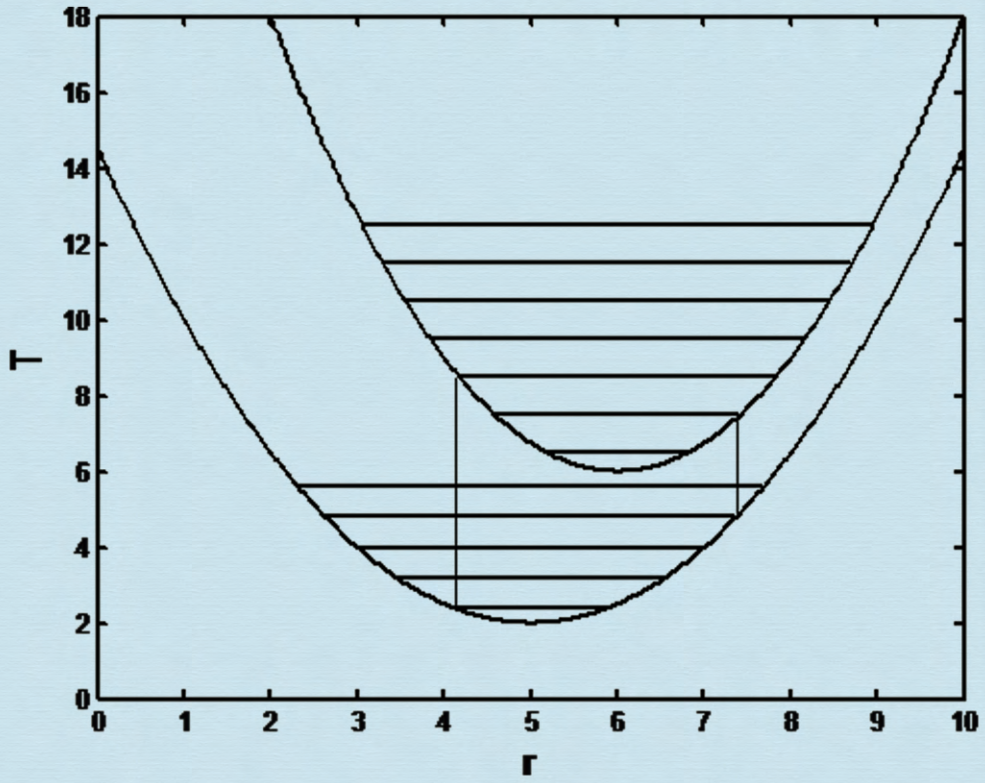
ISSN: 2153-1196 Volume 9, Number 9, August 2018

Scientific Research Publishing

Journal of Modern Physics



ISSN: 2153-1196
9 772153 119007 09

www.scirp.org/journal/jmp

Journal Editorial Board

ISSN: 2153-1196 (Print) ISSN: 2153-120X (Online)
http://www.scirp.org/journal/jmp

Scientific Research Publishing

Journal of Modern Physics, 2018, 9, 1713-1865
http://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Table of Contents

Volume 9 Number 9 August 2018

The figure on the front cover is from the article published in Journal of Modern Physics, 2018, Vol. 9, No. 9, pp. 1735-1760 by Jeremy B. Tatum.

Journal of Modern Physics (JMP)
Journal Information

SUBSCRIPTIONS

The *Journal of Modern Physics* (Online at Scientific Research Publishing, www.SciRP.org) is published monthly by Scientific Research Publishing, Inc., USA.

Subscription rates:
Print: $89 per issue.
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: jmp@scirp.org

Journal of Modern Physics, 2018, 9, 1713-1723
http://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Two-Level Block Decompositions for Solving Helmholtz Equation via Chebyshev Pseudo Spectral Method

Hsin-Chu Chen

Department of Computer and Information Science, Clark Atlanta University, Atlanta, GA, USA
Email: hchen@cau.edu

How to cite this paper: Chen, H.-C. (2018) Two-Level Block Decompositions for Solving Helmholtz Equation via Chebyshev Pseudo Spectral Method. *Journal of Modern Physics*, **9**, 1713-1723.
https://doi.org/10.4236/jmp.2018.99107

Received: June 14, 2018
Accepted: July 29, 2018
Published: August 1, 2018

Abstract

In this paper, we consider solving the Helmholtz equation $\nabla^2 u + \omega^2 u = f(x, y)$ in the Cartesian domain $\Omega = [-1,1] \otimes [-1,1]$, subject to homogeneous Dirichlet boundary condition, discretized with the Chebyshev pseudo-spectral method. The main purpose of this paper is to present the formulation of a two-level decomposition scheme for decoupling the linear system obtained from the discretization into independent subsystems. This scheme takes advantage of the homogeneity property of the physical problem along one direction to reduce a 2D problem to several 1D problems via a block diagonalization approach and the reflexivity property along the second direction to decompose each of the 1D problems to two independent subproblems using a reflexive decomposition, effectively doubling the number of subproblems. Based on the special structure of the coefficient matrix of the linear system derived from the discretization and a reflexivity property of the second-order Chebyshev differentiation matrix, we show that the decomposed submatrices exhibits a similar property, enabling the system to be decomposed using reflexive decompositions. Explicit forms of the decomposed submatrices are derived. The decomposition not only yields more efficient algorithm but introduces coarse-grain parallelism. Furthermore, it preserves all eigenvalues of the original matrix.

Keywords

Helmholtz Equation, Chebyshev Pseudo-Spectral Method, Chebyshev Differentiation Matrix, Coarse-Grain Parallelism, Reflexive Matrix

1. Introduction

In this paper, we consider the numerical solution to the Helmholtz equation

$\nabla^2 u + \omega^2 u = f(x,y)$ in the Cartesian domain $\Omega = [-1,1] \otimes [-1,1]$ with homogeneous Dirichlet boundary condition, where ∇^2 is the Laplacian operator, ω is a real parameter representing the wavenumber, u is the amplitude associated with wave propagation, and $f(x,y)$ is a forcing function. Without loss of generality we assume $u = 0$ on the boundary, *i.e.*,

$$\nabla^2 u + \omega^2 u = f(x,y), (x,y) \in \Omega \text{ and } u = 0 \text{ on } \partial\Omega \tag{1}$$

We present a highly parallel numerical algorithm for solving the Helmholtz equation discretized by the Chebyshev pseudospectral scheme. Chebyshev pseudospectral methods have long been used to numerically solve partial differential equations [1] [2] [3]. This discretization yields a linear system with coefficient matrix of a special block structure that allows for a block diagonalization which decouples the original matrix, along with some proper permutation of the decoupled matrix into several independent submatrices. By exploiting a reflexivity property that is inherent in the Chebyshev differentiation matrix, it is possible to further decompose each of the submatrices into two independent submatrices when the boundary conditions at both ends of the decoupled 1D subproblems are the same. This second level of decomposition effectively doubles the number of independent submatrices, yielding a higher degree of course-grain parallelism.

This paper is organized as follows. In Section 2, we present the general structure of the linear system derived from the Chebyshev collocation method discretized independently in x and y direction. In Section 3, we address the two-level block decomposition. This decomposition was proposed by the author in [4] for solving Poisson equations with identical grids in either direction. In this paper, we reformulate the decomposition to allow for the number of grid points in the x direction to be different from that in the y direction and derive the explicit forms for the decomposed submatrices, making applications of this approach more flexible. The computational procedure and a numerical example to demonstrate its usefulness are presented in Sections 4 and 5, respectively.

2. Linear System from the Chebyshev Collocation Method

In this section, we briefly describe the Chebyshev collocation method for solving the Helmholtz equation in two dimensions. We assume that the problem is discretized with a tensor product grid (x_i, y_j), $0 \le i \le N_x, 0 \le j \le N_y$, where x_i and y_j are the Chebyshev points in x and y direction, respectively. Let D_x, indexed from 0 to N_x, be the $(N_x+1)*(N_x+1)$ Chebyshev spectral differentiation matrix associated with the x direction. The entries of D_x are given as [1] [2] [3]

$$(D_x)_{00} = \frac{2N_x^2+1}{6}, \ (D_x)_{N_x N_x} = -\frac{2N_x^2+1}{6},$$

$$(D_x)_{jj} = \frac{-x_j}{2(1-x_j^2)}, \ j = 1,2,\cdots,N_x - 1$$

$$\left(D_x\right)_{ij} = \frac{c_i\left(-1\right)^{i+j}}{c_j\left(x_i - x_j\right)},\ i \neq j, i, j = 0,1,\cdots,N_x$$

where $c_i = \begin{cases} 2 & \text{for } i = 0 \text{ or } N_x \\ 1 & \text{otherwise} \end{cases}$. The Chebyshev spectral differentiation matrix D_y associated with the y direction can be obtained in a similar may.

Now, let A_x be the stripped matrix of D_x^2 by removing the first and last rows and columns of D_x^2 and A_y be the stripped matrix of D_y^2. We have

$$\begin{aligned} \left(A_x\right)_{ij} &= \left(D_x^2\right)_{ij}, i, j = 1,2,\cdots,N_x - 1, \\ \left(A_y\right)_{ij} &= \left(D_y^2\right)_{ij}, i, j = 1,2,\cdots,N_y - 1. \end{aligned} \tag{2}$$

With the grid just described, the Chebyshev collocation method yields, after removing the boundary nodes whose values are already known from the system, the following linear system

$$Ku = f, K = I_p \otimes A_y + A_x \otimes I_q + \omega^2\left(I_p \otimes I_q\right) \tag{3}$$

where $p = N_x - 1$, $q = N_y - 1$, and I_k represents the identity matrix of dimension k.

Many solution schemes can be employed to solve Equation (3), including Gaussian eliminations, Gauss-Seidel iterations, ADI iterations, and matrix diagonalizations [5] [6] [7] [8]. In this paper, the block version of the matrix diagonalization method will be employed to first decompose the matrix K into a block diagonal matrix consisting of q diagonal blocks. Each diagonal block will then be further decomposed into two smaller diagonal sub-blocks using reflexive decompositions, yielding a total of $2q$ diagonal blocks. Note that q is the number of internal grid points along the y direction. This two-step block decomposition scheme allows for the linear system Equation (3) to be decoupled into $2q$ linear subsystems which can then be solved in parallel with course-grain parallelism using multiprocessors or networked computers.

3. Two-Level Block Decompositions

3.1. Level-1 Decomposition: Block Diagonalization

The diagonalization scheme has been used elsewhere [7] [8] [9]. In this section, however, we address the decomposition of the matrix K into q diagonal blocks using a different representation. To begin with, let Q_y be such a matrix that the similarity transformation $Q_y^{-1}A_yQ_y$ diagonalizes A_y into Λ_y:

$$Q_y^{-1}A_yQ_y = \Lambda_y$$

A natural choice of Q_y is to have its columns formed from the eigenvectors of A_y so that Λ_y is a diagonal matrix that consists of the eigenvalues of A_y. Now split the coefficient matrix K in Equation (3) into three parts K_1, K_2, and K_3:

$$K_1 = I_p \otimes A_y, K_2 = A_x \otimes I_q, K_3 = \omega^2\left(I_p \otimes I_q\right)$$

and let $Q = I_p \otimes Q_y$. We have the transformed matrix

$$\tilde{K} = Q^{-1}KQ = \tilde{K}_1 + \tilde{K}_2 + \tilde{K}_3 \tag{4}$$

with

$$\tilde{K}_1 = Q^{-1}K_1Q, \tilde{K}_2 = Q^{-1}K_2Q, \tilde{K}_3 = Q^{-1}K_3Q, \text{ where } Q^{-1} = I_p \otimes Q_y^{-1}.$$

We now show that the transformed matrix $\tilde{K}$ can be permuted to yield a block diagonal matrix. First, we observe that both $\tilde{K}_1$ and $\tilde{K}_3$ are diagonal matrices and $\tilde{K}_2$ is a diagonal block matrix since

$$\tilde{K}_1 = Q^{-1}K_1Q = \left(I_p \otimes Q_y^{-1}\right)\left(I_p \otimes A_y\right)\left(I_p \otimes Q_y\right) = I_p \otimes \left(Q_y^{-1}A_yQ_y\right) = I_p \otimes \Lambda_y,$$

$$\tilde{K}_3 = Q^{-1}K_3Q = \left(I_p \otimes Q_y^{-1}\right)\left(\omega^2 I_p \otimes I_q\right)\left(I_p \otimes Q_y\right) = \omega^2\left(I_p \otimes I_q\right)$$

and

$$\tilde{K}_2 = Q^{-1}K_2Q = \left(I_p \otimes Q_y^{-1}\right)\left(A_x \otimes I_q\right)\left(I_p \otimes Q_y\right) = A_x \otimes I_q.$$

Accordingly,

$$\tilde{K} = Q^{-1}KQ = I_p \otimes \Lambda_y + A_x \otimes I_q + \omega^2\left(I_p \otimes I_q\right). \tag{5}$$

Let a_{ij} be the $(i,j)^{th}$ element of A_y, $i, j = 1, 2, \cdots, q$. In its matrix form, $\tilde{K}$ can be expressed as

$$\tilde{K} = \begin{bmatrix} \left(a_{11}+\omega^2\right)I + \Lambda_y & a_{12}I & \cdots & a_{1p}I \\ a_{21}I & \left(a_{22}+\omega^2\right)I + \Lambda_y & \cdots & a_{2p}I \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1}I & a_{p2}I & \cdots & \left(a_{pp}+\omega^2\right)I + \Lambda_y \end{bmatrix}$$

where I is of dimension q. Apparently, $\tilde{K}$ is also a diagonal block matrix because both I and Λ_y are diagonal matrices. The diagonal block matrix $\tilde{K}$ can be rearranged to yield a block diagonal matrix $\hat{K}$:

$$\hat{K} = \begin{bmatrix} \hat{K}_1 & 0 & \cdots & 0 \\ 0 & \hat{K}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \hat{K}_q \end{bmatrix}$$

via appropriate permutations of rows and columns as shown below. Let $\tilde{K}_{ij}$ be the $(i,j)^{th}$ block of $\tilde{K}$ and P be such a permutation matrix that the transformation $P^{\mathrm{T}}\tilde{K}P$ rearranges the entries of $\tilde{K}$ in such a way that

$$\hat{K}_l(i,j) = \tilde{K}_{i,j}(l), i, j = 1, 2, \cdots, p$$

where $\hat{K}_l(i,j)$ is the $(i,j)^{th}$ element of $\hat{K}_l$ and $\tilde{K}_{i,j}(l)$ is the l^{th} diagonal element of $\tilde{K}_{i,j}$. Then we have

$$P^{\mathrm{T}}\tilde{K}P = \hat{K} = \hat{K}_1 \oplus \hat{K}_2 \oplus \cdots \oplus \hat{K}_q, \tag{6}$$

where

$$\hat{K}_l = A_x + \left(\omega^2 + \left(\lambda_l\right)_y\right)I_p, l = 1, 2, \cdots, q$$

with $(\lambda_l)_y$ being the l^{th} eigenvalue of A_y. Accordingly,

$$\hat{K} = P^{\mathrm{T}}\tilde{K}P = P^{\mathrm{T}}Q^{-1}KQP$$

since $\tilde{K} = Q^{-1}KQ$. In other words, the matrix K has been transformed to a block diagonal matrix consisting of q blocks. This decomposition takes the advantage of the homogeneity property of the problem with the boundary conditions mentioned in Equation (1) along the y direction to decouple the 2D problem into q independent 1D problems. A similar decomposition can be performed by using the homogeneity property of the problem along the x direction.

3.2. Level-2 Decomposition: Reflexive Decomposition

As seen from the previous section, the matrix K can be transformed to a block diagonal matrix, implying that the original problem can be decomposed into independent subproblems which can then be solved in parallel using multi-processors. Although each diagonal block $\hat{K}_l$ in Equation (6) can be further decomposed into a diagonal matrix using a single transformation matrix Q_x that consists of the eigenvectors of A_x, this diagonalization yields little advantage in general, especially a multiprocessing environment, unless the eigenpairs are readily available, which is not the case for the matrix A_x.

In this section, we shall employ the reflexivity property of the diagonal blocks $\hat{K}_l$ to further decompose each of them into two independent sub-blocks using another decomposition technique called the reflexive decomposition [10].

Let J_n be the cross identity matrix of dimension n: $J_n = \begin{bmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{bmatrix}$. It is trivial to see that $J_n J_n = I_n$. Now, we note that the Chebyshev differentiation matrix D_x is anti-reflexive and D_x^2 is reflexive, with respect to J_{N+1}. In other words,

$$D_x = -J_{N+1}D_xJ_{N+1} \text{ and } D_x^2 = J_{N+1}D_x^2J_{N+1}.$$

Accordingly, the $p \times p$ matrix A_x, which is the stripped matrix of D_x^2 (Equation (2)), is reflexive with respect to J_p. In fact, it is a centro-symmetric matrix [11] [12]. Recall that $\hat{K}_l = A_x + \left(\omega^2 + (\lambda_l)_y\right)I_p$. Therefore, $\hat{K}_l$ satisfy the reflexivity property

$$\hat{K}_l = J_p\hat{K}_lJ_p, l = 1, 2, \cdots, q,$$

indicating that each $\hat{K}_l$ can be decomposed via reflexive decompositions into two block diagonal submatrices of almost equal size, depending on whether p is even or odd. The decompositions can be done through orthogonal transformations which have been shown in [10]. Here we just present the results.

First consider the case when p is even. Assume $p = 2k$ for some $k > 0$, and A_x is evenly partitioned into 2×2 sub-blocks: $A_x = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$. Taking the advantage of the reflexivity property of $\hat{K}_l$ by applying the orthogonal

transformation $X_A^{\mathrm{T}}\hat{K}_l X_A$ to $\hat{K}_l$ where $X_A = \frac{1}{\sqrt{2}}\begin{bmatrix} I & -J_k \\ J_k & I \end{bmatrix}$, one can easily decompose $\hat{K}_l$ into two decoupled diagonal subblocks of equal size:

$$\begin{aligned} D_l &= X_A^{\mathrm{T}}\hat{K}_l X_A = X_A^{\mathrm{T}}\left(A_x + \left(\omega^2 + (\lambda_l)_y\right)I\right)X \\ &= X_A^{\mathrm{T}}A_x X_A + \left(\omega^2 + \lambda_l\right)I = \begin{bmatrix} D_{l1} & 0 \\ 0 & D_{l2} \end{bmatrix} \end{aligned} \tag{7}$$

where $D_{l1} = A_{11} + A_{12}J_k + \left(\omega^2 + (\lambda_l)_y\right)I_k$ and $D_{l2} = A_{22} - A_{21}J_k + \left(\omega^2 + (\lambda_l)_y\right)I_k$

In the case when p is odd, say $p = 2k+1, k > 0$, we consistently partition A_x and J as

$$A_x = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & a_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}, J = \begin{bmatrix} & & J_k \\ & 1 & \\ J_k & & \end{bmatrix}.$$

By taking the transformation matrix X_A to be $X_A = \frac{1}{\sqrt{2}}\begin{bmatrix} I & 0 & -J_k \\ 0 & \sqrt{2} & 0 \\ J_k & 0 & I \end{bmatrix}$, the matrix $\hat{K}_l$ can be decoupled via the orthogonal transformation $X_A^{\mathrm{T}}\hat{K}_l X_A$ as

$$\begin{aligned} D_l &= X_A^{\mathrm{T}}\hat{K}_l X_A = X_A^{\mathrm{T}}\left(A_x + \left(\omega^2 + (\lambda_l)_y\right)I\right)X_A \\ &= X_A^{\mathrm{T}}A_x X_A + \left(\omega^2 + (\lambda_l)_y\right)I = \begin{bmatrix} D_{l1} & 0 \\ 0 & D_{l2} \end{bmatrix} \end{aligned}$$

where

$$D_{l1} = \begin{bmatrix} A_{11} + A_{13}J_k & \sqrt{2}A_{12} \\ \sqrt{2}A_{21} & a_{22} \end{bmatrix} + \left(\omega^2 + (\lambda_l)_y\right)$$

and

$$D_{l2} = \left(A_{33} - A_{31}J_k\right) + \left(\omega^2 + (\lambda_l)_y\right)I_k.$$

In either case, let $X = I_q \otimes X_A = X_A \oplus X_A \oplus \cdots \oplus X_A$. The orthogonal transformation $X^{\mathrm{T}}\hat{K}X$ yields

$$\begin{aligned} X^{\mathrm{T}}\hat{K}X &= X_A^{\mathrm{T}}\hat{K}_1 X_A \oplus X_A^{\mathrm{T}}\hat{K}_2 X_A \oplus \cdots \oplus X_A^{\mathrm{T}}\hat{K}_q X_A \\ &= D_1 \oplus D_2 \oplus \cdots \oplus D_q \\ &= \left(D_{11} \oplus D_{12}\right) \oplus \left(D_{21} \oplus D_{22}\right) \oplus \cdots \left(D_{q1} \oplus D_{q2}\right), \end{aligned}$$

which obviously consists of $2q$ independent diagonal blocks since each D_l has two independent diagonal blocks.

4. Computational Procedure

The numerical solution to the Helmholtz Equation (1) discretized by the Chebyshev pseudospectral method yields the linear system

$$Ku = f, K = I_p \otimes A_y + A_x \otimes I_q + \omega^2 \left(I_p \otimes I_q \right)$$

where I_p is the identity matrix of dimension p, A_y is a $q \times q$, A_x a $p \times p$, and K a $pq \times pq$ matrix. Here p (q) is the number of internal grid points in the x (y) direction. The two-level decomposition scheme described in this paper yields the following transformed linear system consisting of $2q$ independent subsystems, each of dimension $\frac{p}{2}$ when p is even.

$$Dv = g, D = \left(D_{11} \oplus D_{12} \right) \oplus \left(D_{21} \oplus D_{22} \right) \oplus \cdots \oplus \left(D_{q1} \oplus D_{q2} \right).$$

The sizes of the sub-blocks in D differ at most by 1 when p is an odd number. Computationally, this decomposition is very attractive. Not only can the decoupled subsystems be solved much more efficiently, they can also be solved in parallel with course-grain parallelism using a computer system with multi-processors. This two-level composition consists of the following three stages.

1) Transform $Ku = f$ to $\tilde{K}\tilde{u} = \tilde{f}$ where $\tilde{K} = Q^{-1}KQ$, $\tilde{u} = Q^{-1}u$, and $\tilde{f} = Q^{-1}f$.

This is a typical similarity transformation that is achieved by premultiplying both sides of $Ku = f$ with Q^{-1} and inseting QQ^{-1}, which is simply an identity matrix, between K and u. The matrix Q has been defined in Section 3.1 and $\tilde{K}$ can be found in Equation (5).

2) Transform $\tilde{K}\tilde{u} = \tilde{f}$ to $\hat{K}\hat{u} = \hat{f}$ where $\hat{K} = P^{\mathrm{T}}\tilde{K}P$, $\hat{u} = P^{\mathrm{T}}\tilde{u}$, and $\hat{f} = P^{\mathrm{T}}\tilde{f}$.

The main purpose of this transformation is to reorder the unknowns of the linear system so that the decoupled coefficent matrix $\tilde{K}$ is converted to a block diagonal matrix ($\hat{K}$) for the sake of computational convenience. The matrix P, which is just a permutation matrix, has been explained in Section 3.1 and $\hat{K}$ can be found in Equation (6).

3) Transform $\hat{K}\hat{u} = \hat{f}$ to $Dv = g$ where $D = X^{\mathrm{T}}\hat{K}X$, $v = X^{\mathrm{T}}\hat{u}$, and $g = X^{\mathrm{T}}\hat{f}$.

This transformation is analogous to that in stage 1 except that X here is an orthogonal matrix. Therefore we replace X^{-1} by X^{T}. The matrices X and D can be found in Section 3.2

The advantage of this approach is that the decomposed form of D is explicitly known and, therefore, the actual decompositions from K to $\tilde{K}$, from $\tilde{K}$ to $\hat{K}$, and from $\hat{K}$ to D are not necessary. Furthermore, the diagonal blocks of D can be obtained without any matrix-matrix multiplications. The computational procedure of this two-level decomposition boils down to the following three typical steps.

1) Decomposition step. In this step, we shall compute D and g. Since each diagonal block of D,

$$D = \left(D_{11} \oplus D_{12} \right) \oplus \left(D_{21} \oplus D_{22} \right) \oplus \cdots \oplus \left(D_{q1} \oplus D_{q2} \right),$$

is independent and explicitly known, D can be obtained in parallel using $2q$ processors, without any difficulty. The main task here is, thus, to obtain g. This

can be achieved by transforming f to $\tilde{f}$ first, which involves solving the linear system $Q\tilde{f} = f$. It is essential to note that Q, $Q = I_p \otimes Q_y$, is a block diagonal matrix and, therefore, $\tilde{f}$ can be obtained by solving p independent linear subsystems, each of size q, instead of solving the linear system as a whole, which is of size pq. This step can be performed in parallel with large granularity. The matrix Q_y, consisting of the eigenvectors of A_y, is in general not known explicitly and, thus, has to be computed from A_y. Fortunately, A_y is only of dimension q. Once $\tilde{f}$ has been computed, $\hat{f}$ and g can be obtained easily because $\hat{f}$ is simply a permuted version of $\tilde{f}$ and $g = X^{\mathrm{T}}\hat{f}$ involves no matrix-vector multiplications due to the special form of X.

2) Solution step. In this step, we solve the linear system $Dv = q$ for v. Note that D consists of $2q$ diagonal blocks: D_{l1} and D_{l2}, $l = 1, 2, \cdots, q$. By consistently partitioning the vectors x and g into $2q$ parts, based on the size of the blocks, the linear system $Dv = g$ can now be expressed as

$$D_{l1}v_{2l-1} = g_{2l-1},$$

$$D_{l2}v_{2l} = g_{2l}$$

for $l = 1, 2, \cdots, q$. Since all of the subsystems are independent, the linear system $Dv = g$ can be solved in parallel, using $2q$ processors, one for each subsystem.

3) Retrieval step. Once v has been obtained, the solution u to the original system can be retrieved from v as follows. We first compute $\hat{u}$ from v using $\hat{u} = Xv$, which again involves no matrix-vector multiplication. We then obtain $\tilde{u}$ by simply permuting $\hat{u}$ using $\tilde{u} = P\hat{u}$. The final solution u can now be computed by the matrix multiplication $u = Q\tilde{u}$. Note that $X = I_q \otimes X_A$ and $Q = I_p \otimes Q_y$, implying that $\hat{u}$ and u can be obtained in parallel with course-grain parallelism.

To end this section, it deserves mentioning that this two-level decomposition is a similarity transformation and therefore, all eigenvalues of the original matrix K are preserved in matrix D. Obtaining eigenvalues from D is far more efficient than from K.

5. Numerical Example

To demonstrate the usefulness of the two-level decomposition scheme, we present a numerical example derived from the Helmholtz equation with $\omega = 3$ over a square domain on $[-1,1]\times[-1,1]$, subject to homogeneous Dirichlet boundary conditions on a grid with $N_x = 5$ and $N_y = 4$. The numerical results presented in this section were conducted using GNU Octave which is software that offers many powerful high-level programming features for scientific computations, similar to Matlab. It suffices to show the final decomposed submatrices to illustrate the advantage of this approach. By excluding the grid points on the boundary and using the lexicographic ordering to number the internal nodes [3], the matrix K, of dimension 12, yields

$$K = I_4 \otimes A_y + A_x \otimes I_3 + \omega^2 \left(I_4 \otimes I_3\right)$$

where

$$A_x = \begin{bmatrix} -31.53 & 12.68 & -3.69 & 2.21 \\ 7.32 & -10.07 & 5.79 & -1.91 \\ -1.91 & 5.79 & -10.07 & 7.32 \\ 2.21 & -3.69 & 12.68 & -31.53 \end{bmatrix} \text{ and } A_y = \begin{bmatrix} -14.0 & 6.0 & -2.0 \\ 4.0 & -6.0 & 4.0 \\ -2.0 & 6.0 & -14.0 \end{bmatrix}.$$

For simplicity, we have rounded off the numbers after the last digit shown. First we observe that both A_x and A_y are reflexive. In fact, they are centro-symmetric matrices. Let the columns of Q_y be the eigenvectors of A_y, of dimension 3, and P^{T} be the following permutation matrix of dimension 12,

$$P^{\mathrm{T}}(k,l) = \begin{cases} 1 & \text{for } k = i + (j-1)*4, l = j + (i-1)*3, 1 \le i \le 4, 1 \le j \le 3 \\ 0 & \text{otherwise.} \end{cases}$$

From the Level-1 decomposition presented in Section 3,

$$P^{\mathrm{T}}\tilde{K}P = \hat{K} = P^{\mathrm{T}}Q^{-1}KQP = \hat{K}_1 \oplus \hat{K}_2 \oplus \hat{K}_3, Q = I_3 \otimes Q_x,$$

we easily obtain the decoupled diagonal blocks $\hat{K}_l$ without the need of any matrix multiplications once the eigenvalues of A_y, $(\lambda_l)_y$, are available:

$$\hat{K}_l = A_x + \left(\omega^2 + (\lambda_l)_y\right) I_4 = \hat{K}_l = A_x + \left(\omega^2 I_4\right) + (\lambda_l)_y I_4$$

$$= \begin{bmatrix} -22.53 + (\lambda_l)_y & 12.68 & -3.69 & 2.21 \\ 7.32 & -1.07 + (\lambda_l)_y & 5.79 & -1.91 \\ -1.91 & 5.79 & -1.07 + (\lambda_l)_y & 7.32 \\ 2.21 & -3.69 & 12.68 & -22.53 + (\lambda_l)_y \end{bmatrix}, 1 \le l \le 3,$$

where $\left\{(\lambda_1)_y, (\lambda_2)_y, (\lambda_3)_y\right\} = \{-19.54, -12.00, -2.46\}$.

The numerical values of $\hat{K}_l$ clearly indicate that each $\hat{K}_l$ is reflexive with respect to J_4. Now, applying the Level-2 decomposition to each $\hat{K}_l$ by taking $X_A = \frac{1}{\sqrt{2}}\begin{bmatrix} I & -J_2 \\ J_2 & I \end{bmatrix}$ yields

$$D_1 = X_A^{\mathrm{T}}\hat{K}_1 X_A = \begin{bmatrix} -39.87 & 8.99 & & \\ 5.41 & -14.82 & & \\ & & -26.40 & 9.22 \\ & & 16.38 & -44.29 \end{bmatrix},$$

$$D_2 = X_A^{\mathrm{T}}\hat{K}_2 X_A = \begin{bmatrix} -32.32 & 8.99 & & \\ 5.41 & -7.28 & & \\ & & -18.86 & 9.22 \\ & & 16.38 & -36.74 \end{bmatrix},$$

$$D_3 = X_A^{\mathrm{T}}\hat{K}_3 X_A = \begin{bmatrix} -22.78 & 8.99 & & \\ 5.41 & 2.27 & & \\ & & -9.31 & 9.22 \\ & & 16.38 & -27.20 \end{bmatrix},$$

where the unshown entries in the matrices are 0. Apparently, each $\hat{K}_l$ has

been further decomposed into two independent diagonal blocks, yielding a total of six independent diagonal blocks from the original matrix *K*. Note that the decomposed matrices $D_l, 1 \le l \le 3$ can be obtained directly from A_x, ω, and $(\lambda_l)_y$ using Equation (7) without the need of computing $\hat{K}_l$. Note also that the eigenvalues of *K*, of dimension 12, can now be easily obtained from the six diagonal blocks, each of dimension only 2, a clear indication of the advantage of this approach.

6. Conclusions

In this paper, we have presented a two-level block decomposition scheme for the numerical solution to the Helmholtz equation $\nabla^2 u + \omega^2 u = f(x, y)$ in the Cartesian domain $\Omega = [-1,1] \otimes [-1,1]$, subject to homogeneous Dirichlet boundary condition, discretized with the Chebyshev pseudo-spectral method. For a computational grid with $N_x + 1$ grid points in the *x* direction and $N_y + 1$ in the *y* direction, our first-level decomposition yields $N_y - 1$ independent 1D subproblems, excluding the known values at boundary points, using the eigenpairs of the 1D second-order Chebyshev differentiation matrix along *y*. The decoupled coefficient matrices are then rearranged to form a block diagonal matrix through a proper permutation. The second level of decomposition further decouples each of the subproblem into two smaller and independent subproblems using reflexive decompositions, yielding a total of $2(N_y - 1)$ independent subproblems.

The general form of the final decoupled submatrices has also been explicitly derived, eliminating the need of performing matrix-matrix multiplications during the decomposition step. A numerical example is presented to illustrate the applicability and to demonstrate the advantage of this two-level decomposition scheme. This scheme leads naturally to a highly parallel numerical algorithm for solving the problem under consideration, since all subproblems can be solved in parallel using multiprocessors.

To close our conclusion, it deserves mentioning that this block decomposition can be readily employed for finding the eigenvalues of the problem as well. This is due to the fact that all eigenvalues of the original matrix are preserved in the decomposed submatrices because the first-level decomposition is a similarity transformation and the second-level decomposition is an orthogonal transformation. The preservation of all eigenvalues in the submatrices makes this block decomposition scheme even more attractive because finding eigenvalues from the decomposed submatrices is obviously much more efficient than from the original matrix, not to mention the computational benefit that can be achieved from the large-grain parallelism induced by the decomposition.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Martinez, J.d.J. and Esperanca, P.d.T.T. (2007) *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, **XXIX**, 317-328. https://doi.org/10.1590/S1678-58782007000300013

[2] Peyret, R. (2002) Applied Mathematical Sciences. Vol. 148, Springer-Verlag, New York, 448 p.

[3] Trefethen, L.N. (2000) Spectral Methods in Matlab. SIAM University City Center, Philadelphia, 165 p.

[4] Chen, H.-C. (2015) *Neural, Parallel, and Scientific Computations*, **23**, 267-275.

[5] Ehrenstein, U. and Peyret, P. (1989) *International Journal for Numerical Methods in Fluids*, **9**, 427-452. https://doi.org/10.1002/fld.1650090405

[6] Golub, G.H. and Van Loan, C.F. (1983) Matrix Computations. The Johns Hopkins University Press, Maryland, 476 p.

[7] Haidvogel, D.B. and Zang, T. (1979) *Journal of Computational Physics*, **30**, 167-180. https://doi.org/10.1016/0021-9991(79)90097-4

[8] Julien, K. and Watson, M. (2009) *Journal of Computational Physics*, **228**, 1480-1503. https://doi.org/10.1016/j.jcp.2008.10.043

[9] Lynch, R.E., Rice, J.R. and Thomas, D.H. (1964) *Numerische Mathematik*, **6**, 185-199. https://doi.org/10.1007/BF01386067

[10] Chen, H.-C. and Sameh, A. (1989) *SIAM Journal on Matrix Analysis and Applications*, **10**, 39-64. https://doi.org/10.1137/0610004

[11] Andrew, A.L. (1973) *Technometrics*, **15**, 405-407. https://doi.org/10.1080/00401706.1973.10489052

[12] Cantoni, A. and Butler, P. (1976) *Linear Algebra and Its Applications*, **13**, 275-288. https://doi.org/10.1016/0024-3795(76)90101-4

Journal of Modern Physics, 2018, 9, 1724-1734
http://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Energy Saving in Construction by Wide Application of High-Quality Insulation Based on Basalt Fibers

Vyacheslav Kremnev[1], Boris Basok[1], Andriy Timoshchenko[1], Sergei Tymchyshyn[2]

[1]Institute of Engineering Thermophysics of the National Academy of Sciences of Ukraine, Kiev, Ukraine
[2]LLC CE "Chernivtsi Plant of Heat-Insulating Materials", Chernivtsi, Ukraine
Email: a_timoshchenko@ukr.net

How to cite this paper: Kremnev, V., Basok, B., Timoshchenko, A. and Tymchyshyn, S. (2018) Energy Saving in Construction by Wide Application of High-Quality Insulation Based on Basalt Fibers. *Journal of Modern Physics*, **9**, 1724-1734.
https://doi.org/10.4236/jmp.2018.99108

Received: May 8, 2018
Accepted: July 29, 2018
Published: August 1, 2018

Abstract

The influence of modern heat insulation materials on the ecological compatibility and fire resistance of residential and public buildings is considered. The nomenclature of fibrous heat insulation materials from basic rocks of volcanic origin is described. For basalt super thin fibers (BSTF), the production technology is described. The factors negatively influencing the engineering-and-economical performance of the duplex technology of production of the BSTF are analyzed. The engineering-and-economical performance of innovative enterprises for the production of basalt heat insulation and measures for its use in construction are given.

