimensional motions is given in [82].

Now, one can see that two canonically conjugate operators, the time operator (1) and the energy operator

$$\hat{E} = \begin{cases} E & \text{in the energy } (E\text{-}) \text{ representation,} \\ -i\hbar\dfrac{\partial}{\partial t} & \text{in the time } (t\text{-}) \text{ representation} \end{cases} \qquad (9)$$

satisfy the typical commutation relation

$$[\hat{E}, \hat{t}] = i\hbar . \qquad (10)$$

Although up to now according to the Stone and von Neumann theorem [88] the relation (10) has been interpreted as holding *only* for the pair of the self-adjoint canonically conjugate operators, in both representations, and it was not directly generalized for *maximal hermitian operators*, the difficulty of such direct generalization has in fact been by-passed by introducing $\hat{t}$ with the help of the *single-valued* Fourier(Laplace)-transformation from the t-axis ($-\infty < t < \infty$) to the E-semi-axis ($0 < E < \infty$) and by utilizing the peculiar mathematical properties of maximal hermitian operators.

Actually, from **Eq.10** the uncertainty relation

$$\Delta E\ \Delta t \geq \hbar/2 \qquad (11)$$

(where the standard deviations are $\Delta a = \sqrt{\mathrm{D}a}$, quantity D$a$ being the variance $\mathrm{D}a = <a^2> - <a>^2$; and $a = E, t$, while $<\ldots>$ denotes an average over t by the measures $W(x,t)\mathrm{d}t$ or $W_{\pm}(x,t)\mathrm{d}t$ in the t-representation or an average over E similar to the right-hand-part of (7a) and (8) in the E-representation) was derived by the simple generalizing of the similar procedures which are standard in the case of self-adjoint canonically conjugate quantities. Moreover, relation (10) satisfies the Dirac "correspondence principle", since the classical Poisson brackets $\{q_0 , p_0\}$, with $q_0 = t$ and $p_0 = -E$, are equal to unity [89]. In [57] (see also [59]) it was also shown that *the differences* between the mean times at which a wave-packet passes through *a pair* of points obey the Ehrenfest correspondence principle; in other words, in [57,59] the Ehrenfest theorem was suitably generalized.

After what precedes, one can state that, for systems with continuous energy spectra, the mathematical properties of the maximal hermitian operators (described, in particular, in [57,59]), like $\hat{t}$ in **Eq.1**, are *sufficient* for considering them as quantum observables: Namely, the *uniqueness* of the "spectral decomposition" (also called spectral function) for operators $\hat{t}$, as well as for $\hat{t}^{\,n}$ ($n>1$) guarantees (although such an expansion is not

orthogonal) the *equivalence* of the mean values of any analytic functions of time, evaluated either in the *t*- or in the *E*-representations. In other words, the existence of this expansion is equivalent to a *completeness relation* for the (formal) eigen functions of $\hat{t}^{\,n}$ ($n \geq 1$), corresponding with any accuracy to *real eigen values* of the continuous spectrum; such eigen functions belonging to the space of the square integrable functions of the energy *E* with the boundary conditions like (2)-(3) (see details in [81,82]).

From this point of view, there is *no practical difference between self-adjoint and maximal hermitian operators* for systems with continuous energy spectra.

Time as a quantum observable in quantum mechanics for systems with discrete spectra. For systems with discrete energy spectra it is natural (following [59,81,82]) to introduce wave packets of the form

$$\psi(x,t) = \sum_{n=0}^{\infty} g_n \varphi_n(x) \exp[-i(\varepsilon_n - \varepsilon_0)t/\hbar] \qquad (12)$$

(where $\varphi_n(x)$ are orthogonal and normalized wave functions of system bound states which satisfy equation $\hat{H}\varphi_n(x) = \varepsilon_n \varphi_n(x)$, $\hat{H}$ being the system Hamiltonian; $\sum_{n=0}^{\infty} |g_n|^2 = 1$; here we factually omitted a non-significant phase factor $\exp(-i\varepsilon_0 t/\hbar)$ as being general for all terms of the sum $\sum_{n=0}^{\infty}$) for describing the evolution of systems in the regions of the purely discrete spectrum. Without limiting the generality, we choose moment $t = 0$ as an initial time instant.

Firstly, we shall consider those systems, whose energy levels are spaced with distances for which the maximal common divisor is *factually existing*. Examples of such systems are *harmonic oscillator, particle in a rigid box* and *spherical spinning top*. For these systems the wave packet (12) is a periodic function of time with the period (*Poincaré cycle time*) $T = 2\pi\hbar/D$, *D* being the maximal common divisor of distances between system energy level.

In the *t*-representation the relevant energy operator $\hat{H}$ is a self-adjoint operator acting in the space of *periodical* functions whereas the function $t\psi(t)$ does not belong to the same space. In the space of periodical functions the time operator $\hat{t}$, even in the eigen representation, has to be also a periodical function of time *t*. This situation is quite similar to the case of azimuth momentum φ, canonically conjugated to angular momentum $\hat{L}_z$ (see, for instance, [90,91]). Utilizing the example and result from [92], let us choose, instead of *t*, a *periodical* function

$$\hat{t} = t - T \sum_{n=o}^{\infty} \Theta\,(t - [2n+1]\,T/2)$$

$$+\; T \sum_{n=o}^{\infty} \Theta\,(-t - [2n+1]\,T/2) \qquad (13)$$

which is the so-called saw-function of *t* (see **Figure 1**).

This choice is convenient because the periodical function of time operator (13) is linear function (one- directional) within each Poincaré interval, *i.e.* time conserves its flowing and its usual meaning of an *order* parameter for the system evolution.

The commutation relation of the *self-adjoint* energy and time operators acquires in this case (discrete energies and periodical functions) the form:

$$[\hat{E}, \hat{t}\,] = i\hbar\{1 - T \sum_{n=0}^{\infty} \delta\,(t - [2n+1]\,T)\}. \qquad (14)$$

Let us recall (see, e.g. [92]) that a generalized form of uncertainty relation holds

$$(\Delta A)^2 \cdot (\Delta B)^2 \geq \hbar^2 [\,\langle N \rangle\,]^2 \qquad (15)$$

for two self-adjoint operators $\hat{A}$ and $\hat{B}$, canonically conjugate each to other by the commutator

$$[\hat{A}, \hat{B}\,] = i\hbar\hat{N}, \qquad (16)$$

$\hat{N}$ being a third self-adjoint operator. One can easily obtain

$$(\Delta E)^2 \cdot (\Delta t)^2 \geq \hbar^2 \left[1 - \frac{\mathrm{T}\,|\psi(\mathrm{T}/2 + \gamma)|^2}{\int_{-\mathrm{T}/2}^{+\mathrm{T}/2} |\psi(t)|^2\, dt}\right], \qquad (17)$$

where the parameter γ (with an arbitrary value between $-T/2$ and $+T/2$) is introduced for the univocality of calculating the integral on right part of (17) over *dt* in the limits from $-T/2$ to $+T/2$, just similarly to the procedure introduced in [90] (see also [92]).