Keywords

Basalt Fiber, Ecological Materials, Heat Insulation, Energy Efficiency, Heat Technological Processes

1. Introduction

Contemporary trends in construction are such that the requirements to heat consumption for residential and public buildings are getting tougher from year to year ref. [1]. The international community agrees that energy-efficient buildings should not exceed the heat consumption of 70 kW·h per square meter per year. However, today in Ukraine, most of the buildings in which there are heating systems do not meet the modern requirements for heat-retaining properties. This situation has developed historically, because for long-time prices for fuel and energy resources have been artificially understated in comparison with the world prices ref. [2]. Consequently, under such conditions, measures aimed at

energy saving in construction did not have economic substantiation because the growth of capital expenditures due to the increase of heat-retaining properties of buildings did not formally result in significant savings in heating costs. However, attempts were made to introduce energy-saving measures in construction. For example, within the framework of prefabricated concrete technology, specialized three-layer concrete panels were constructed that had two outer layers of heavy concrete and a cavity intermediate, which was filled with mineral wool insulation under the conditions of a prefabricated reinforced concrete plant. Such a concrete panel was quite energy-efficient and quite fire-safe. In particular, it was made from such panels that the external walling constructions of the four buildings of the Department of heat and mass transfer processes and equipment of the Institute of Engineering Thermophysics of the National Academy of Sciences of Ukraine, built in the 70's of last century. These buildings are relatively energy-efficient. At present, the technology of precast concrete is almost not used in the construction. It was superseded by the technology of monolithic construction, on which the bearing frame of the building is executed from solid monolithic reinforced concrete, and external fencing wall constructions—mainly from brick. Unfortunately, the release of three-layer reinforced concrete panels is accordingly discontinued. With monolithic construction, modern normative heat-retaining properties of structures are achieved by arranging an external heat-insulating layer. Recently, regulatory requirements for heat-retaining properties of buildings have been significantly increased (**Table 1**).

In these conditions, existing buildings are thermo-modernized (external surfaces are insulated), and in the new construction, the insulation layers are foreseen at the design stage.

But in our view, these processes are of a spontaneous nature and are accompanied by a significant reduction in environmental friendliness and, worse, the fire safety of structures.

Table 1. Minimum permissible values of resistance to heat transfer.

Number in order	Item name	Resistance to heat transfer construction, [(square meter·C)/W]	
		R0 required, ref. [3], Table 1(a)*	Rq min, ref. [4], Table 1
1	Residential buildings		
	– walls	1.6	3.3
	– attic constructions (4000 [degrees·day])	2.2	4.95
2	Administrative and household buildings		
	– walls	1.4	3.3
	– attic constructions (4000 [degrees·day])	2.0	4.95

2. Qualitative Evaluation of Proposals in the Market of Heat Insulation

Most of the thermal insulation materials are completely synthetic, or contain in their composition binders of synthetic origin. The location of synthetic insulating materials and materials on synthetic binders inside the wall constructions, or on the interior surfaces of the enclosing structures, is prohibited by sanitary-hygienic and fire regulations. Such sanitary-hygienic requirements are related to the possibility of diffusion of substances harmful to humans (allergens) at the temperature conditions that occur in residentials. As for the fire protection requirements, even the removal of the insulation layer outside is not effective and leads to a significant reduction in fire resistance of structures, especially many-storied buildings. It is through thermal insulation materials the fires take catastrophic character and are accompanied by death of people, primarily because of poisoning with gaseous substances coming from heat-insulating materials, even if they are not burned, but only heated. In this case, for the poisoning of people with gases, it is sufficient to have only synthetic binders in the composition of heat-insulating materials. In recent times, cases of powerful fires in new many-storied buildings have become more frequent. Through natural draught, the flame quickly spread along the facades through a layer of thermal insulation through the entire height of the building. In addition, it should be borne in mind that the internal space of the heat insulation layer is 95% -99% composed of air, and the rest composed of the fibers, that is, in the volume of the heat insulation layer, oxygen is constantly present in an amount sufficient to burn the synthetic binder and poison people with gases.

Such fires occurred in Grozny (Russia) ref. [5], Kiev (Ukraine) ref. [6] and other cities. The last powerful fire under the scenario described was recently held in London (UK) ref. [7]. A few years ago, the fire completely destroyed the main assembly conveyor of KAMAZ PTC (Russia) several kilometers long ref. [8]. The main factor of the catastrophic fire scenario was the thermal insulation layer of the roofing, which was made of foam plastic.

In our opinion regulatory measures are needed from the state aimed at the fact that technical solutions for heat insulation of facilities should not cause a decrease in their fire resistance, at least if it concerns residential or other facilities with people staying.

3. Effective Heat Insulation Materials from Basic Rocks of Volcanic Origin

At the same time, heat insulation materials are capable, when used in buildings, not only not to reduce their fire resistance, but, on the contrary, significantly increase their fire resistance. It is to such materials that the heat insulation based on staple basalt fibers belongs.

The raw material for the production of basalt fibers is the basic rocks of magmatic origin-basalts, diabases, andesites and the like. Today, products from these

rocks are called basalt.

With the help of modern innovative technologies, from basalts produce fibers that are thinner than human hair and have unique useful qualities. These fibers are the primary form of commercial products and on its basis "soft" and "hard" heat insulating materials are produced.

Depending on the diameter, the staple basaltic fibers are divided into coarse (more than 20 μm), thin (from 4 to 20 μm), super-thin (from 1 to 3 μm), ultra-thin (0.5 to 1 μm), microthin (less than 0.5 μm). The diameter of elementary fibers significantly affects their properties-active surface, strength, flexibility, density, elasticity, coefficient of heat conductivity, etc., and thereby determines areas of their scope of use. So, micro- and ultrathin fibers are used in the production of heat-resistant paper-like materials (particularly in aircraft building), in the production of multi-layer heat- and electrical insulation, are used in a number of special technologies. Thin fibers are used in filters and composite materials, the coarse fibers are used for reinforcing concrete products ref. [9].

A promising basis for heat insulation materials and products are super-thin basalt fibers (*BSTF*). "Soft" heat insulation based on basaltic superthin fibers is a nonwoven stitched material in the form of the canvas, which consists of a set of staple basalt glass fibers with a length of 10 to 100 mm. In addition to "soft" materials on this basis, also "hard" products are produced in the form of slabs of different thickness and rigidity (from cardboards 5 - 20 mm thick to slabs 30 - 100 mm thick). As a binder they use bentonite clay.

In the field of developing technologies for the production of basalt fibers, materials and products based on them, as well as equipment for the hardware design of these technologies, the USSR took the leading place in the world, and the main role was played by scientists and industrialists of the Ukrainian SSR, in particular, the Institute for Material Science Problems of the Ukrainian SSR Academy of Sciences, Research Institute of fiberglass and fiberglass, etc. One of the reasons that it is Ukrainian science and industry that created this elite product it is the unique properties of Ukrainian basalts. These basalts are extremely promising for the production of super-thin fibers because the temperature range of their melt, at which stretching of the super-thin thread without its breakage is possible, is much higher than the similar index of basalts from other countries. That is, Ukrainian basalts are extremely valuable for the organization of production of super-thin fibers. The first industrial enterprises were created around Kiev (Ukraine).

Heat insulation products based on BSTF have practically no alternative in such industries as: energy (thermal, nuclear) shipbuilding and aircraft construction; wide selection of modern domestic heating boilers of low power; specialized systems in the construction, aimed at increasing the fire resistance of structures (cabins and elevator shafts, passive fire protection systems). Thermophysical properties, durability, resistance to low and high temperatures, vibrations, environmental friendliness and fire resistance, heat insulation products based on BSTF significantly exceed analogues. *On the way of expanding their*

application is a high cost, in which part of the energy carriers is more than 80%. *Thus, the cost of production depends mainly on the economic efficiency of the heat-technological processes*.

4. Duplex Technology for Production of Staple Super-Thin Basalt Fibers. Ways and Potentials of Intensification

BSTF is obtained as follows. The raw material is the basalt chippings of fractions of 5 - 20 mm, which is mined in quarries. Basalt chippings is dosed in the basalt furnace, the furnace heated by burners on natural gas. With the help of furnaces, the following technological tasks are solved: melting of basalt chippings, which ensures their transition from the solid state to the liquid, and also the homogenization of the melt, which consists in removing microcrystals from the melt and converting the melt into an amorphous glassy liquid. This is necessary in order to avoid the fragility of the fibers, which in the presence of microcrystals appear in the subsequent stages of processing. So the homogenized amorphous glassy liquid passes gravity into the so-called feeder device, which is a ceramic container with a perforated bottom (the so-called spinneret), which is made of platinum-rhodium or special refractory alloys. To prevent the cooling of the glass melt, the spinneret is heated by an electric current. The melt streams are forcibly removed through the spinneret into the environment. Thus, due to cooling in air, so-called solid continuous primary filaments are formed in the form of cylinders with a diameter from 150 to 250 μm. Primary filaments are routed to the area of action of a flat high-temperature jet of natural gas combustion products, which has a speed of about 300 m/s. When the ensemble of primary filaments interacts with the high-temperature jet, there is a repeated melting and bloating to super-thin fibers that are deposited from the gas stream on the screen filter. A canvas of staple super-thin fibers is formed on the filter. Canvas is continuously or periodically removed from the filter. The basalt raw material undergoes melting two times, so the described multi-stage technology for obtaining super-thin fibers is called "duplex technology".

5. Features of Duplex Technology, Which Negatively Affect the Engineering-and-Economical Performance of Production

Integral negative indicator that interferes with the wide distribution of high-quality heat insulation is its high prime cost. The high prime cost consists of a combination of *costs*, some of which can be considered *objectively inherent in technology* and such that it has no prospects for a significant reduction, and *costs that can be reduced through intensification* based on the introduction of innovative technical solutions for the technology and equipment.

The *objective costs* for the production of basalt fiber by duplex technology are the following.

Expenses for expert labor. The operators of the technological plant manually remove breakage of coarse fibers with the help of a metal device. Operators ma-

nually draw out the fiber from the "bulb" and direct it into the drawing out rollers. This work requires special attention, speed of response, high sensory sensitivity, exceptional technical safety in servicing the equipment of the technological plant. It should be noted that these duties are performed in conditions of high-power heat radiation fluxes from the spinnerets that have a temperature > 1300˚C, high dust content of air, noise and vibrations. All this allows us to regard operators as workers' elite, and the working conditions are extremely difficult. It can be predicted that in the future the duties of operators will not change significantly. It is hardly possible to replace them with robots.

The cost of electrical energy, which is supplied to the spinneret. They should be considered justified, since attempts to replace electricity with other energy carriers (natural gas) are not advisable because of the need for a very fine melt temperature control to ensure the ductile state that is necessary for the stable drawing out of coarse fibers.

Expenses for limited productivity of the spinneret. The high temperature level of the process of formation of coarse fibers does not allow many times to increase its productivity, using the traditional method for manufacturing of the glass fibers-increasing the geometric dimensions of the spinnerets and, accordingly, the number of fibers produced. Actions that in the adjacent industry lead to a multiple increase in the productivity of the coarse fiber production unit (2 - 6 times) in the case of duplex technology are not applied due to the impossibility of organizing a long stable process of forming and drawing out fibers on the spinneret. One of the main reasons that causes the instability of the formation of coarse fibers is the creep phenomenon of the spinneret material. Thus, it is thermodynamically possible to reduce the specific energy consumption per unit mass of basalt fibrous materials, but this is technologically limited to the existing level of material science development, which causes over-expenditure of energy resources. In addition, spinnerets require periodic replacement due to destruction, and the costs of replacing them are beyond the competence of thermophysicists and constitute a task for materials science specialists.

To the expenditure that can be significantly reduced through the use of innovative technical solutions include the following:

- *For raw materials* (*basalt chippings*);
- *On structural materials* (*ceramic materials of furnaces*, *materials of spinnerets*);
- *On the supply of compressed air to the burners* (*furnaces and blowing chambers*);
- *On water for cooling*;
- *For natural gas*.

IET NASU together with IG NASU and LLC CE "Chernivtsi plant of heat-insulating materials" developed a new energy-saving duplex technology for the production of BSTF and the technological plant for its hardware implementation, which enabled the first stage of technology (production of the basalt glass) to reduce the specific consumption of natural gas in 4 times, at the second stage

(melting and blowing of the coarse basalt fibers)—1.2 times. In relation to the final product-"soft" heat insulation—the specific consumption of natural gas has been reduced by a factor of 1.5. The prime cost of finished products is reduced by 1.2 times in relation to products manufactured in the technological plant of the first generation.

A new technology and equipment for the production of the slabs, in which BSTF act as raw material, has also been developed. In this technology, the specific consumption of natural gas per cubic meter of slabs is reduced by 1.3 - 1.5 times.

Conducted fundamental researches and carried out with the application of their results the developments provide the technical and technological basis for mass production of "soft" and "hard" heat insulation products based on basalt fibers. But for the wide application of such heat insulation in the construction and reconstruction of buildings, it is equally *important to use the optimal technical solutions for the arrangement of the basalt heat insulation layer in the conditions of the construction site with the obligatory provision of its mechanical and hydrotechnical protection for the entire service life*.

6. Engineering-and-Economical Performance of Innovative Enterprises for the Production of Basalt Heat Insulation and Measures for Its Use in Construction

The practical solution of the problem of wide introduction in the construction of basalt insulation, which is very important, requires a systematic approach and the convergence of various fields of science, technology and industry with mandatory regulatory influence on the part of the state. It is necessary to coordinate the efforts of heat engineering, materials science specialists, construction heat engineers, architects, economists and the like. The feasibility study must necessarily cover the entire "chain" from the extraction of basalt crushed stone into the operation of finished heat-insulated structures, where the final result is achieved to save energy for heating. In the course of such a feasibility study it will be possible to optimize the entire process aimed at the final result-energy saving. So, for example, the rise in price of "hard" heat insulating materials in comparison with "soft" materials can be compensated by a reduction in the cost of work on the installation of heat insulation in building conditions, etc. It is important to note that Ukraine has significant advantages in organizing the production of basalt heat insulation due to the fact that *the local basalts are exceptionally suitable for the production of fibers due to the ability of their melt to be in the ductile state in the wide temperature range*. *This is extremely important for the stable technological process of drawing out basalt fibers*.

The new technologies and equipment allow the creation of innovative standard enterprises in a modular manner with the ability to regulate their productivity in accordance with local demand. Enterprises should be created with the understanding that it is much more economical to transport basalt raw materials with a density of 1600 kg per cubic meter than heat insulation with a density of 40 -

50 kg per cubic meter. Complete developments already made allow to achieve the cost of basalt heat insulation at such enterprises-due to a significant reduction in energy consumption-at a level that does not exceed the cost of flammable heat insulation is widely used today. Below are the planned technical and economic indicators of the second generation workshops for the production of basalt heat insulation (Table 2), the specific indicators of the cost of heat insulation (Table 3) and the current information for assessing the economic efficiency of measures for thermomodernization of the enclosing structures of residential and administrative buildings (Table 4).

7. Economic-Organizing Systemic Measures That Promote the Production and Wide Application of Heat Insulation Based on Basalt Fibers

In order to reduce the cost of energy consumption by significantly increasing heat-retaining properties with the use of long-term, ecological heat insulation,

Table 2. Capital costs for the creation of workshops for the production of basalt heat insulation.

Number in order	Item name	Unit of measurement	Value
1	Workshop for production heat insulation-sew-knit mats without facing МТПБа (ТУ У В.2.7-23.9-00292818-001: 2012) ref. [10]	[thousand dollars US]	860
	including:		
	building of workshop	[thousand dollars US]	270
	technological equipment (including installation, starting-up and adjustment, personnel training)	[thousand dollars US]	590
1.1	Annual capacity of the workshop for finished products	[thousand cubic meters]	to 30.0
1.2	Operating mode of equipment	[hours per day]	16
		[hours per year]	4800
1.3	Service life of equipment and structures	[years]	20
2	Workshop for production heat-insulating-slabs ПМТБ-2 (ТУ У В.2.7-23.9-00292818-001:2012) ref. [10]	[thousand dollars US]	990
	including:		
	building of workshop	[thousand dollars US]	270
	technological equipment (including installation, starting-up and adjustment, personnel training)	[thousand dollars US]	720
2.1	Annual capacity of the workshop for finished products	[thousand cubic meters]	to 5.7
2.2	Operating mode of equipment	[hours per day]	24
		[hours per year]	8040
2.3	Service life of equipment and structures	[years]	20

Table 3. Planned specific indicators of the prime cost of heat insulation products-variable costs.

Number in order	Item name	Unit of measurement	Value
1	Sew-knit mats without facing МТПБа (ТУ У В.2.7-23.9-00292818-001: 2012) ref. [10]		
1.1	Volume of finished products (annual)	thousand cubic meters	to 30.0
1.2	Consumption of natural gas (annual)	million cubic meters	to 2.6
1.3	Electricity consumption (annual)	million kW•h	to 4.5
1.4	Prime cost one cubic meter heat insulation	dollars US	48.0
	including:		
	– natural gas	dollars US	27.9
	– electricity	dollars US	7.2
	– wages fund	dollars US	10.2
	– raw materials	dollars US	0.6
	– amortization costs	dollars US	0.1
	– current repairs	dollars US	2.0
2	Soft heat-insulating slabs ПМТБ-2 (ТУ У В.2.7-23.9-00292818-001: 2012) ref. [10]		
2.1	Volume of finished products (annual)	thousand cubic meters	to 5.7
2.2	Consumption of natural gas (annual)	million cubic meters	to 0.12
2.3	Electricity consumption (annual)	million kW•h	to 0.5
2.4	Prime cost one cubic meter heat insulation	dollars US	259.8
	including:		
	– natural gas	dollars US	135.8
	– electricity	dollars US	23.5
	– wages fund	dollars US	31.4
	– raw materials	dollars US	52.3
	– amortization costs	dollars US	16.8

which simultaneously increases the fire resistance of structures, the following *economic-organizing system measures* are proposed that will stimulate both the production of such products and its use in construction.

Measures to stimulate production, for legal entities wishing to create new production capacities for the production of basalt heat insulation, are proposed:

- Facilitate long-term loans with low interest rates;
- Facilitate the receipt of land for the construction of production capacities in the regions;
- To apply forms of interaction of entrepreneurs with the state on the basis of legislation on state-private partnership (this form plays an important role in the development of modern China);

Table 4. Up-to-date information for assessing the economic efficiency of actions for the thermomodernization of constructions of residential and administrative buildings.

Number in order	Item name	Unit of measurement	Value	
			walls	attic constructions (4000 degrees·day)
1	The minimum value of the thermal resistance, which must be achieved by additional heat insulation of the construction	(square meter·C)/W	1.7	2.75
2	Reduction of the average heat flow density after thermomodernization	W/square meter	6.2	4.8
3	Sew-knit mats without facing МТПБа (ТУ У В.2.7-23.9-00292818-001: 2012) ref. [10]			
3.1	The required thickness of the heat insulation layer	meter	0.05	0.08
3.2	Cost of heat insulation materials	dollars US/square meter	2.4	4.0
4	Soft heat-insulating slabs ПМТБ-2 (ТУ У В.2.7-23.9-00292818-001: 2012) ref. [10]			
4.1	The required thickness of the heat insulation layer	meter	0.065	0.11
4.2	Cost of heat insulation materials	dollars US/square meter	24.9	41.6

– To create such a subindustry of heat insulation production in the form of franchising, which will allow to maintain the constant connection of scientific institutions with enterprises by producers and quickly transfer practical experience and innovations.

Measures to stimulate production:

– With the aim of promoting broad use in construction and stimulating demand from users of the housing, it is proposed at the stage of state expertise of projects for buildings in which ecological heat insulation materials are used that do not contain organic substances and increase the fire resistance of structures, establish high safety and quality indicators;
– Private developers who use such heat insulation materials provide targeted multi-year soft loans.

8. Conclusions

Regulatory measures are needed from the state aimed at the fact that technical solutions for heat insulation of facilities should not cause a decrease in their fire resistance, at least if it concerns residential or other facilities with people staying.

A promising basis for heat insulation materials and products is super-thin basalt fiber (BSTF).

On the way of expanding their application is a high cost, in which part of the energy carriers is more than 80%. Thus, the cost of production depends mainly on the economic efficiency of the heat-technological processes.

For the wide application of such heat insulation in the construction and reconstruction of buildings, it is equally important to use the optimal technical solutions for the arrangement of the basalt heat insulation layer in the conditions of the construction site with the obligatory provision of its mechanical and hydrotechnical protection for the entire service life.

Engineering-and-economical performance of innovative enterprises for the production of basalt heat insulation and measures for its use in construction are presented.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Bogoslovskiy, V.N. and Skanavi, A.N. (1991) Otopleniye: Ucheb. dlya vuzov. [Heating: Educational for Universities]. Stroyizdat, 735 s. (In Russian)

[2] https://www.calc.ru/dinamika-Gas.html

[3] SNiP II-3-79* Stroitel'naya teplotekhnika [Construction Heat Engineering]. (s izm. No. 4 [With Changes #4]). (In Russian)

[4] DBN V.2.6-31-2006 Teplova izolyatsiya budivel' [Thermal Insulation of Buildings] (zi zminoyu [With Changes]). (In Ukraine)

[5] https://life.ru/t/новости/112431

[6] http://www.novoteka.ru/seventexp/6950507/15263358

[7] http://www.bbc.com/russian/news-40366259

[8] http://fire-truck.ru/encyclopedia/pozhar-na-kamaze-14-aprelya-1993-goda-faktyi-i-kommentarii.html

[9] Dzhigiris, D.D. and Makhova, M.F. (2002) Osnovy proizvodstva bazal'tovykh volokon i izdeliy [Basics of Production of Basalt Fibers and Articles]. Teploenergetik [Thermal Power Engineer], Moscow, 416 p. (In Russian)

[10] http://rotys.com

Journal of Modern Physics, 2018, 9, 1735-1760
http://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Theory of the Condon Locus

Jeremy B. Tatum

Department of Physics and Astronomy, University of Victoria, Victoria, Canada
Email: jtatum@uvic.ca

How to cite this paper: Tatum, J.B. (2018) Theory of the Condon Locus. *Journal of Modern Physics*, **9**, 1735-1760.
https://doi.org/10.4236/jmp.2018.99109

Received: May 8, 2018
Accepted: July 30, 2018
Published: August 2, 2018

Abstract

The Condon locus for a diatomic molecule is the locus, in the (υ',υ'') plane, of the strongest bands in an electronic band system. The form of the locus depends upon the form of the potential energy function of the electronic states involved. We show how the locus depends on the potential energy function for simple harmonic and anharmonic oscillators, first from a classical point of view, and then from a quantum mechanical point of view. One phenomenon of interest is that, in the case of anharmonic oscillators, the upper branch of the Condon locus traces much stronger bands than the lower branch. Another phenomenon, predicted by quantum mechanics but not by classical mechanics, is the existence of secondary nested Condon loci.

Keywords

Diatomic Molecules, Franck-Condon Factors, Condon Locus, Vibrational Constants

1. Introduction

The Franck-Condon factors are factors that give the relative strengths of the many bands in an electronic band system of a diatomic molecule. A band is a transition between the vibrational quantum number υ' of an upper electronic state and the vibrational quantum number υ'' of a lower electronic state. Thus a band is defined by the number pair (υ',υ''). There is no rigorous selection rule governing such transitions between different electronic states. However, not all such transitions are equally strong. If the Franck-Condon factors are displayed in a table with rows of constant υ' and columns of constant υ'', the strongest Franck-Condon factors are often seen to lie on a locus (called the Condon locus) which is roughly a parabola whose symmetry axis makes an angle of about 45˚ to the υ' and υ'' axes. This angle may be a little more than 45˚, or a little less. The parabola may be quite narrow (small latus rectum) or quite

open (large latus rectum).

A qualitative interpretation for the Franck-Condon loci was originally described by Franck [1], and a quantum mechanical explanation was developed by Condon [2]. Their physical explanation was described with customary clarity by Herzberg [3]. Roughly it is as follows. At any given instant of time, a molecule is likely to be at its condition of maximum extension or compression, since near these positions in harmonic motion, the speed is slowest. Therefore transitions are most likely to take place from a condition of maximum extension or compression in one electronic state, to a similar condition in the other electronic state. The time taken for an electronic transition is assumed to be very much smaller than the period of vibration of the molecule.

Figure 1 represents the potential energy curves of two electronic states and their quantized total energy levels. It will be seen that a vertical line can easily be drawn from $\upsilon' = 1$ in the upper state to $\upsilon'' = 3$ in the lower state. Thus the $(1,3)$ band is likely to be strong. Likewise the (2, 0) band is also likely to be strong.

In this paper, we examine the form of the Condon locus at four levels of sophistication:

1) Simple harmonic oscillations, classical treatment.

2) Anharmonic oscillations, classical treatment.

3) Simple harmonic oscillations, wave mechanical treatment.

4) Anharmonic oscillations, wave mechanical treatment.

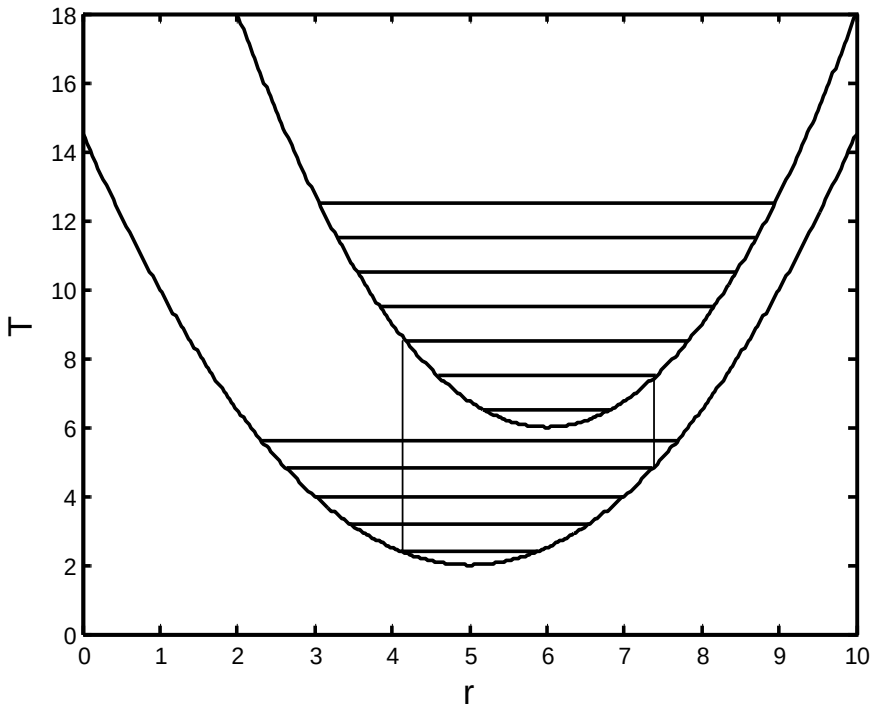


Figure 1. Illustrating two transitions between the vibrational levels of two electronic states that are likely to be strong. They take place between conditions of greatest extension or compression, which is where the vibrational motion is slowest.

In addition to the accounts by Condon and Herzberg cited above, one of the most illuminating papers on this topic of which I am aware is that of Nicholls [4], and there will be some inevitable overlap between that paper and this. However, I have not seen it treated systematically and quantitatively with quite the approach that I am adopting here. Nicholls described the phenomenon of nested Condon loci, which are not predicted classically but which are a result of the wave-mechanical nature of the molecular system.

2. Simple Harmonic Oscillations, Classical Treatment

In the simple harmonic treatment, the potential energy V as a function of internuclear separation r is given by an equation of the form

$$V = V_e + \frac{1}{2}k\left(r - r_e\right)^2. \tag{1}$$

It is customary in spectroscopic practice to express energies as "term values" T, which are the energies divided by hc, so that the term values are then dimensionally similar (L^{-1}) to wavenumbers. Thus Equation (1) would customarily be written as

$$T = T_e + \frac{k}{2hc}\left(r - r_e\right)^2. \tag{2}$$

Here T_e is the electronic contribution to the potential energy. The second term on the right hand side is the elastic contribution (of the vibrating molecule) to the potential energy, r is the internuclear distance, and r_e is its equilibrium value. The symbol k is the force constant, related to the molecular vibrational constant ω_e by

$$k = 4\pi^2 mc^2\omega_e^2 \tag{3}$$

where m is the "reduced mass" $\frac{m_1 m_2}{m_1 + m_2}$ of the molecule.

Mathematically the problem is to draw a horizontal line to intersect the upper curve of **Figure 1**; then drop vertical lines from the two points of intersection; and find the two places where these vertical lines intersect the lower curve. This will result in a relation for the strongest Franck-Condon factors in the form of an equation relating the term values T' in the upper electronic state to the term values T'' in the lower electronic state.

This relation can then be transferred to the (υ', υ'') plane via the relations

$$T'' = T_e'' + \left(\upsilon'' + \frac{1}{2}\right)\omega_e'' \quad \text{and} \quad T' = T_e' + \left(\upsilon' + \frac{1}{2}\right)\omega_e' \tag{4}$$

The derivation and other details of the analysis are given in Hefferlin *et al.* [5] and are not repeated here. Suffice it to say that the Condon locus in the (υ', υ'') plane is a parabola whose equation in a form that is convenient to compute can be written in the form

$$\Omega''\left(\upsilon'' + \frac{1}{2}\right) = \left(\Delta \pm \sqrt{\Omega'\left(\upsilon' + \frac{1}{2}\right)}\right)^2. \tag{5}$$

In this equation I make use of a quantity

$$L = \frac{\hbar}{\pi mc} \tag{6}$$

having the dimensions of a length. This is merely to avoid having to repeat $\hbar/(\pi mc)$ in subsequent equations. If m is expressed in amu, L has the numerical value $\frac{6.743052\times10^{-17}}{m}$ metres. Then

$$\Omega'' = \frac{1}{L\omega_e''}, \quad \Omega' = \frac{1}{L\omega_e'}, \quad \Delta = \frac{r_e'' - r_e'}{L}. \tag{7}$$

The properties of the Condon parabola can be traced using any good text on the conic sections (I used my trusty Loney [6]). The axis of the parabola makes an angle θ with the υ'' axis given by

$$\tan\theta = \frac{\omega'}{\omega''} \tag{8}$$

and the length of the latus rectum is

$$2l = \frac{4\left(r_e'' - r_e'\right)^2 \omega_e''^2 \omega_e'^2}{L\left(\omega_e''^2 + \omega_e'^2\right)^{3/2}}. \tag{9}$$

The parabola is tangent to the lines $\upsilon' = \frac{1}{2}$ and $\upsilon'' = \frac{1}{2}$.

If the equilibrium internuclear separations in the two electronic states are equal, the parabola degenerates into a straight line. A wide Condon parabola indicates that the internuclear separations in the two electronic states are rather different. I show, in **Table 1** and **Figure 2**, two examples, in one of which, CN $B^2\Sigma^+ - X^2\Sigma^+$, the internuclear distances in the two states are not very different and the Condon parabola is consequently rather narrow; and in the other, AlO $A^2\Pi_r - X^2\Sigma^+$, the internuclear distances are rather different, and the Condon parabola is broad.

In the classical model, a simple harmonic oscillator is most likely to be found, at some instant of time, at one of the extrema of its motion. The probability that its distance ξ from its equilibrium position will be between ξ and $\xi + \mathrm{d}\xi$ is proportional to the reciprocal of its speed (by which I mean $\mathrm{d}r/\mathrm{d}t$ rather than the speed of one of the atoms), which I shall call its "slowness", s. The time spent in traversing a distance $\mathrm{d}\xi$ is $s\mathrm{d}\xi$. Indeed the probability that the position of the system will, at some instant of time, be in the interval $\mathrm{d}\xi$ is $2s\mathrm{d}\xi/P = \omega s\mathrm{d}\xi/\pi$. Here P is the period of the motion, and ω (not to be confused with the vibrational constant ω_e) is $2\pi/P$. The factor 2 on the left hand side of the equation arises because the displacement from equilibrium passes through ξ twice per period. And, since the speed in simple harmonic motion of amplitude a is $\omega\sqrt{a^2 - \xi^2}$, the probability $\mathrm{d}\Phi$, at some instant, that the displacement will be between ξ and $\xi + d\xi$ is

$$\mathrm{d}\Phi = \frac{2\mathrm{d}\xi}{\pi\sqrt{a^2 - \xi^2}}. \tag{10}$$

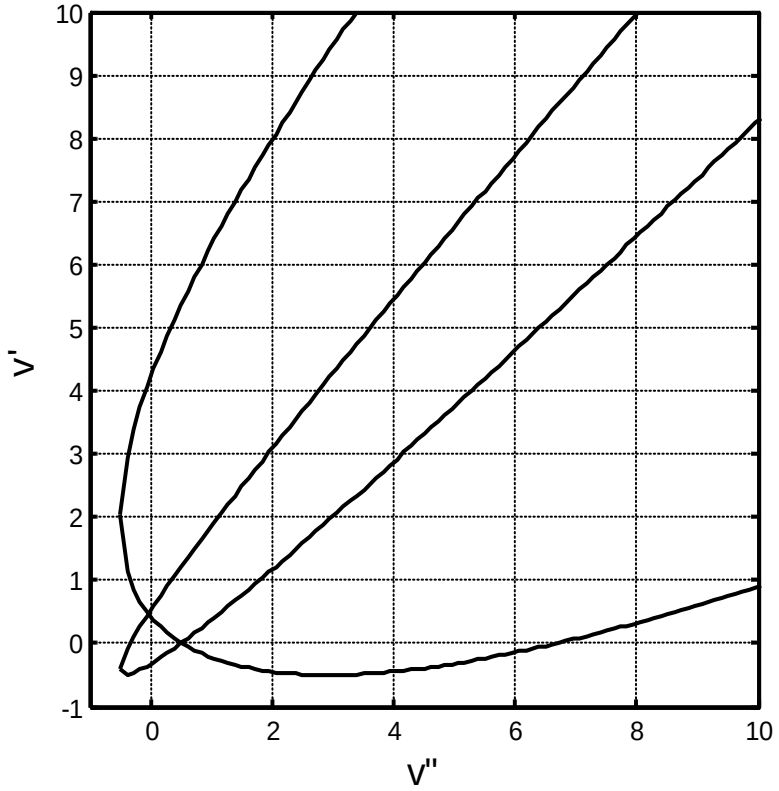


Figure 2. Two Condon parabolas with markedly different latera recta. The narrow parabola (for CN) results from transitions between electronic states of nearly equal internuclear distances. The broad parabola (for AlO) results from transitions between electronic states of rather different internuclear distances.

Table 1. Molecular constants for CN and AlO used in the calculations for **Figure 2**.

CN $B^2\Sigma^+ - X^2\Sigma^+$	AlO $A^2\Pi_r - X^2\Sigma^+$
$m = 6.46427$ amu	10.041 amu
$L = 1.04331\times10^{-17}$ m	0.35537×10^{-17} m
$r_e' = 1.1506\times10^{-10}$ m	1.7708×10^{-10} m
$r_e'' = 1.1718\times10^{-10}$ m	1.6179×10^{-10} m
$\omega_e' = 2.16413\times10^{5}$ m^{-1}	7.285×10^{4} m^{-1}
$\omega_e'' = 2.068705\times10^{5}$ m^{-1}	9.7923×10^{4} m^{-1}
$\Delta = 2.0320\times10^{5}$	-4.303×10^{6}
$\Omega' = 4.42898\times10^{11}$	6.85268×10^{10}
$\Omega'' = 4.63328\times10^{11}$	7.16877×10^{10}
$2I = 0.1361$	3.8982
$\theta = 46.29^\circ$	36.65°
vertex: $(v',v'') = (-0.4775, -0.4743)$	(0.5510, −0.0672)
focus: $(v',v'') = (-0.4529, -0.4508)$	(1.1327, 0.7147)

The coefficient of $d\xi$, which I may call the probability density, is shown graphically in Figure 3 for $a = 1$. The integral between $\xi = 0$ and $\xi = a$ is unity, as befits a probability. The function goes to infinity at $\xi = \pm a$, but of course $d\Phi$ is everywhere finite (zero) as $d\xi \to 0$.

3. Anharmonic Oscillations, Classical Treatment

The vibration of a real molecule will not be simple harmonic, and the curve representing its potential energy as a function of internuclear distance will not be a parabola. For small internuclear distances, when the molecule is compressed, there will be a strong Coulomb repulsion between the nuclei, so the potential energy curve there is steep and negative. For large internuclear distances, the molecule will tend to dissociate, so the potential energy curve asymptotically approaches a dissociation limit. Qualitatively the potential energy function would be expected to look somewhat similar to one of the curves shown in Figure 4. (The small difference between these curves will be described later, following Equation (15)).

The same principles apply in forming the Condon locus as in the simple harmonic case, except that the Condon locus will no longer be a parabola. Transitions are most likely to take place near the stationary points (greatest extension or compression) of the vibration as before. However, if one were to imagine a particle sliding without friction to and fro in one of the potential wells of Figure 4, it is easy to conclude that the particle will spend more time at large r than at

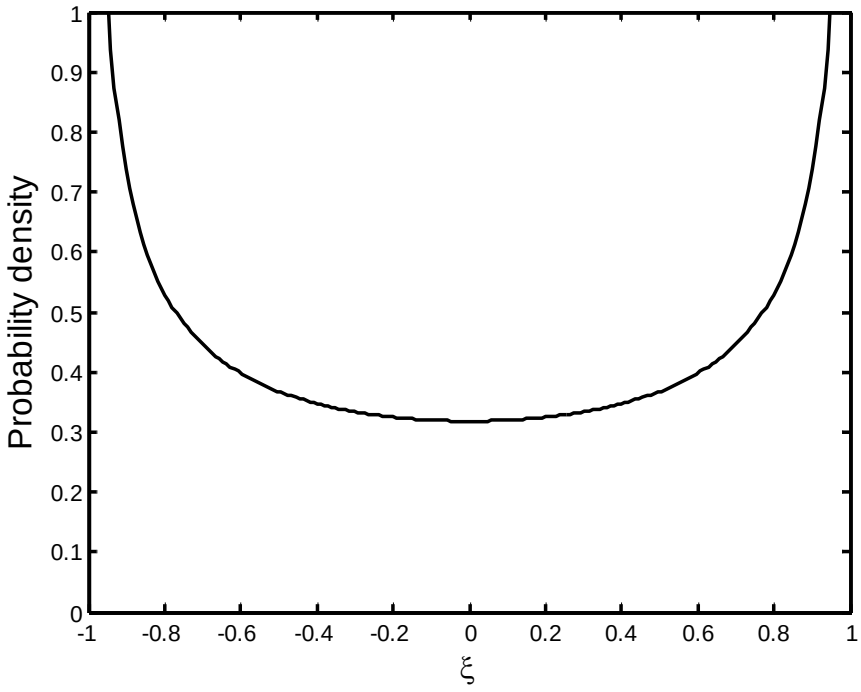


Figure 3. Showing that, at any instant of time during a vibration, a molecule is most likely to be at greatest extension or compression, when it is moving at its slowest, and least likely to be at its equilibrium nuclear separation, when it is moving most rapidly.

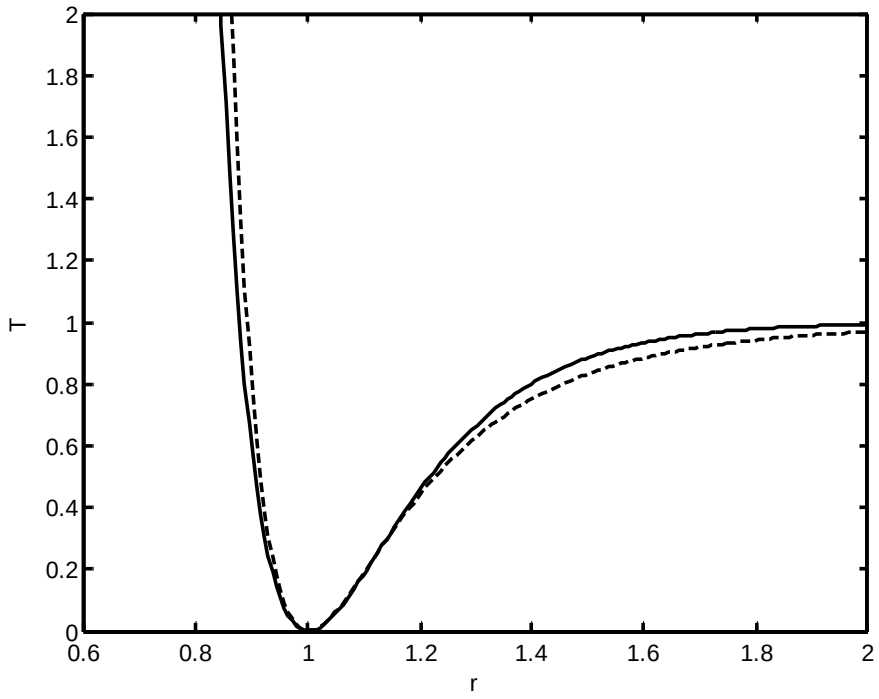


Figure 4. Potential energy of a vibrating molecule as a function of internuclear distance. The continuous curve is a Morse function. The dashed curve is a Lennard-Jones (12-6) function. The two curves are drawn so as to have the same width at $T = 0.5$.

small r. That means that transitions are more likely to occur when the molecule is at greatest extension than at greatest compression. This leads to the conclusion, by qualitative argument alone and without any numerical calculation, that the *upper branch of the Condon locus traces much stronger bands than the lower branch.* We shall see later that the wavemechanical treatment leads to the same conclusion.

Various attempts can be made to devise empirical equations that mimic the expected qualitative potential curve. Two of them are the *Morse potential* [7] and the *Lennard-Jones potential* [8].

The Morse potential is

$$V = D_e\left[1-\mathrm{e}^{-(r-r_e)/a}\right]^2. \tag{11}$$

V = potential energy as a function of the internuclear distance r. r_e = equilibrium internuclear distance. D_e = dissociation energy. The parameter a has the dimensions of length. What little geometric meaning we can give to a is such that when $r - r_e = a$,

$$V = \left(1-\mathrm{e}^{-1}\right)^2 D_e = 0.399576 D_e \approx 0.4 D_e. \tag{12}$$

That is to say the extension of the molecule from its equilibrium separation is a when the potential energy is about 40 percent (*i.e.* 0.399576) of the dissociation energy. Alternatively, when the potential energy is half of the dissociation energy, the extension or compression is

$$r - r_e = \ln\left(2 \pm \sqrt{2}\right) = 1.2279a \text{ or } -0.5348a. \tag{13}$$

There is another formal solution, namely when $r_e - r = a$,

$$V = (1-e)^2 D_e = 2.952 D_e, \tag{14}$$

but this is not a physically interesting solution, because if $V > D_e$, the molecule is unstable. If the molecule is compressed by an amount $r_e - r = a$, it will bounce back and dissociate.

The Lennard-Jones potential is

$$V = D_e\left[1 + \left(\frac{r_e}{r}\right)^m - 2\left(\frac{r_e}{r}\right)^n\right]. \tag{15}$$

In Figure 4, the dashed line is a Lennard-Jones potential with $m = 12$ and $n = 6$, while the continuous line is a Morse potential with $a = 0.1772$. Some slightly tedious algebra will show that the Morse and Lennard-Jones potentials (with $m = 12$ and $n = 6$) will have the same full widths at half minimum (FWHm) for

$$a = \frac{\left(2+\sqrt{2}\right)^{1/6} - \left(2-\sqrt{2}\right)^{1/6}}{\ln\left(3+\sqrt{8}\right)} = 0.177212908, \tag{16}$$

which is the reason why I chose that value for the Morse parameter in preparing Figure 4.

Although the two curves look somewhat similar, only the Lennard-Jones function has the physically desirable characteristic of going to infinity as $r \to 0$. However, the Morse function is very steep for small r, and, in the example of Figure 4, it reaches the respectable value of 7.9×10^4 at $r = 0$, so that the function does not give a wholly unreasonable representation of a real potential function. Another attractive feature of the Morse function is that, when it is inserted into the Schrödinger equation, the eigenfunctions can be written explicitly in terms of algebraic functions (Laguerre polynomials), and, especially, the eigenvalues (vibrational energy levels) are given as

$$T = T_e + G(\upsilon) = T_e + \omega_e\left(\upsilon + \frac{1}{2}\right) - \omega_e x_e\left(\upsilon + \frac{1}{2}\right)^2, \tag{17}$$

with no higher powers of $\left(\upsilon + \frac{1}{2}\right)$. For these reasons I use the Morse function in the present analysis of anharmonicity. (In Equation (17), T is the energy divided by hc. That is to say, it is the *term value* (in m^{-1}) of the level.)

Further comparisons between these two potential functions can be found in Lim [9].

Let us introduce the dimensionless variables

$$\xi = (r - r_e)/a. \tag{18}$$

and

$$U = V/D_e \tag{19}$$

Then the Morse function is

$$U = \left(1 - \mathrm{e}^{-\xi}\right)^2. \tag{20}$$

The Taylor expansion of this to ξ^2 is just $U = \xi^2$. Thus, to order ξ^2, the Morse potential is the same as the simple harmonic oscillator potential for which

$$V = \frac{1}{2}k\left(r - r_e\right)^2$$

where k is the force constant, and hence

$$a = \sqrt{2D_e/k}, \quad k = 2D_e/a^2. \tag{21}$$

The fundamental frequency is

$$\nu_0 = \frac{1}{2\pi}\sqrt{\frac{k}{m}} = \frac{1}{2\pi a}\sqrt{\frac{2D_e}{m}}, \tag{22}$$

and the following relations are also of interest:

$$\omega_e = \frac{\nu_0}{c}, \quad \omega_e x_e = \frac{h\nu_0^2}{4D_e c}. \tag{23}$$

For large a (small force constant), the graph of the potential energy *versus* $r - r_e$ has a wide and shallow minimum. For small a, the graph of the potential energy *versus* $r - r_e$ has a sharp and steep minimum. Figure 5 shows a Morse curve with its energy levels.

In Figure 5, ξ is $r - r_e$ in units of a, and, for reference, the energies of the v th levels and the limits of the motion are given in Table 2.

The Condon locus resulting from transitions between two electronic states whose potential energies are given by Morse functions can be found by the same

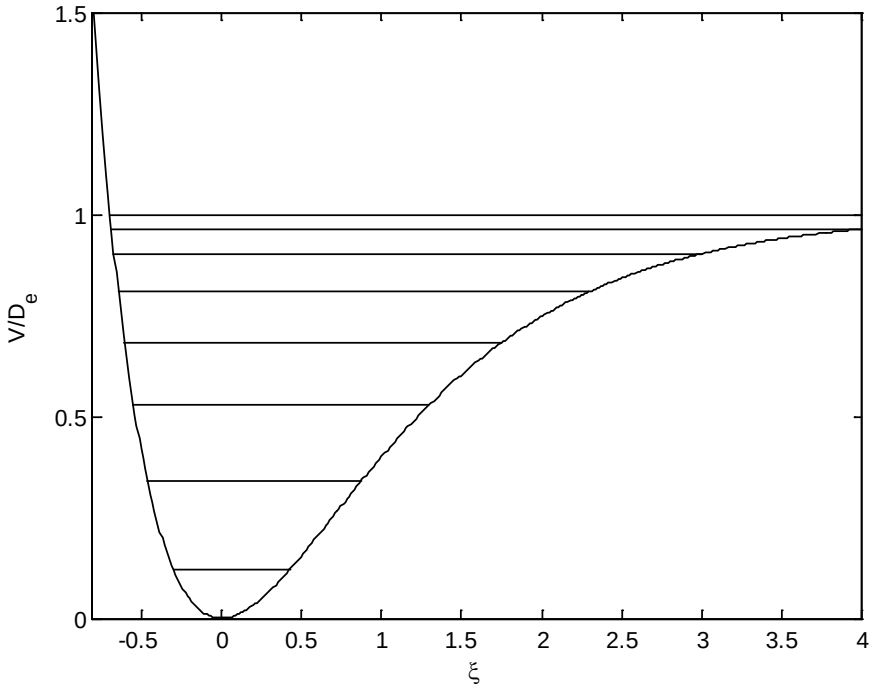


Figure 5. A Morse potential function and its vibrational energy levels, which bunch up closely towards the dissociation limit.

Table 2. Limits of the motion for the vibrational levels of Figure 5.

v	E_v/D_e	ξ_{left}	ξ_{right}
0	0.121093750	−0.298611087	0.427688129
1	0.339843750	−0.459297264	0.874575994
2	0.527343750	−0.545913410	1.295300309
3	0.683593750	−0.602564316	1.753292606
4	0.808593750	−0.641442439	2.294799585
5	0.902343750	−0.667787197	2.994088817
6	0.964843750	−0.684239930	4.032192797
7	0.996093750	−0.692169184	6.237346629

procedure as for simple harmonic functions, as described following Equation (3) in Section 2. Because of the transcendental nature of the equations, it is not possible to arrive at a simple, explicit equation for the Condon locus (which is no longer a parabola). It is straightforward, in any particular case, to carry out the procedure numerically by computer. Transformation from the (T',T'') plane to the (υ',υ'') plane is performed by inversion of Equation (17). The electronic contributions to the energies of the two states do not come into the Condon locus in the (υ',υ'') plane, and can both conveniently be taken to be zero.

I have done the calculations for several cases below, in Figure 6, where we can see how the nature of the Condon locus varies with the molecular constants of the two states. Except for the anharmonicity constants $\omega_e'x_e'$ and $\omega_e''x_e''$, I have used the molecular constants for the CN $B^2\Sigma^+ - X^2\Sigma^+$ tabulated in Section 2, Table 1. For illustrative purposes I have added various purely fictional values of $\omega_e'x_e'$ and $\omega_e''x_e''$, shown in Table 3 in order to see their effect on the Condon locus. The dashed curves show the Condon parabola in the simple harmonic approximation, in which $\omega_e'x_e'$ and $\omega_e''x_e''$ are both zero. The full lines are the Condon loci when anharmonicity is added. I have drawn the upper arm of the Condon locus in the anharmonic case with a thicker line than the lower arm, to reflect the fact that, as explained in the second paragraph of this Section, the upper arm delineates stronger Franck-Condon factors than the lower arm. The values chosen for the anharmonic constants are indicated in Table 3. Since a picture is worth a thousand words, I leave it to the reader to discern the trends that arise from various choices of the anharmonicity constants.

As in the simple harmonic case, the probability that the position of the system will, at some instant of time, be in the interval $\mathrm{d}\xi$ is $2s\mathrm{d}\xi/P$, where s is the slowness and P is the period, though determining this quantity is slightly less easy that in the simple harmonic case. For illustrative purposes I shall consider a Morse potential of the form given by Equation (11) and I shall determine expressions for the period P of the motion and the slowness s as a function of ξ.

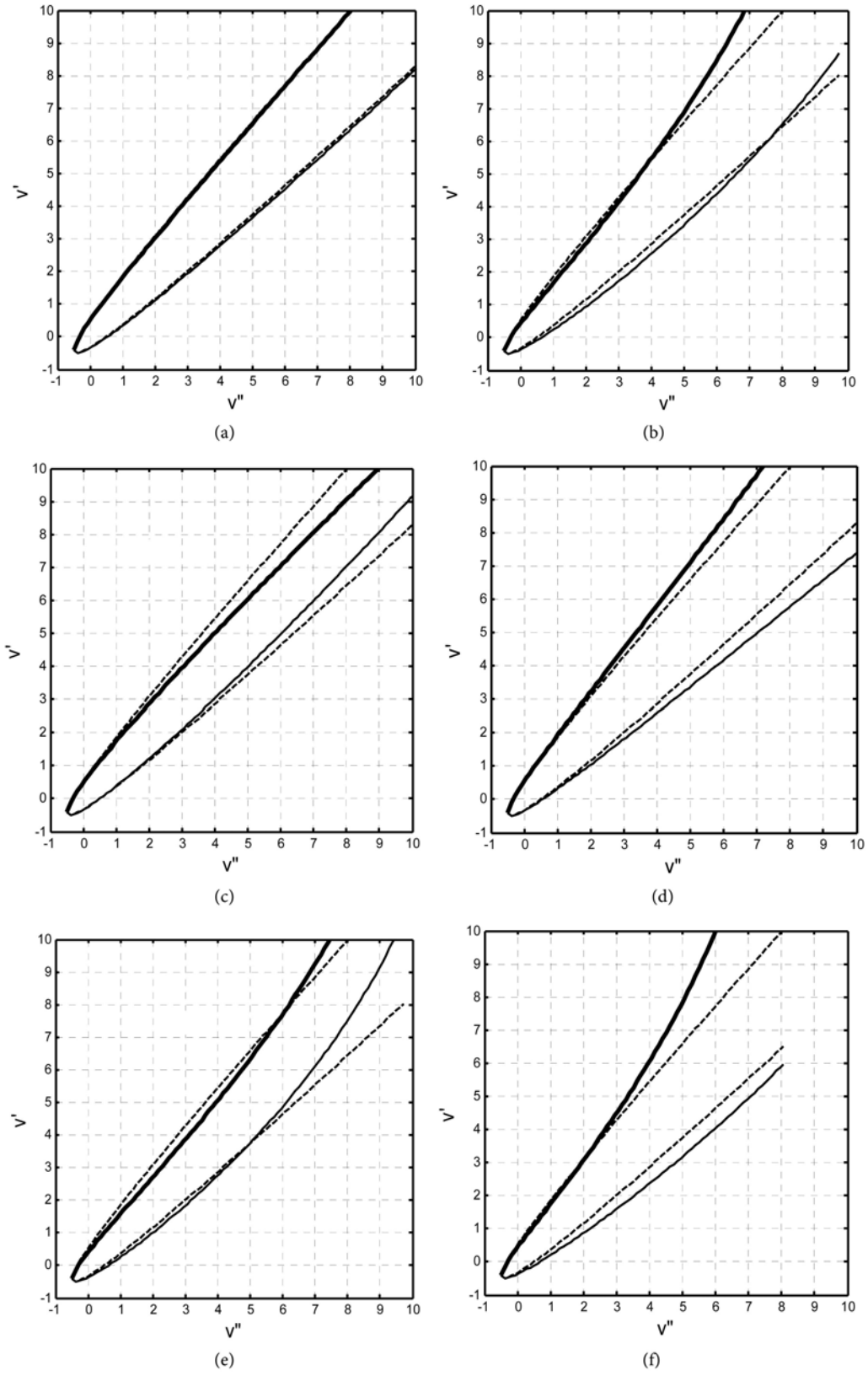


Figure 6. The dashed curves show the Condon parabola for the simple harmonic approximation. The full curves show the effect of anharmonicity upon the Condon locus (no longer a parabola). The anharmonicity constants used are shown in **Table 3**. As explained in the text, in the anharmonic case the upper arm of the locus is much stronger than the lower arm.

Table 3. Anharmonicity constants used for the calculations of Figure 6.

	$\omega_e''x_e''$ (m^{-1})	$\omega_e'x_e'$ (m^{-1})
(a)	500	500
(b)	5000	5000
(c)	600	300
(d)	300	600
(e)	6000	5000
(f)	5000	6000

The period P of the motion, in units of $a\sqrt{m/D_e}$ (a-Morse parameter, m = reduced mass of the molecule, D_e = dissociation energy), is

$$P=\sqrt{2}\left[\int_{\xi\text{left}}^{0}\frac{\mathrm{d}\xi}{\sqrt{E_\upsilon-\left(1-\mathrm{e}^{-\xi}\right)^2}}+\int_{0}^{\xi\text{right}}\frac{\mathrm{d}\xi}{\sqrt{E_\upsilon-\left(1-\mathrm{e}^{-\xi}\right)^2}}\right], \tag{24}$$

where E_v is the total energy in the vth vibrational level. The first integral pertains to the time when the molecule is compressed; the second integral pertains to the time when the molecule is extended. The integration results in

$$P=\sqrt{\frac{8}{1-E_\upsilon}}\left[\tan^{-1}\left(\frac{1-\sqrt{E_\upsilon}}{\sqrt{1-E_\upsilon}}\right)+\tan^{-1}\left(\frac{1+\sqrt{E_\upsilon}}{\sqrt{1-E_\upsilon}}\right)\right]. \tag{25}$$

Table 4 shows, for the first eight vibrational levels, the time P_1, in units of $a\sqrt{m/D_e}$, during which the molecule is compressed; the time P_2 during which it is extended; the total period P; and the ratio P_2/P_1. It will be seen that this ratio increases with vibrational quantum number. This means that our prediction that the upper arm of the Condon locus delineates stronger Franck-Condon factors than the lower arm is more pronounced at larger quantum numbers, and less pronounced at lower quantum numbers.

The speed (*i.e.* $\mathrm{d}r/\mathrm{d}t$) in units of $\sqrt{D_e/m}$ as a function of internuclear distance is given by (nonrelativistic) energy considerations to be

$$\frac{\mathrm{d}r}{\mathrm{d}t}=\sqrt{2\left[E_\upsilon-\left(1-\mathrm{e}^{-\xi}\right)^2\right]}, \tag{26}$$

where $\xi=\left(r-r_e\right)/a$, and the slowness s is the reciprocal of this. The probability density $2s/P$ is shown in Figure 7 as a function of ξ for the $v=7$ level. The area under the curve is unity. Table 4 and Figure 7 show that the molecule spends more time in extension $(\xi>1)$ than in compression $(\xi<1)$, with the consequence that the upper arm of the Condon locus is stronger than the lower arm; and that for large v the molecule spends *much* more time in extension than in compression, with the consequence that in practice the lower arm of the Condon locus is likely to be observed only for low v.

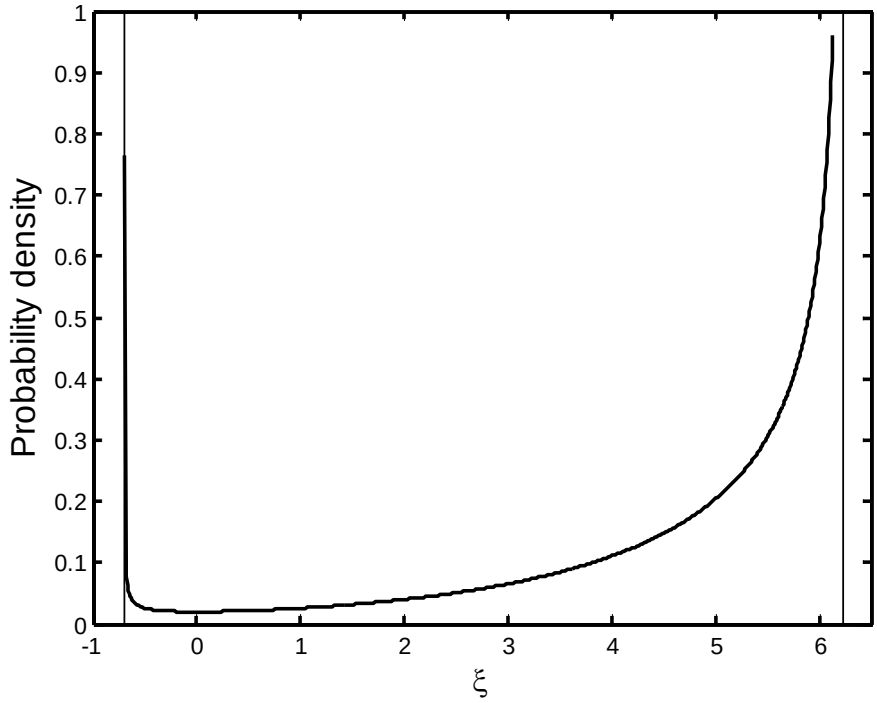


Figure 7. Showing that, at any instant of time during an anharmonic vibration, a molecule is more likely to be extended than compressed, with consequences to the Condon loci that are illustrated in **Figure 6**.

Table 4. Comparison of the time spent in compression (P_1) and expansion (P_2) in each level of **Figure 5**.

v	P_1	P_2	P	P_2/P_1
0	1.833386651	2.905688483	4.739075134	1.584874899
1	1.650805553	3.817358063	5.468163616	2.312421385
2	1.559318591	4.903056592	6.462375183	3.144358453
3	1.501973790	6.396484767	7.898458557	4.258719301
4	1.463724273	8.691436728	10.155161002	5.937892052
5	1.438305896	12.778919506	14.217225402	8.884702163
6	1.422633874	22.272741796	23.695375670	15.655990062
7	1.415135897	69.670991114	71.086127011	49.232721242

4. Simple Harmonic Oscillations, Wave Mechanical Treatment

In the wave mechanical model, the probability density, denoted by the symbol ψ^2, is a wavefunction with several nodes (zeroes) and antinodes (maxima), and which extends slightly beyond the classical limits of the motion. At a given instant of time, the extension of a molecule from its equilibrium position is most likely to be at one of the maxima of ψ^2. And, since there are several maxima, this gives rise to the possibility that there will be, in the (υ',υ'') plane, *several*

nested Condon loci delineating strong Franck-Condon factors. Examples of such nested Frank-Condon factors will be found in Nicholls [4].

For simple harmonic potentials the wavefunctions are given in many standard texts, such as that of Eyring, Walter and Kimbal [10]. Normalized to unit area they are

$$\psi^2 = \frac{1}{2^{\upsilon}\upsilon!\sqrt{\pi}} \frac{H_{\upsilon}^2(\xi)\mathrm{e}^{-\xi^2}}{l} \tag{27}$$

where

$$l = \sqrt{\frac{\hbar}{\sqrt{mk}}} \tag{28}$$

and the H_{υ} are the Hermite polynomials. The constant l has dimensions of length. (Recall that $\psi^2 \mathrm{d}r$ must be dimensionless, which verifies that Equation (27) balances dimensionally.) Samples of these, for $\upsilon = 0,5,10$ are shown in Figures 8-10, where the vertical lines are the classical limits of the motion.

Several features are worthy of comment. For large v the locus of the maxima closely follow the probability density for the classical case illustrated in Figure 3, and this is often cited as an example of the Bohr correspondence principle. The molecule, at some instant, is most likely to be found near (but not exactly at) the extrema of the motion. The subsidiary maxima are rather lower than the main maxima, which means that the Franck-Condon factors that they delineate will not be as strong as those delineated by the principal maxima. For intermediate v these characteristic are not at all as pronounced, while for $v = 0$ (no vibration)

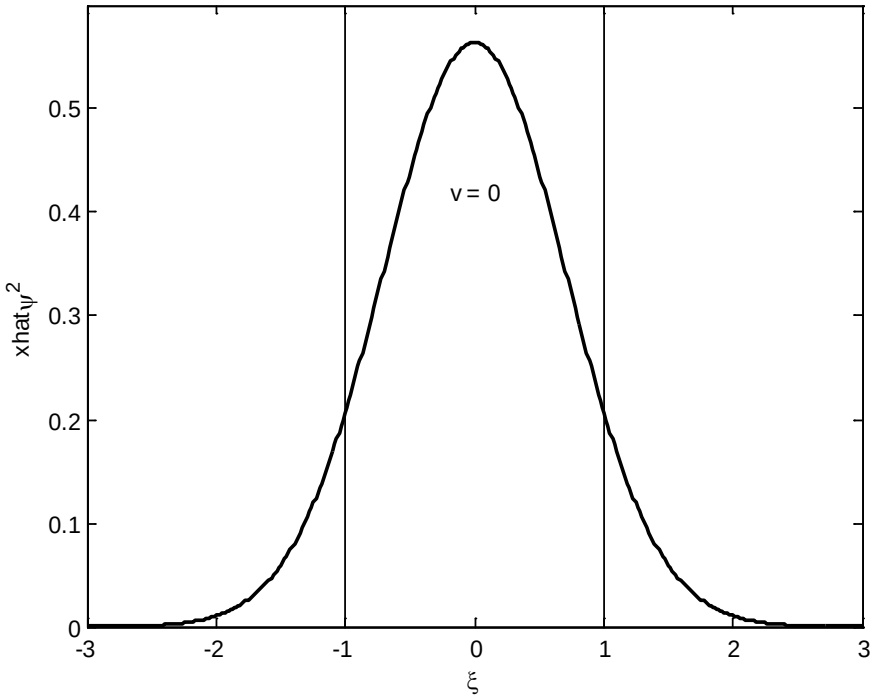


Figure 8. Probability density for $v = 0$. The most likely configuration of the molecule at any instant is, of course, its equilibrium configuration.

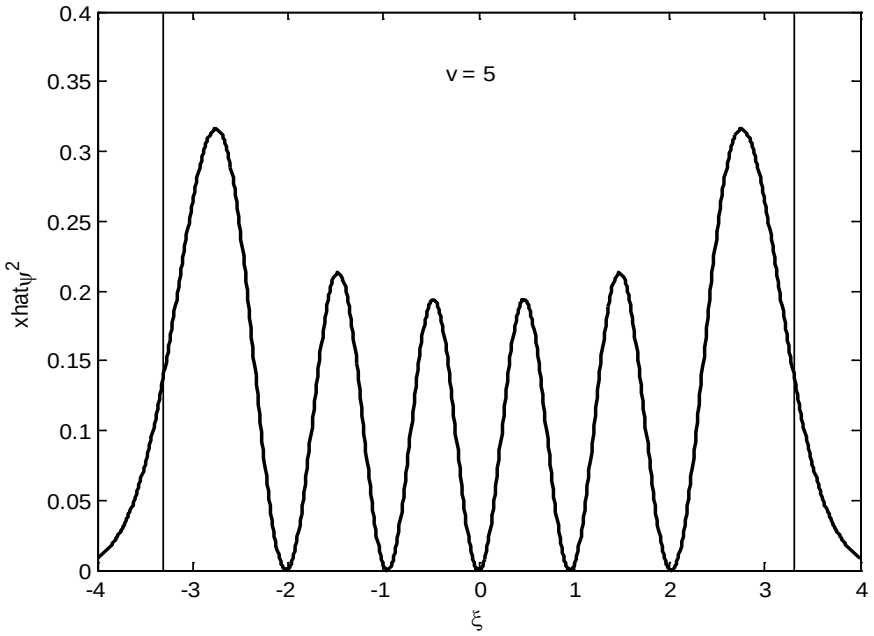


Figure 9. Probability density for $v = 5$. The most likely configuration of the molecule at any instant is a little less than greatest extension or compression.

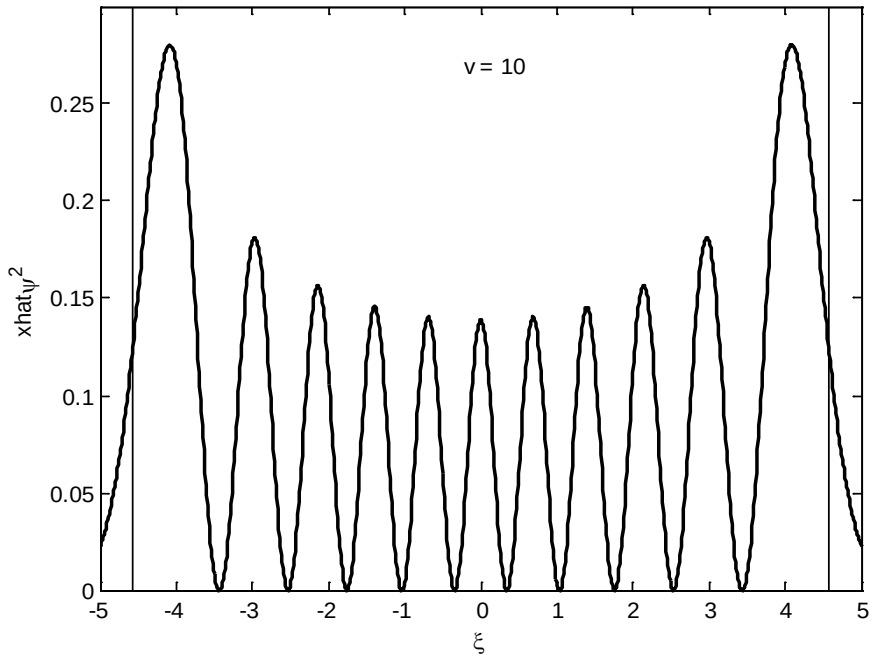


Figure 10. Probability density for $v = 10$. The figure is beginning to look remarkably similar to that for the classical model of **Figure 3**.

the most likely condition of the molecule is not at an extremum of the potential well, but it will most likely be at its equilibrium position, which will surprise no one.

It is evident that, in order to calculate the Condon loci, we need to know the positions of the maxima of these wavefunctions. These are given in **Table 5**.

Table 5. Classical limits, and positions and heights of the maxima of the probability densities (ψ^2) for the first 11 vibrational levels of a simple harmonic oscillator.

	Classical limits	ψ^2 max abscissa	ψ^2 max ordinate
$\nu = 0$	±1.000000	0.000000	0.564190
$\nu = 1$	±1.732051	±1.000000	0.415107
$\nu = 2$	±2.236068	±1.581139	0.370492
		0.000000	0.282095
$\nu = 3$	±2.645751	±2.034075	0.345616
		±0.602114	0.245552
$\nu = 4$	±3.000000	±2.417686	0.328801
		±1.074613	0.225993
		0.000000	0.211571
$\nu = 5$	±3.316625	±2.756849	0.316282
		±1.475241	0.213132
		±0.476251	0.193521
$\nu = 6$	±3.605551	±3.062508	0.306401
		±1.828611	0.203748
		±0.881604	0.181731
		0.000000	0.176309
$\nu = 7$	±3.872983	±3.344197	0.298291
		±2.147928	0.196458
		±1.239870	0.173179
		±0.406782	0.165096
$\nu = 8$	±4.123106	±3.606369	0.291445
		±2.441238	0.190550
		±1.563978	0.166570
		±0.767093	0.156984
		0.000000	0.154271
$\nu = 9$	±4.358899	±3.852560	0.285543
		±2.713869	0.185615
		±1.861876	0.161237
		±1.093513	0.150730
		±0.361030	0.146450
$\nu = 10$		±4.085357	0.280369
		±2.969559	0.181399
		±2.138862	0.156801
		±1.393823	0.145697
		±0.688554	0.140428
		0.000000	0.138844

By way of example of the principles involved, I draw in Figure 11 two simple harmonic potential energy curves for a fictitious molecule. I am taking the potential energy of the two electronic states involved to be

$$V = V_e + \frac{1}{2}k\left(r - r_e\right)^2, \tag{29}$$

where for the lower state,

$$V_e = 0,\ \frac{k}{2hc} = 2.447\times10^{25}\,\mathrm{m}^{-3},\ r_e = 3.9\times10^{-10}\,\mathrm{m} \tag{30}$$

and for the upper state,

$$\frac{V_e}{hc} = 2\times10^{5}\,\mathrm{m}^{-1},\ \frac{k}{2hc} = 1.625\times10^{25}\,\mathrm{m}^{-3},\ r_e = 4.0\times10^{-10}\,\mathrm{m}. \tag{31}$$

In Figure 11 I have drawn the first eleven vibrational energy levels in each electronic state, and I have indicated by dots the positions of the maxima of the squares of the wavefunctions. The heights of these maxima can be found from Table 5 and from Figures 8-10. At a given instant of time the most probable condition of the molecule is not that of greatest extension or compression, but at a separation corresponding to the position of one of the dots. A likely transition (*i.e.* a large Franck-Condon factor) is one for which a dot on a level in the lower state is vertically beneath a dot on a level in the upper state (on the supposition that the time taken for an electronic transition is much shorter than the vibrational period of the molecule).

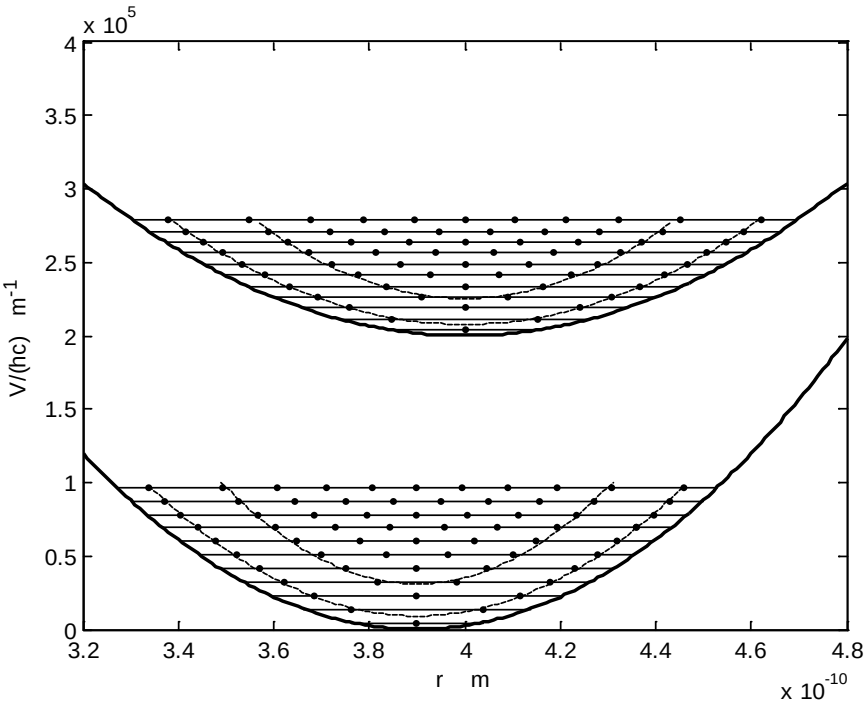


Figure 11. Two simple harmonic electronic states, with the positions of the maxima of the squares of the eigenfunctions (probability densities) marked by dots on each vibrational level. Vibrational transitions (bands) are most likely where a dot in one electronic state is vertically above or below a dot in the other.

I have also indicated, by dashed curves in the figure the approximate loci of the largest and the second largest maxima. These loci are not exact parabolas, although in the drawing I have indicated the "best" parabolas (quadratic least squares regressions of r upon V) through the dots.

These loci can now be used to calculate Condon loci in the (υ',υ'') plane, which are shown in Figure 12. The full curve is the Condon locus for simple harmonic oscillations, classical treatment. The dashed curves are the loci of the most prominent bands according to the wavemechanical calculations. In addition to large Franck-Condon factors delineated along the Condon loci, there may also be a few strong bands randomly distributed. For example, reference to Figure 11 shows that there will be a strong band joining a secondary maximum at $\upsilon'=8$ to a principal maximum at $\upsilon''=7$.