From (17) it follows that when $\Delta E \to 0$ (*i.e.* when $|g_n| \to \delta_{nn'}$) the right part of (17) tends to zero since $|\psi(t)|^2$ tends to a constant. In this case the distribution of time instants of wavepacket passing through point *x* in the

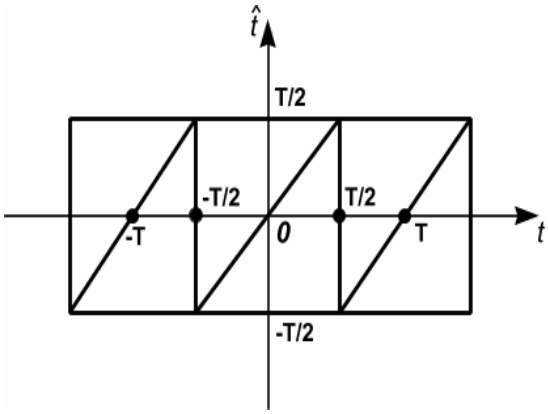


Figure 1. The periodical saw-tooth function for time operator in the case of (13).

limits of one Poincaré cycle becomes uniform. When $\Delta E >> D$ and $|\psi(T+\gamma)|^2 << T^{-1} \cdot \int_{-T/2}^{T/2} |\psi(t)|^2 \, dt$, the periodicity condition may be inessential for $\Delta t << T$, *i.e.* (17) *passes to uncertainty relation* (11), which is just the same one as for systems with continuous spectra.

One can also obtain the expression for the time operator (13) in energy representation [59,82].

In general cases, for excited states of nuclei, atoms and molecules, *level distances in discrete spectra have not strictly defined the maximal common divisor and hence they have not the strictly defined time of the Poincaré cycle*. And also there is no strictly defined passage from the discrete part of the spectrum to the continuous part. Nevertheless, even for those systems one can introduce an approximate description (and with any desired degree of the accuracy within the chosen maximal limit of the level width, let us say, γ_{lim}) by quasi-cycles with quasi-periodical evolution and for sufficiently long intervals of time the motion inside such systems (however, less than $\hbar/\gamma_{lim}$) one can consider as *a periodical motion also with any desired accuracy*. For them one can choose (define) a time of the Poincare' cycle with any desired accuracy, including in one cycle as many quasi-cycles as it is necessary for demanded accuracy. Then, with the same accuracy the *quasi-self-adjoint time operator* (13) can be introduced and all time characteristics can be defined.

In the degenerate case when at the state (12) the sum $\sum_{n=0}^{\infty}$ contains only one term ($g_n \rightarrow \delta_{nn'}$), the evolution is absent and the time of the Poincare' cycle is equal formally to infinity.

For systems with continuous and discrete regions of the energy spectrum, one uses both forms: (1) for the continuous energy spectrum and (13) for the discrete energy spectrum. As a concluding remark, it is possible to state that the *mathematical* properties of $\hat{t}$ and $\hat{t}^{\,n}$ ($n>1$) are quite enough for considering *time* as *a quantum-mechanical observable* (like for energy, momentum, spatial coordinates,...) *without having to introduce any new physical postulates*.

Time analysis of quantum processes, based on time operator. Let us limit ourselves here by only some known results and perspectives:

1) Now there are certain foundations to accept [81,82] that energy-time uncertainty relations (11) with (17) can help to attenuate endless debates on their interpretation, originated in [44,45].

2) Time analysis of the motion of the non-relativistic particles and photons revealed not only the similarity in the motion of particles and photons [93-95] (see also [66(2 and 3),78,81,82]) but did also brought to the introduction of the maximal hermitian time operator for quantum electrodynamics (at least, for the 1D photon motion [78, 81,82]).

3) There are already known two measures of averaging on time in quantum mechanics. Earlier it was reported on the first measure which is related with the particle passing through space point or interval (volume). The second measure is related with the particle accumulation or dwelling (or sojourning) inside the limited space interval (volume) during passing through it (it is described, in particular in reviews [78,81,82]).

4) Actually, the time operator (1) has been rather fruitfully used in the case of the tunnelling times and, generally, in the time analysis of tunnelling processes. It had established that practically all earlier known particular tunnelling times appear to be the special cases of the mean tunnelling time or of the square root of the variance in the tunnelling-time distribution (or pass into them under some boundary conditions), defined within the general quantum-mechanical approach. It had been carried out in some reviews (in particular, in [78,81,82], see also [96]). Then, a lot of other interesting results concerning time behaviour of tunnelling particles and photons inside a barrier had been revealed, including the experimental revealing of the superluminal group velocities of tunnelling photons [97-99].

It is meaningful to stress also that, although any direct classical limit for particle tunnelling through potential barrier with sub-barrier energies is really absent, *there is the direct classical limit for the wave-packet tunnelling*. Let us recall *real evanescent and anti-evanescent waves*, well-known in *classical* optics and in *classical* acoustics (as it was, for instance, mentioned in [78 (conclusions IV and V)] and, in a more detailed manner, in [82 (conclusion 4)]).

5) An actual perspective for the nearest future is opened for generalizing the time analysis of quantum processes for more complicated particle and photon motions (for instance, such as along helixes and motions through two-dimensional and three-dimensional (including non-spherical) potential barriers *etc*).

6) Similar derivations and conclusions with quite evident generalization can be carried out for time operator in relativistic quantum mechanics (the Klein-Gordon case and the Dirac case). It is rather perspective (but, of course, not always simple) to develop the analysis of four-position operators for other relativistic cases, especially to analyze the localization problems. A review of the preliminary results on this topic is already appeared [100].

7. MORE DETAILED COMMENTS ON THE COMPLEX OF PROBLEMS CONNECTED WITH THE ORIGIN OF THE BIOLOGIC LIFE

Now let us analyze, in a condensed way, one of the great

natural problems marked in [13] – *the problem of the reductionism of biology to physics* (including, first of all, *the problem of the physical and chemical explanation of the origin of the biologic life*).