5. Anharmonic Oscillations, Wave Mechanical Treatment

The Morse potential is given by Equation (11). When this is inserted into the Schrödinger equation, it is well known that the eigenvalues (energy levels) can be written as a series in $\upsilon+\frac{1}{2}$ containing no powers higher than the second. The energy levels are given by

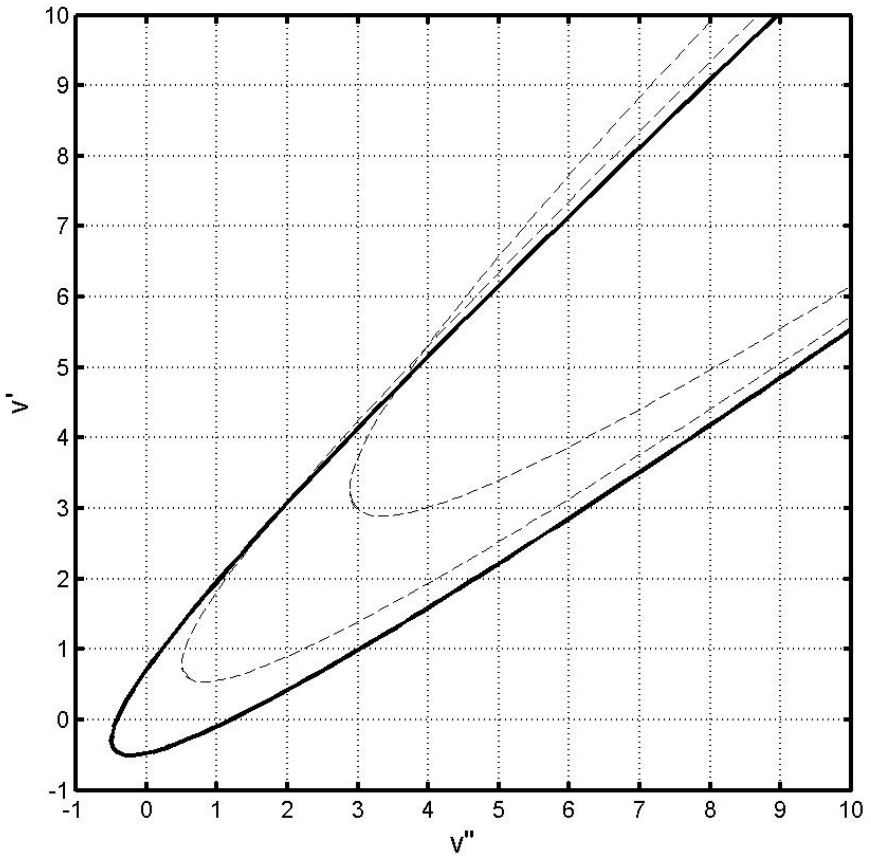


Figure 12. The full curve shows the Condon parabola calculated on the classical model for simple harmonic oscillations. The dashed curves show the principal and the secondary Condon loci from wavemechanical calculations.

$$E_\upsilon = h\nu_0\left[\left(\upsilon+\frac{1}{2}\right)-\frac{h\nu_0}{4D_e}\left(\upsilon+\frac{1}{2}\right)^2\right] \tag{32}$$

In spectroscopic practice, this is usually written as vibrational term values $G(\upsilon)$ (energy divided by *hc*) and the vibrational and anharmonicity constants ω_e and $\omega_e x_e$:

$$G(\upsilon) = \omega_e\left(\upsilon+\frac{1}{2}\right)-\omega_e x_e\left(\upsilon+\frac{1}{2}\right)^2 \tag{33}$$

from which we see that

$$\omega_e = \frac{\nu_0}{c} \tag{34}$$

and

$$\omega_e x_e = \frac{h\nu_0^2}{4D_e c} \tag{35}$$

Another way of writing equations (32) or (33) is

$$\frac{E_\upsilon}{D_e} = b\left[\left(\upsilon+\frac{1}{2}\right)-\frac{1}{4}\left(\upsilon+\frac{1}{2}\right)^2\right] \tag{36}$$

where *b* (dimensionless) is

$$b = \frac{h\nu_0}{D_e} \tag{37}$$

Figure 5 was drawn with an arbitrary value of $b=\frac{1}{4}$.

I also introduce, for convenience, some more quantities as follows:

$$\hat{E} = \frac{h^2}{2ma^2} = \frac{\pi^2c^2\hbar^2\omega_e^2}{D_e} = \frac{c^2h^2\omega_e^2}{4D_e} = \frac{1}{4}b^2D_e. \tag{38}$$

$$\beta = \frac{1}{b} = \frac{D_e}{h\nu_0} \tag{39}$$

$$z = 4\beta e^{-\xi} \tag{40}$$

$$P = \upsilon!\left[\frac{4\beta-\upsilon-1}{\Gamma(\upsilon+1)\Gamma(4\beta-\upsilon)}\right]^{1/2} \tag{41}$$

$$Q = z^{2\beta-\left(\upsilon+\frac{1}{2}\right)}e^{-z/2} \tag{42}$$

$$R = L_\upsilon^{4\beta-2\upsilon-1}(z) \tag{42}$$

where the *L* are the Laguerre polynomials, generated by

$$L_\upsilon^n(z) = \frac{z^{-n}e^z}{\upsilon!}\frac{d^\upsilon}{dz^\upsilon}\left(z^{n+\upsilon}e^{-z}\right). \tag{43}$$

Of these quantities, $\hat{E}$ has dimensions of energy, while the others are dimensionless.

The wavefunctions for the Morse potential are then given by

$$\psi_{\upsilon}(z) = PQRa^{-1/2}. \tag{44}$$

The quantity P is the normalization factor chosen to ensure that $\int_0^{\infty} \psi^2 \mathrm{d}r = 1$. In this paper I am concerned only with the Condon loci and hence only with the positions and *relative* heights of the maxima of the wavefunctions. Thus I shall be concerned only with the ξ-dependent part of the wavefunctions, namely the product QR. The Morse function of Figure 5 was drawn with $b = \frac{1}{4}, \beta = 4$, and I tabulate below, in perhaps a more comprehensible form than Equation (44), the unnormalized wave functions (*i.e.* the product QR) for the first eight vibrational levels.

$$\psi_0 = z^{7.5}\mathrm{e}^{-z/2} \tag{45}$$

$$\psi_1 = (14 - z)z^{6.5}\mathrm{e}^{-z/2} \tag{46}$$

$$\psi_2 = \frac{1}{2}\left(156 - 26z + z^2\right)z^{5.5}\mathrm{e}^{-z/2} \tag{47}$$

$$\psi_3 = \frac{1}{6}\left(1320 - 396z + 36z^2 - z^3\right)z^{4.5}\mathrm{e}^{-z/2} \tag{48}$$

$$\psi_4 = \frac{1}{24}\left(7920 - 3960z + 660z^2 - 44z^3 + z^4\right)z^{3.5}\mathrm{e}^{-z/2} \tag{49}$$

$$\psi_5 = \frac{1}{120}\left(30240 - 25200z + 7200z^2 - 900z^3 + 50z^4 - z^5\right)z^{2.5}\mathrm{e}^{-z/2} \tag{50}$$

$$\begin{aligned}\psi_6 = \frac{1}{720}\big(&60480 - 90720z + 45360z^2 - 10080z^3\\ &+1080z^4 - 54z^5 + z^6\big)z^{1.5}\mathrm{e}^{-z/2}\end{aligned} \tag{51}$$

$$\begin{aligned}\psi_7 = \frac{1}{5040}\big(&40320 - 141120z + 141120z^2 - 58800z^3\\ &+11760z^4 - 1176z^5 + 56z^6 - z^7\big)z^{0.5}\mathrm{e}^{-z/2}\end{aligned} \tag{52}$$

By way of example I draw, in Figures 13-15, the squares of the unnormalized wavefunctions (Q^2R^2) for ν = 0, 4 and 7. The vertical dashed lines indicate the classical limits of the motion.

As in the classical model, we see that, for large ν, the molecule spends *much* more time in extension than in compression, so that the lower arm of the Condon locus is likely to be observed only for very small ν. And, as for the simple harmonic case, for large ν, the locus of the maxima becomes more and more similar to the classical "slowness" curve of Figure 7, providing another example of the correspondence principle.

The strongest bands in a band system depend immediately on the positions of the maxima of the squares of the wavefunctions. Accordingly I provide, for the first eight vibrational levels, these positions in Table 6, in which the entries in normal font are the values of ξ at which ψ^2 is a maximum, and **boldface** entries are the classical limits of the motion.

As with the simple harmonic case of Section 4, I draw as Figure 16, by way of example of the principles involved, two anharmonic potential energy curves for

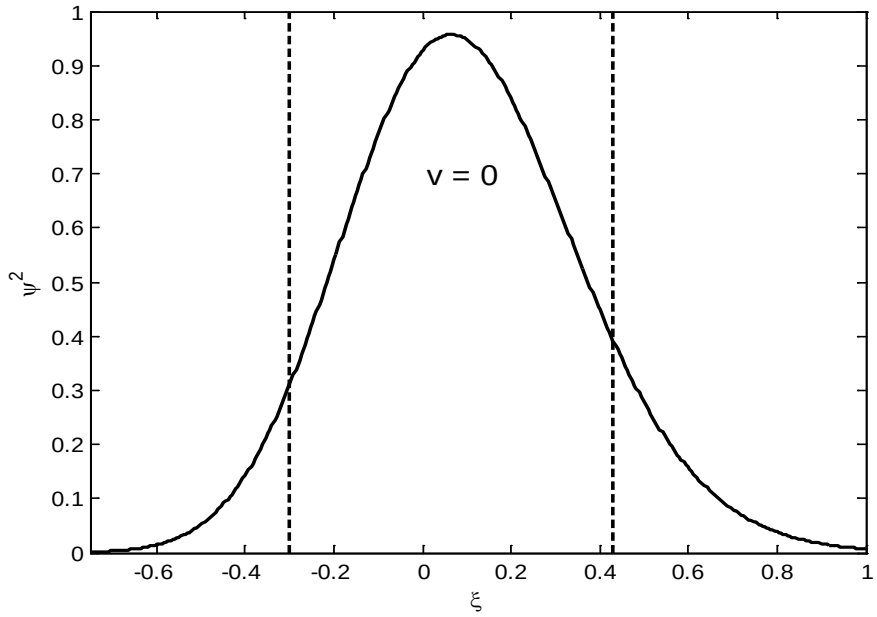


Figure 13. Probability density for $v = 0$. The most likely configuration of the molecule at any instant is not quite at its equilibrium configuration. The vertical dashed lines represent the classical limits of the motion.

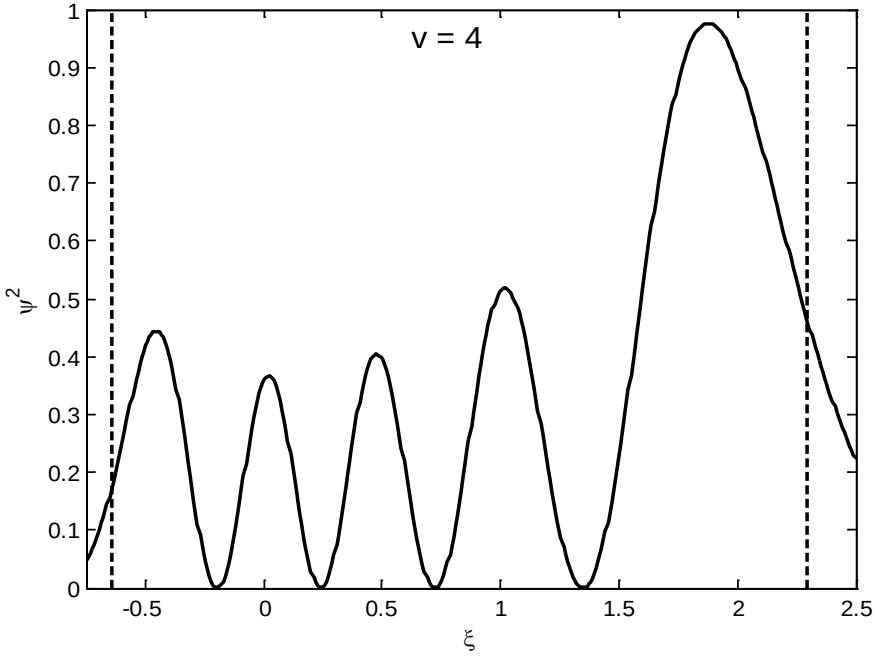


Figure 14. Probability density for $v = 4$. The most likely configuration of the molecule at any instant is near to, but not quite at, greatest extension. It is much less likely to be near greatest compression.

a fictitious molecule. The potential energy curves are calculated with the (arbitrary but realistic) values given in **Table 7**.

The dots in **Figure 16** indicate the positions of the maxima of the squares of the wavefunctions. These dots give the most probable extension or compression

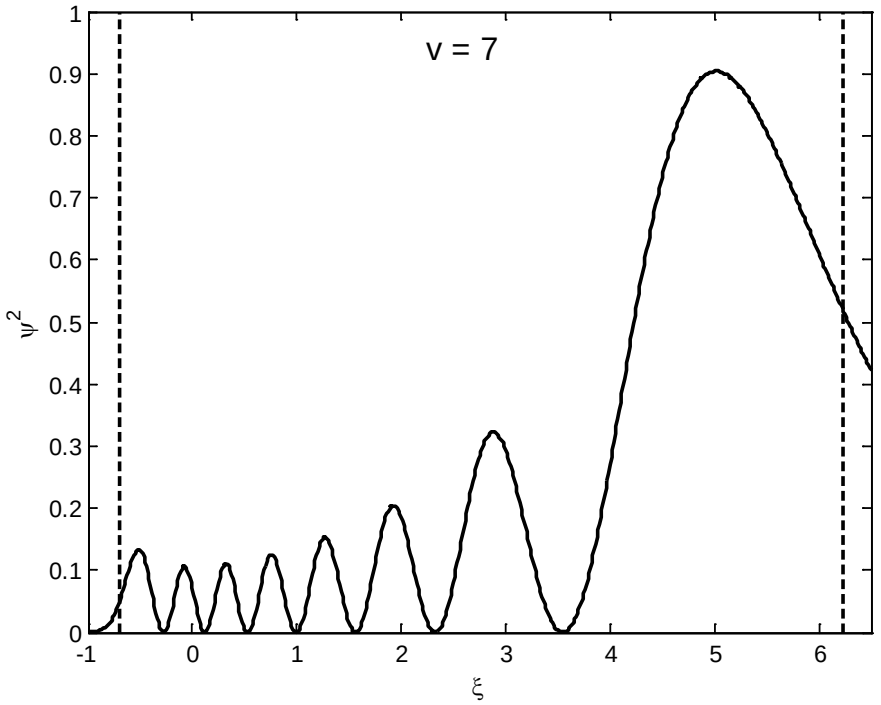


Figure 15. Probability density for $v = 7$. Compare this with the classical case shown in **Figure 7**.

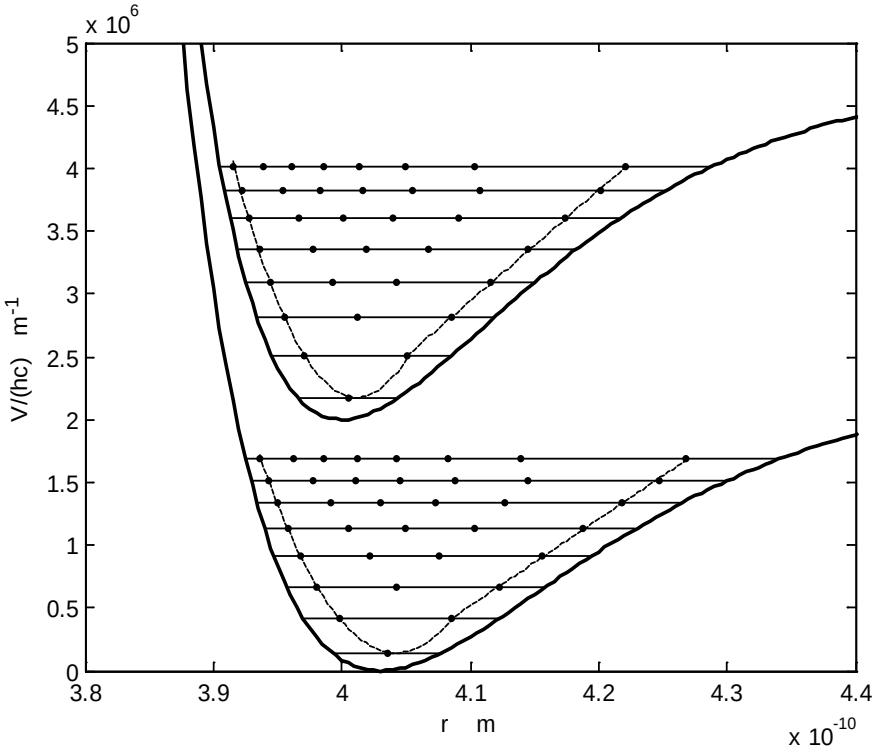


Figure 16. Two anharmonic electronic states, with the positions of the maxima of the squares of the eigenfunctions (probability densities) marked by dots on each vibrational level. Vibrational transitions (bands) are most likely where a dot in one electronic state is vertically above or below a dot in the other.

Table 6. Classical limits, and positions of the maxima of the probability densities (ψ^2) for the first eight vibrational levels of an anharmonic (Morse) oscillator.

$v = 0$	$v = 1$	$v = 2$
−0.298611087	**−0.459297264**	**−0.545913410**
−0.064538521	−0.213854184	−0.331747632
+0.427688129	+0.555024941	+0.239376541
	+0.874575994	+0.962385978
		+**1.295300309**
$v = 3$	$v = 4$	$v = 5$
−0.602564316	**−0.641442439**	**−0.667787197**
−0.404447021	−0.452953895	−0.485304849
+0.103244511	+0.021148216	−0.030959327
+0.620563556	+0.475860785	+0.392400221
+1.388382250	+1.018545638	+0.869710180
+1.753292606	+1.877286233	+1.474643142
	+**2.294799585**	+2.488684223
		+2.994088817
$v = 6$	$v = 7$	
−0.684239930	**−0.692169184**	
−0.505321775	−0.514921216	
−0.062338833	−0.077179042	
+0.344304585	+0.322030283	
+0.791542138	+0.756684695	
+1.328614250	+1.268855283	
+2.047561579	+1.924187023	
+3.355078812	+2.880975731	
+4.032192797	+5.015474119	
	+6.237346629	

Table 7. Molecular constants used for the calculation of Figure 16.

	T_e (m^{-1})	D_e (m^{-1})	r_e (m)	a (m)	ω_e (m^{-1})	$\omega_e x_e$ (m^{-1})
Upper	2×10^6	2.82×10^6	4.00×10^{-10}	1.546×10^{-11}	3.5×10^5	1.085×10^4
Lower	0	2.42×10^6	4.03×10^{-10}	1.728×10^{-11}	2.9×10^5	8.689×10^3

at any instant of time. The magnitudes of ψ^2 at each of these position are given in Table 6 and Figures 13-15. The loci of the strongest maxima are shown by dashed curves.

The Condon loci have to be determined numerically by the following procedure. We start with some value of the upper vibrational quantum number υ'

and determine the corresponding term value $T' = V'/hc$ from Equation (17). We then determine the two values of the internuclear distance corresponding to this term value from the dashed loci in the upper state of Figure 16. Then determine the term values of the lower state corresponding to these internuclear distances from the dashed curves of the lower state. Finally calculate the corresponding values of υ'' from the converse of Equation (17).

In Figure 17 I show the principal Condon locus as a dashed curve. The full curve shows the "classical" locus for Morse (anharmonic) potential functions-*i.e.* the locus calculated on the assumption that the most likely internuclear separation at any given instant is that of full compression or full extension, rather than at the maxima of the eigenfunctions.

I have not shown any secondary loci, because the possibilities are almost endless, quantum numbers of the strong and weak bands being almost random. Strong bands occur wherever the internuclear separation corresponding to a maximum in the eigenfunction in the upper state corresponds with a maximum in the lower state at the same internuclear separation. In terms of Figure 16, and expressed more simply, there is a strong band wherever a dot in the upper part of Figure 16 is vertically above a dot in the lower part.

It will be understood from this that small differences in the shapes and positions of the potential curves (*i.e.* in the equilibrium internuclear distances and the vibrational constants) will result in differences in the positions of the dots

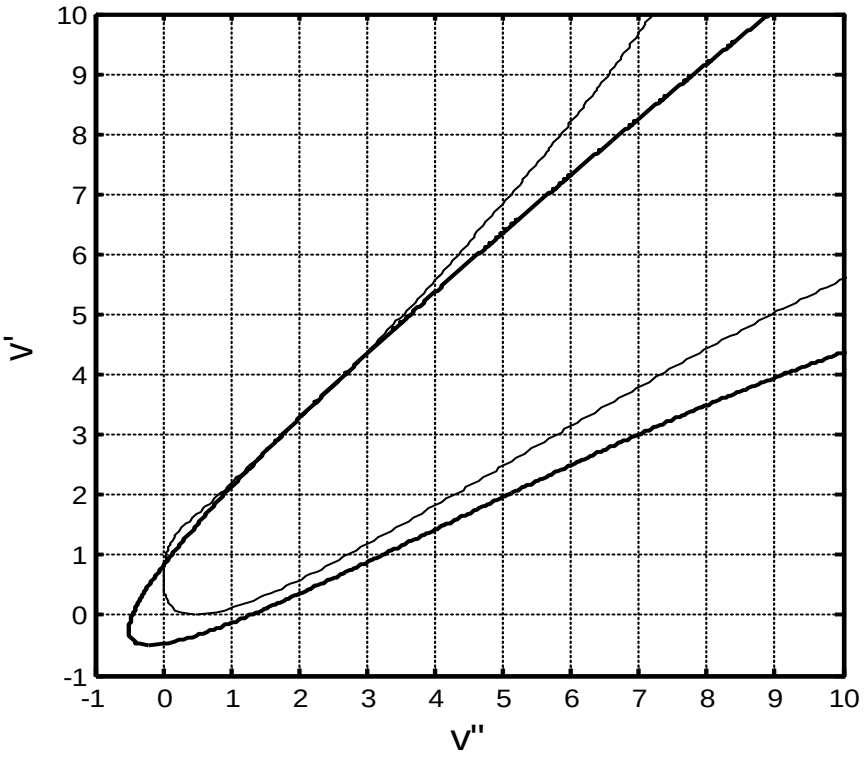


Figure 17. The heavy curve shows the Condon parabola calculated on the classical model for anharmonic oscillations. The lighter curve shows the principal Condon locus from wavemechanical calculations.

and hence in the designations of the bands that are likely to be strong or weak. It will also be understood that if the two potential curves are quite similar, secondary and tertiary Condon loci will be much more likely. Indeed, it could be advanced that, if a secondary Condon locus is very evident, it is likely that the internuclear separations of the two electronic states are not very different. Unfortunately in that case the entire electronic band system is likely to be weak because of the small difference in electric dipole moments of the two electronic states, and also, as noted in Section 2, the Condon locus is likely to be narrow and the strongest bands are those in which Δv is small.

6. Comparison with Observation

The thrust of this paper has been primarily theoretical. However, calculations of the predicted Condon loci, including in the simple harmonic case the latera recta of the Condon parabolas and the inclinations of their axes, have been carried out for 47 electronic band systems and compared with observations. The results for these are given in Hefferlin *et al.* [5].

Acknowledgements

I would like to thank Rev. Dr Ray Hefferlin of Southern Adventist University, Tennessee, for his encouragement and for exciting my interest in this problem during his continuing studies of the systematics of molecular spectroscopic constants.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Franck, J. (1925) *Transactions of the Faraday Society*, **21**, 536-542. https://doi.org/10.1039/tf9262100536

[2] Condon, E.U. (1928) *Physical Review*, **32**, 858-872. https://doi.org/10.1103/PhysRev.32.858

[3] Herzberg, G. (1950) Molecular Spectra and Molecular Structure. I. Spectra of Diatomic Molecules, Van Nostrand, Princeton, 194-204.

[4] Nicholls, R.W. (1982) *Journal of Quantitative Spectroscopy & Radiative Transfer*, **28**, 481-492. https://doi.org/10.1016/0022-4073(82)90014-0

[5] Hefferlin, R., Sackett, J. and Tatum, J.B. (2012) Systematics and Prediction in Franck-Condon Factors, in Quantum Systems in Chemistry and Physics. Springer, Dordrecht, 179-191. https://doi.org/10.1007/978-94-007-5297-9_8

[6] Loney, S.L. (1895) The Elements of Coordinate Geometry. MacMillan, London, 359-363.

[7] Morse, P.M. (1929) *Physical Review*, **A34**, 57-64. https://doi.org/10.1103/PhysRev.34.57

[8] Lennard-Jones, J.E. and Dent, B.M. (1926) *Proceedings of the Royal Society*, **A112**,

230-234. https://doi.org/10.1098/rspa.1926.0107

[9] Lim, T.-C. (2003) *Zeitschrift für Naturforschung*, **58a**, 615-617. https://doi.org/10.1515/zna-2003-1104

[10] Eyring, H., Walter, J. and Kimball, G.E. (1944) Quantum Chemistry. Wiley, New York, 75-79.

Journal of Modern Physics, 2018, 9, 1761-1772
http://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Correlation-Induced Electron-Electron Attraction

Hubert Klar

University of Freiburg, Freiburg, Germany (Retired)
Email: HubKlar@aol.com

How to cite this paper: Klar, H. (2018) Correlation-Induced Electron-Electron Attraction. *Journal of Modern Physics*, **9**, 1761-1772.
https://doi.org/10.4236/jmp.2018.99110

Received: June 17, 2018
Accepted: July 30, 2018
Published: August 2, 2018

Abstract

The multidimensional potential surface of a few (2 - 4)-electron atom shows near equilibrium configurations ridge structures. The electron wave diffraction from such a ridge is shown to induce a novel fictitious force which manifests itself as temporary electron-electron attraction. In contrast to a Cooper pair our pair does not need a lattice vibration and may be regarded as elementary prototype pair in an isolated atom. Also electron triples and quadruples are discussed.

Keywords

Cooper Pairs, Dominant Correlation, Wannier Theory, Multiple Excitation

1. Introduction

Electronic structure of atoms and molecules is often well understood within the shell model. Each electron moves in a potential field generated by all other electrons. Principal quantum numbers and angular momentum quantum numbers describe approximately the state of individual electrons.

That traditional picture, however, breaks down at high multiple excitation, particularly near thresholds of multiple escape. The reason for that breakdown is that the electrons move no longer in potential valleys. This may be seen in the following simple model.

To this end, we consider a two-electron-atom in a collinear configuration, the nucleus located between the electrons. The electron-electron distance may be

$$\rho = r_1 + r_2 \tag{1}$$

the r_i, $i = 1, 2$, being the electron-nucleus separations. The condition (1) may be satisfied by the parametrization

$$r_1 = \rho\varepsilon$$
$$r_2 = \rho(1-\varepsilon) \tag{2}$$

with 0 < ε < 1. The electrostatic potential of this atom with nuclear charge Z

$$V = -\frac{Z}{r_1} - \frac{Z}{r_2} + \frac{1}{\rho} \tag{3}$$

reads then as function of the parameters ε and ρ

$$V = \frac{1}{\rho}\left(-\frac{Z}{\varepsilon} - \frac{Z}{1-\varepsilon} + 1\right) \tag{4}$$

As long as $\varepsilon \approx 0$ or $\varepsilon \approx 1$ one electron is close the nucleus and experiences a Coulomb field. The electrons experience however no hydrogen-like force at all if

$$\frac{\mathrm{d}V}{\mathrm{d}\varepsilon} = 0 \tag{5}$$

which has the solution $\varepsilon = 1/2$. At that ε-value the electrons have equal distances from the nucleus and are trapped in an equilibrium configuration. The Taylor expansion of (4) at $\varepsilon = 1/2$ reads

$$V \approx -\frac{4Z-1}{\rho} - \frac{16Z}{\rho}\left(\varepsilon - \frac{1}{2}\right)^2 \tag{6}$$

Equation (6) identifies this equilibrium as unstable because $\frac{\partial^2 V}{\partial \varepsilon^2} < 0$. Along the coordinate ε the potential shows an antioscillator, in the (ρ, ε)-scattering plane (6) represents a ridge. The importance of potential ridges for electronic excitation has been stressed recently [1].

The above model of a synthetic atom is not entirely unrealistic. Wannier [2] has studied the threshold ionization of H by electron impact. His model has been as follows. He assumes that the projectile electron transfers its kinetic energy onto the target electron. We are then confronted with two extremely slow electrons in the field of the nucleus. We meet here exactly the situation described above; the projectile electron moves near a potential ridge. Within the framework of classical Lagrange equations Wannier extracted from the decay of this negative complex the slope of the total ionization cross section. He found a power law with fractional exponent,

$$\sigma \propto \left(E - E_0\right)^{1.127\cdots} \tag{7}$$

where the fractional exponent stems from the curvature of the potential surface along the unstable equilibrium.

The present paper goes fare beyond the classical Wannier work, and also beyond synthetic models. We correctly employ quantum theory to a real atom, and intend to investigate the electrons motion near unstable equilibria.

The paper is organized as follows. In Section 2, we discuss the choice of suitable coordinates for two active electrons. Coordinates are suitable if the ridge structure of the potential can easily be described, and if the expression of the ki-

netic energy is not too difficult. We have here to fight against two mathematical problems. 1) In consequence of correlation three/more-body Coulomb problems are not separable in any coordinate system. 2) Even if correlation is properly taken into account by suitable coordinates we arrive at non-adiabatic couplings [3] [4]. The present paper presents a coordinate frame which treats properly correlation near thresholds for multiple ionization, and eliminates at least in first-order non-adiabatic coupling. This repairs a fundamental short-coming of the familiar Born-Oppenheimer method.

The method mentioned above has been recently outlined in a broader context [1]. Section 3 presents a detailed description in the example of two active electrons. Section 4 compares our quantum treatment with the corresponding classical one. It is usually believed that Wannier's cross section Formula (7) is an exact result although Wannier's paper has nothing to do with quantum theory. Section 5 finally shows that the phenomenon of wave diffraction from a ridge is a classical as well as a quantum effect. Section 6 discusses the formation of few-electron complexes like pairs and triples. Section 7 stresses breathing motions of electron charge clouds in contrast to Bohr orbits in an atom.

2. Alternative Coordinates for 2-Electron Atoms

Single electron coordinates are certainly not suitable in the spectral range of high multiple excitation where correlation is dominant. We start with the trivial remark that a three-body system (nucleus + 2 electrons) needs in the centre-of mass system 6 space coordinates. We do only a small mistake if we put the centre of mass into the nucleus. Three of the six coordinates may be used to describe the orientation of the whole atom in space. Euler angles serve for that. We put therefore

$$\boldsymbol{r}_i = D(\alpha,\beta,\gamma)\boldsymbol{b}_i \tag{8}$$

i = 1, 2. Here the vectors $\boldsymbol{r}_i$ are space-fixed electron positions, the $\boldsymbol{b}_i$ are the body-fixed coordinates, and the $D(\ldots)$ is an orthogonal rotation matrix depending on Euler angles. For S-states treated below we have $\boldsymbol{r}_i = \boldsymbol{b}_i$.

Hyperspherical coordinates are most convenient in situations where correlation is important or even dominant. One coordinate is the hyperradius given by

$$R = \sqrt{r_1^2 + r_2^2} \tag{9}$$

where r_i (i = 1, 2) are the electron nucleus separations.. Two body-fixed hyperspherical angles may be expressed in terms of moments of inertia. We denote the principal axes by X, Y ynd Z such that X and Y ly in the particles plane. Previous papers by the author [3] have employed the inertia asymmetry given by

$$\psi = \frac{1}{2}\cos^{-1}\frac{I_{yy} - I_{xx}}{I_{zz}} \tag{10}$$

as one angle where the principal axes system has been oriented according to $I_{xx} \le I \le I_{zz}$ for the principal moments. The value $\psi = 0$ describes then all col-

linear particle configurations. Radial correlation we describe following Sommerfeld [5] by the angle

$$\varphi = \tan^{-1} \frac{r_1^2 - r_2^2}{4\boldsymbol{r}_1 \cdot \boldsymbol{r}_2} \tag{11}$$

which basically is the difference of single electron moments of inertia.

The key point is now that the inertia sphere is spanned by the vector

$$\boldsymbol{c} = R \begin{pmatrix} \cos 2\psi \cos\varphi \\ \cos 2\psi \sin\varphi \\ \sin 2\psi \end{pmatrix} \tag{12}$$

i.e. the angle ψ $\left(0 \le \psi \le \pi/4\right)$ is a latitude spherical angle[1] and φ $\left(0 \le \varphi \le 2\pi\right)$ is the corresponding azimuth. In consequence of that, φ is a cyclic coordinate in the kinetic energy which makes the integration of the wave equation relatively simple. Alternative hyperspherical angles like

$$\alpha = \tan^{-1} \frac{r_1}{r_2}$$

do not show that convenient property. In terms of our inertia coordinates the study of radial correlation in the collinear configuration reduces exactly to a 2-dimensional problem in the equatorial plane of the inertia sphere.

The electrostatic potential of the system electron-nucleus-electron

$$V = -\frac{Z}{r_1} - \frac{Z}{r_2} + \frac{1}{r_{12}} \tag{13}$$

where r_{12} is the electron-electron separation and Z being the nuclear charge shows a saddle point located at $\psi = 0$ and $\varphi = \pi$. The potential (13) has in that saddle point the following Taylor expansion

$$V = \frac{1}{R}\left[-C_0 + \frac{C_1}{2}\psi^2 - \frac{C_2}{2}\left(\varphi - \pi\right)^2\right] \tag{14}$$

the coefficients given by

$$\begin{aligned} C_0 &= \frac{4Z-1}{\sqrt{2}} \\ C_1 &= \frac{1}{\sqrt{2}} \\ C_2 &= \frac{12Z-1}{4\sqrt{2}} \end{aligned} \tag{15}$$

We fall to the collinear configuration by fixing $\psi = 0$.

The expansion (14) has been obtained as follows. We express the potential (13) in terms of the coordinates given in (18) and perform a Taylor expansion in the two variables φ and ψ.

The Laplacian reads in our coordinates

$$\Delta = R^{-5} \frac{\partial}{\partial R} R^5 \frac{\partial}{\partial R} - \frac{\Lambda^2}{R^2} \tag{16}$$

[1]Mathematians usually count the latitude by the angle $\vartheta = \pi/2 - 2\psi$.

with the grand angular momentum in the collinear configuration given by

$$\Lambda^2 = 4\frac{\partial^2}{\partial\varphi^2} \tag{17}$$

Equation (17) confirms that φ is a cyclic coordinate.

These coordinates (R, ψ, φ) are fare from new. They were already used long ago by Sommerfeld [5], see also [6]. The single electron positions read in these body-fixed inertia coordinates

$$\begin{aligned} \boldsymbol{b}_1 &= R\begin{pmatrix} \cos\psi\cos\frac{1}{2}\left(\varphi+\frac{3\pi}{2}\right) \\ \sin\psi\sin\frac{1}{2}\left(\kappa+\frac{3\pi}{2}\right) \\ 0 \end{pmatrix} \\ \boldsymbol{b}_2 &= R\begin{pmatrix} \cos\psi\cos\frac{1}{2}\left(\varphi-\frac{3\pi}{2}\right) \\ \sin\psi\sin\frac{1}{2}\left(\varphi-\frac{3\pi}{2}\right) \\ 0 \end{pmatrix} \end{aligned} \tag{18}$$

We are now arrived probably at the most convenient way to study an electron motion near a potential ridge. The ridge occurs in the (R,φ) plane. The curvature across the ridge is $-C_2/2R$. The potential on the top of the ridge reads $-C_0/R$. In the ridge region we arrive at the following reduced stationary wave equation.

$$\left[-\frac{1}{2}\frac{\partial^2}{\partial R^2}-\frac{4}{2R^2}\frac{\partial^2}{\partial\varphi^2}+\frac{15}{8R^2}-\frac{C_0}{R}-\frac{C_2}{2R}(\varphi-\pi)^2-E\right]\left(R^{\frac{5}{2}}\Psi\right)=0 \tag{19}$$

where the centrifugal term emerges from the elimination of the first derivative $\frac{\partial}{\partial R}$.

This is a convenient way to start the diffraction study of an electron from a potential ridge. (19) describes a two-dimensional wave propagation in the (R, φ) plane. Geometrically we are confronted with the diffraction of an electron from a potential ridge. Correlation has been properly taken into account by the use of the above inertia coordinates. But we have paid a high price for that achievement since (19) is not separable.

3. Non-Adiabatic Solution of the Wave Equation (19)

The wave equation for two-electron atoms in hyperspherical coordinates has been investigated using an adiabatic approach by Macek [3] and by Klar *et al.* [4], *i.e.* we treat at first the angular motion at fixed values of R. That step generates a potential which controls the radial motion. Macek [3] has used these angles (α,Θ) for that, Klar *et al.* [4] did it using the convenient angles $(\psi.\varphi)$.