Explanation of the origin of the biologic life in terms of physics and, in general, of natural sciences (chemistry *etc*., including also mathematics) ↔ there is a problem of the origin of the genetics, genetic code (or at least a small set of several codes) which is unique for all the terrestrial biosphere, and the defense mechanisms for the defense of the organism development during cell reproduction,…

↔ there is an inevitable choice (dilemma): *either* a natural event (or process) like a certain jump which is similar to some kind of phase transition (or like to synergetic process, or even like the irrational many-world interpretation), *or* a supreme intelligent design of a super-human creative basis (or a Creator).

Any attempt of the natural origin is failed. And not only because the self-origin of only one self-reproducing cell has not a scientifically reliable explanation in the limits of modern physics (the probability of the chance formation of the protein configuration, containing still 500 nucleotides, is extremely small, *i.e.* near $1/10^{950}$, and for the cell formation it is necessary at least 250 different proteins). There are no scientific explanations yet even for the following facts and no answers for the following problems:

How a numerous quantity of the chemical reactions could take place in a very limited space volume for create one protein molecule?

How there were created the conditions, which were necessary for uniting some components and at the same time were unfavorable for uniting other components, and how then the successive creation of a protein (or RNA or DNA) molecule can happen?

If even a principal possibility of the formation of the simplest protein components (DNA) had been shown in the known Oparin, Miller (*etc*) experiments under the special laboratory conditions, all the same it is very remote from the conditions of the primordial earth or of the unstable cosmos. So, no terrestrial or cosmic origin of cells (moreover, with the genetic structure) are impossible!

And how one can explain that

a) The genetic information in the DNA can be read only by the specific ferments, for the creation of which the special information is also coded in the DNA.

b) The biochemical process of the protein synthesis is the most complicated process between all known biochemical processes in the cell, and also some protein is already necessary for the protein production. Then, the genetic code is beforehand required for the information transfer from the DNA to the protein, and such code is almost universal for the whole terrestrial biosphere.

c) And finally, the genetic code has the vitally necessary control system, which is, in its turn, is coded in the DNA.

It is impossible to explain all these facts in the natural way.

d) And how one (or almost one) main genetic code for the whole terrestrial biosphere had been originated?

Nobody could elaborate somehow working model of the origin of even one self-reproducing cell yet.

The first main part of this problem of the origin consists in the absence of the answer to the following question: *how* had been originated the conditions, which are vitally necessary for living systems now, during that time when the life had been absent but which are created by only living systems! So, it is absolutely unclear: what had been earlier – habitat with is necessary for the life, or the living organisms in the medium which had not supported the life.

The second main part of this problem consists in the mystery of the origin of the enormous quantity of the coded genetic information.

Finally, there is no doubt that *the whole terrestrial biosphere is a wonderfully balanced eco-system of the irreducible complexity and integrity*. The interaction of all its components (flora, fauna, micro-organisms and habitat) is such that the disappearance of even if one of them will bring to the disappearance of the whole biosphere.

So, it is not surprising that during the last ten (or somewhat more) years the number of scientific papers dedicated to the critics of the natural evolutional biologic and pre-biologic theories has become to increase [101-104].

There some, may be, naturalists who do still hope that certain synergetic processes can initiate the self-organization of the non-living matter into the living organisms. But now it is known (see, for instance, [105]) that *all concrete macroscopic systems with known history of their origin, which are more highly ordered than their environment, were created not by rare occasional fluctuations, but under the direct influence of external forces or as a result of bifurcations caused by some non-linearity and external forces in the open systems*. Moreover, I.Prigogine denied that revealed by him processes of local decreasing of entropy can explain the origin of the alive from the non-alive [106]: "The point is that in a non-isolated system there exists a possibility for formation of ordered, low-entropy structures at sufficiently low temperatures. This ordering principle is responsible for the appearance of ordered structures such as crystals as well as for the phenomena of phase transitions. *Unfortunately this principle cannot explain the formation of biological structures*."

Returning to the *direct* analysis of the problem of the reductionism of biology to physics in the narrow sense ("if the biology (at least molecular biology and genetics) can be totally explained in terms of physics (and chem-

istry)"), I can recommend to pay a particular attention to the discussion on the special problem of the principal possibility of the explanation of the cell self-reproduction in terms of quantum mechanics, initiated by E.Wigner [107], then continued by M.Eigen [18] and afterwards analyzed by M.V.Vol'kenstein [109]. Firstly, E.Wigner had simply demonstrated that really in the stochastic quantum-mechanical description the process of the cell self-reproduction cannot be explained by quantum mechanics. Then M.Eigen had shown that the possibility of the cell self-reproduction can be explained by quantum mechanics if and only if the evolution matrix (the *S*-matrix of the process) is *specially instructed* for this aim. Further M.V.Vol'kenstein in his analytic review [109] had expressed his expectation that M.Eigen in his future study of the pre-biologic evolution can find the possibility of such special instruction. But up to now nobody had revealed such possibility! As to me, I can see only a certain similarity (of course, partial) between two kinds of processes (with are more intellectual than naturalistic, by the way): between the process of the human writing of certain scientific files in modern computer devices and the process of the supreme-Intelligence-design writing of certain genetic programs (including the genetic program of the cell reproduction) in cells of alive organisms.

8. MORE DETAILED COMMENTS ON THE COMPLEX OF PROBLEMS CONNECTED WITH THE UNIVERSE ORIGIN

1) Earlier, after Enlightenment till approximately 1920, scientists in the natural sciences did usually consider the Universe as *eternally existing and eternally moving.*

Now the most convincing arguments against the model of the eternally existing Universe are:

a) *the second law of thermodynamics which does inevitably bring to the heat death of the Universe,*

b) *the observed cosmic microwave background* .

The most surprising conclusion of the revealed non-stationary state of the Universe is *the existence of the* "*beginning*"*, under which the majority of physicists understand the beginning of the Universe expansion.*

The cosmologic problem as the problem of the origin and evolution of the Universe has initiated to be analyzed by A.Einstein (after 1917) and now it is connected with papers of many other physicists. The first several authors had been G.Lemaître (who proposed what became known as the Big Bang theory of the origin of the Universe, although he called it his "hypothesis of the primeval atom"), A.Friedman and G.Gamow.

And what namely had been in the "beginning"? Gamow had assumed in 1921 that the expansion had initiated from the super-condensed hot state as a result of the *Big Bang*, to which he and others had ascribed the time moment $t = 0$, *i.e.* the beginning of the Universe history. The initial state in this model is *postulated*. However, *the nature* of the initial super-condensed hot Universe state *is not known*. Such initial point (or super-small region), in which the temperature, pressure, energy density *etc* had reached the anomalous huge (almost infinite) values, can be considered as a particular point, where The "physical" processes cannot be described by physical equations and in fact are excluded from the model analysis.

Strictly speaking, namely in the region of this point (from $t = 0$ till $\sim t_0 = 10^{-44}$ sec, where t_0 is the Planck time) is arising the general problem of the world origin and also the choice dilemma: the beginning of the Universe formation from vacuum ("nothing") is *either* a result of the irrational randomness after passing from other space-time dimensions or from other universe, caused by some unknown process, *or* a result of the creation of the expanding Universe (together with the laws of its functioning) by the supreme intelligent design from *nigilo*.

The framework for the standard cosmologic model relies on Albert Einstein's general relativity and on simplifying assumptions (such as homogeneity and isotropy of space). There are even non-standard alternative models. Now there are many supporters of Big Bang models. The number of papers and books on standard and non-standard versions of the cosmologic Big Bang models is too enormous for citing in this not very large paper (it is possible to indicate, only for instance, [110-113] for initial reading in cosmology of the Universe and in the different quasi-classical and quantum approaches in cosmology for description of the creation and the initial expansion of the Universe). However, there is no well-supported model describing the action prior to 10^{-15} seconds or so. Apparently a new unified theory of quantum gravitation is needed to break this barrier. Understanding this earliest of eras in the history of the Universe is currently one of the greatest unsolved problems in physics.

Moreover, it is worth to underline that many physicists consider that the second law of thermodynamics is universal for all closed systems, including also our Universe as a whole (which is closed in naturalistic one-world view). Therefore the heat death is inevitable (see, for instance, [13] and especially [114]). Finally, to-day a lot of attention of researchers is dedicated to the problems of the hypothesis of dark mass and of dark energy.

2) From 1973 (and particularly after eighties) the term "*anthropic principle*", introduced by B.Carter, has become to acquire in the science and out of the science a certain popularity [115,116]. Carter and other authors had been noted that physical constants must have values in the very narrow interval in order the existence of the biologic life can become possible, and that the measured

values of these constants are really found in this interval. In other words, the Universe seems to be exactly such as it is necessary for the origin of the life. If physical constants would be even slightly other, then the life could be impossible.

After meeting such testimonies, a number of scientists had formulated several interpretations of anthropic principle each of which brings the researchers to the world-view choice in its peculiar way. We shall consider here two of them.

According to *the weak anthropic principle* (WAP), the observed values of physical and cosmological constants caused by the necessary demand that the regions, where the organic life would be developed, ought to be possible. And in the context of WAP there is the possibility of choice between two alternatives:

1) *Either* someone does irrationally believe that there are possible an infinity of universes, in the past, in the present and in the future, and we exist and are sure in the existence of our Universe namely because the unique combination of its parameters and properties could permit our origin and existence.

2) *Or* someone does (also irrationally) believe that our unique Universe is created by Intelligent Design of a Creator (or God) and the human being is also created by Creator in order to govern the Universe.

According to *the strong anthropic principle* (SAP), the Universe has to have such properties which permit earlier or later the development of life. This form of the anthropic principle does not only state that the universe properties are limited by the narrow set of values, compatible with the development of the human life, but does also state that this limitation is *necessary* for such purpose. So, one can interpret such tuning of the universe parameters as the testimony of the supreme intelligent design of a certain creative basis. There is also a rather unexpected interpretation of SAP, connected with the eastern philosophy, but it is not widely known.

9. CONCLUSIONS

It is proposed firstly a novel division of the different classes of natural sciences (and in some degree of all sciences) with different objects and paradigms: a) the entirely natural sciences, b) the natural sciences with the essential role of the human factor, or with the human intelligent design, in their objects and c) the sciences, implicated in the origin and the subsequent history of such natural "meta-systems" as the whole Universe and the whole (terrestrial) biosphere.

Several reasons caused here to formulate a new retrospective view of the science history (especially in the field of natural sciences) in XX-XXI: Firstly, under the influence of scientific and technological progress it has been intensified such direction of the science philosophy as the scientific realism (*i.e.* the correspondence of the science to the reality), which has in turn changed three forms: from the naïve realism to the usual realism and then to the critical scientific realism (the last one had been developed under the strong influence of sharp discussions in quantum mechanics). Secondly, some big problems of physics and natural sciences a) sharp problems and paradoxes *revealed in the development of quantum mechanics and quantum theory of measurements*, b) a huge complex of the problems connected with the *Universe origin and the expansion after the Big Bang*, c) the open problem of the origin of the biological life) have been gradually concentrated the attention of the researchers, if not scientifically but at least philosophically, to those problems as to the grand or great problems. And thirdly, the analysis of mathematics in different sciences, beginning from physics, shows that mathematics did now become the branch of the natural sciences (namely of theoretical physics) and in fact generated the final solution of the old problem of time as a quantum observable.

The interpretation questions in the considered here grand and great problems of natural sciences are practically inevitably connected with the world-views of the researchers. Therefore, it is quite clear that the strong divergences in the various interpretations and even paradigms of various researchers, especially relating to these grand and great open problems, can be caused by the incompatibility of their world-views.

Such phenomena, as 1) the enhancement of the philosophy of the critical scientific realism, 2) the problems, the paradoxes and the variety of interpretations in quantum theory, 3) the open problems of origin of the Universe and 4) the unresolved problem of the origin of the biosphere, 5) the clear extension of the role of mathematics in physics and other sciences, 6) the competition of various interpretations and even of the worldviews of researchers in the study of the great and grand problems, are the important peculiarities of the history of natural sciences in XX-XXI, which in many respects define and pre-determine the further science history.

REFERENCES

[1] Boyd, R. (1983) On the current status of the issue of scientific realism. *Erkenntnis*, 45-50.

[2] Bird, A. (1998) Philosophy of science, UCL Press, Montreal, McGill-Queen's University Press, London, 124.

[3] Niiniluotto, I. (2002) Critical scientific realism. Oxford University. Press.

[4] Quine, W.V.O. (1968) Ontological Relativity. *Journal of Philosophy*, LXV, **7**, 185-212.