Although the latter method has led to some computational advantages, but both methods failed at high double excitation where correlation becomes dominant.

Nevertheless, these methods are in principle exact provided we could go beyond the adiabatic approximation. That would however require the treatment of an infinity of coupled differential equations. A truncated set would not solve the problem.

For the less experienced reader we demonstrate how an adiabatic solution of (19) looks like. We determine at first a socalled channel function, *i.e.* the angular eigenfunction at constant values of R. The channel equation

$$-\frac{2}{R^2}\frac{\partial^2\Phi}{\partial\varphi^2}-\frac{C_2}{2R}\left(\varphi-\pi\right)^2\Phi=U\Phi \tag{20}$$

has the eigenfunction

$$\Phi_{\text{adiabatic}}=\exp\left(i\kappa\left(\varphi-\pi\right)^2\right) \tag{21}$$

to which belongs the wave number

$$\kappa_{\text{adiabatic}}=\frac{1}{4}\sqrt{C_2R} \tag{22}$$

and the eigenvalue

$$U_{\text{adiabatic}}-iR^{-3/2}\sqrt{C_2} \tag{23}$$

It is evident that (20) is not Hermitian, and has therefore no real eigenvalues. Equation (22) indicates where the difficulty is located. We observe that the wave number κ depends explicitely on the coordinate R. Therefore the derivative of the channel function $\frac{\partial\Phi}{\partial R}$ must be taken into account in the radial wave equation.

A recent paper [1] has shown how to go beyond the above shortly outlined adiabatic approach. We still rely on a product form of the wave function,

$$\Psi\left(r,\varphi\right)=R^{5/2}F\left(R\right)\Phi\left(\varphi;R\right) \tag{24}$$

Because all potential terms, see (14), scale smoothly with $1/R$, and also because Wannier's threshold ionization law is a classical result, we employ for the radial function an Eikonal Ansatz. An improved channel function may still have the structure (21) except thar the wavenumber must be properly determined. It was pointed out by Wannier [2] that near threshold $E=0$ we have to distinguish two different regions of the hyperradius R. The Coulomb zone extends from the core radius R_0 to infinity and the reaction zone with $0\le R\le R_0$ where the core is limited by the centrifugal barrier, see (19).

To this end we try to find an improved solution of (19) with help of the Ansatz

$$R^{-5/2}\Psi=\exp\left(i\int^R K\left(R'\right)\mathrm{d}R'\right)\times\exp\left(i\kappa\left(\varphi-\pi\right)^2\right) \tag{25}$$

We take powers of φ up to the quadratic term $\left(\varphi-\pi\right)^2$ into account but dis-

regard higher order power terms. To this end we arrive at two conditions for the wave numbers

$$K^2 = 2\left(E - \frac{15}{8R^2} + \frac{C_0}{R} + \frac{i}{2}\frac{\mathrm{d}K}{\mathrm{d}R} + \frac{4i\kappa}{R^2} \right) \tag{26}$$

$$\frac{8}{R^2}\kappa^2 = \frac{C_2}{2R} - K\frac{\mathrm{d}\kappa}{\mathrm{d}R} + \frac{i}{2}\frac{\mathrm{d}^2\kappa}{\mathrm{d}R^2} \tag{27}$$

Equations (26) and (27) constitute a set of two coupled Riccati equations. This is an exact result within our frame of taking into account terms up to φ^2 in the wave equation. We stress that we have taken into account the first and the second derivative with respect to R of the angular function $\exp\left(i\kappa(\varphi-\pi)^2\right)$.

Near threshold ($E \cong 0$) we can do a big simplification. Since all potential terms scale as $1/R$ we are close to the classical limit of quantum mechanics. For that reason we had already employed an Eikonal approach in (25). Remember that the Eikonal approximation is exact for Coulomb potentials. Within a WKB spirit we are then allowed to omit the derivative $\frac{\mathrm{d}K}{\mathrm{d}R}$ in (26). Finally, for our purpose (27) must be solved only in the Coulomb zone which extends at threshold to infinity $R = \infty$. Then we are allowed to omit the second derivative of κ in (27).

To this end we arrive at the simpler equations

$$K^2 = 2\left(E - \frac{15}{8R^2} + \frac{C_0}{R} + \frac{4i\kappa}{R^2} \right) \tag{28}$$

$$\frac{8\kappa^2}{R^2} = \frac{C_2}{2R} - K\frac{\mathrm{d}\kappa}{\mathrm{d}R} \tag{29}$$

recently presented in [1].

4. Comparison with Classical Mechanics

We investigate here the corresponding stationary Hamilton-Jacobi-equation for the same problem, *i.e.*

$$\frac{1}{2}\left(\frac{\mathrm{d}S}{\mathrm{d}R}\right)^2 + \frac{1}{2R^2}\left(\frac{\mathrm{d}S}{\mathrm{d}\varphi}\right)^2 - \frac{C_0}{R} - \frac{C_2}{2R}(\varphi-\pi)^2 = E \tag{30}$$

and put for the action

$$S = KR + \kappa(\varphi-\pi)^2 \tag{31}$$

As above we allow the wave numbers K and κ to depend on the coordinate R, *i.e.* $K = K(R)$ and $\kappa = \kappa(R)$; but we disregard the derivative $\frac{\mathrm{d}K}{\mathrm{d}R}$ and reject powers higher than $(\varphi-\pi)^2$. That leads us to the simple equation

$$K^2 = 2\left(E + \frac{C_0}{R} \right) = 0 \tag{32}$$

being different from the above (28). We conclude that both the centrifugal barrier and the imaginary part in (28) are manifestly quantum contributions. In

contrast to that observation the classical equation for the wavenumber κ is the same as above (29).

The derivation of Wannier's threshold ionization law has used the quantum Equations (28) and (29) in the Coulomb zone, *i.e.* it was sufficient to take into account contributions of the order $1/R$. Therefore the centrifugal term $\propto 1/R^2$ did not contribute. The **Appendix** of this paper shows $\kappa \propto \sqrt{R}$; therefore in the Coulomb zone we get $\frac{\kappa}{R^2} \propto \frac{1}{R^{\frac{3}{2}}}$, again negligible in the fare Coulomb zone. We conclude that Wannier has been correctly employed classical mechanics for the threshold law. However, the above conclusions are not suffficient for the formation and decay of electron pairs. The quantum terms play then key roles as will be shown in Section 5.

5. Formation and Decay of Electron Pairs

Our development given in Section 3 deserves several comments regarding the pair production of electrons. We start our analysis with Equation (29). Multiplied with the angular term $(\varphi-\pi)^2$ we observe that the Coulomb potential term [see (14)] has been replaced according to

$$\frac{C_2}{2R}(\varphi-\pi)^2 \Rightarrow \frac{C_2}{2R}(\varphi-\pi)^2 - K\frac{\mathrm{d}\kappa}{\mathrm{d}R}(\varphi-\pi)^2 \tag{33}$$

Equation (33) may be regarded as gap in the potential energy, *i.e.* the potential ridge has been pressed down. An alternative statement claims that the static potential surface has been deformed. Note that the potential deformation causes a force given by

$$\boldsymbol{F} = \nabla\left(K\frac{\mathrm{d}\kappa}{\mathrm{d}R}(\kappa-\pi)^2\right) \tag{34}$$

whose radial component reads

$$F_R = \frac{\partial}{\partial R}\left(K\frac{\mathrm{d}\kappa}{\mathrm{d}R}\right)(\varphi-\pi)^2 \tag{35}$$

and its angular component is given by

$$F_\varphi = \frac{1}{R}\frac{\partial}{\partial\varphi}\left(K\frac{\mathrm{d}\kappa}{\mathrm{d}R}(\varphi-\pi)^2\right) = \frac{2}{R}K\frac{\mathrm{d}\kappa}{\mathrm{d}R}(\varphi-\pi) \tag{36}$$

That angular force, obviously a momentum dependent fictitious force, is directed across the ridge, and resembles formally to a Lorentz force for a charged particle in a magnetic field given by

$$\boldsymbol{F} = Q[\boldsymbol{v}\times\boldsymbol{B}] \tag{37}$$

The similarity becomes immediately visible because (36) may be rewritten in the vectorial form

$$\boldsymbol{F}_\varphi = \frac{2}{R}\frac{\mathrm{d}\kappa}{\mathrm{d}R}(\varphi-\pi)[\boldsymbol{K}_\parallel\times\hat{n}] \tag{38}$$

where $\boldsymbol{K}_\parallel$ is the component of the radial pair momentum parallel to the ridge

and $\hat{n}$ is a unit vector orthogonal to the scattering plane. Thus, our radial pair momentum corresponds to the single particle velocity in (37), the normal vector $\hat{n}$ corresponds to the magnetic field $\boldsymbol{B}$, and the factor in front of the vector product in (38) corresponds to the charge Q in (37).

From (38) it is evident that the fictitious force changes sign if the radial momentum $\boldsymbol{K}_{\|}$ does. In consequence of that an incoming wave front is directed towards to the top of the ridge, whereas an outgoing wave mode is turned away from the ridge top. This compares favorably with Wannier's [2] finding of two classical trajectories, a converging one and a diverging one. He did not pay further attention to the converging one because he needed only the diverging one for his threshold law. We conclude that the angular force component is the key tool for electronic excitation, in the present study for the pair production.

We are now able to describe the whole process of pair production. When the incoming electron experiences the ridge its curvature is reduced according the Riccati Equation (29). The target electron as well as the projectile electron are pushed towards the top of the ridge, *i.e.* the electron-electron attraction due to the fictitious force works. Arrived on the top of the ridge the static attraction given by the potential on the ridge $-C_0/R$ attracts the pair as a whole towards the nucleus. The pair runs against the centrifugal barrier given by $15/8R^2$, see (19) or (26), where reflection occurs, and a collaps of the negative ion has been avoided. The pair is now reflected into the outgoing wave mode. The fictitious force has now changed sign, *i.e.* the electron-electron interaction has now become repulsive. The pair decays, one electron falls back to a Rydberg orbital whereas the other one escapes. Above threshold both electrons escape. This latter case was treated by Wannier [2]. From highly accurate experimental data we know that the range of validity of the threshold ionization law is a few eV. From that we conclude a lifetime of the pair of the order of $\tau \approx 10^{-15}$ sec.

6. Outlook

We have investigated also other atoms with more than two active electrons for equilibrium configurations. To this end we generalize the hyperradius to more than two electrons,

$$R = \sqrt{\sum r_i^2} \tag{39}$$

and study the total electrostatic potential restricted to a hypersphere of radius R for stationary points, *i.e.*

$$W = -\sum_i \frac{Z}{r_i} + \sum_{i<j} \frac{1}{r_{ij}} + \lambda\left(R^2 - \sum_i r_i^2\right) \tag{40}$$

must become stationary.

Above the indice $i, j \in [1, \cdots, N]$; N being the number of active electrons. We have found the following configurations [7]:

1) $N = 2$ corresponds to the pair formation treated above. The electrons sit the ends *of* a dumbbell.

2) $N = 3$. The electrons sit in the corners of an equilateral triangle,; the nucleus i*s* located at the electronic centre of charge [7].

3) $N = 4$. The electrons sit at the corners of a tetrahedron; the nucleus is again in the centre of electronic charge [8].

4) $N > 4$. No equilibrium configurations seem to be known.

Analogous to the above treatment of two electrons we can find hypersherical coordinates such that one angle describes the fragmentation symbolically referred to as $N \to (N-1)+1$, and with help of an eventual rotation we can arrive at a cyclic angle in the Laplacian along the hyperspherical equator. We arrive then at a wave equation like (19) except that the constants C_0 and C_2 depend on the number of equilibrium electrons.

Applied to the example of three or four active electrons we predict beyond pairs now triples and quadruples. A prototype of a triple configuration may be generated for instance from an alkali earth atom with two valence electrons in a $(ns^2)S$ state. We hit that target with an electron to the threshold of double escape of the target (usually a few eV). In that energy range the projectile electron is attracted by the target electron pair; the three electrons enter into the equilateral triangle configuration. We stress that the electrons do not move on Bohr orbits but the triangle shrinking in size is electrostatically attracted by the nucleus. That motion occurs on the top of the generalized ridge due to the Coulomb potential $-C_0/R$. The triple runs now against the generalized centrifugal barrier now given by $6/R^2$, see [1]. The triple returns now under the influence of the repulsive fictitious force, and decays. One electron escapes whereas the other two electrons become trapped into a double Rydberg state of the target atom.

In both cases of complexes, pair or triple, the emitted electron from the complex decay may hit another target atom and creates there a further complex. In this way a complex may travel through a solid without suffering from inelastic collisions; *i.e.* the complex transports a current without any resistance.

The description of small complexes runs quite generally along the same lines outlined above except that the constants C_0 and C_2 have different numerical values for pairs, triples and quadruples. Note however, one difference may not be overlooked: complexes with an odd number of electrons are fermions whereas an even number of electrons allows only for bosons.

7. Conclusions

Atomic structure has been traditionally described within the shell model, *i.e.* each electron moves in a static potential generated by all other electrons moving on circular orbits. The motion takes place in attractive potential valleys. Each electron can be described by a principal quantum number and angular momentum quantum numbers.

That model based on hydrogen-like atoms becomes shaky at multiple excitation and breaks entirely down near thresholds of multiple escape. In that spectral range several electrons move dominantly correlated, and form pairs, triples

and/or quadruples as we have seen in this paper. The electrons move no longer on approximately circular orbits, but a cloud of highly excited electrons performs a breathing motion caused by diffraction of an electron wave from a potential ridge. The breathing is an entirely novel and unexpected electron motion in an atom. In the spectral range near a threshold of multiple escape a resonance would be located due to the long-range Coulomb interaction in an infinity of one-electron continua. The Fano resonance theory [9] does not apply to that crucial situation. Actually the embedding into a huge number of continua produces the new kind of motion as described here. The main result of the present study has been that quite generally the particle wave diffraction from a potential ridge induces a fictitious force between the constituents of the system under consideration. Moreover, that force may be attractive or repulsive depending on the mode of motion. A few-electron atom is perhaps one of the simplest examples for that situation. We believe that potential surfaces of big molecules show multidimensional ridge structures leading to similar surprising effects.

Acknowledgements

Financial support during the start of this work by DFG under contract SFB 276 is gratefully acknowledged.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Klar, H. (2018) *Physical Review Letters*, **120**, 053401. https://doi.org/10.1103/PhysRevLett.120.053401

[2] Wannier, G. (1953) *Physical Review*, **90**, 817. https://doi.org/10.1103/PhysRev.90.817

[3] Macek, J.H. (1968) *Journal of Physics*, **B1**, 831. https://doi.org/10.1088/0022-3700/1/5/309

[4] Klar, M. and Klar, H. (1980) *Journal of Physics*, **b13**, 1057. https://doi.org/10.1088/0022-3700/13/6/014

[5] Sommerfeld, A. (1944) Atombau und Spektrallinien (Vieweg Braunschweig). Vol. 2, 627ff.

[6] Klar, H. (2017) *Journal of Modern Physics*, **8**, 1029. https://doi.org/10.4236/jmp.2017.87065

[7] Schlecht, W. and Klar, H. (1076) *Journal of Physics*, **B9**, 1699.

[8] Schlecht, W. (1975) Thesis University Freiburg.

[9] Fano, U. (1967) *Physical Review*, **124**, 1866. https://doi.org/10.1103/PhysRev.124.1866

Appendix

We show here how to solve Equations (28) and (29) at threshold $E = 0$.

We need a solution only in the Coulomb zone which extends at zero energy up to infinity $R = \infty$. To this end we get from (28)

$$K = \sqrt{\frac{C_0}{R}} \tag{A1}$$

To solve (29) we put $\kappa = \bar{\kappa}\sqrt{R}$ and get after multiplication with R the algebraic equation

$$8\bar{\kappa}^2 = \frac{C_2}{2} - \frac{\sqrt{C_0}}{2}\bar{\kappa} \tag{A2}$$

with the solutions

$$\bar{\kappa}_{1,2} = -\frac{\sqrt{C_0}}{32} \pm \frac{\sqrt{C_0 + 64C_2}}{32} \tag{A3}$$

This solution justifies that we have disregarded the imaginary portions in (27) and (28) since $\frac{\mathrm{d}^2\kappa}{\mathrm{d}R^2} \propto R^{-3/2}$ and $\frac{\kappa}{R^2} \propto R^{-3/2}$ are negligible in the fare Coulomb zone.

Journal of Modern Physics, 2018, 9, 1773-1784
http://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Predictions of Electronic, Transport, and Structural Properties of Magnesium Sulfide (MgS) in the Rocksalt Structure

Uttam Bhandari[1,2], Cheick Oumar Bamba[1], Yuriy Malozovsky[1], La Shounda Franklin[1], Diola Bagayoko[1]

[1]Department of Mathematics and Physics, Southern University and A & M College, Baton Rouge, LA, USA
[2]Department of Physics and Astronomy, Texas Tech University, Lubbock, TX, USA
Email: bagayoko@aol.com

How to cite this paper: Bhandari, U., Bamba, C.O., Malozovsky, Y., Franklin, L.S. and Bagayoko, D. (2018) Predictions of Electronic, Transport, and Structural Properties of Magnesium Sulfide (MgS) in the Rocksalt Structure. *Journal of Modern Physics*, **9**, 1773-1784.
https://doi.org/10.4236/jmp.2018.99111

Received: June 28, 2018
Accepted: August 5, 2018
Published: August 8, 2018

Abstract

We report results from *ab-initio*, self-consistent density functional theory (DFT) calculations of electronic, transport and bulk properties of rock salt magnesium sulfide (MgS). In the absence of experimental data on these properties, except for the bulk modulus, these results are predictions. Our calculations utilized the Ceperley and Alder local density approximation (LDA) potential and the linear combination of Gaussian orbitals (LCGO). The key difference between our computations and other previous *ab-initio* DFT ones stems from our use of successively larger basis sets, in consecutive, self-consistent calculations, to attain the ground state of the material. We predicted an indirect (Γ-X) band gap of 3.278 eV for a room temperature lattice constant of 5.200Å. We obtained a predicted low temperature indirect (Γ-X) band gap of 3.512 eV, using the equilibrium lattice constant of 5.183Å. We found a theoretical value of 79.76 GPa for the bulk modulus; it agrees very well with the experimental finding of 78 ± 3.7 GPa.

Keywords

Density Functional Theory (DFT), Local Density Approximation (LDA), Linear Combination of Atomic Orbitals (LCAO), Band Gap, Band Structure, The Bagayoko, Zhao and Williams (BZW) Method

1. Background and the Motivation for This Work

Magnesium sulfide (MgS) belongs to the group of alkaline earth sulfides. The interests of scientists in alkaline earth sulfides have been increasing due to their potential applications in multicolor thin-films electroluminescent and magne-

to-optical devices [1]. Due to its large band gap, MgS has wide applications in blue-light emitting diodes and optical storage devices [2] [3]. It is a favorable material for active solar blind UV detection [4]. It is also useful in lithium-ion batteries as electrode material [5]. MgS exists in three structures (rock salt, zinc blende, and wurtzite), with the rock salt structure as the most stable one. Many theoretical [1] [6]-[20] studies have been carried out on MgS. **Table 1** below lists both calculated and experimental band gap values for MgS. The Hartree-Fock method has been applied by Pandey *et al.* [1] to compute the electronic band structure. Their study found an indirect band gap value of 6.48 eV. Drief *et al.* [6] reported first principle calculations for rock salt MgS, employing a local density functional approximation (LDA) potential and the full potential linearized augmented plane wave (FP-LAPW) method. They predicted an indirect band gap of 2.208 eV. There are several other calculated, [7] [8] [9] [10] [11], indirect band gap values of rock salt MgS, obtained with *ab-initio* LDA potentials. Five (5) of these values were within the range of 2.5 to 2.7 eV. Recently, Tairi *et al.* [12] studied the structural and electronic properties of rock salt MgS by using

Table 1. Theoretical band gaps of rock salt magnesium sulfide (MgS) in the literature.

Computational formalism and method	Potentials (DFT and others)	Band gap, E_g (eV)
Full potential-linearized augmented plane wave (FP-LAPW)	LDA	2.208 (indirect)[a]
Full potential linear muffintin orbitals (FP-LMTO)	LDA	2.657 (indirect)[b]
Linear augmented plane wave (LAPW)	LDA	2.69 (indirect)[c]
Plane-wave pseudo potential approach	LDA	2.56 (indirect)[d]
FP-LAPW	LDA	2.6 (indirect)[e]
Pseudo potential plane wave basis	LDA	2.65 (indirect)[f]
FP-LAPW	GGA	2.745 (indirect)[g]
FP-LAPW	GGA	2.794 (indirect)[h]
Pseudo-potential plane wave	GGA	2.76 (indirect)[i]
Plane-wave pseudo potential approach	GGA	2.76 (indirect)[j]
FP-LAPW	GGA	2.64 (indirect)[k]
Tight binding LMTO (TB-LMTO)	LDA	2.7 (indirect)[l]
OLCAO	LDA	4.59 (indirect)[m]
Projector augmented wave (PAW)	TB-mBJ	4.00 (indirect)[n]
FP-LAPW	mBJ	4.248[h]
PAW	GW	4.80 (indirect)[o]
Pseudo potential plane wave basis	GW	4.15[f]
Hartree-Fock method		6.48 (indirect)[p]

[a]Reference [6]; [b]Reference [7]; [c]Reference [8]; [d]Reference [9]; [e]Reference [10]; [f]Reference [11]; [g]Reference [12]; [h]Reference [13]; [i]Reference [14]; [j]Reference [15]; [k]Reference [16]; [l]Reference [17]; [m]Reference [18]; [n]Reference [19]; [o]Reference [20]; [p]Reference [1].

the full potential linearized augmented plane wave method. They obtained an indirect band gap of 2.745 eV and 4.248 eV with a generalized gradient approximation (GGA) and mBJ potentials, respectively. Several other groups [13] [14] [15] [16] have studied the properties of rock salt MgS by utilizing *ab-initio* GGA potentials. Their calculated band gap values range from 2.6 - 2.79 eV. An indirect band gap of 2.7 eV was found by Kalpana *et al.* [17] utilizing tight binding linear muffin tin orbitals (LMTO), with GGA potentials. Ching *et al.* [18] calculated the band structure of MgS by applying the self-consistent orthogonalized linear combination of atomic orbital (OLCAO) method. Their calculated, indirect band gap was 4.59 eV. The indirect band gap range of MgS was found to be 4.0 - 4.80 eV, [11] [12] [19] [20] using hybrid potentials.

Experimental investigations performed on this material are very limited. Taleatu *et al.* [21] experimentally determined a direct band gap for MgS thin films (3.54 eV, 3.73 eV, and 3.14 eV) using X-ray diffraction technique. They have suggested that there is difficulty in the determination of the nature and value of the band gap of MgS. They noted difficulties stemming from the spontaneous oxidation of MgS in air and its reactivity with water.

The disagreement among theoretical results indicates the need for further investigations of the band gap and related properties of MgS. The above disagreement and current and potential applications of MgS motivated our *ab-initio*, self-consistent calculations of electronic, transport, and bulk properties of MgS. Our computational method has been successful in correctly describing and predicting properties of semiconductors in the past [22]. Hence, this work is expected to provide an accurate DFT description of MgS, as most previous, *ab-initio* DFT calculations are presumed to have underestimated the band gap of MgS. Further, in the case of MgS, a reliable experimental value of the band gap of bulk MgS is not available. Indeed, the experimental values noted above pertain to thin films and are subject to significant uncertainties, according to the authors. We describe below our method that led to highly accurate results in dozens of previous *ab-initio* DFT calculations.

2. Computational Method

We recall the fundamental theorems of DFT, along with the correct understanding of them, in order to facilitate the discussions on our method. For a system of N electrons in a box, subject to an external potential v(**R**), the first theorem states that v(**R**) is a unique functional of the charge density n(**R**), except for an additive constant. A corollary of this theorem is that the energy content of the Hamiltonian $E(\Psi) = \langle \Psi | H | \Psi \rangle$ is a unique functional of the charge density n(**R**). This energy content of the Hamiltonian is simply the sum of the occupied energies. Hence, a second corollary of the first theorem is that the spectrum of the Hamiltonian is a unique functional of the charge density [22].

The second DFT theorem states that $E(\Psi') = \langle \Psi' | H | \Psi' \rangle$ reaches its minimum if $\Psi' \equiv \Psi$ of the correct ground state, for arbitrary variations of Ψ' that keep the

number of particles constant [22]. Specifically, the second theorem means that $E(\Psi) > E(\Psi)$, *where* Ψ *is for the ground state, for all* $\Psi \neq \Psi$. For over 50 years, the community at large understood that the minimization of $E(\Psi')$, using a single basis set Ψ', provides the DFT solution for the system under study. *It does not.* The minimization resulting from self-consistency with a single basis set only leads to a stationary solution among an infinite number of stationary solutions. Indeed, every reasonable basis set (without linear dependency) leads to a stationary solution upon the attainment of self-consistency. The generalized minimization unveiled by Bagayoko [22] is required to attain, verifiably, the ground state of the system. This generalized minimization entails several self-consistent calculations with successively augmented basis sets until the occupied energies totally cease to decrease. Then, and only then, do we have the ground state of the system, and hence, the DFT description of the material. The first basis set to produce this description, indubitably, generates the ground state charge density upon reaching self-consistency.

Our calculations employed a method that ensures the attainment of the ground state of a material without the use of very large basis sets that are over-complete for the description of the ground state [22]-[31]. The Ceperley and Alder [32] LDA potential was parameterized by Vosko *et al.* [33]. We used the linear combination of atomic orbitals (LCAO) and the Bagayoko, Zhao, and Williams (BZW) method, [23] [24] [25] [29] as enhanced by Ekuma and Franklin (BZW-EF) [28]. We employed the electronic structure package developed at the US Department of Energy Laboratory in Ames, Iowa [34] [35].

Our calculations started with a basis set slightly larger, by one orbital, than the minimum basis set. We performed self-consistent Calculation I with this small basis set. The second self-consistent calculation was done with a basis set that consists of the first basis set, from Calculation I, augmented by one orbital corresponding to an excited state in the ionic species of the system. A comparison of the occupied energies from the two calculations shows that the occupied energies from Calculation II were lower than or equal to their corresponding values from Calculation I. This comparison follows the setting of the Fermi level to zero in both sets of eigenvalues. This lowering of occupied energies, from their values from Calculation I, indicates that the basis set used in Calculation I does not lead to the ground state. After augmenting the basis set of Calculation II with one orbital, we performed Calculation III and compared the resulting occupied energies with those from Calculation II. The successive, self-consistent calculations, with augmented basis sets, were stopped only after three consecutive ones led to the same occupied energies.

This situation signifies that we reached the lowest possible values of the occupied energies, *i.e.*, the ground state of the material, as required by the second DFT theorem.

The above generalized minimization of the energy, one that is far more than the minimization resulting from self-consistent iterations with a single basis set, is required by the second DFT theorem [22]. Calculations with augmented basis

sets larger than that of the third of the referenced three (3) calculations, also led to the same occupied energies. While the above three (3) successive calculations, and others with larger, augmented basis sets, led to the same occupied energies, they generally produced some different (*i.e.*, lower), unoccupied energies, including some of the lowest ones. Generally, the larger the basis set, the lower were some unoccupied energies as compared to their counterparts obtained with the first of the calculations leading to the ground state. Clearly, the question arises as to which of the calculations producing the ground state provides the DFT description of the material. Bagayoko [22] showed, using a corollary of the first DFT theorem, that the first of the three (3) consecutive calculations leading to the ground state, with the smallest basis set, is the one producing the DFT description of the material. The basis set of this calculation is referred to as the *optimal basis set*. Indeed, the corollary simply states that the spectrum of the Hamiltonian is, like the external potential, a unique functional of the charge density. Thus, the occupied energy content of the Hamiltonian is a unique functional of the charge density [22]. The sum of the occupied energies in this spectrum is the energy content of the Hamiltonian. In light of this corollary, unoccupied energies different (generally lower) from their corresponding values obtained with the optimal basis set cannot belong to the spectrum of the Hamiltonian, a unique functional of the charge density.

There is another way to prove that the first of the calculations producing the ground state is the calculation providing the DFT description of the material. This proof rests on a theorem for eigenvalue due to Rayleigh. The theorem applies to eigenvalues of the same equation obtained with two different basis sets where the larger one includes entirely the small one. In this condition, the theorem states that the ordered eigenvalues obtained with the larger basis set are lower than or equal to their corresponding values obtained with the smaller basis set. *All the calculations producing the ground state lead to the same self-consistent charge density and the same Hamiltonian*, within computational uncertainties. We note that the same Hamiltonian does not mean the same Hamiltonian matrix, as the dimension of the matrix varies with the size of the basis set. Hence, after the first calculation producing the ground state, other calculations leading to the ground state and some different (lowered) unoccupied energies do not provide the DFT description of the material. Indeed, the unoccupied energies lowered from their corresponding values obtained with the optimal basis set result from a mathematical artifact stemming from the Rayleigh theorem; they are not physically correct eigenvalues of the Hamiltonian which did not change from its value obtained with the optimal basis set.

Computational details in this work follow. Magnesium sulfide has a cubic structure. It is in the space group $Fm\bar{3}m$. The positions of the Mg and S ions in the irreducible zone are (0, 0, 0) and (1/2, 1/2, 1/2), respectively. The lattice constant we utilized in our self-consistent calculations is 5.200Å [36]. Our first self-consistent calculations were for Mg^{2+} and S^{2-}; they provided the input orbitals for the study of solid MgS. We expanded the radial parts of the wave function

in terms of Gaussian functions with even-tempered Gaussian exponents. The numbers of Gaussian orbitals utilized for the s, p, and d orbitals for Mg^{2+} were 18, 16, and 16, respectively, and 18, 18, and 16, respectively, for S^{2-}. The smallest and largest exponents were respectively 0.1822 and 0.11×10^6, for the Mg^{2+}, and 0.1489 and 0.44×10^5 for the S^{2-}. Our criterion for self-consistency is to have no more than a difference of 10^{-5} between potentials from two consecutive iterations. It typically took 60 iterations to satisfy this condition. We utilized 81 k points in the irreducible Brillouin zone for the production of the final, self-consistent bands. Extensive testing, over the years, has established the 20 divisions for each of the intervals L-Γ, Γ-X, X-K, and K-Γ that are amply adequate for the description of this structure.

3. Results

The valence orbitals for the successive, self-consistent calculations for MgS are shown in Table 2, along with the resulting band gaps. As per the above explanation of our method, Calculation III is the one that provides the DFT description of MgS. While Calculations IV-VI also led to the absolute minima of the occupied energies (*i.e.*, the ground state), they produced some low-laying unoccupied energies lower than their corresponding values obtained in Calculation III, with the optimal basis set. The electronic energies from Calculation III are shown in Table 3. The electronic energy bands, densities of states, effective masses, and bulk modulus discussed below are from Calculation III.

Figure 1 compares the results from Calculations III (solid lines) and IV (dashed lines). The basis set of Calculation IV consists of that of Calculation III plus the $4s^0$ on S^{2-}. The occupied bands from the two calculations are the same. Calculations V and VI also produced the same occupied energies as Calculation III.

Figure 2 and Figure 3 respectively show the calculated, total density of states (DOS) and the partial densities of states (pDOS), derived from the bands from Calculation III. We found a total valence bandwidth of 12.597 eV. From the

Table 2. The successive calculations of the BZW-EF method, with the valence orbitals of Mg^{2+} and S^{2-} in Columns 2 and 3, respectively. Column 4 and 5 show the total number of valence functions and the band gap, respectively. The employed, room temperature experimental lattice constant was 5.200Å. The superscript zero indicates an empty orbital. The optimal basis set is that of Calculation III. The corresponding, indirect band gap for bulk MgS, from Γ to X, is 3.278 eV, at room temperature.

Calculation no.	Magnesium (Mg^{2+}) ($1s^2$-core)	Sulfur (S^{2-}) ($1s^22s^22p^2$-core)	No. of valence functions	Energy Gap (eV)
I	$2s^22p^6\mathbf{3p^0}$	$3s^23p^6$	22	6.590 (Γ-L)
II	$2s^22p^63p^0$	$3s^23p^6\mathbf{4p^0}$	28	6.475 (Γ-L)
III	$\mathbf{2s^22p^63p^03s^0}$	$\mathbf{3s^23p^64p^0}$	**30**	**3.278 (Γ-X)**
IV	$2s^22p^63p^03s^0$	$3s^23p^64p^0\mathbf{4s^0}$	32	3.337 (Γ-X)
V	$2s^22p^63p^03s^0\mathbf{4p^0}$	$3s^23p^64p^04s^0$	38	3.370 (Γ-X)
VI	$2s^22p^63p^03s^04p^0\mathbf{4s^0}$	$3s^23p^64p^04s^0$	40	3.19 (Γ-X)

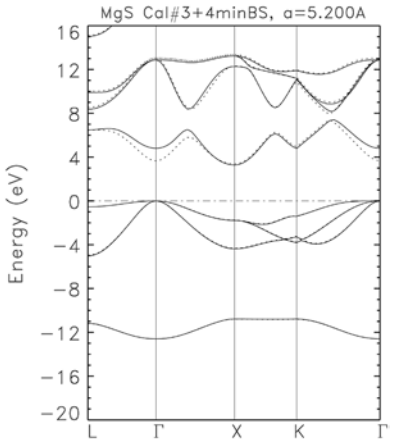


Figure 1. Calculated, electronic energy bands of MgS, from Calculations III (solid lines) and IV (dashed lines). The perfect superposition of the occupied bands from Calculations III, IV, and IV signifies the attainment of the absolute minima of the occupied energies, *i.e.*, the ground state, by Calculation III.

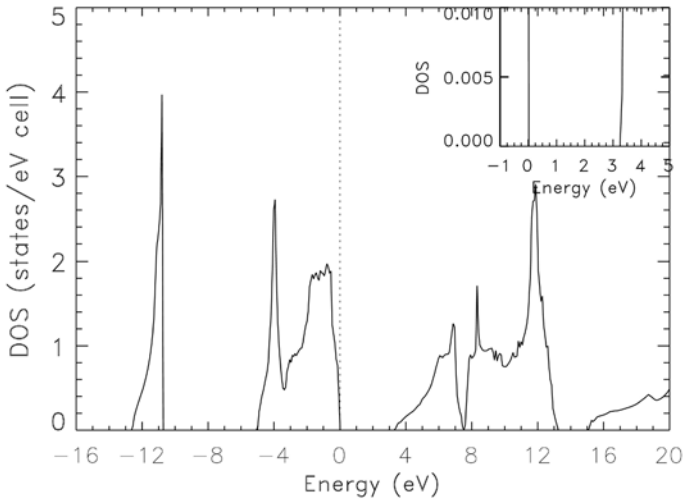


Figure 2. Total density of states (DOS) for rock salt MgS, derived from the bands from Calculation III. Zero, on the horizontal axis, indicates the position of the Fermi energy.

Table 3. Electronic energies at high symmetry points in the Brillouin zone for bulk rock salt MgS, as obtained from Calculation III, with the optimal basis of the LDA-BZW-EF method. The lattice constant is 5.200 Å, at room temperature.

L-point	Γ-point	X-point	f K-point
15.022	38.257	26.933	20.186
9.893	12.884	13.231	11.901
8.362	12.884	13.231	11.186
8.362	12.884	12.269	11.102
6.476	4.809	3.278	4.785
−0.541	0.000	−1.780	−1.408
−0.541	0.000	−1.780	−3.269
−5.018	0.000	−4.369	−3.809
−11.179	−12.597	−10.782	−10.781

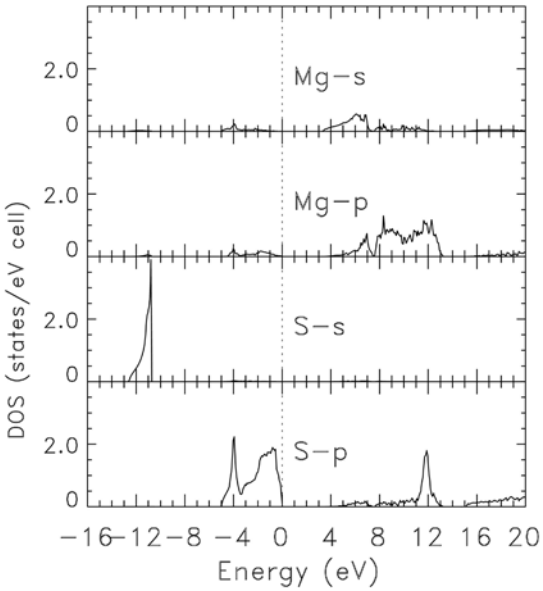


Figure 3. Results for partial densities of states for rock salt MgS, derived from the bands from Calculation III. The zero on the horizontal axis indicates the position of the Fermi energy.

DOS figure, the width of the lowest laying valence band is 1.85 eV, while that of the group of upper valence bands is about 5.0 eV. As per the partial densities of states in **Figure 3**, Mg-s and Mg-p dominate in the lowest, conduction band. There is a small contribution from S-p. In contrast, S-p strongly dominates upper valence bands, with minimal contributions from Mg-p and Mg-s. Sulfur s (S-s) clearly dominates in the lowest laying valence band. These results are in qualitative agreement with findings of Drief *et al.* [6] for the valence states.

Effective masses are important quantities in the theoretical calculation of transport properties; such properties include electrical conductivities and the Seebeck coefficient [37]. As per the content of **Table 4**, we have performed calculations of electron effective masses around the minimum of conduction band, at X, and around the lowest conduction bands at Γ. The effective masses of the light and heavy holes were derived from the uppermost valence bands at Γ. The calculated effective masses are shown in **Table 4**, for the various directions, in units of the mass of the electron (m_o). The effective masses of heavy Hole 1 and heavy Hole 2 are equal, except in the $(\Gamma\text{-K})_{110}$ direction. Their difference in that direction is due to the splitting of the bands in the $(\Gamma\text{-K})_{110}$ direction by the Coulomb crystal field. The quasi-isotropic nature of the electron effective masses at Γ is in contrast to that of the clearly anisotropic ones at X. The calculation performed by Ghebouli *et al.* [14] reported the electron effective mass for rock salt MgS, at the X point, to be 0.2179 m_o. This value is smaller than our predicted ones of 0.324 and 0.319 m_o at X, in the transverse direction. Our result is 0.503 for the longitudinal direction, at X. We found no experimental values for these effective masses. We expect future measurements to confirm our predictions in **Table 4**.

Figure 4 shows the total energy versus the lattice constant. The range of the lattice constant over which we calculated the total energy was from 5.10 to 5.30 Å. Our predicted equilibrium lattice constant, at the minimum of the total energy, is 5.183Å. With this lattice constant, we predicted a zero temperature band gap of 3.512 eV, larger than the room temperature value by 0.234 eV. For our purposes, no general fitting of the total energy to an equation of state is necessary; we fitted the curve in the immediate vicinity of the minimum in order to calculate the bulk modulus. Our calculated bulk modulus of 79.76 GPa is in excellent agreement with the experimental value of 78.9 ± 3.7 GPa [36].

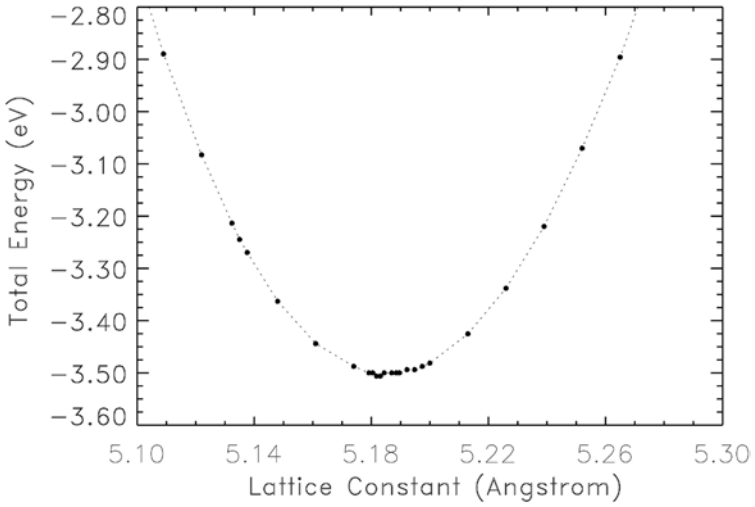


Figure 4. The graph of the total energy versus the lattice constant, for rock salt MgS, as obtained from calculations using the optimal basis set.

Table 4. Calculated, effective masses for rock salt MgS, in units of free electron mass (m_o): M_e indicates an electron effective mass at the X-point and Γ-point, in the conduction band; M_{hh} and M_{lh} represent the heavy and light holes effective masses, respectively.

Types and Directions of Effective Masses	Values of Effective Masses (m_o)
M_e(X-Γ) Longitudinal	0.503
M_e(X-U) Transverse	0.324
M_e(X-W) Transverse	0.319
$M_e(\Gamma\text{-L})_{111}$	0.493
$M_e(\Gamma\text{-X})_{100}$	0.420
$M_e(\Gamma\text{-K})_{111}$	0.433
$M_{hh1}(\Gamma\text{-L})_{111}$	3.116
$M_{hh1}(\Gamma\text{-X})_{100}$	1.056
$M_{hh1}(\Gamma\text{-K})_{110}$	1.620
$M_{hh2}(\Gamma\text{-L})_{111}$	3.116
$M_{hh2}(\Gamma\text{-X})_{100}$	1.056
$M_{hh2}(\Gamma\text{-K})_{110}$	1.204
$M_{lh}(\Gamma\text{-L})_{111}$	0.220
$M_{lh}(\Gamma\text{-X})_{100}$	0.300
$M_{lh}(\Gamma\text{-K})_{110}$	0.270

4. Discussion

Our calculated, indirect band gap is 3.278 eV. This value is higher than the predicted value from calculations with *ab-initio* LDA and GGA potentials, with ranges of 2.208 - 2.69 eV, for LDA, and 2.64 - 2.79 eV, for GGA. One cannot draw conclusions from results of calculations with ad hoc potentials, given their variability with the adjustable parameters. The Green function and dressed Coulomb approximation (GWA) is not DFT theory. For this reason, we do not make comparisons between our LDA findings and those from GWA (or GW) calculations. We suggest the following explanation for the differences between our results and the previous, *ab-initio* DFT ones in Table 1.

To our knowledge, previous DFT calculations mostly employed single basis sets. The chance for a self-consistent calculation with a single basis to lead to the ground state of the material is practically zero. Rather, such a calculation leads to one of the potentially infinite number of stationary solutions. Our calculations in Table 2 produced such stationary solutions. The progressive increase of the basis set allowed us to identify the first one among them that 1) yields the ground state energy and 2) is not over-complete for the description of the ground state. Basis sets that contain the optimal one and are over-complete produce the same occupied energies obtained with the optimal basis set, but they generally produce some unoccupied energies, including lowest laying ones, that are lower than their counterparts obtained with the optimal basis set. Consequently, these basis sets produce band gaps that are smaller than corresponding, experimental ones. This phenomenon, we suggest, explains the differences between our calculated band gap and those from previous *ab-initio* DFT (*i.e.*, LDA or GGA) calculations as shown in Table 1.

Similarly, our calculated electron effective masses are larger than the previous results discussed above. In general, a lowering of unoccupied bands results in a decrease of their degree of flatness. A decrease of the degree of flatness around the conduction band minimum clearly leads to a decrease of the electron effective mass. Hence, the above unphysical lowering of unoccupied energies is responsible for the underestimation of the electron effective masses in a way that is linked to the underestimation of the band gap.

5. Conclusion

We have employed the LCAO formalism and an LDA potential to perform *ab-initio*, self-consistent calculations to predict electronic and related properties of rock salt MgS. By searching for and by verifiably attaining the ground state of the material, as required by the second DFT theorem, our results have the full, physical content of DFT and agree with available experimental results. Our calculated, room temperature indirect band gap is 3.278 eV. Our predicted equilibrium lattice constant and zero temperature band gap are 5.183Å and 3.512 eV, respectively. The calculated bulk modulus (79.76 GPa) agrees very well with the measured value of 78.9 ± 3.7 GPa. It is hoped that our predictions will spur new

experimental investigations of properties of MgS. As noted in the background and motivation section, we only found bang gap measurements for thin films of MgS. Possible methods for measuring band gaps include optical absorption, optical transmittance, spectroscopic ellipsometry, and UV to visible and near infrared absorption, with diffuse reflectance.

Acknowledgements

The funding sources for this work are the US Department of Energy (DOE), National Nuclear Security Administration (NNSA) (Award No. DE-NA0003679), the National Science Foundation (NSF, Award No. 1503226DE), LaSPACE, and LONI-SUBR.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Pandey, R., Jaffe, J.E. and Barry Kunz, A. (1991) *Physical Review B*, **43**, 11.

[2] Haase, M.A., Qiu, J., DePuydt, J.M. and Cheng, H. (1991) *Applied Physics Letters,* **59**, 1272. https://doi.org/10.1063/1.105472

[3] Albin, S., Satira, J.D., Livingston, D.L. and Shull, T.A. (1992) *Japanese Journal of Applied Physics*, **31**, 715-719. https://doi.org/10.1143/JJAP.31.715

[4] Lai, Y., Cheung, W., Lok, S., Wong, G.K.L., Ho, S., Tam, K. and Sou, I. (2012) *AIP Advances*, **2**, 012149. https://doi.org/10.1063/1.3690124

[5] Wang, M., Li, X., Gao, M., Pan, H. and Liu, Y. (2014) *Journal of Alloys and Compounds*, **603**, 158-166. https://doi.org/10.1016/j.jallcom.2014.03.052

[6] Drief, F., Tadjer, A., Mesri D. and Aourag, H. (2004) *Catalysis Today,* **89**, 343-355. https://doi.org/10.1016/j.cattod.2003.12.013

[7] Rached, D., Benkhettou, N., Soudini, B., Abbar, B., Sekkal N. and Driz, M. (2003) *Physica Status Solidi B*, **240**, 565-573. https://doi.org/10.1002/pssb.200301889

[8] Stepanyuk, V.S., Grigorenko, A.A., Katsnelson, A.A., Farberovich, O.V., Szasz, A. and Mikhailin, V.V. (1992) *Physica Status Solidi,* **174**, 289. https://doi.org/10.1002/pssb.2221740129

[9] Duman, S., Bagcl, S., Tutuncu, H.M. and Srivastava, G.P. (2006) *Physical Review B*, **73**, 205201. https://doi.org/10.1103/PhysRevB.73.205201

[10] de Boer, P.K. and de Groot, R.A. (1998) *Journal of Physics*: *Condensed Matter*, **10**, 10241-10248. https://doi.org/10.1088/0953-8984/10/45/011

[11] Zhu, X., Zhang, J., Cheng, X. and Zheng, Y. (2017) *Journal of Atomic and Molecular Physics*, **34**, 2.

[12] Tairi, L., Touam, S., Boumaza, A., Boukhtouta, M., Meradji, H., Ghemid, S., Omran, S., Hassan, F. and Khenata, R. (2017) *A Multinational Journal*, **90**, 929-941.

[13] Madu, C.A. and Onwuagba, B.N. (2005) The Abus Salam International Centre for Theoretical Physic (Limited Distribution).

[14] Ghebouli, M.A., Choutri, H. and Bouarissa, N. (2014) *Materials Science in Semiconductor Processing*, **18**, 71-79. https://doi.org/10.1016/j.mssp.2013.10.029

[15] Wu, H., Chen, Y., Zhou, P., Han, X. and Liu, Z. (2014) *Zeitschrift für Naturforschung*, **69a**, 403-410.

[16] Yogeswari, M., Umamaheswari, R. and Kalpana, G. (2013) *Advanced Materials Research*, **665**, 22-28.

[17] Kalpana, G., Palanivel, B., Thomas, R.M. and Rajagopalan, M. (1996) *Physica B*, **222**, 223-228. https://doi.org/10.1016/0921-4526(96)00014-2

[18] Ching, W.F., Gan, F. and Huang, M. (1995) *Physical Review B*, **52**, 1596.

[19] Zhao, C., Duan, Y., Gao, J. and Dong, H. (2017) *Journal of Applied Physics*, **121**, Article ID: 235705. https://doi.org/10.1063/1.4986929

[20] Lany, S. (2014) *Proceedings of SPIE*, **8987**, 89870K.

[21] Taleatu, B.A., Omotoso, E., Arbab, E.A.A., Lasisi, R.A., Makinde, W.O. and Mola, G.T. (2015) *Applied Physics A*, **118**, 539-545. https://doi.org/10.1007/s00339-014-8753-0

[22] Bagayoko, D. (2014) *AIP Advances*, **4**, Article ID: 127104. https://doi.org/10.1063/1.4903408

[23] Bagayoko, D. and Franklin, L. (2005) *Journal of Applied Physics*, **97**, Article ID: 123708. https://doi.org/10.1063/1.1939069

[24] Bagayoko, D., Franklin, L. and Zhao, G.L. (2004) *Journal of Applied Physics*, **96**, 4297-4301. https://doi.org/10.1063/1.1790064

[25] Bagayoko, D., Zhao, G.L., Fan, J.D. and Wang, J. (1998) *Journal of Physics*: *Condensed Matter*, **10**, 5645-5655. https://doi.org/10.1088/0953-8984/10/25/014

[26] Zhao, G.L., Bagayoko, D. and Williams, T.D. (1998) *Bulletin of the American Physical Society*, **43**, 846.

[27] Ekuma, E., Franklin, L., Zhao, G.L., Wang, J. and Bagayoko, D. (2011) *Canadian Journal of Physics*, **89**, 319-324. https://doi.org/10.1139/P11-023

[28] Franklin, L., Ekuma, C., Zhao, G.L. and Bagayoko, D. (2013) *Journal of Physics and Chemistry of Solids*, **74**, 729-736. https://doi.org/10.1016/j.jpcs.2013.01.013

[29] Zhao, G.L., Bagayoko, D. and Williams, T. (1999) *Physical Review B*, **60**, 1563. https://doi.org/10.1103/PhysRevB.60.1563

[30] Stewart, A., Hart, D., Khamala, B., Malovzowsky, Y. and Bagayoko, D. (2015) *Journal of Advances in Physics*, **9**, 1.

[31] Ekuma C.E. and Bagayoko, D. (2011) *Japanese Journal of Applied Physics*, **50**, Article ID: 101103. https://doi.org/10.7567/JJAP.50.101103

[32] Ceperley, D.M. and Alder, B.J. (1980) *Physical Review Letter*, **45**, 566-569.

[33] Vosko, S.H., Wilk, L. and Nusair, M. (1980) *Canadian Journal of Physics*, **58**, 1200-1211. https://doi.org/10.1139/p80-159

[34] Feibelman, P.J., Appelbaum, J.A. and Hamann, D.R. (1979) *Physical Review B*, **20**, 1433-1443. https://doi.org/10.1103/PhysRevB.20.1433

[35] Harmon, B.N., Weber, W. and Hamann, D.R. (1982) *Physical Review B*, **25**, 1109. https://doi.org/10.1103/PhysRevB.25.1109

[36] Peiris, S.M., Campbell, A.J. and Heinz, D.L. (1994) *Journal of Physics and Chemistry of Solids*, **55**, 413-419. https://doi.org/10.1016/0022-3697(94)90166-X

[37] Ekuma, C.E., Jarrell, M., Moreno, J. and Bagayoko, D. (2012) *AIP Advances*, **2**, Article ID: 012189. https://doi.org/10.1063/1.3700433

Journal of Modern Physics, 2018, 9, 1785-1792
http://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

One Application of the Quantized Electromagnetic Field outside the High-Dimensional Static Gauss-Bonnet Black Holes

Ming Zhang

Faculty of Science, Xi'an Aeronautical University, Xi'an, China
Email: zhangming@xaau.edu.cn

How to cite this paper: Zhang, M. (2018) One Application of the Quantized Electromagnetic Field outside the High-Dimensional Static Gauss-Bonnet Black Holes. *Journal of Modern Physics*, **9**, 1785-1792.
https://doi.org/10.4236/jmp.2018.99112

Received: July 19, 2018
Accepted: August 5, 2018
Published: August 8, 2018

Abstract

In present paper, we calculated the response rate of a static charge outside *d*-dimensional Gauss-Bonnet black hole in the low-frequency regime, by using the quantization of an electromagnetic field in the background of static spherically symmetric *d*-dimensional spacetime in the Boulware vacuum, which can be expressed as the summation of hypergeometric functions.

Keywords

Response Rate, Gauss-Bonnet Black Hole, High-Dimensional Spacetime

1. Introduction

An important prediction in this field is the thermal evaporation of black hole [1]. This nontrivial effect was soon realized to be closely associated with the existence of an event horizon in Schwarzschild spacetime. One of the difficulties in studying fields in Schwarzschild [2] and other black hole spacetime, even when the fields are non-interacting, stems from the fact that the solutions to the field equations are functions whose properties are not well known. In the low-frequency regime, however, the situation is much simpler and the mode functions of the massless scalar field are well known [3]. Recently, Crispino *et al.* [4] suggested a scheme to quantize the free quantum electrodynamics in static spherically symmetric *d*-dimensional spacetime and gave out the response rate of a static charge outside the four-dimensional Schwarzschild black hole.

Following the advent of string theory, extra dimensions were promoted from an interesting curiosity to a theoretical necessity since superstring theory

requires an eleven-dimensional spacetime to be consistent from a quantum point of view ([5] [6] [7] [8]). Among the higher curvature gravities, the most extensively studied theory is the so-called Gauss-Bonnet gravity ([9]-[15]), which naturally emerges when we want to generalize Einstein's theory in higher dimensions by keeping all characteristics of usual general relativity excepting the linear dependence of the Riemann tensor. Therefore, it is necessary to study the quantization of the free electromagnetic field with the high-dimensional spacetime and its specific application outside a Gauss-Bonnet black hole.

In this paper, under the quantized electromagnetic field, we calculate the response rate of a static charge outside the the *d*-dimensional GB black hole in the Unruh vacuum [16]. The response rate is a quantum concept with no natural analog in classical physics though it is possible to define a corresponding classical quantity mathematically and represents the number of times the source responds to the field per unit time.

The paper is organized as follows. In Section 2, we review the basic concepts of the electromagnetic field in the arbitrary dimensional spacetime of a spherically symmetric black hole in a modified Feynann gauge. Section 3 devotes to calculate the response rate of a static charge outside a GB black hole of arbitrary-dimensions. In Section 5, we summarize the main results.

2. Gupta-Bleuler Quantization in a Modified Feynman Gauge

In this section, we follow the notation of Ref. [4] to study the solutions of field equations for electromagnetic field in an asymptotic flat and static spherically symmetric $(p+2)$-dimensional spacetime. The quantization of electromagnetic field will be carried out in the frame of Gupta-Bleuler formalism in a modified Feynmann gauge.

The line element under considered takes the form

$$\mathrm{d}\tau^2 = f(r)\mathrm{d}t^2 - h(r)\mathrm{d}r^2 - r^2\mathrm{d}s_p^2 \tag{1}$$

with the line element of a unit *p*-sphere $\mathrm{d}s_p^2$. We assume that both $f(r)$ and $h(r)^{-1}$ have a zero at $r = r_h$ and positive for $r > r_h$.

The Lagrangian density for electromagnetic field in a modified Feynman gauge is

$$\mathcal{L}_F = \sqrt{-g}\left[-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2}G^2\right], \tag{2}$$

and G stands for the modified Feynman gauge

$$G = \nabla^\mu A_\mu + K^\mu A_\mu. \tag{3}$$

Here the vector K^μ is independent on electromagnetic field A_μ, and takes the form

$$K^\mu = \left(0, f'/(fh), 0, 0\right). \tag{4}$$

Under this choice, the gauge condition changes into

$$G=\frac{1}{f}\partial_t A_t-\sqrt{\frac{f}{h}}\frac{1}{r^p}\partial_r\left[\frac{r^p}{\sqrt{fh}}A_r\right]-\frac{1}{r^2}\nabla^i A_i \tag{5}$$

From the Lagrangian density for electromagnetic field, the equations of motion are

$$\begin{cases}-\frac{1}{f}\partial_t^2 A_t+\sqrt{\frac{f}{h}}\frac{1}{r^p}\partial_r\left[\frac{r_p}{\sqrt{fh}}\partial_r A_r\right]+\frac{1}{r^2}\nabla^2 A_t=0 & \text{(6a)}\\ -\frac{1}{f}\partial_t^2 A_r+\frac{1}{f}\partial_r\left[\sqrt{\frac{f}{h}}\frac{f}{r^p}\partial_r\left(\frac{r_p}{\sqrt{fh}}A_r\right)\right]+\frac{1}{r^2}\nabla^2 A_r+\frac{1}{f}\partial_r\left(\frac{f}{r^2}\right)\nabla^i A_i=0 & \text{(6b)}\\ -\frac{1}{f}\partial_t^2 A_i+\frac{r^{2-p}}{\sqrt{fh}}\partial_r\left(\sqrt{\frac{f}{h}}r^{p-2}\partial_r A_i\right)-\frac{r^2}{fh}\partial_r\left(\frac{f}{r^2}\right)\partial_i A_r & \\ \quad+\frac{1}{r^2}\left[\nabla^j\left(\nabla_j A_i-\nabla_i A_j\right)+\partial_i\left(\nabla^j A_j\right)\right]=0,\ \ (j=1,\cdots,p) & \text{(6c)}\end{cases}$$

Here ∇_i is the covariant derivative on S^p.

We denote the complete set of solutions of Equation (6) by $A_\mu^{(\lambda n;\omega lm)}$, and call a non-physical modes for $\lambda=0$, physical modes for $\lambda=1,2,3,\cdots,p$ and a pure-gauge mode for $\lambda=p+1$. The label n represents modes incoming from the past null infinity $(n=\leftarrow)$ and coming out from the past horizon $(n=\rightarrow)$.

2.1. Physical Modes

For other independent solutions $\lambda=1,2,\cdots,p$, which represent physical degrees of freedom, the time-component can be taken as zero. They are the linear independent solution of Equation (6) with a gauge condition $G=0$, and are classified into two types.

Physical Modes 1

The "physical modes 1" solution can be written as [4]

$$\begin{cases}A_t^{(1n;\omega lm)}=0 & \text{(7a)}\\ A_r^{(1n;\omega lm)}=R_{\omega l}^{(1n)}(r)Y_{lm}\mathrm{e}^{-i\omega t},(l\geq 1) & \text{(7b)}\\ A_i^{(1n;\omega lm)}=\frac{r^{2-p}}{l(l+p-1)}\sqrt{\frac{f}{h}}\frac{\mathrm{d}}{\mathrm{d}r}\left[\frac{r^p}{\sqrt{fh}}R_{\omega l}^{(1n)}(r)\right]\partial_i Y_{lm}\mathrm{e}^{-i\omega t} & \text{(7c)}\end{cases}$$

where $i=1,2,3,\cdots,p$ and:

$$\left[\frac{\omega^2}{f}-\frac{l(l+p-1)}{r^2}\right]R_{\omega l}^{(1n)}(r)+\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left[\sqrt{\frac{f}{h}}r^{2-p}\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{r^p}{\sqrt{fh}}R_{\omega l}^{(1n)}(r)\right)\right]=0 \tag{8}$$

2.2. Quantization

Using Gupta-Bleuler quantization, we impose the equal-time commutation relations on the field $\hat{A}_\mu$ and momentum $\hat{\Pi}^{t\mu}$ operators

$$\left[\hat{A}_\mu(t,\mathbf{x}),\hat{A}_\nu(t,\mathbf{x}')\right]=\left[\hat{\Pi}^{t\mu}(t,\mathbf{x}),\hat{\Pi}^{t\nu}(t,\mathbf{x}')\right]=0 \tag{9}$$

$$\left[\hat{A}_\mu(t,\mathbf{x}),\hat{\Pi}^{t\nu}(t,\mathbf{x}')\right]=\frac{i\delta_\mu^\nu}{\sqrt{-g}}\delta^{p+1}(\mathbf{x}-\mathbf{x}') \tag{10}$$

where $\mathbf{x}$ and $\mathbf{x}'$ represent all spatial coordinates. The field $\hat{A}_\mu$ can be expressed in terms of $A_\mu^{(\xi)}$

$$\hat{A}_\mu(t,\mathbf{x}) = \sum_\rho \int_{-\infty}^{+\infty} \frac{\mathrm{d}\omega}{\sqrt{4\pi|\omega|}} A_\mu^{(\omega\rho)}(t,\mathbf{x}) a_{\omega\rho}, \tag{11}$$

where $A_\mu^{(-\omega\rho)} \equiv \overline{A_\mu^{(\omega\rho)}}$, $a_{-\omega\rho} \equiv a_{\omega\rho}^\dagger$ and ρ labels all quantum numbers.

3. Response Rate of a Static Charge Outside a D-Dimensional Gauss-Bonnet Gravity Black Hole

In this section, we will calculate the response rate of a static charge outside a d-dimensional $(d = p+2)$ Gauss-Bonnet black hole by following the procedure of Ref. [4]. In this case, the black hole is characterized by $f(r) = h(r)^{-1} = 1-(r_h/r)^{(p-1)}$.

The action of pure GB can be written as

$$S = \int \mathrm{d}^{p+2}x \sqrt{-g}\alpha\mathcal{L}_2 \tag{12}$$

where the coupling constant α can be regarded as the inverse of string tension and be assumed $\alpha > 0$ in this paper. The Gauss-Bonnet term is given by

$$\mathcal{L}_2 = R^2 - 4R_{ab}R^{ab} + R_{abcd}R^{abcd}. \tag{13}$$

The line element for the exterior region of the d-dimensional Gauss-Bonnet gravity is given by

$$\mathrm{d}s^2 = f(r)\mathrm{d}t^2 - f(r)^{-1}\mathrm{d}r^2 - r^2\mathrm{d}s_p^2 \tag{14}$$

The $f(r)$ in the d-dimensional GB gravity is

$$f(r) = 1-\sqrt{\frac{2M}{\alpha r^{d-5}}} \equiv 1-\frac{B}{r^{\frac{d-5}{2}}}, \tag{15}$$

The horizon radius $r_h = \left(\frac{2M}{\alpha}\right)^{\frac{1}{d-5}} = B^{\frac{2}{d-5}}$.

In order to avoid the indefinite results ([17] [18]), we use the formula suggested by Crispino *et al.* [19] and assume the static charge located at (r_0,θ_0) with a current density j^μ

$$\begin{cases} j^\mu = \left(j^t, j^r, 0, \cdots, 0\right) & \text{(16a)} \\ j^t = \dfrac{\sqrt{2}q\cos Et}{\sqrt{-g}}\delta(r-r_0)\delta(\theta_1-\theta_{10})\cdots\cdots\delta(\theta_p-\theta_{p0}) & \text{(16b)} \\ j^r = \dfrac{\sqrt{2}qE\sin Et}{\sqrt{-g}}\Theta(r-r_0)\delta(\theta_1-\theta_{10})\cdots\cdots\delta(\theta_p-\theta_{p0}) & \text{(16c)} \end{cases}$$

The step function $\Theta(x)$ is defined by $\Theta(x) = 1, (x>0)$ and vanishing for $x \le 0$.

Such current interacts with vector potential A_μ through the Lagrangian $\sqrt{-g}\, j^\mu A_\mu$. Since $a_{\omega lm}^{((p+1)n)\dagger}|\text{phys}\rangle$ is non-physical states, which excludes the interaction with the pure-gauge particles created by $a_{\omega lm}^{((p+1)n)\dagger}$. We will neglect it.

However, the current does interact with the states created by $a_{\omega lm}^{(0n)\dagger}$ but the contribution to physical probabilities maybe taken as zero once the non-physical modes are appropriately chosen. Furthermore, there is no interacting term between the current and physical modes 2 due to $A_t = A_r = 0$. Therefore, we only need to consider the physical modes 1. To make the counting process more concise, we will limit in the spherical coulomb gauge. The mode function can be written as [4]

$$A_t^{1'n;\omega lm} = \frac{i\omega r^{2-p}}{l(l+p-1)}\sqrt{\frac{f}{h}}\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{r^p}{\sqrt{fh}}R_{\omega l}^{1n}\right)Y_{lm}\mathrm{e}^{-i\omega t} \tag{17}$$

$$A_r^{1'n;\omega lm} = \frac{\omega^2 r^2}{l(l+p-1)}\frac{1}{f}R_{\omega l}^{1n}Y_{lm}\mathrm{e}^{-i\omega t}. \tag{18}$$

In the limit $E \to 0$, the proper response rate of the charge can be written [4]

$$\frac{R_{0lm}}{\sqrt{f(r_0)}} = 4\pi\lim_{E\to 0}\frac{\left|\mathcal{T}_{Elm}^{\rightarrow}\right|^2}{\sqrt{f(r_0)}\beta E} \tag{19}$$

where $\mathcal{T}_{Elm}^{\rightarrow}$ is the index $n=\rightarrow$ in transition amplitude $\mathcal{T}_{\omega lm}^{n}$ which has the form

$$\mathcal{T}_{\omega lm}^{n} \equiv \frac{1}{2\pi\delta(\omega - E)}\int \mathrm{d}^{p+2}x\sqrt{-g}\,j^{\mu}\left\langle 1n;\omega lm\left|\hat{A}_{\mu}\right|0\right\rangle \tag{20}$$

Now, we return to calculate $\mathcal{T}_{Elm}^{\rightarrow}$. The Equation (8) can be written

$$\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left[fr^{2-p}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^pR_{\omega l}^{(1\to)}(r)\right)\right] - \frac{l(l+p-1)}{r^2}R_{\omega l}^{(1\to)}(r) + \frac{\omega^2}{f}R_{\omega l}^{(1\to)}(r) = 0. \tag{21}$$

After introducing the Wheeler tortoise coordinate and a function φ

$$R_{\omega l}^{(1\to)}(r) \equiv \frac{\sqrt{l(l+p-1)}}{\omega}r^{-\frac{p}{2}-1}\varphi_{\omega l}^{(1\to)}(r), \tag{22}$$

the Equation (21) changes into

$$\left(\omega^2 + \frac{\mathrm{d}^2}{\mathrm{d}r^{*2}} - V_1(r^*)\right)\varphi_{\omega l}^{(1\to)}(r) = 0 \tag{23}$$

with

$$V_1\left[r^*(r)\right] = f\frac{l(l+p-1)}{r^2} + f^2\frac{p(p-2)}{4r^2} - ff'\frac{(p-2)}{2r} \tag{24}$$

For the small ω and the condition $\left(r - r_h \ll \omega^2 r_h^3, \left|\omega r^*\right| \ll 1\right)$, the wave coming from the past horizon H^- is almost completely reflected back by the potential toward the horizon

$$\varphi_{\omega l}^{(1\to)} \approx -2\omega r^* + \text{const} \tag{25}$$

Generally, it is hard to find the analytic expression for the Wheeler tortoise coordinate r^*. Fortunately, what we need is just the behavior of r^* near horizon. Substituting the expression of $f(r)$, the leading term of the Wheeler tortoise coordinate can be written

$$r^* \approx \frac{2}{p-3}\left(\frac{2M}{\alpha}\right)^{\frac{1}{p-3}} \ln(z-1) \tag{26}$$

by using the transition

$$z = \frac{2}{B} r^{\frac{p-3}{2}} - 1. \tag{27}$$

The boundary condition of $R_{\omega l}^{(1\to)}$ reads

$$R_{\omega l}^{(1\to)} \approx -\frac{\sqrt{l(l+p-1)}}{p-3} 2^{\frac{3p-4}{p-3}} B^{-\frac{p}{p-3}} \ln(z-1),\ \left(r - r_H \ll \omega^2 r_H^3, \left|\omega r^*\right| \ll 1\right) \tag{28}$$

In terms of variable z, Equation (21) can be written

$$\begin{aligned} &\frac{1}{4}(z^2-1)\frac{\mathrm{d}^2 q(z)}{\mathrm{d}z^2} + \frac{1}{4}\left[\frac{3p-1}{p-3}(z-1)+2\right]\frac{\mathrm{d}q(z)}{\mathrm{d}z} \\ &+ \frac{1}{(p-3)^2}\left[-l(l+p-1) + \frac{p(p-5)}{z+1} + p + \omega^2 \frac{z+1}{z-1}\left(\frac{B}{2}(z+1)\right)^{\frac{4}{p-3}}\right] q(z) = 0 \end{aligned} \tag{29}$$

and this equation can be solved explicitly for small ω limit. Combining the asymptotic behavior $R_{\omega l}^{(1\to)} \to 0$ as $z \to +\infty$, the solution of Equation (29) is

$$\begin{aligned} R_{\omega l}^{(1\to)} = &\frac{\sqrt{l(l+p-1)}}{p-3} 2^{\frac{3p-4}{p-3}} B^{-\frac{p}{p-3}} \\ &\times \mathrm{F}\left(\left[\frac{2l}{p-3}, \frac{-2l-2p+2}{p-3}\right], \left[\frac{-2p+2}{p-3}\right], \frac{z+1}{2}\right)(z+1)^{\frac{2p}{p-3}} \end{aligned} \tag{30}$$

where the coefficient has been appropriately chosen to agreement with the boundary condition Equation (28).

Thus, the expression of the proper response rate of the charge is

$$\frac{R_{0lm}}{\sqrt{f(r_0)}} = \frac{(p-3)q^2(z_0+1)^{-\frac{2p+2}{p-3}} B^{\frac{2-2p}{p-3}} 2^{\frac{6}{p-3}}}{\pi l(l+p-1)\sqrt{f(r_0)}} \left(\frac{z_0-1}{z_0+1}\right)^2 \left[\frac{\mathrm{d}}{\mathrm{d}z_0}\mathrm{F}_l(z_0)\right]^2 |Y_{lm}|^2 \tag{31}$$

ubstituting the all parameters for d-dimensional GB black hole

$$z_0 = \frac{2}{B} r_0^{\frac{p-3}{2}} - 1 \tag{32}$$

$$f(r_0) = 1 - \frac{B}{r_0^{\frac{d-5}{2}}} \tag{33}$$

$$\mathrm{F}_l(z_0) = \mathrm{F}\left(\left[\frac{2l}{p-3}, \frac{-2l-2p+2}{p-3}\right], \left[\frac{-2p+2}{p-3}\right], \frac{z+1}{2}\right) \tag{34}$$

the total transition probability per proper time of the charge is given by

$$\begin{aligned} R^{\mathrm{tot}} &= \sum_{l}^{+\infty} \sum_{m=-l}^{l} \frac{R_{0lm}}{\sqrt{f(r_0)}} \\ &= \sum_{l}^{+\infty} \frac{(p-3)q^2(z_0+1)^{-\frac{2p+2}{p-3}} B^{\frac{2-2p}{p-3}} 2^{\frac{6}{p-3}}}{\pi l(l+p-1)\sqrt{f(r_0)}} \left(\frac{z_0-1}{z_0+1}\right)^2 \left[\frac{\mathrm{d}}{\mathrm{d}z_0}\mathrm{F}_l(z_0)\right]^2 \frac{G(l)}{\Omega_p} \end{aligned} \tag{35}$$

We can give the main contribution of the result of the Schwarzschild black hole and GB black hole,

$$R^{\text{tot}} \sim \left(\frac{r_h}{r_0}\right)^{p-3}, (\text{Schwarzschild}) \tag{36}$$

$$R^{\text{tot}} \sim \left(\frac{r_h}{r_0}\right)^{\frac{3p-1}{2}}. (\text{GB}) \tag{37}$$

Assuming that there is a charge outside an unknown black hole, the radius of the horizon (r_h), the mass of the black hole (M) and the location of the charge (r_0) are known. Meanwhile, we can measure the response rate of the charge outside this unknown black hole. Compare the measured value with Equation (36) and Equation (37), then we can tell the black hole is the GB type or not.

4. Conclusions

In this paper, we quantized the free electrodynamics in static spherically symmetric spacetime of arbitrary dimensions in a modified Feynman gauge. Then we examined the Gupta-Bleuler quantization in this modified Feynman gauge. The results obtained were applied to compute the total response rate of a static charge outside the d-dimensional GB black hole in the Unruh vacuum.

For the Einstein-Gauss-Bonnet gravity, one can follow the same procedure. The Wheeler tortoise coordinate r^* has a asymptotic behavior $r^* \approx f'(r_h)^{-1}\ln(r-r_h)$ as near horizon. The boundary condition Equation (25) changes into $\varphi_{\omega l}^{(1\to)} \approx -2\omega/f'(r_h)\ln(r-r_h)$. One problem is that we do not find a new variable to simplify the Equation (21) and its analytic solution. An applicable way is to find a series solution of Equation (21), the results can not be expressed in terms of familiar special functions and we neglect it here.

Such an outcome is not only a simple promotion work for what Crispino *et al.* [4] have done. Having the specific form of the free quantum electrodynamics in static spherically symmetric spacetime of arbitrary dimensions, It may provide us a chance for further investigating quantum field theory in high-dimensional curved spacetime. For instance, some authors ([19] [20] [21]) have researched whether or not a quantum version of the equivalence principle could be formulated and show some equivalence for low-frequency quantum phenomena in flat and curved spacetime. The same problem could be reconsidered in high-dimensional spacetime and discuss the dimension dependence of the results.