[5] Polkinghorne, J. (1991) Reason, and Reality, The Rela-

tionship Between Science and Theology, Trinity Press Internat, Philadelphia.
[6] McGrath, A.I. (2004) The science of God, T&T Clark International, London, 139-152.
[7] Kuhn, T.S. (1962; 2nd edition, 1970; 3rd edition, 1996) The Structure of Scientific Revolutions, University of Chicago Press, Chicago, IL.
[8] Einstein, A. (1944) Remarks on Russel's theory of knowledge, Philosophy of Bertrand Russel, Schlipp, A., Muschalek, H., Ed. (1950, 2nd ed.) Dio e gli scienziati [Translation in Italian from German (Original title: Gottbekenntnisse moderner Naturforscher) by Valeria Cremona], Alba: Paoline, , **421**, 30-31.
[9] von Wolfgang Sartorius von Waltershausen (1856, repr. 1965) Gauss zum Gedächtniss, Sändig Reprint Verlag H. R. Wohlwend.
http://www.amazon.de/Gauss-Gedächtnis-Wolfgang-Sartorius-Waltershausen/dp/3253017028
http://en.wikipedia.org/wiki/Mathematics. von Wolfgang Sartorius von Waltershausen (Autor)
[10] Popper, K.R. (1995) In Search of a Better World: Lectures and Essays "On knowledge", from Thirty Years. Routledge.
[11] Bogolyubov, N.N., Medvedev, B.V. and Polivanov, M.K. Ed., Dispersion relations, transactions Problems of modern Physics, N2 [in Russian: Боголюбов Н.Н., Медведев Б.В., and Поливанов М.К. (редакторы), (1957) Дисперсионные соотношения, статьи в сб. Проблемы современной физики, **2**]; Arnold V.I., What is mathematics?, Moscow, MCNMO-press < 3 [in Russian: Арнольд В.И., Что такое математика? М., изд-во МЦНМО, 2008, стр.3].
[12] Galilei, G. (1623) II Saggiatore (in Italian) (Rome)
[13] Ginzburg, V.L. (1999) What problems of physics and astrophysics seem now to be especially important and interesting (30 years later, already on the verge of XXI century), Physics – Uspekhi, **42**, 353-272; (2002) On some advances in physics and astronomy over the past 3 years, **45**, 205-211.
[14] Einstein, A., Podolsky, B. and Rosen, N. (1935) Can Quantum-Mechanical Description of Physical Reality be Considered Complete? *Physical Review*, 47, 777-780.
[15] Bohm, D. (1952) A Suggested Interpretation of the Quantum Theory in Terms of "Hidden Variables" I, *Physical Review*, **85**, 66-179; (1952) A Suggested Interpretation of the Quantum Theory in Terms of "Hidden Variables" II, *Physical Review*, **85**, 80-193; Bohm,D., Aharonov, Y. (1957) Discussion of Experimental Proof for the Paradox of Einstein, Rosen, and Podolsky, *Physical Review*, **108**, 1070-1076; Aharonov, Y. and Bohm D., (1959) Significance of electromagnetic potentials in the quantum theory. *Physical Review*, **115**, 485-491; Bohm, D. and Aharonov, Y. (1960) Further discussion of possible experimental tests for the paradox of einstein, Podolsky and Rosen, N.C. **17**, 964; Aharonov, Y. and Bohm D. (1962) Remarks on the possibility of quantum electrodynamics without potentials, *Physical Review*, **125**, 192; Aharonov, Y. and Bohm D. (1963) Further discussion of the role of electromagnetic potentials in the quantum theory. *Physical Review*, **130**, 1625.
[16] Bell, J.S. (1964) On the Einstein-Podolsky-Rosen paradox. *Physics*, 195-200 [Bell, J.S. (1965)On the Einstein-Poldolsky-Rosen paradox, *Physics 1*, 195-200] ,Bell, J.S, (1987) Speakable and unspeakable in quantum mechanics. Cambridge University Press.
[17] Pais, A. (1979) Einstein and the quantum theory. *Reviews of Modern Physics*, **51**, 863-914.
[18] Popper, K. (1982) A critical note on the greatest days of quantum theory. *Foundations of Physics*, **12**, 971-976.
[19] Holland, P.R. (1993) The quantum theory of motion: an account of the de broglie-bohm causal interpretation of quantum mechanics. Cambridge University Press, Cambridge.
[20] Mermin, N.D. (1993) Hidden variables and the two theorems of john bell. *Reviews of Modern Physics*, **65**, 803-815.
[21] Paty, M. (1995) The nature of Einstein's objections to the copenhagen interpretation of quantum mechanics. *Found. Phys.*, **25**, 183-204.
[22] Dürr, D., Goldstein, S. and Zanghì, N. (1997) Bohmian Mechanics and the Meaning of the Wave Function, in Cohen, R.S., Horne, M., and Stachel, J., Eds., Experimental Metaphysics — Quantum Mechanical Studies for Abner Shimony, **1**; Boston Studies in the Philosophy of Science,**193**, Boston: Kluwer Academic Publishers; Dürr, D. (2001) Bohmsche mechanik als grundlage der quantenmechanik. Springer-Verlag. Berlin
[23] Hardy, L. (1993) Non-locality for 2 particles without inequalities for almost all entangled states. *Physical Review Letters*, **71**, 1665-1668; Sakurai, J.J. (1994) Modern Quantum Mechanics. Addison-Wesley, USA, (see 174-187, 223-232).
[24] Vaidman, L. (1994) Teleportation of quantum states. *Physical Review, A*, **49**, 1473-1476.
[25] Brassard, G., Braunstein, S., and Cleve, R. (1998) Teleportation as a quantum computation. *Physical D*, **120**, 43-47.
[26] Bouwmeester, D., Pan, J.-W., Mattle, K., Eibl, M., Weinfurter, H., and Zeilinger, A. (1997) Experimental quantum teleportation. *Nature*, **390(6660)**, 575-579.
[27] Boschi, D., Branca, S., De Martini, F., Hardy, L. and Popescu, S. (1998) Experimental realization of teleporting an unknown pure quantum state via dual classical and Einstein-Podolsky-Rosen channels, *Physical Review Letter*, **80(6)**, 1121-1125.
[28] Kilin, S.Y. (2001) Quanta and information, Progress in optics, **42**, 1-90.
[29] Riebe, M., Häffner, H., Roos, C. F., Hänsel, W., Ruth, M., Benhelm, J., Lancaster, G. T., Körber, T. W., Becher, C., Schmidt-Kaler, F., James, D.F.V., and Blatt, R., (2004) Deterministic quantum teleportation with atoms. *Nature*, **429**, 734-737.
[30] Ursin, R., *et al.* (2004) Quantum teleportation link across the danube. *Nature*, **430**, 849.
[31] Olmschenk, S., *et al.* (2009) Quantum teleportation between distant matter qubits. *Science*, **323**, 486.
[32] Everett, H. (1957) *Relative State* Formulation of quantum mechanics. *Review of Modern Physics*, **29**, 454-462.
[33] De Witt, B.S.M. (1970) Quantum mechanics and reality. *Physics Today*, **23**, 30-35.
[34] Everett, H. (1973) The theory of the universal wave function. De Witt, B. and Graham, N. Eds., *The Many-Worlds Interpretation of Quantum Mechanics*, Princeton NJ: Princeton University Press.