Acknowledgements

This work has been supported by the Scientific Research Program Funded by Shaanxi Provincial Education Department under Program No.16JK1394.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Hawking, S.W. (1975) *Communications in Mathematical Physics*, **43**, 199. [(1976) Erratum-Ibid, **46**, 206]. https://doi.org/10.1007/BF02345020

[2] Jensen, B.P. and Candelas, P. (1986) *Physical Review D*, **33**, 1590. [(1987) Erratum-Ibid, **D35**, 4041]. https://doi.org/10.1103/PhysRevD.33.1590

[3] Candelas, P. (1980) *Physical Review D*, **21**, 2185. https://doi.org/10.1103/PhysRevD.21.2185

[4] Crispino, L.C.B., Higuchi, A. and Matsas, G.E.A. (2001) *Physical Review D*, **63**, 124008. [(2009) Erratum-Ibid, **D80**, 029906]. https://doi.org/10.1103/PhysRevD.63.124008

[5] Horava, P. and Witten, E. (1996) *Nuclear Physics B*, **475**, 94 [hep-th/9603142]. https://doi.org/10.1016/0550-3213(96)00308-2

[6] Lukas, A., Ovrut, B.A. and Waldram, D. (1999) *Physical Review D*, **60**, 086001 [hep-th/9806022]. https://doi.org/10.1103/PhysRevD.60.086001

[7] Randall, L. and Sundrum, R. (1999) *Physical Review Letters*, **83**, 3370 [hep-ph/9905221]. https://doi.org/10.1103/PhysRevLett.83.3370

[8] Randall, L. and Sundrum, R. (1999) *Physical Review Letters*, **83**, 4690 [hep-th/9906064]. https://doi.org/10.1103/PhysRevLett.83.4690

[9] Lanczos, C. (1938) *Annals of Mathematics*, **39**, 842. https://doi.org/10.2307/1968467

[10] Lovelock, D. (1971) *Journal of Mathematical Physics*, **12**, 498. https://doi.org/10.1063/1.1665613

[11] Zou, D.-C., Yang, Z.-Y., Yue, R.-H. and Yu, T.-Y. (2011) *Chinese Physics B*, **20**, 100403 [arXiv:1107.2741 [gr-qc]]. https://doi.org/10.1088/1674-1056/20/10/100403

[12] Yue, R.-H., Zou, D.-C., Yu, T.-Y. and Yang, Z.-Y. (2011) *Chinese Physics B*, **20**, 050401 [arXiv:1107.2743 [gr-qc]].

[13] Zhou, K., Yang, Z.-Y., Zou, D.-C. and Yue, R.-H. (2012) *Chinese Physics B*, **21**, 020401 [arXiv:1107.2732 [gr-qc]]. https://doi.org/10.1088/1674-1056/21/2/020401

[14] Boulware, D.G. and Deser, S. (1985) *Physical Review Letters*, **55**, 2656. https://doi.org/10.1103/PhysRevLett.55.2656

[15] Wheeler, J.T. (1986) *Nuclear Physics B*, **273**, 732. https://doi.org/10.1016/0550-3213(86)90388-3

[16] Unruh, W.G. (1976) *Physical Review D*, **14**, 870. https://doi.org/10.1103/PhysRevD.14.870

[17] Higuchi, A., Matsas, G.E.A. and Sudarsky, D. (1992) *Physical Review D*, **45**, 3308. https://doi.org/10.1103/PhysRevD.45.R3308

[18] Higuchi, A., Matsas, G.E.A. and Sudarsky, D. (1992) *Physical Review D*, **46**, 3450. https://doi.org/10.1103/PhysRevD.46.3450

[19] Crispino, L.C.B., Higuchi, A. and Matsas, G.E.A. (1998) *Physical Review D*, **58**, 084027 [gr-qc/9804066]. https://doi.org/10.1103/PhysRevD.58.084027

[20] Candelas, P. and Sciama, D.W. (1983) *Physical Review D*, **27**, 1715. https://doi.org/10.1103/PhysRevD.27.1715

[21] Kleinert, H. (1996) Quantum Equivalence Principle. Cargese, Functional Integration, 67-92 [Quant-ph/9612040].

Journal of Modern Physics, 2018, 9, 1793-1806
http://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

On Properties of Torsional Metric Spaces

Hector Alejandro Trejo-Mandujano[1], George H. Goedecke[2]

[1]Universidad Autónoma de Ciudad Juárez, Chihuahua, México
[2]Physics Department, New Mexico State University, Las Cruces, NM, USA
Email: htrejo@uacj.mx, ggoedeck@nmsu.edu

How to cite this paper: Trejo-Mandujano, H.A. and Goedecke, G.H. (2018) On Properties of Torsional Metric Spaces. *Journal of Modern Physics*, **9**, 1793-1806.
https://doi.org/10.4236/jmp.2018.99113

Received: July 12, 2018
Accepted: August 6, 2018
Published: August 9, 2018

Abstract

We apply general tensor calculus to arbitrary nonrelativistic classical Lagrangian systems and derive general relationships between the torsion tensor and other quantities associated with the coordinate configuration space, such as the metric, the Christoffel symbols, the Euler-Lagrange equations, the affine connections, and the curvature tensor. Using Euler angle metric spaces as examples of spaces with nonzero torsion, we calculate these quantities for nonrelativistic rigid rotators of arbitrary structure. For free rotations, we show that the Euler-Lagrange equations agree with the manifestly correct Euler equations.

Keywords

Torsion, Rigid Rotators, Euler Angle Metrics

1. Introduction

In 1922, Élie Cartan included torsion [1] in Einstein's general relativity [2] [3]. Work continued on the Einstein-Cartan theory attempts to associate particle spin and/or rotating matter with torsion [4] [5]. Other investigations considered general effects of torsion in space-time [6] and in a gravitational model of the strong interaction [7]. In addition, several authors (including Einstein) investigated the torsion concept in an attempt to unify electromagnetism with gravity [8] [9] [10]. Trautman [11] also rewrote the Einstein field equations to include nonzero torsion and solved the system for a model consisting of a spinning cloud of dust previous to the big bang. Shapiro [12] investigated effects of including a torsion background in quantized fields. Despite these and many other investigations, the physical role of torsion is still unclear, and modern general relativity is still torsion-free.

In what follows, in Section 2, we apply general tensor calculus to classical

nonrelativistic Lagrangian systems. We establish general relations between the torsion tensor and other quantities in the coordinate configuration space, such as basis vectors, the metric, the affine connections, the Christoffel symbols, the Euler-Lagrange equations, and the curvature tensor. In Section 3, we consider the often-used "zyz" Euler angle spaces associated with rigid rotators as examples. Very few authors have pursued these examples, e.g. [13] [14]. We outline the development of a symbolic computer algorithm that calculates the essential geometrical quantities for anisotropic rigid rotators. The theory and the computational results, as well as hand calculations, show that the associated Euler angle 3-spaces are indeed non-Euclidean flat spaces with torsion, and that for free rotators the Euler-Lagrange equations agree with the Euler equations derived from conservation of angular momentum. In Section 4, we discuss our results and provide a list of conclusions and suggestions for future work in this area. In **Appendix**, we report the results of calculations for the popular yaw, pitch, and roll angles.

2. General Theory

We represent the N generalized coordinates of a classical system by the set $x=\left[x^1,\cdots,x^N\right]$, where each x^p is an independent real continuous variable that may have any physical dimension and any range. Typical generalized coordinates are Cartesian and radial variables having dimension length, and spherical polar and Euler angles having no dimension. A nonrelativistic system moves on a trajectory in this N-dimensional configuration space (N-space) given by

$$\left[x^p=X^p(t),p=(1,\cdots,N)\right]$$

where t is the time. The space may or may not be a Euclidean space E_N or a Riemannian space R_N [15].

2.1. Euler-Lagramge Equations

For nonrelativistic non-dissipative classical systems, the Lagrangian functional L has the form

$$L=(1/2)mg_{pq}\dot{X}^p\dot{X}^q+mu_p(X,t)\dot{X}^p-w(X,t). \tag{1}$$

Here, the first term is the kinetic energy, m is an overall mass parameter that may be chosen at will, $g_{pq}=g_{qp}$ is called the metric of the N-space, and an overdot indicates a time derivative. We follow the extended Einstein summation convention that repeated indices are summed over from 1 to N. The $u_p(X,t)$ are the covariant components of an N-vector potential, and $w(X,t)$ is an N-scalar potential energy. We insert the mass parameter so that the distance ds between infinitesimally separated points is given by the ususal expression, $\mathrm{d}s^2=g_{pq}(x)\mathrm{d}x^p\mathrm{d}x^q$. If the system contains particles having different masses and/or different moments of inertia, those quantities appear in the metric, ratioed to m. The conjugate momenta are

$$P_r = \partial L / \partial \dot{X}^r = m g_{rq} \dot{X}^q + m u_r. \tag{2}$$

The Euler-Lagrange (EL) equation is

$$\mathrm{d}P_r / \mathrm{d}t = \partial L / \partial X^r. \tag{3}$$

This equation is *assumed* to be the correct physical motion equation. When applied to Equations (2) and (1) it yields the detailed EL motion equation

$$g_{rq} \ddot{X}^q + [r, pq] \dot{X}^p \dot{X}^q = -\partial_t u_r - m^{-1} \partial_r w + \left(\partial_r u_p - \partial_p u_r \right) \dot{X}^p, \tag{4}$$

where $\partial_r = \partial / \partial X^r$. The quantities $[r, pq] = [r, qp]$ are the Christoffel symbols of the first kind, defined by

$$[r, pq] = (1/2)\left(\partial_p g_{rq} + \partial_q g_{rp} - \partial_r g_{pq} \right). \tag{5}$$

Equation (4) reduces to the geodesic equation if the potentials u_p and w are constants.

2.2. Basis Vectors

We use the coordinate basis vector approach to general tensor calculus in metric spaces; see *e.g.* [16] [17]. In this work, we need only the covariant (subscripted) basis vectors $\boldsymbol{e}_p(x)$, which are tangents to the coordinate curves [18]. In general, they are not unit vectors and they carry physical dimensions; for example, the covariant basis vectors in an Euler angle 3-space have the dimension of length, as shown below. The generalized dot or inner products among these basis vectors are defined so that

$$\boldsymbol{e}_p \cdot \boldsymbol{e}_q = g_{pq}. \tag{6}$$

We also need the inverse matrix g^{pq} such that $g^{pq} g_{qr} = \delta_r^p$, the Kronecker delta.

These generalized-coordinate basis vectors at any point x may be expressed as a linear combinations (LC) of a set of N Cartesian unit basis vectors $\hat{\boldsymbol{e}}_i, i = (1, \cdots, N)$ that are independent of position x in the N-space:

$$\boldsymbol{e}_p(x) = \Lambda_{pi}(x) \hat{\boldsymbol{e}}_i, \tag{7}$$

where the Cartesian basis vectors satisfy $\hat{\boldsymbol{e}}_i \cdot \hat{\boldsymbol{e}}_j = \delta_{ij}$, the Kronecker delta metric. In many cases the generalized coordinates x^p may be obtained by a global coordinate transformation $x^p(\tilde{x}^i) \leftrightarrow \tilde{x}^i(x^p)$ from global Cartesian coordinates $\tilde{x}^i$, *e.g.*, consider spherical polar coordinates $[x^p] = [r, \theta, \phi]$ in Euclidean 3-space. Then the matrix (Λ) is given by $\Lambda_{pi} = \partial \tilde{x}^i / \partial x^p$. In such cases, the matrix elements satisfy the integrability condition $\partial_q \Lambda_{pi} = \partial_p \Lambda_{qi} = \partial^2 \tilde{x}^i / \partial x^p \partial x^q$. In other cases, such as torsional spaces, this condition is not satisfied, and the space is non-Euclidean. Such matrices (Λ) may always be found. They provide the easiest route to detailed expressions for the affine connections and the torsion tensor for Euler angle and other isolated spaces.

2.3. Affine Connections and Torsion

In this work, we consider only *isolated* spaces, as defined recently by Goedecke [18]. In such spaces, the coordinate derivatives of the N-space basis vectors at

any point x may be written as LC's of those same basis vectors:

$$\partial_p \boldsymbol{e}_q(x) = \Gamma^r_{pq}(x)\boldsymbol{e}(x). \tag{8}$$

(This relation is not valid in non-isolated spaces such as some subspaces of higher-dimensional ones, *e.g.*, a 2-sphere embedded in Euclidean 3-space [18].) The coefficients Γ^r_{pq} are exactly the conventional affine connections. In terms of the matrix Λ defined by Equation (7), the connections are given by

$$\Gamma^r_{pq} = \left(\partial_p \Lambda_{qi}\right)\left(\Lambda^{-1}_{ir}\right). \tag{9}$$

In isolated spaces, the matrix Λ is square, and its inverse exists.

These connections themselves do not constitute a mixed tensor of third rank [1]. However, their antisymmetric parts,

$$T^r_{pq} = (1/2)\left(\Gamma^r_{pq} - \Gamma^r_{qp}\right) = -T^r_{qp} \tag{10}$$

do form a tensor of third rank, called the torsion tensor. In terms of the matrix Λ,

$$T^r_{pq} = (1/2)\left(\partial_p \Lambda_{qi} - \partial_q \Lambda_{pi}\right)\left(\Lambda^{-1}_{ir}\right). \tag{11}$$

Evidently, in any Euclidean space, the torsion tensor is zero in all coordinate systems.

A pertinent general relationship exists among the Christoffel symbols, the symmetrized affine connections, and the torsion tensor. To derive it, we apply Equations (6) and (8), and obtain

$$\boldsymbol{e}_r \cdot \partial_p \boldsymbol{e}_q = g_{rs}\Gamma^s_{pq} \equiv \Gamma_{r,pq}; \quad \boldsymbol{e}_q \cdot \partial_p \boldsymbol{e}_r = \Gamma_{q,pr},$$

and add these to get

$$\partial_p g_{qr} = \Gamma_{r,pq} + \Gamma_{q,pr}. \tag{12}$$

Then, using this equation and Equations (5) and (10), we obtain

$$[r, pq] = \Gamma_{r,(pq)} + T_{qpr} + T_{pqr}, \tag{13}$$

where the notation $\Gamma_{r,(pq)}$ means the symmnetric part, *i.e.*

$\Gamma_{r,(pq)} \equiv (1/2)\left(\Gamma_{r,pq} + \Gamma_{r,qp}\right)$.

2.4. Curvature

We assume an isolated space, as above, so that Equation (8) holds. We follow Goedecke's method [18] by calculating

$$\left(\partial_p\partial_q - \partial_q\partial_p\right)\boldsymbol{e}_r = R^t_{rpq}\boldsymbol{e}_t, \tag{14}$$

where

$$R^t_{rpq} \equiv \partial_p\Gamma^t_{qr} - \partial_q\Gamma^t_{pr} + \Gamma^s_{qr}\Gamma^t_{ps} - \Gamma^s_{pr}\Gamma^t_{qs} \tag{15}$$

is the usual 4th rank curvature tensor of the space. But since $\left(\partial_p\partial_q - \partial_q\partial_p\right)(\text{anything}) = 0$ provided the second partials exist, which they do almost everywhere, then $R^t_{rpq} = 0$ for an isolated space, so such a space is by definition flat. For example, isolated Euler angle 3-spaces are indeed flat, despite

being torsional and thus non-Euclidean, as we verify by direct calculation of Equation (15) in the next section.

Note that in a torsional space, the affine connections that appear in the curvature tensor are *not* equal to the Christoffel symbols of the second kind. Instead, they satisfy $\Gamma^{s}_{pq} = \Gamma^{s}_{(pq)} + T^{s}_{pq}$, and $\Gamma_{s,(pq)}$ satisfies Equation (13), whereby

$$\Gamma^{s}_{pq} = \left[{}^{s}_{pq}\right] + T^{s}_{pq} - g^{sr}\left(T_{qpr} + T_{pqr}\right) \tag{16}$$

is a general relation valid in all N-spaces. This relation and Equation (13) were derived by Schouten [15], using quite different nomenclature and notation.

3. Rotations of Rigid Bodies

This familiar topic provides calculable examples of the general relations presented above for torsional spaces. Leonhard Euler (1707-1783) introduced the angles named for him as coordinates to describe all possible rotations of any rigid body about a center point, usually the center of mass. These angles expressed as functions of time connect a Cartesian triad of basis vectors fixed in a rotating rigid body to a non-rotating Cartesian reference triad, both with origins at the body center of mass, which itself may be translating nonuniformly under the influence of various force fields. In this section we investigate the properties of Euler angle metric spaces associated with a rigid body having arbitrary structure.

3.1. Euler Angles and Angular Velocities

We choose the Euler angles $\left(\alpha^1, \alpha^2, \alpha^3\right)$ corresponding to the (α, β, γ) used by Arfken [19]. This choice is called the "zyz" set [20], corresponding first to a rotation by α^1 about the non-rotating Cartesian system z-axis, then a rotation by α^2 about the new y-axis, then a rotation by α^3 about the final z-axis. Altogether there are 12 sets of Euler angles that may be used, but the principal results of our investigation will be the same for all sets. (In **Appendix**, we present these results for another popular set of Euler angles.)

Let the $\hat{\bar{\boldsymbol{e}}}_i$ be the unit basis vectors of the "barred" Cartesian coordinate system that rotates with the object, and the $\hat{\boldsymbol{e}}_i$ be the same for the non-rotating ("unbarred") system, where $i = 1, 2, 3$. These basis vectors are related by

$$\hat{\bar{\boldsymbol{e}}}_i = R_{ij}\left(\alpha(t)\right)\hat{\boldsymbol{e}}_j \tag{17}$$

where

$$R(\alpha) = R^z\left(\alpha^3\right)R^y\left(\alpha^2\right)R^z\left(\alpha^1\right) \tag{18}$$

is the rotation matrix for given set $\left[\alpha^b\right]$. The constituent matrices R^z and R^y are given by

$$R^z(\mu) = \begin{pmatrix} \cos\mu & \sin\mu & 0 \\ -\sin\mu & \cos\mu & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad R^y(\mu) = \begin{pmatrix} \cos\mu & 0 & -\sin\mu \\ 0 & 1 & 0 \\ \sin\mu & 0 & \cos\mu \end{pmatrix}. \tag{19}$$

We may obtain the angular velocity 3-vector $\boldsymbol{\omega}$ from the defining equation

$$\frac{\mathrm{d}\hat{\bar{\boldsymbol{e}}}_i(t)}{\mathrm{d}t} \equiv \boldsymbol{\omega} \times \hat{\bar{\boldsymbol{e}}}_i(t). \tag{20}$$

With $\boldsymbol{\omega} = \hat{\boldsymbol{e}}_i \omega_i = \hat{\bar{\boldsymbol{e}}}_i \bar{\omega}_i$, we also obtain expressions for the nonrotating and rotating Cartesian components of the angular velocity from Equations (20) and (19):

$$\omega_i = a_{ib}\dot{\alpha}^b; \quad \bar{\omega}_i = b_{ib}\dot{\alpha}^b. \tag{21}$$

These equations imply that

$$b_{ib} = R_{ik}a_{kb} \to (b) = (R)(a), \tag{22}$$

where (b) and (a) are 3×3 matrices given by

$$(a) = \begin{pmatrix} 0 & -\sin\alpha^1 & \sin\alpha^2\cos\alpha^1 \\ 0 & \cos\alpha^1 & \sin\alpha^2\sin\alpha^1 \\ 1 & 0 & \cos\alpha^2 \end{pmatrix}; \quad (b) = \begin{pmatrix} -\sin\alpha^2\cos\alpha^3 & \sin\alpha^3 & 0 \\ \sin\alpha^2\sin\alpha^3 & \cos\alpha^3 & 0 \\ \cos\alpha^2 & 0 & 1 \end{pmatrix} \tag{23}$$

The relations in this subsection are independent of the structure of the rigid body that rotates with the barred system.

3.2. Kinetic Energy and Metric

The rotational kinetic energy for a general rotating rigid body is

$$T = \frac{1}{2}\left[\bar{I}_1\bar{\omega}_1^2 + \bar{I}_2\bar{\omega}_2^2 + \bar{I}_3\bar{\omega}_3^2\right] = \frac{m}{2}g_{bc}\dot{\alpha}^b\dot{\alpha}^c, \tag{24}$$

where $\bar{I}_i$ are the three principal moments of inertia in the rotating (barred) body-fixed system; we have chosen a principal axis rotating Cartesian system, as we may always do. Combining Equations (21) and (24) yields

$$g_{bc} = m^{-1}\bar{I}_i b_{ib}b_{ic} = B_{ib}B_{ic}, \tag{25}$$

where for convenince we have defined the quantities

$$B_{ib} = l_i b_{ib}, \tag{26}$$

with principal axis radii of gyration l_i given by

$$l_i = \sqrt{\frac{\bar{I}_i}{m}}. \tag{27}$$

Applying Equation (23) and Equations (25)-(27) yields the following metric components:

$$\begin{aligned} g_{11} &= l_3^2\cos^2\alpha^2 + l_1^2\cos^2\alpha^3\sin^2\alpha^2 + l_2^2\sin^2\alpha^3\sin^2\alpha^2 \\ g_{12} &= \left(l_2^2 - l_1^2\right)\cos\alpha^3\sin\alpha^2\sin\alpha^3 = g_{21} \\ g_{13} &= l_3^2\cos\alpha^2 = g_{31} \\ g_{22} &= l_2^2\cos^2\alpha^3 + l_1^2\sin^2\alpha^3 \\ g_{23} &= 0 = g_{32} \\ g_{33} &= l_3^2 \end{aligned} \tag{28}$$

The basis vectors $\boldsymbol{e}_b$ of the Euler angle 3-space yield the correct metric, Equation (25):

$$\boldsymbol{e}_b = B_{ib}\hat{\boldsymbol{e}}_i \leftrightarrow \hat{\boldsymbol{e}}_i = B_{bi}^{-1}\boldsymbol{e}_b \tag{29}$$

3.3. Affine Connections

Using Equations (29) and (17), we obtain

$$\boldsymbol{e}_b = \Lambda_{bj}\hat{\boldsymbol{e}}_j, \tag{30}$$

where

$$\Lambda_{bj} = B_{ib}R_{ij}. \tag{31}$$

Then, using Equation (30) and its inverse yields $\partial_c \boldsymbol{e}_b = \left(\partial_c \Lambda_{bj}\right)\left(\Lambda_{jd}^{-1}\right)\boldsymbol{e}_d$, from which we infer an expression for the affine connection

$$\Gamma_{cb}^{d} = \left(\partial_c \Lambda_{bj}\right)\left(\Lambda_{jd}^{-1}\right). \tag{32}$$

Equations (30) and (32) correspond to the general Equations (7) and (9). While it is possible to hand-calculate detailed expressions for these affine connections and Christoffel symbols in terms of the Euler angles, the algebra is difficult for a general anisotropic rotator. In Section 3.6, we outline our construction of a symbolic computer algorithm that produces the desired expressions.

3.4. Euler Equations

The well-known Euler equations [20] for a *torque-free* rigid rotator follow from the constancy of the body's intrinsic angular momentum 3-vector $\boldsymbol{S}$. A simple derivation that takes advantage of rotating basis vectors begins with the expression

$$\boldsymbol{S} = \hat{\boldsymbol{e}}_i \overline{I}_i \overline{\omega}_i. \tag{33}$$

We apply Equation (20) to obtain

$$\mathrm{d}\boldsymbol{S}/\mathrm{d}t = 0 = \hat{\boldsymbol{e}}_k \overline{I}_k \dot{\overline{\omega}}_k + \boldsymbol{\omega} \times \hat{\boldsymbol{e}}_i \overline{I}_i \overline{\omega}_i. \tag{34}$$

Expanding the cross-product using $\boldsymbol{\omega} \times \hat{\boldsymbol{e}}_i = \hat{\boldsymbol{e}}_k \epsilon_{kji} \overline{\omega}_j$ yields the Euler equations for $\overline{\omega}_k$:

$$\overline{I}_k \dot{\overline{\omega}}_k + \epsilon_{kji} \overline{I}_i \overline{\omega}_j \overline{\omega}_i = 0. \tag{35}$$

We expect these equations to be equivalent to the geodesic equation applied to an anisotropic free rotator:

$$g_{ad}\ddot{\alpha}^d + [a,bc]\dot{\alpha}^b\dot{\alpha}^c = 0. \tag{36}$$

We may investigate the equivalence of Equations (35) and (36) as follows. First, substitute Equations (21) and (27) into Equation (35), then multiply the whole equation by b_{ka} and use Equation (25), to obtain

$$g_{ad}\ddot{\alpha}^d + \tilde{\Gamma}_{a,(bc)}\dot{\alpha}^b\dot{\alpha}^c = 0, \tag{37}$$

where

$$\tilde{\Gamma}_{a,bc} = b_{ka}\left[I_k^2 \partial_c b_{kb} + I_i^2 \epsilon_{kji} b_{jc} b_{ib}\right] \tag{38}$$

are defined as the "Euler affine connections", with symmetrized forms defined by

$$\tilde{\Gamma}_{a,(bc)} = (1/2)\left(\tilde{\Gamma}_{a,bc} + \tilde{\Gamma}_{a,cb}\right). \tag{39}$$

These results show that the relation

$$\tilde{\Gamma}_{a,(bc)} = [a,bc] \tag{40}$$

must be satisfied in order that the Euler equations for a free rotator are indeed equivalent to the geodesic equations. The hand calculations of the $\tilde{\Gamma}_{a,(bc)}$ quantities are also quite difficult for anisotropic rotators. We verified Equation (40) for such rotators using the algebraic computer algorithms described in Section 3.6 below and also using hand calculations.

3.5. Isotropic Rotators

We make hand calculations for an isotropic rotator, defined as any rigid body for which $I_1 = I_2 = I_3 = I$. The metric reduces to

$$g = I^2 \begin{pmatrix} 1 & 0 & \cos\alpha^2 \\ 0 & 1 & 0 \\ \cos\alpha^2 & 0 & 1 \end{pmatrix}. \tag{41}$$

The Christoffel symbols of the first kind are easily calculated:

$$[1,23] = [3,12] = -[2,13] = -(1/2)I^2 \sin\alpha^2, \tag{42}$$

with $[a,bc] = [a,cb]$ and with all others zero.

One fairly efficient way to obtain expressions for the affine connections in this isotropic case is to begin with Equation (30), put $I_i = I$, note that $R_{ij}R_{ik} = \delta_{jk}$, and find that

$$\boldsymbol{e}_b = I a_{ib}\hat{\boldsymbol{e}}_i, \tag{43}$$

where a_{ib} are the elements of the matrix (a) given by Equation (23). Then, continue by using the derivatives and the inverse of Equation (43), and obtain $\partial_c \boldsymbol{e}_b = \left(\partial_c a_{ib}\right)\left(a^{-1}_{di}\right)\boldsymbol{e}_d$, which reveals that the affine connections are given by

$$\Gamma^d_{cb} = \left(\partial_c a_{ib}\right)\left(a^{-1}_{di}\right). \tag{44}$$

If we then apply $\Gamma_{a,cb} = g_{ad}\Gamma^d_{cb}$, we obtain

$$\Gamma_{1,23} = \Gamma_{3,12} = -\Gamma_{2,13} = -I^2 \sin\alpha^2, \tag{45}$$

with all others zero. Therefore, the symmetrized affine connections are simply equal to

$$\Gamma_{1,(23)} = \Gamma_{3,(12)} = -\Gamma_{2,(13)} = -\left(I^2/2\right)\sin\alpha^2, \tag{46}$$

with of course $\Gamma_{1,(32)} = \Gamma_{1,(23)}$, etc., and with all others zero.Then, using $T_{abc} = (1/2)\left(\Gamma_{a,bc} - \Gamma_{a,cb}\right)$, we obtain the covariant torsion tensor,

$$T_{abc} = -\left(I^2/2\right)\sin\alpha^2 \epsilon_{abc}, \tag{47}$$

where ϵ_{abc} is the completely antisymmetric Levi-Civita 3-index symbol. As indicated by Equation (13), Equation (47) verifies that the symmetrized affine connections are equal to the Christoffel symbols for an isotropic nonrelativistic

rotator.

It remains to be shown that the intrinsic angular momentum is conserved for the isotropic free rotator. The condition for that conservation is Equation (40). From Equation (39) with $l_i = l$ one obtains easily

$$\tilde{\Gamma}_{a,(bc)} = \left(l^2/2\right) b_{ka} \left(\partial_c b_{kb} + \partial_b b_{kc}\right),$$

because the terms in ϵ_{kji} in Equation (39) cancel out. Then, using Equation (23), it is a simple matter to hand calculate $\tilde{\Gamma}_{a,(bc)}$ and show that indeed Equation (40) is valid.

These results verify that this Euler angle metric space is indeed a space with torsion, whereby it is not a Euclidean space. It must also be a flat space because it is an isolated space. Using the values of the affine connections $\Gamma^d_{cb} = g^{da}\Gamma_{a,cb}$ that can be obtained easily from Equation (45), it is straightforward if a bit laborious to hand-calculate all the components of the curvature tensor R^a_{bcd} defined by Equation (15) and show that they are indeed zero.

3.6. Anisotropic Rotators

In this section we investigate properties of the "zyz" Euler angle 3-spaces for rigid rotators of arbitrary shape and structure.

3.6.1. Equations in Matrix Form

The computer program that we developed for the general algebraic calculations requires the equations to be in matrix form. In matrix notation, Equation (25) for the metric is simply

$$(g) = (B)^{\mathrm{T}} (B), \tag{48}$$

where $(B)^{\mathrm{T}}$ is the transpose of (B). In order to calculate the affine connections, we first define the matrix (M_c) by specifying its elements

$$(M_c)_{bd} = \Gamma^d_{cb}. \tag{49}$$

Then Equation (32) in matrix form is

$$(M_c) = \left(\partial_c (\Lambda)\right)(\Lambda)^{-1}, \tag{50}$$

where (Λ) is the matrix with elements Λ_{bj} defined by Equation (31). Equation (50) is very convenient computationally.

3.6.2. Method and Results

Our computer algorithm calculates the affine connections, the Euler affine connections, and the torsion tensor in any Euler angle metric space for any rigid rotator, starting from Equation (50). The algorithm also calculates the corresponding curvature tensor, using Equation (15), and the Christoffel symbols, using Equations (28) and (5). We selected MATLAB [21], [22] to perform the symbolic calculations. Equations (52)-(54) below summarize the results of our calculations for the "zyz" Euler angles. The results are written using the following

abbreviations:

$$\sin\alpha^i \to s_i; \quad \cos\alpha^i \to c_i; \quad 2\alpha^3 \to 2a3. \tag{51}$$

Equations (52) show the nonzero Christoffel symbols $[a,bc]$, which as expected are symmetric under interchange of their last two indices, and were computationally found to be equal to the corresponding symmetrized Euler affine connections $\tilde{\Gamma}_{a,(bc)}$. This result means that the Euler Equation (35) are indeed equivalent to the geodesic equations for a torque-free rotator. Equations (53) present the nonzero symmetrized affine connections $\Gamma_{a,(bc)}$. Comparison with (52) reveals that these symmetrized connections are *not* equal to the Christoffel symbols. Equations (54) list the covariant torsion tensor T_{abc}. Note that it is *not* proportional to the Levi-Civita symbol ϵ_{abc}. Also, note that in the special case of isotropic rotators, for which all $l_i = l$, the quantities in all these equations reduce to their values calculated in Subsection 3.5.

$$\begin{aligned}
[1,12] &= c_2 s_2 \left(l_1^2 c_3^2 - l_2^2 c_3^2 + l_2^2 - l_3^2\right) = [1,21] \\
[1,13] &= -s_2^2 \sin(2a3)\left(l_1^2 - l_2^2\right)/2 = [1,31] \\
[1,22] &= -c_2 c_3 s_3 \left(l_1^2 - l_2^2\right) \\
[1,23] &= -s_2 \left(\cos(2a3) l_1^2 - \cos(2a3) l_2^2 + l_3^2\right)/2 = [1,32] \\
[2,11] &= -c_2 s_2 \left(l_1^2 c_3^2 - l_2^2 c_3^2 + l_2^2 - l_3^2\right) \\
[2,13] &= s_2 \left(-\cos(2a3) l_1^2 + \cos(2a3) l_2^2 + l_3^2\right)/2 = [2,31] \\
[2,23] &= \sin(2a3)\left(l_1^2 - l_2^2\right)/2 = [2,32] \\
[3,11] &= s_2^2 \sin(2a3)\left(l_1^2 - l_2^2\right)/2 \\
[3,12] &= -s_2 \left(-\cos(2a3) l_1^2 + \cos(2a3) l_2^2 + l_3^2\right)/2 = [3,21] \\
[3,22] &= -\sin(2a3)\left(l_1^2 - l_2^2\right)/2
\end{aligned} \tag{52}$$

$$\begin{aligned}
\Gamma_{1,(12)} &= \sin(2a2)(l_1 + l_2 + l_3)\left(l_2 - l_3 + l_1 c_3^2 - l_2 c_3^2\right)/4 = \Gamma_{1,(21)} \\
\Gamma_{1,(13)} &= -s_2^2 \sin(2a3)(l_1 - l_2)(l_1 + l_2 + l_3)/4 = \Gamma_{1,(31)} \\
\Gamma_{1,(22)} &= -l_3 c_2 c_3 s_3 (l_1 - l_2) \\
\Gamma_{1,(23)} &= -s_2 \left(-l_1^2 s_3^2 + l_1^2 - l_1 l_2 - l_3 l_1 s_3^2 + l_3 l_1 + l_2^2 s_3^2 + l_3 l_2 s_3^2\right)/2 = \Gamma_{1,(32)} \\
\Gamma_{2,(11)} &= -c_2 s_2 \left(l_1 l_2 - l_1 l_3 + l_1 l_3 c_3^2 - l_2 l_3 c_3^2\right) \\
\Gamma_{2,(12)} &= -c_2 c_3 s_3 (l_1 - l_2)(l_1 + l_2 - l_3)/2 = \Gamma_{2,(21)} \\
\Gamma_{2,(13)} &= s_2 \left(l_1^2 s_3^2 - l_1 l_2 + l_3 l_1 s_3^2 - l_2^2 s_3^2 + l_2^2 - l_3 l_2 s_3^2 + l_3 l_2\right)/2 = \Gamma_{2,(31)} \\
\Gamma_{2,(23)} &= \sin(2a3)(l_1 - l_2)(l_1 + l_2 + l_3)/4 = \Gamma_{2,(32)} \\
\Gamma_{3,(11)} &= l_3 s_2^2 \sin(2a3)(l_1 - l_2)/2 \\
\Gamma_{3,(12)} &= -l_3 s_2 \left(l_3 - l_1 \cos(2a3) + l_2 \cos(2a3)\right)/2 = \Gamma_{3,(21)} \\
\Gamma_{3,(22)} &= -l_3 \sin(2a3)(l_1 - l_2)/2
\end{aligned} \tag{53}$$

$$
\begin{aligned}
T_{121} &= \frac{c_2}{2}\left(l_1^2 c_3^2 - l_1 l_2 - l_1 l_3 c_3^2 + l_1 l_3 - l_2^2 c_3^2 + l_2^2 + l_2 l_3 c_3^2 - l_3^2\right) = -T_{112} \\
T_{131} &= c_3 s_3 \left(c_2^2 - 1\right)\left(l_1 - l_2\right)\left(l_1 + l_2 - l_3\right)/2 = -T_{113} \\
T_{132} &= s_2 \left(l_1^2 s_3^2 - l_1^2 + l_1 l_2 - l_3 l_1 s_3^2 + l_3 l_1 - l_2^2 s_3^2 + l_3 L_2 s_3^2\right)/2 = -T_{123} \\
T_{221} &= -c_2 c_3 s_3 \left(l_1 - l_2\right)\left(l_1 + l_2 - l_3\right)/2 = -T_{212} \\
T_{232} &= s_3 c_3 \left(l_1 - l_2\right)\left(l_1 + l_2 - l_3\right)/2 = -T_{223} \\
T_{213} &= \frac{s_2}{2}\left(-l_1^2 s_3^2 + l_1 l_2 + l_3 l_1 s_3^2 + l_2^2 s_3^2 - l_2^2 - l_3 l_2 s_3^2 + l_3 l_2\right) = -T_{231} \\
T_{312} &= -l_3 s_2 \left(l_1 + l_2 - l_3\right)/2 = -T_{321}
\end{aligned}
\tag{54}
$$

In addition, we calculated the Riemann-Christoffel curvature tensor for an anisotropic rotator using Equation (15) and the calculated values of the affine connections. We found that all components are zero, as predicted.

4. Conclusions and Discussion

4.1. Conclusions

The principal goals of this work were to investigate properties of affinely connected metric spaces with torsion, and to provide applications to the familiar Euler angle spaces that describe rigid body rotations. The results allow the following conclusions:

1) In any affinely connected metric space, the symmetrized affine connections are equal to the corresponding Christoffel symbols if and only if the covariant torsion tensor is completely antisymmetric or zero (see Equation (13));

2) Any Euler angle 3-space is a flat space with torsion. The covariant torsion tensor is completely antisymmetric if and only if the rigid rotator being described is isotropic;

3) For torque-free rotations of all rigid bodies, the Euler-Lagrange equations that follow from the kinetic energy expressed in terms of the Euler angles agree with the Euler equations that follow from constancy of the intrinsic angular momentum vector;

4) Torsion has no direct effect on the coordinate motions of any non-relativistic Lagrangian system, because the torsion tensor does not appear in the Euler-Lagrange motion Equations (4);

5) Torsion does have an effect on the detailed form of the Riemann-Christoffel curvature tensor in *any* space, because the affine connections that appear therein are not equal to the corresponding Christoffel symbols if the space is torsional; see Equations (15) and (16).