[35] Deutsch, D. (1986) Three experimental implications of the Everett interpretation, in Penrose, R. and Isham, C.J., Ed., *Quantum Concepts of Space and Time*, The Clarendon Press, Oxford, 204-214.
[36] Tipler, D. (1986) The many-worlds interpretation of quantum mechanics in quantum cosmology. in Penrose, R. and Isham, C.J. Eds., *Quantum Concepts of Space and Time*, The Clarendon Press, Oxford, 204-214.
[37] Albert, D. and Loewer, B. (1988) Interpreting the many worlds interpretation. *Synthese*, **77**, 195-213.
[38] Barvinsky, A.O. and Kamenshchik, A.Y. (1995) Preferred basis in quantum theory and the problem of classicalization of the quantum universe. *Physical Review D*, **52**, 743-757.
[39] Deutsch, D. (1996) The fabric of reality, The Penguin Press, New York.
[40] Lockwood, M., Brown, H.R., Butterfield, J., Deutsch, D., Loewer, B., Papineau, D. and Saunders, S. (1996) Symposium: The "many minds" interpretation of quantum theory. *British Journal for the Philosophy of Science*, **47**, 159-248.
[41] Barrett, J.A. (1999) The quantum mechanics of minds and worlds. Oxford: University Press.
[42] Zeh, H.D. (1970) On the interpretation of measurement in quantum theory. *Foundations of Physics*, **1**, 69-76; Zeh, H.D. (1973) Toward a quantum theory of observation. *Foundations of Physics*, **3**, 109-116.
[43] Pauli, W. (1926) in: Handbuch der Physik, **5(1)**, 60, ed. by Fluegge, S. (Berlin),; see also: Pauli, W. General principles of quantum theory. (Springer; Berlin, 1980).
[44] Aharonov, Y. and Bohm, D. (1961) Time in the quantum theory and the uncertainty relation for time and energy, *Physical Review*, **122**, 1649-1658; Aharonov, Y. and Bohm, D. (1964) Answer to fock concerning the time energy indeterminacy relation. *Physical Review B*, **134**, 1417-1418.
[45] Krylov, N.S. and Fock, V.A. (1947) On two main interpretations of energy-time uncertainty. *Sov. J. Zhetf*, **17**, 93-99; Fock, V.A. (1962) On the energy-time uncertainty and on an attempt to refute it. *Sov. J. Zhetf*, **42**, 1135-1140.
[46] Paul, H. (1962) Über quantenmechanische Zeitoperatoren. *Annalen der Physik*, **9**, 252-261.
[47] Engelman, F. and Fick, E. (1963) Quantentheorie der Zeitmessung. *Zeitschrift für Physikalische A*, **175**, 271-282; (1964) Quantentheorie der Zeitmessung-II. *Zeitschrift für Physikalische A*, **178**, 551-562.
[48] Lippmann, B.A., (1966) Operator for time delay induced by scattering. *Physical Review*, **151**, 1023-1024.
[49] Razavy, M. (1969) Quantum-mechanical conjugate to the Hamiltonian operator. *Nuovo Cimento, B*, **63**, 271-308.
[50] Gien, T.T., (1969) On the operators for time of motion and delay time induced by scattering. Can. *Journal of Physics*, **47**, 278-289; (1970) Delay time and phase of the initial state. *Canadian Journal of Physics*, **48**, 639-652.
[51] Allcock, G.R. (1969) The time of arrival in quantum mechanics. *Annals of Physics* (N.Y.), **53**, 253-285.
[52] Rosenbaum, D.M. (1969) Super hilbert space and the quantum-mechanical time operators. *Journal of Mathematical Physics*, **10**, 1127-1144.
[53] Olkhovsky, V.S. and Recami, E. (1968) Space-time shifts and cross sections in collisions between relativistic wane packets. *Nuovo Cimento*, **A53**, 610-624; (1969) About collision-process lifetimes and causality. *Nuovo Cimento A*, **63**, 814-826; (1970) About a space-time operator in collision descriptions. *Nuovo Cimento*, **4**, 1165-1173.
[54] Olkhovsky, V.S. (1973) On the problem of time operator and collision duration, Ukrainskiy Fiz. Zhurnal [in Ukrainian and Russian], **18**, 1910-1913; Olkhovsky, V.S., Recami, E. and Gerasimchuk, A.I. (1974) Time operator in quantum mechanics – non-relativistic case. *Nuovo Cimento A*, **22**, 263-278.
[55] Recami, E. (1977) A time operator and the time-energy uncertainty relation. in *The Uncertainty Principle and Foundation of Quantum Mechanics* (J.Wiley; London), 21-28; "An operator for observable time", in Proceeding of the XIII Winter School in Theor. *Physics*, (Wroclaw; 1976), **2**, 251-256.
[56] Holevo, A.S. (1978) Estimation of shift parameters of a quantum state. Re Math. Phys., **13**, 379-399; Holevo, A.S. (1982) Probabilistic and statistical aspects of quantum theory, Amsterdam.
[57] Olkhovsky, V.S. (1984) To the investigation of nuclear reactions and decays by analysis of their durations. *Soviet Journal Nuclear Physics*, **15**, 130-148.
[58] Jaworski, W. and Wardlaw, D.M. (1988) Time delay in tunneling: Transmission and reflection time delays. *Physical Review A*, **37**, 2843-2854.
[59] Olkhovsky, V.S. (1990) Non-stationary characteristics in study of nuclear reaction mechanism and kinetics and compound-nucleus properties. *Nukleonika*, **35**, 99-144; (1992) Time analysis of nuclear collisions and decays. *Atti Accademia Peloritana dei Pericolanti, Sci. Mat. Nat.* **70**, 21-40; (1998) in Mysteries, Puzzles and Paradoxes in Quantum Mechanics, Bonifacio, R. Ed., (AIP Conference Proceeding, Amer. Institute of Physics, Woodbury, NY, USA), 272-276.
[60] Olkhovsky, V.S. and Recami, E. (1992) Recent developments in the time analysis of tunnelling processes. *Physics Reports*, **214**, 339-356; Olkhovsky, V.S., Recami, E., Raciti, F. and Zaichenko, A.K. (1995) More about tunnelling times, the dwell time and the Hartman effect. *Journal de Physique*, (France) I, **5**, 1351-1365.
[61] Busch, Grabowski, M. and Lahti, J. (1994) Time observables in quantum theory. *Physics Letters A*, **191**, 357-361.
[62] Kobe, D.H. and Aguilera-Navarro, V.C., (1994) Derivation of the energy-time uncertainty relation. *Physical Review A*, **50**, 933-938.
[63] Blanchard, and Jadczyk, A. (1996) Timeof events in quantum theory. *Helvetica Physica Acta*, **69**, 613-635.
[64] Grot, N., Rovelli, C. and Tate, R.S. (1996) Time of arrival in quantum mechanics. *Physical Review A*, **54**, 4676-4690.
[65] Olkhovsky, V.S. and Agresti, A. (1997) Developments in time analysis of particle and photon tunnelling, in proceeding of the adriatico research conference on tunnelling and its implications (World Sci.; Singapore), 327-355.
[66] Olkhovsky, V.S. (1997) Time analysis of particles and photons. *Physics of the Alive*, **5**, 19-37; (1998) Developments in examining time as a quantum-physical observable. *Physics of the Alive*, **6**, 17-29; Olkhovsky, V. S. (1998-1999) Recent developments on time as a quantum-physical observable, *Atti dell'Academia di*

Pericolanti, classe di scienze fis. mat. e natur., Universita' di Messina, v.LXXVI-LXXVII, 193-209.

[67] Leo'n, J. (1997) Time-of-arrival formalism for the relativistic particle. *Journal of Physics* A, **30**, 4791-4801.

[68] Giannitrapani, R. (1997) Positive-operator-valued time observable in quantum mechanics. *International Journal of Theoretical Physics*, **36**, 1575-1584.

[69] Aharonov, Y., Oppenheim, J., Popescu, S., Reznik, B. and Unruh, W. (1998) Measurement of time of arrival in quantum mechanics. *Physical Review A*, **57**, 4130-4139.

[70] Atmanspacher, H. and Amann, A. (1998) Positive-operator-valued measures and projection-valued measures of noncommutative time operators. *International Journal of Theoretical Physics*, **37**, 629-660.

[71] Toller, M. (1999) Localization of events in space-time. *Physical Review A*, **59**, 960-970.

[72] Kijowski, J. (1999) Comment on "arrival time in quantum mechanics" and "time of arrival in quantum mechanics". *Physical Review A*, **59**, 897-899.

[73] Delgado, V. (1999) Quantum probability distribution of arrival times and probability current density. *Physical Review A*, **59**, 1010-1020.

[74] Muga, J., Papao, J. and Leavens, C. (1999) Arrival time distributions and perfect absorption in classical and quantum mechanics. *Physics Letters A*, **253**, 21-27.

[75] Kochànski, and Wòdkievicz, K. (1999) Operational time of arrival in quantum phase space. *Physical Review A*, **60**, 2689-2699.

[76] Kobe, D.H., Iwamoto, H., Goto, M. and Aguilera-Navarro, V.C., (2001) Tunneling time through a barrier using the local value of a "time operator". *Physical Review A*, **64**, Article ID 022104, 8.

[77] Muga, J., Egusquiza, I., Damborenea, J. and Delgado, V. (2002) Bounds and enhancements for negative scattering time delays. *Physical Review A*, **66**, Article ID 042115, 8 pages.

[78] Olkhovsky, V.S., Recami, E. and Jakiel, J. (2004) Unified time analysis of photon and particle tunneling. *Physical Review*, **398**, 133-178.

[79] Góźdź A. and Dębicki M. (2007) Time operator and quantum projection evolution. *Physics of Atomic Nucle*, **70**, 529-536.

[80] Wang, Z.Y. and Xiong, C.D. (2007) How to introduce time operator. *Annals of Physics*, **322**, 2304-2314.

[81] Olkhovsky, V.S. and Recami, E. (2007) Time as a quantum observable. *International Journal of Modern Physics A*, **22**, 5063-5087; Olkhovsky, V.S. and Recami, E. (2008) New developments in the study of time as a quantum observable. *International Journal of Modern Physics B*, **22**, 1877-1897.

[82] Olkhovsky, V.S. (2009) Time as a quantum observable, canonically conjugated to energy, and foundations of self-consistent time analysis of quantum processes. *Advances in Theoretical and Mathematical Physics*, **2009**, article ID 859710, 83.

[83] Naimark, M.A. (1940) Spectral functions of a symmetric operators. Izvestiya Akademii Nauk SSSR, seriya matematicheskaya [in Russian, partially in English], **4**, 277-318; we stress that it is used here the Carleman type of the spectral function of (a maximal and not only maximal) symmetrical operator, which was cited from the reference [Carleman, T. (1923) Sur les e'quations i'ntegrales a' noyau re'el et syme'trque (Uppsala)] and was utilized as a tool of the proof, based on the approximation of the (maximal and not only maximal) symmetric operator H by such succession of the bounded self-adjoint operators, the spectral functions the spectral functions of which do weakly converge to the spectral function $E(\lambda)$ of the operator H (This circumstance was kindly indicated to the author by Holevo, A.S.).

[84] Aharonov, Y. and Bohm, D. (1961) Time in the quantum theory and the uncertainty relation for time and energy. *Physical Review*, **122**, 1649-1658.

[85] Naimark, M.A. (1943) Positive definite operator functions on a commutative group, Izvestiya Akademii Nauk SSSR. *seriya matematikcheskaya*, **7**, 237-244.

[86] Akhiezer, N.I. and Glazman, I.M. (1981) The Theory of Linear Operators in Hilbert Space, Pitman; Boston, Mass.

[87] Neuman, J. (1932) von, Mathematischen Grundlagen del Quantum Mechanik (Hizzel, Leipzig).

[88] Stone, M.H. (1930) Proceeding Nat. Acad. Sci. USA, **16**, N1.

[89] ter Haar, D. (1971) Elements of hamiltonian mechanics, Oxford.

[90] Judge, D. and Levis, J.L. (1963) On the commutator [L_z, φ].*Physics Letters*, **5**, 190-195.

[91] Carruthers and Nieto, M.M. (1968) Phase and angle variables in quantum mechanics. *Reviews of Modern Physics*, **40**, 411-440.

[92] Davydov, A.S. (1976) (1982) Quantum Mechanics (Pergamon, Oxford).

[93] Chiao, R.Y., Kwiat, G. and Steinberg, A.M. (1991) Analogies between electron and photon tunneling: A proposed experiment to measure photon tunneling times. *Physica B*, **175**, 257-262.

[94] Martin, T. and Landauer, R. (1992) Time delay of evanescent electromagnetic waves and the analogy to particle tunneling. *Physical Review*, **A45**, 2611-2617.

[95] Steinberg A.M. (1995) Conditional probabilities in quantum theory, and the tunneling time controversy. *Physical Review A*, **52**, 32-42.

[96] Abolhasani, M. and Golshani, M. (2000) Tunneling times in the Copenhagen interpretation of quantum mechanics. *Physical Review A*, **62**, 012106, 7.

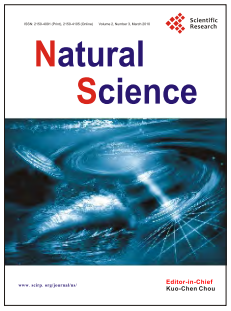
[97] Enders, A. and Nimtz, G. (1992) On superluminal barrier traversal, *Journal of Physics*-I (France), **2**, 1693-1698; (1993) Zero-time tunneling of evanescent mode packets. *Journal of Physics*-I (France), **3**, 1089-1092; (1993) Photonic-tunneling experiments. *Physical Review B*, **47**, 9605-9609; (1993) Evanescent-mode propagation and quantum tunneling. *Physical Review E*, **48**, 632-634; G.Nimtz, in: (1997) Tunneling and its applications. *World Scientific*,Singapore, 223-237.

[98] Steinberg, A.M., Kwiat, G. and Chiao, R.Y. (1993) Measurement of the single-photon tunneling time. *Physical Review Letter*, **71**, 708-711; Chiao, R.Y., Kwiat G., and Steinberg, A.M. (1993) Faster Than Light? Scient.Am., **269**, 38-52; Chiao, R.Y. and Steinberg, A.M. (1997) Tunneling times and superluminality, in progress in optics, by Wolf, E., Ed., **37**, Elsevier Sci., Amsterdam, 346-405.

[99] Longhi, S., *et al.* (2002) Measurements of superluminal optical tunneling times in double-barrier photonic band

gaps. *Physical Review E*, **65**, 046610.
[100] Recami, E., Olkhovsky, V.S. and Maydanyuk, S.P. (2010) On non-self-adjoint operators for observables in quantum mechanics and quantum field theory. accepted and to be published in Internat. *International Journal of Modern Physics A.*
[101] Macnab, R. (1978) Bacterial motility and chemotaxis - molecular-biology of a behavioral system. *CRC Critical Reviews in Biochemistry*, **5**, 291-341; Moorhead, S. and Kaplan, M.M. Ed. (1967) Mathematical challenges to the neo-darwinian interpretation of evolution. Wistar Institute Press ,Philadelphia.
[102] Behe, M.J. (1996) Darwin's Black Box. The biochemical challenge to evolution, the Free Press.
[103] Junker, R. and Scherer, S. Evolution: Ein kritisches Lehrbuch, 4th ed. Giessen (Germany): Weyel Verlag, 1998; 5th ed. Giessen (Germany): Weyel Verlag, 2001; 6th ed. Giessen (Germany), Weyel Verlag, 2006.
[104] Olkhovsky, V.S. (2001) Comparison of the faith postulates in evolutionism and creationism with respect to modern scientific data. *Physics of the Alive*, **9**, 108-121.
[105] Prigogine, I. and Stengers, I. (1984) Order out of chaos. man's new dialogue with nature, Heinemann, London; Nicolis, G. and Prigogine, I. Exploring complexity, Freeman, W. and Co, N.Y.,1989.
[106] Prigogine, I., Nicolis, G. and Babloyants, A. (1972) Thermodynamics of evolution. *Physics Today*, **25**, 23.
[107] Symmetries and reflections, scientific essays of Eugen Wigner, Indiana University Press, Bloomington–London, 1970; essay 11 ("The possibility of existence of a self-reproducing system").
[108] Eigen, M. (1971) Self-organization of matter and the evolution of biological macromolecules. *Naturwiss*, **58**, 465-523.
[109] Vol'kenstein, M.V. (1973) Physics and biology. *Soviet Physics Uspekhi*, **16**, 207-216; see also: Vol'kenstein, M.V. (1988) Complementarity, physics and biology. *Soviet Physics Uspekhi*, **31**, 140-150.
[110] Hartle, J.B. and Hawking, S.W. (1983) Wave function of the Universe. *Physical Review D*, **28**, 2960-2975.
[111] Vilenkin, A. (1994) Approaches to quantum cosmology. *Physical Review D*, **50**, 2581-2594.
[112] Kragh, H. (1996) Cosmology and controversy, Princeton (NJ), Princeton University Press.
[113] Peacock, J. (1999) Cosmological physics, Cambridge University Press.
[114] Adams, F.C. and Laughlin, G. (1997) A dying universe: the long-term fate and evolution of astrophysical objects. *Reviews of Modern Physics*, **69**, 337-372.
[115] Carter, B. (1974) Large number coincidences and the anthropic principle in cosmology. IAU Symposium 63: Confrontation of Cosmological Theories with Observational Data, Dordrecht, Reidel.
[116] Barrow, J.D. and Tipler, F.J. (1986) The anthropic cosmological principle. Clarendon Press, Oxford.

Call for Papers



Natural Science

A Journal Published by Scientific Research Publishing, USA
www.scirp.org/journal/ns

Natural Science is an international journal dedicated to the latest advancement of natural sciences. The goal of this journal is to provide a platform for scientists and academicians all over the world to promote, share, and discuss various new issues and developments in different areas of natural sciences. All manuscripts must be prepared in English, and are subject to a rigorous and fair peer-review process. Accepted papers will immediately appear online followed by printed hard copy. The journal publishes original papers including but not limited to the following fields:

- **Astronomy & Space Sciences**
  - Astronomy
  - Astrophysics
  - Atmospheric Science
  - Space Physics
- **Earth Science**
  - Geography
  - Geology
  - Geophysics/Geochemistry
  - Oceanography
- **Chemistry**
  - Analytical Chemistry
  - Biochemistry
  - Computational Chemistry
  - Inorganic Chemistry
  - Organic Chemistry
  - Physical Chemistry
- **Life Science**
  - Cell Biology
  - Computational Biology
  - Genetics
  - Immunology
  - Medicine/Diseases
  - Microbiology
  - Molecular Biology
  - Neuroscience
  - Pharmacology/Toxicology
  - Physiology
  - Psychology
  - Virology
- **Physics**
  - Applied Physics
  - Atomic, Molecular, and Optical Physics
  - Biophysics
  - High Energy/Particle Physics
  - Material Science
  - Plasma Physics
- **Others**
  - Education
  - History of Science
  - Science and Innovations

We are also interested in: 1) Short Reports—2-5 page papers where an author can either present an idea with theoretical background but has not yet completed the research needed for a complete paper or preliminary data; 2) Book Reviews—Comments and critiques.

Notes for Intending Authors

Submitted papers should not be previously published nor be currently under consideration for publication elsewhere. Paper submission will be handled electronically through the website. For more details, please access the website.

Website and E-Mail

http://www.scirp.org/journal/ns ns@scirp.org

Natural Science, 2010, 2, 131-245
Published Online March 2010 in SciRes (http://www.scirp.org/journal/ns).

TABLE OF CONTENTS

Volume 2, Number 3, March 2010

ISSN: 2150-4091

9772150409001 06