As emphasized, the first and last of these conclusions apply in any N-space, whereby they apply to relativistic systems. Extension of these results to nonrelativistic systems of many rigid bosies is trivial: We simply have three Euler angles and three CM coordinates for each body in the system. Extension to relativistic many-body systems should be pursued, but it may not be simple, and is beyond the scope of this work.

4.2. Discussion

One topic that merits discussion is the significance of torsion. We should consider two types of significance, geometrical and physical.

The principal geometrical significance follows directly from the definition of the affine connection in terms of coordinate basis vectors, Equation (13), which provides the insight that $\partial_p \boldsymbol{e}_q = \partial_q \boldsymbol{e}_p$ in any torsion-free space. At first glance it seems rather unlikely that these partial derivatives would be the same for $p \neq q$. However, as shown earlier, they *are* equal in all curvilinear coordinate systems x in any Euclidean space. A corollary insight is also well-known: An isolated N-space with torsion does not admit any global Cartesian or pseudo-Cartesian coordinate systems, and is thus a non-Euclidean space. Another geometrical significance was noted in conclusion v) just above.

The physical significance of torsion is still unknown. Several possible interpretations and applications have been proposed, as discussed in the Introduction; we will not repeat that discussion here. We have shown that nonzero torsion always accompanies rigid body rotations.

Acknowledgements

Part of this work is based on part of the Ph.D. dissertaion research of H. T-M.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] McConnell, A.J. (1957) Applications of Tensor Analysis. Dover, New York.

[2] Cartan, E. (1922) *Comptes Rendus de l'Académie des Sciences* (*Paris*), **174**, 593-595.

[3] Trautman, A. (2006) *Encyclopedia of Mathematical Physics*, **2**, 189-195. https://doi.org/10.1016/B0-12-512666-2/00014-6

[4] Friedrich, W., Von Der Heyde, P., Kerlick, D. and Nester, J. (1976) *Reviews of Modern Physics*, **48-3**, 393-416.

[5] Petti, R.J. (1986) *General Relativity and Gravitation*, **18**, 441-460. https://doi.org/10.1007/BF00770462

[6] Capozziello, S., Lambiase, G. and Stornaiolo, C. (2001) *Annalen der Physik*, **513**, 713-727. https://doi.org/10.1002/1521-3889(200108)10:8<713::AID-ANDP713>3.0.CO;2-2

[7] Fedosin, S.G. (2009) *Journal of Vectorial Relativity*, **4-1**, 1-24.

[8] Horie, K. (1994) Geometric Interpretation of Electromagnetism in a Gravitational Theory with Space-Time Torsion. ArXiv Eprints hep-th/9409018.

[9] Suhendro, I. (2008) *Progress in Physics*, **1**, 31-37.

[10] Popławski, N. (2010) *International Journal of Theoretical Physics*, **49-7**, 1481-1488.

[11] Trautman, A. (1973) *Nature Physical Science*, **242**, 7-8. https://doi.org/10.1038/physci242007a0

[12] Shapiro, I. (2002) *Physics Reports*, **357-2**, 113-213.

[13] Goedecke, G. (2014) Statistical Approach to Quantum Mechanics II: Nonrelativistic Spin. ArXiv e-Prints 1408.1721.

[14] Fiziev, P. and Kleinert, H. (1995) Motion of a Rigid Body in Body-Fixed Coordinate Systems—For Autoparallel Trajectories in Spaces with Torsion. ArXiv e-Prints, hep-th/9503075.

[15] Schouten, J. (1989) Tensor Analysis for Physicists. 2d Edition, Dover, New York.

[16] Lichnerowicz, A. (1962) Elements of Tensor Calculus. Methuen, London.

[17] Hartle, J. (2003) Gravity: An Introduction to Einstein's General Relativity. Addison-Wesley, Reading.

[18] Goedecke, G. (2011) *The European Physical Journal Plus*, **126**, 32-41. https://doi.org/10.1140/epjp/i2011-11032-x

[19] Arfken, G. (1985) Mathematical Methods for Physicists. 3d Edition, Academic Press, Orlando.

[20] Goldstein, H. (1980) Classical Mechanics. 2nd Edition, Addison-Wesley, Reading.

[21] (2010) Matlab. The MathWorks Inc.

[22] Handelman, D. and Littlefield, B. (2005) Mastering Matlab 7. Pearson/Prentice-Hall, Englewood Cliffs.

Appendix: Other Euler Angles

In Equations (A1) and (A2) below we present the results of calculations for anisotropic rigid bodies using "xyz" rotations, with Euler angles known as yaw, pitch, and roll angles. The results have the same character as those in Equations (52) and (54), and they reduce to the correct values for isotropic bodies.

$$
\begin{aligned}
&[1,12] = c_2 s_2 \left(l_1^2 + l_2^2 c_3^2 - l_2^2 - l_3^2 c_3^2\right) = [1,21] \\
&[1,13] = c_2^2 \sin(2a3)\left(l_2^2 - l_3^2\right)/2 = [1,31] \\
&[1,22] = -c_2 c_2 s_3 \left(l_2^2 - l_3^2\right) \\
&[1,23] = -c_2 \left(l_1^2 - \cos(2a3) l_2^2 + \cos(2a3) l_3^2\right)/2 = [1,32] \\
&[2,11] = -c_2 s_2 \left(l_1^2 + l_2^2 c_3^2 - l_2^2 - l_3^2 c_3^2\right) \\
&[2,23] = -\sin(2a3)\left(l_2^2 - l_3^2\right)/2 = [2,32] \\
&[2,31] = c_2 \left(l_1^2 + \cos(2a3) l_2^2 - \cos(2a3) l_3^2\right)/2 = [2,13] \\
&[3,11] = -\left(c_2^2 \sin(2a3)\left(l_2^2 - l_3^2\right)\right)/2 \\
&[3,12] = -c_2 \left(l_1^2 + \cos(2a3) l_2^2 - \cos(2a3) l_3^2\right)/2 = [3,21] \\
&[3,22] = \sin(2a3)\left(l_2^2 - l_3^2\right)/2
\end{aligned}
\tag{A1}
$$

$$
\begin{aligned}
&T_{121} = c_2 s_2 \left(l_1^2 - l_1 l_2 c_3^2 + l_1 l_3 c_3^2 - l_1 l_3 + l_2^2 c_3^2 - l_2^2 + l_2 l_3 - l_3^2 c_3^2\right)/2 = -T_{112} \\
&T_{131} = c_2^2 \sin(2a3)\left(l_2 - l_3\right)\left(l_2 - l_1 + l_3\right)/4 = -T_{113} \\
&T_{132} = c_2 \left(l_2^2 c_3^2 - l_2^2 + l_2 l_3 - l_1 l_2 c_3^2 + l_1 l_2 - l_3^2 c_3^2 + l_1 l_3 c_3^2\right)/2 = -T_{123} \\
&T_{221} = -\left(c_3 s_2 s_3 \left(l_2 - l_3\right)\left(l_2 - l_1 + l_3\right)\right)/2 = -T_{212} \\
&T_{231} = -c_2 \left(-l_2^2 c_3^2 + l_2 l_3 + l_1 l_2 c_3^2 + l_3^2 c_3^2 - l_3^2 - l_1 l_3 c_3^2 + l_1 l_3\right)/2 = -T_{213} \\
&T_{232} = -\left(\sin(2a3)\left(l_2 - l_3\right)\left(l_2 - l_1 + l_3\right)\right)/4 = -T_{223} \\
&T_{321} = l_1 c_2 \left(l_2 - l_1 + l_3\right)/2 = -T_{312}
\end{aligned}
\tag{A2}
$$

Journal of Modern Physics, 2018, 9, 1807-1813
http://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Impact of Degrees of Freedom of Polyatomic Ideal Gas on Efficiency of Three-Stroke-Cycle Heat Engines

Haiduke Sarafian

University College, The Pennsylvania State University, York, PA, USA
Email: has2@psu.edu

How to cite this paper: Sarafian, H. (2018) Impact of Degrees of Freedom of Polyatomic Ideal Gas on Efficiency of Three-Stroke-Cycle Heat Engines. *Journal of Modern Physics*, **9**, 1807-1813.
https://doi.org/10.4236/jmp.2018.99114

Received: July 16, 2018
Accepted: August 6, 2018
Published: August 9, 2018

Abstract

It is customary to evaluate the efficiency of four-stroke-cycle heat engines such as Carnot, Otto and Diesel with a working ideal gas [1] [2] [3]. Here we consider various three-stroke-cycle heat engines that are composed of three out of four special processes: adiabatic, isothermal, isobaric and isochoric. We deviate from the customary norm considering the working material to be a polyatomic ideal gas with various degrees of freedom. We develop analytic expressions for the efficiency of each design as a function of degrees of freedom. For suitable practical values of relevant (V, P, T), we evaluate the corresponding efficiencies.

Keywords

Efficiency of Heat Engine, Three-Stroke-Cycle, Adiabatic, Isothermal Processes

1. Introduction

Because of the essential conceptual and practical importance of the Carnot heat engine [1] that is conducive to introducing fundamental quantities such as *entropy* and *efficiency*, one wrongfully has the misconception that a four-stroke-cycle is the keystone for evaluating the latter. Consequently, a systematic investigation of characteristics of three-stroke-cycle engines is overlooked [1] [2] [3] [4] [5]. As far as the efficiency of engines is concerned, one would think the simplest cycle is the one that is composed of three interlocked rather than four states that are sequentially connected via various thermodynamic processes. Recognizing the void, in this note we consider various three-stroke-cycles composed of three out of four special practical thermodynamic processes namely: adiabatic, iso-

thermal, isobaric and isochoric. For the working material, we consider polyatomic ideal gases. Given the fact that working materials have different degrees of freedom, we formulate the efficiency of each engine as a function of the corresponding properties. Properties of each gas are specified by: volume, pressure, temperature and degrees of freedom, (V, P, T, f), respectively. For a set of practical parameters, we evaluate their efficiencies. This article is composed of three sections. In addition to Introduction, in Section 2, we outline the physics of the problem providing its solution. Section 3 is the Conclusions with comments on how the investigation may be extended.

2. Physics of the Problem and Its Solution

In a V-P diagram a heat engine is shown by a closed loop, see Figure 1. The efficiency is the ratio of the mechanical work done by the cycle *i.e.* the value of the enclosed area to the intake heat, $\tilde{Q}$; direction of the arrow indicates intake. The simplest physically feasible closed loop is formed using three points/states connected via three processes. Throughout this article for the sake of simplicity for a given cycle frequently we interchange the word "leg" and "process". Because we are interested in four specific processes namely: adiabatic, isothermal, isobaric and isochoric, as essential ancillaries we evaluate their corresponding mechanical work and intake heat. For polyatomic ideal gases the states are specified by (V, P, T); *i.e.* volume, pressure and temperature, respectively; degrees of freedom, f, is implicitly included.

2.1. Case 1

We consider a three-stroke-cycle that is composed of a sequence of isothermal expansion followed by an isobaric compression completing the cycle via an isochoric compression. The cycle is shown in the V-P diagram, Figure 2.

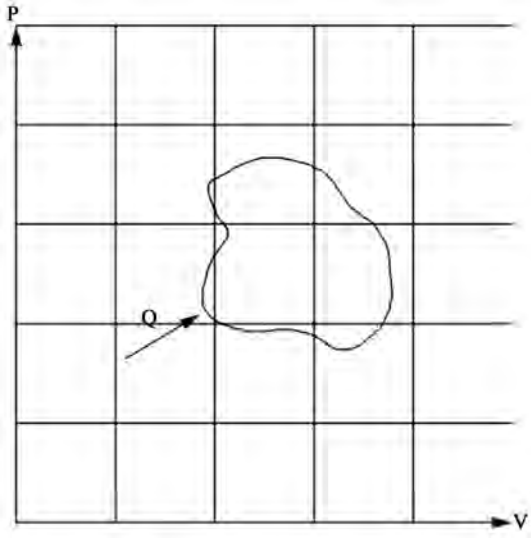


Figure 1. In the shown V-P diagram the closed contour represents a thermodynamic cycle. The enclosed area is the mechanical volumetric work; Q is the intake heart.

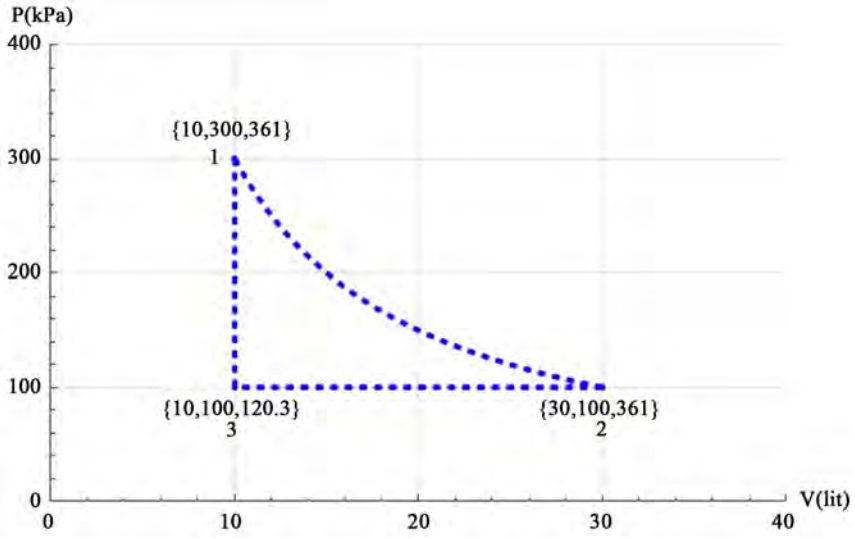


Figure 2. Display of a three-stroke-cycle. Cycle 1231 is composed of isothermal-isobaric-isochoric processes. Units of the axes are Volume in liters (lit) and Pressure in Kilo Pascals (kPa).

In Figure 2 the corners of the cycle are labeled {1, 2, 3}. Each corner represents the state of the gas according to $\{V, P, T\}$ format. For ideal gases these three quantities are related via Equation of state. The equation of state for one mole of an ideal gas is, $PV = RT$, R is the ideal gas constant. In Figure 2 specific numeric values are used. For instance, vertex 1 is labeled {10, 300, 361}; these stand for 10 lit volume, 300 kPa pressure and 361 K temperature, respectfully. These values may be adjusted on a need basis. The net mechanical work for the shown cycle is,

$$W_{1231} = W_{12} + W_{23}$$

where, the volumetric mechanical work is, $W_{12} = \int_1^2 P\mathrm{d}V$. Utilizing the ideal gas equation of state yields, $W_{12} = RT_1 \ln(V_2/V_1)$.

Similarly, for the change of state from 2 to 3 we have, $W_{23} = P_2(V_3 - V_2)$. The net volumetric work for the cycle is,

$$W_{1231} = P_1V_1 \ln\left(\frac{V_2}{V_1}\right) + \left(P_3V_3 - P_2V_2\right), \tag{1}$$

The cycle requires two different intake heats, $\bar{Q}_{12}$ and $\bar{Q}_{31}$. For the isothermal and isochoric legs, we have

$$Q_{12} = P_1V_1 \ln\left(V_2/V_1\right), \text{ and } Q_{31} = C_v\Delta T \equiv fR/2\left(T_1 - T_3\right)$$

where for the molar heat capacity of the ideal gas we applied $C_v = fR/2$, f is the degrees of the freedom. Hence the net intake heat is,

$$Q_{1231} = P_1V_1 \ln\left(\frac{V_2}{V_1}\right) + \frac{1}{2} f\left(P_1V_1 - P_3V_3\right), \tag{2}$$

Efficiency of the cycle is $e = W_{1231}/Q_{1231}$. Utilizing (1) and (2) we arrive at,

$$e = \frac{\ln\left(\frac{V_2}{V_1}\right) + \frac{P_3}{P_1}\left(1 - \frac{V_2}{V_1}\right)}{\ln\left(\frac{V_2}{V_1}\right) + \frac{1}{2} f \left(1 - \frac{P_3}{P_1}\right)}. \quad (3)$$

That is the efficiency, it is written as an explicit function of the degrees of freedom and the relevant numeric parameters that can be read off the V-P diagram, Figure 2. As an example, utilizing numeric values of the vertices in Figure 2 for two different values of degrees of freedom, monatomic and polyatomic with degrees of freedom, $f = 3$ and 8, respectively Equation (3) yields, $e_{f=3} = 20.5\%$, and $e_{f=8} = 11.5\%$. That is the working materials with polyatomic gases with higher degrees of freedom result less efficiencies. This is because polyatomic gases with higher degrees of freedom do require greater intake heat per degree temperature change. One also notes that the contour of the cycle and consequentially the value of the mechanical work is independent of the degrees of freedom but the intake heat depends on *f*.

2.2. Case 2

We consider a three-stroke-cycle that is composed of a sequence of adiabatic expansion followed by an isobaric compression completing the cycle with an isochoric compression. The cycle is shown in the *V*-*P* diagram, Figure 3.

There are differences between the cycles shown in Figure 3 vs. the one shown in Figure 2. For the former state 1 and 2 are connected via an isothermal process where for the latter the states are connected adiabatically. As noted the isothermal process is independent of the degrees of freedom, where the adiabatic

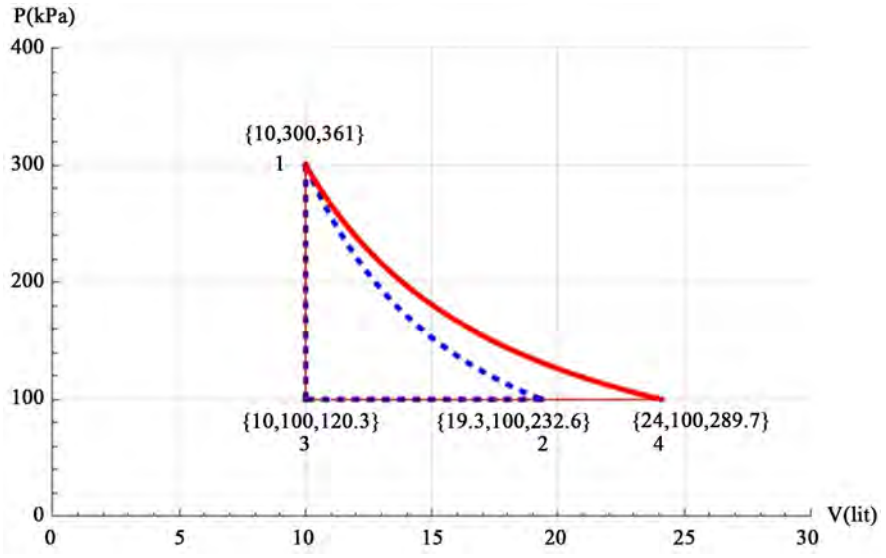


Figure 3. Display of two three-stroke-cycles; cycle 1231 and 1431.Both cycles are composed of adiabatic-isobaric-isochoric processes. Dashed curve corresponds to three degrees of freedom, $f = 3$, the solid curve corresponds to eight degrees of freedom, $f = 8$. Units of the axes are the same as Figure 2, *i.e.* Volume in liters (lit) and Pressure in Kilo Pascals (kPa).

process depends on f. Here we evaluate the volumetric value of the mechanical work done via adiabatic process. For ideal gases subject to adiabatic processes it is known [1] the states are related via PV^{γ} = const, here γ is the adiabatic index and is, $\gamma = C_P/C_V$ *i.e.* the ratio of the molar heat capacity at constant pressure to the molar heat capacity at constant volume. Applying Mayer [1] relationship, $C_P - C_V = R$, and knowing that, $C_V = R(1/2f)$, one gets the γ in terms of the degrees of freedom namely, $\gamma = 1 + 2/f$. Applying these relationships in conjunction with equation of state of ideal gases for states related via adiabatic process one obtains additional useful relationships such as: $TV^{1-\gamma} = \frac{C}{R}$ and $TP^{\frac{1}{\gamma}-1} = \frac{c^{\frac{1}{\gamma}}}{R}$. For the volumetric work $w_{12} = \int_1^2 P\mathrm{d}V$ applying above equations we arrive at, $w_{12} = \frac{C}{1-\gamma}(T_2 - T_1)$. Utilizing w_{23} from Equation (1) the net volumetric work for the cycle shown in **Figure 3** is written as,

$$w_{1231} = \frac{1}{2} f (PV)_1 + (PV)_3 - \left(\frac{1}{2} f + 1\right)(PV)_2, \tag{4}$$

On the other hand, the intake heat of the cycle is given by Equation (2). Utilizing these we arrive at the efficiency of the cycle,

$$e = \frac{1}{P_1 - P_3}\left[P_1 + \frac{2}{f} P_3 - \left(1 + \frac{2}{f}\right) P_2 \left(\frac{V_2}{V_1}\right)\right], \tag{5}$$

As one expects the volumetric mechanical work is a function of the degrees of freedom, yet the intake heat is independent of f. Consequently, depending on value of f the contour of the cycle changes accordingly. Two such contours are depicted in **Figure 3**. The steep dashed curve between state (1) and (2) corresponds to $f = 3$, the soft solid curve connecting state (1) to (4) corresponds to $f = 8$. For two values of degrees of freedom such as, $f = 3$ and 8 Equation (5) yields two values of efficiencies, $e_{f=3} = 22.5\%$ and $e_{f=8} = 12.5\%$.

By comparing the efficiencies of the cycles depicted in **Figure 2** and **Figure 3** we realize their comparative values. It is interesting noting that the impact of the value of f in these two cycles are inverted. For the former the contour of the cycle i.e the volumetric work is f independent, where for the latter it dependents on f. On the other hand, the intake heat has baring on the former and has no impact on the latter.

For comprehensive understanding one needs to compare **Figure 2** and **Figure 3**.

2.3. Case 3

We consider a three-stroke-cycle that is composed of a sequence of adiabatic expansion followed by an isothermal compression closing the cycle via an isochoric compression. One such cycle is shown in the V-P diagram, **Figure 4**.

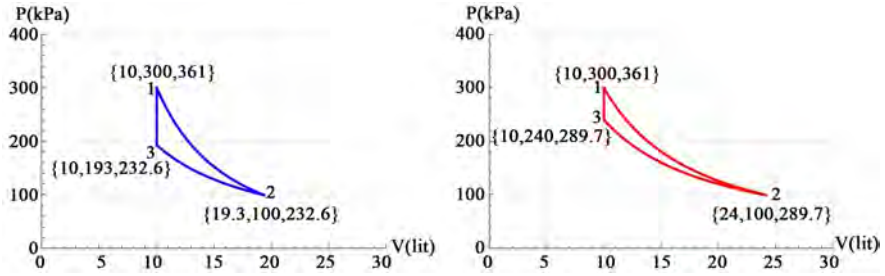


Figure 4. Left cycle clockwise 1231 is composed of adiabatic-isothermal-isometric processes with three degrees of freedom, $f = 3$. The right cycle is like the left diagram corresponding to eight degrees of freedom, $f = 8$. Units of the axes are the same as **Figure 2**, *i.e.* Volume in liters (lit) and Pressure in Kilo Pascals (kPa).

Evaluation of the efficiency of the cycle shown in **Figure 4** as a function of degrees of freedom is straight forward. The volumetric mechanical work and intake heat of each process leg of the cycle is given in the previous cases. Utilizing these expressions, the efficiency of the given cycle yields,

$$e = \frac{(PV)_1 - (PV)_2 + \frac{2}{f}(PV)_2 \ln\left(\frac{V_3}{V_2}\right)}{V_1(P_1 - P_3)}. \quad (6)$$

Applying Equation (6) for the numeric values of the vertices shown in **Figure 4**, we arrive at the efficiencies, namely, $e_{f=3} = 21.0\%$ and $e_{f=8} = 12.4\%$.

3. Conclusions

Our investigation fills in the missing systematic information on the efficiency of three-stroke-cycle heat engines. To form a comprehensive understanding, we narrowed our investigation to practical cycles that are composed of three out of four thermodynamic processes: adiabatic, isothermal, isobaric and isobaric. We derived analytic expressions for the efficiency of each designed three-stroke-heat engine. As expected the efficiencies are explicit functions of degrees of freedom of the polyatomic gases. This underlines the fact that working material of the designed engine plays a substantial role. This contrasts with the famous Carnot heat engine where its efficiency is independent of the working material [6]. Applying derived formulations namely, Equations (3) and (5) and (6) to the set of relevant practical numeric $\{V, P, T\}$ values shown in **Figures 1-4** we calculate the efficiencies for two distinct values of degrees of freedom. Analyzing these values, we deduce the fact that for our designed three-stroke-heat engines the efficiencies of high and low degrees of freedom are related via, $e_{f=8} \sim 0.5 e_{f=3}$. While investigating we consider a few more three-stroke-cycles. However, we decided to report the ones that we consider most interesting.

Our investigation sets the stage for extending systematic analysis for four and five-stroke-cycle heat engines conducive to posing questions such as: 1) What is the relationship between the efficiency and the number of the strokes of a cycle? and 2) Is it possible to design a multiple-stroke-cycle heat engine that its efficiency is independent of the working material?

A note to the readers: All the figures and the implicit calculations in this article are carried out utilizing *Mathematica* resources [7] [8].

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Longair, M.S. (1994) Theoretical Concepts in Physics. Cambridge University Press.

[2] Wolfgang, D. and Westfall, G. (2011) University Physics with Modern Physics. McGraw-Hill.

[3] Halliday, D., Resnick, R. and Walkers, J. (2011) Fundamentals of Physics. 9th Edition, John Wiley & Sons.

[4] Marion, J. and Hornyak, W. (1984) Principle of Physics. Saunders College Publishing.

[5] Tipler, P. and Mosca, G. (2008) Physics for Scientists and Engineers. 6th Edition, W.H. Freeman and Company.

[6] Sarafian, H. (IMS2003) *Proceedings of the* 5*th International Mathematica Symposium*, Imperial College London Press, London.

[7] Wolfram, S. (1996) Mathematica Book. 3rd Edition, Cambridge University Press, Cambridge.

[8] Sarafian, H. (2015) Mathematica Graphics Example Book for Beginners. Scientific Research Publishing.

Journal of Modern Physics, 2018, 9, 1814-1826
http://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

The Universe, Origen and Destiny

Juan L. Lartigue

UNAM, México City, México
Email: jmlg5@hotmail.com

How to cite this paper: Lartigue, J.L. (2018) The Universe, Origen and Destiny. *Journal of Modern Physics*, **9**, 1814-1826.
https://doi.org/10.4236/jmp.2018.99115

Received: June 19, 2018
Accepted: August 12, 2018
Published: August 15, 2018

Abstract

In a previous paper [1] [2] [3], the author has proposed the following concepts: the Hubble field, the constant value of acceleration of the Universe expansion, a De Sitter-Hubble gravitational equation, and the possibility of an imaginary time, all of them mainly oriented to a hypothesis about the Universe destiny. This work mentions some theories about the origin of the Universe and suggests the existence of a limit for the physical Universe. Accordingly to a relevant author [4], there are two main ways in considering these themes: those methods based on physical theories and those founded in metaphysical concepts of philosophy and religion. After some comments on such theories, this paper agrees that, in the physical search, the Big Bang seems to be the most probable origin of the physical Universe, space and time. On the metaphysical approach, the *creation ex nihilo* seems to be the commonest hypothesis for such origin. Besides, it is assumed that the Creator would have set two driving and successive conditions: transcendence and ethics. Reference is made to the arrow of time and to the proper time, a physical concept aimed to define both frozen and imaginary times, which would correspond to a space without time, and another space with imaginary time. A diagram is proposed for the Euclidic space, including a world-line throughout the space-like.

Keywords

Hubble Field

1. The Ancient Cosmologies

In the oldest religions and philosophies, the origin of the Universe was attributed to divinities. Five relevant examples are the following:

1) The pharaonic period of Egypt from (VI B.C.) to (I A.C.) centuries. Its cosmology assumed that, initially, there existed only the Nun, identified as the nothingness, though it contained a Demiurge who, finally, did create the Un-

iverse. Besides, there existed the god Nether, who was rather assumed as having a cosmic function [5].

2) The Toltec religion (X b. C.) had not many gods; it just assumed the existence of a supreme divinity as the creator of the Universe (**Figure 1**), an unnamed and amorphous entity who is invisible and powerful, some deity of the art of "living in equilibrium"; it was represented, much later, by the cyclic feathered snake figure (Quetzalcoatl). Unfortunately, the internal wars and the external conquest caused losses of main relics [6].

3) The Genesis, written about (X b. C.) states that the creation of the Universe was decided by God and being made by periods. Tributary religions to that cosmology are Judaism, Christianity and the Islam.

4) The ancient Greek philosophers had any cosmological approach. Aristotle (IV b. C.) and other eminent Greek philosophers were strongly adhered to the geocentric hypothesis, at such an extent that the later heliocentric model, proposed by Aristarchus, remained out of their reach for a long time. Incredibly, Ptolomeus developed a big work to locate more than one thousands of stars (II a. C.) and he wrote a metaphysical text (Almagesto); by that time, Euclid stated the principles of Geometry that has been the basis for modern scientific theories. The Greek religion was always polytheist, with Zeus as the main divinity. Yet, it must be said that the Aristotle principle about "the first cause" or prime motor was the basis for creationism in subsequent religions, like Aquinas's doctrine in the catholic theology [7].

5) The Roman Empire inherited the Greek culture, including the Pantheon of Gods. However, both Romans and Greeks had a common adoption (II a. C), the Iranian deity Mithras, who was assumed to be the creator of the Universe; that belief began to decline at the time of Constantine [8].

Figure 1. The toltec universe [9].

2. Cosmology in the Middle Ages and Renaissance

In these epochs Cosmology had an important development, founded initially on Religious beliefs, based after on scientific criteria that were often faced to a church opposition.

1) The Islam religion was founded by Mahoma (575-632). His God, Alá, is the origin and creator of the Universe.

2) In the Middle Ages, the physical Cosmology made almost no progress because of Catholic Church opposition (**Figure 2**) [9]. Search was cautiously oriented to Astronomy as it occurred with the works of the Greek Ptolomeus (II A.C.), whose books, such as Almagesto, were adopted by astronomers of the Middle Age and translated to the Arab, so arriving to a final version by Al-Magisti [10].

3) Saint Agustin de Hipona (Algery, 354-430) mentions three themes in his Theodicea: the third one was the Creation problem[1] where he declares that "man has three properties: memory, understanding and intention, which correspond to the tree persons of the Trinity" [11].

4) In Italy, the astronomer G. Bruno (XVI a. C.) did conceive a Universe which is infinite in size and time. Besides, he said that it could exist some inhabited worlds in the universe, an assertion that took him to death.

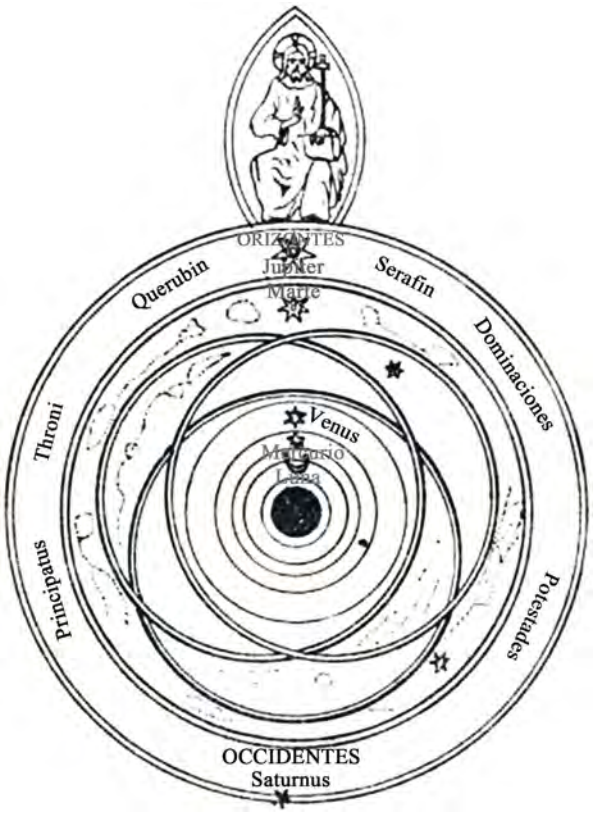


Figure 2. The universe in the middle age [9].

[1]Dios creó al mundo según modelos que son las ideas inmutables y eternas que existían en la mente del creador.

5) N. Copernicus and G. Galilei (The Father of Science, XVI a. C.) remained loyal to the heliocentric theory; their works in Astronomy and Physics have exerted a long-standing influence, yet though their ways of thinking were the objects of dispute with the Church [12].

6) J. Kepler (1571-1630) and Tycho Brahe were two important astronomers. The three laws of Kepler are basics in Astronomy, thanks to the support of Tycho Brahe [13].

7) I. Newton (1642-1772), the greatest astronomer and natural philosopher of the Renaissance, was born in England. He assumed a static yet infinite Universe and deduced three basic laws of mechanics as well as the equation of gravitational attraction [14].

8) R. Descartes (1596-1650), a French philosopher and scientist, considered to be the father of Modern Philosophy [15]. He was the founder of Rationalism ("I think, so I exist") and wrote that "God is the only cause of events". He worked many topics, including the Cartesian coordinates.

9) S. Laplace (1749-1827) did work in the mathematical field with Lagrange and developed the transform between real and imaginary variables. In his "System of the World", he proposed a nebular theory for the origin of the Universe [16].

3. Modern Cosmologies

After the period of development of the today known as classical physics and cosmology, the XIX and XX centuries were specially rich in the arise of new theories on these and related topics such as: modern mathematics, experimental tools, chemistry, thermodynamics, astronomy and the theory of relativity. A brief review of the last one follows.

1) A. Einstein (1879-1955), a German scientist, he was the creator of the Theory of Relativity (Special in 1905 and General in 1915). He developed many new physical concepts and obtained the Nobel Prize due to his work on the photoelectric effect. It must be pointed out that Einstein was based on the works of Geometry of Euclide, Lorentz, Riemann, Minkowski, and Lobachewski, as well as on eventual communication with other scientist such as Poincaré, Sommerfeld, and Weyl. The Special Theory is aimed to demonstrate the relativity of movements and times as functions of the reference frames, *i.e.* the relative velocities of masses and observers; so, he defined the transcendental concept of proper time. After the Special Theory, Einstein tried to incorporate the acceleration factor; then, he developed the space-time concept and replaced the Newtonian definition of force by the space-time curvature, *i.e.* "Space-time tells mass how to move and mass tells space-time how to curve" [17]. His gravitational equation has two terms: the left one refers to space-time as a function of the metric; the right one expresses the energy and momentum in the Universe. Einstein field equations gave not static but dynamical solution; so, to obtain a static one according to his own time, he introduced a cosmological constant, Λ in the left side. In that concerning, Einstein did assume a not-expanding Universe and he

maintained such a view till knowing the A. Friedmann and E. Hubble discoveries.

2) W. De Sitter (1872-1934). A German physicist. He worked with Einstein in 1932 though he later modified the former's gravitational equation by assuming a Universe empty of matter and energy [18]; so, he cancelled the right side of the equations, probably based on the low density of the Universe.

3) A. Friedmann (1888-1925). A Russian mathematician and meteorologist who proposed to eliminate the Λ parameter of Einstein, in two new basic cosmological equations. A problem with one of them was that it affords negative values for the Universe acceleration [1]. Besides, he introduced the curvature parameter in a later form of the RWFL equation [19].

4) G. Lemaitre (1889-1934). A Belgian priest, he was the first in predicting the Universe expansion and calculating a very important parameter value in 1927, two years before than Hubble's confirmation. Lemaitre was also the first one proposing the primeval atom as the origin of the Big Bang, so called by Gamow in 1942. Besides, he cooperated with several scientists in defining new concepts, as the coordinate system of RWFL, necessary to study several Cosmological models [20].

5) E. Hubble (1889-1953) was an American astronomer who provided many fresh concepts to modern Cosmology. He showed that some distant nebula were galaxies in fact. Based on the variable luminosity of the cepheids, he determined distances [21] and velocities between them, and, finally, he deduced his famous law that includes the very important parameter H (s^{-1}).

4. Recent Cosmological Theories

Several cosmological theories have been developed after those of Einstein, Friedmann, and Lemaitre. Some of them are the following:

1) The static Universe model, equivalent to those of the Greek geocentric philosophy age and subsequently to the Middle Age beliefs, developed many centuries later. As mentioned above, the main trouble for Einstein was the acceptance of the Universe expansion.

2) E. Milne (1896-1950) proposed a Universe having infinite time and size, the Static Universe theory, which had many critics [22].

3) In 1948 some theories of the Steady State [23] Universe, as those of Bondi, Gold, Hoyle and Narlikar, among others, reappeared. Theirs were not a Static Universe since it may grow. This theory does not accept the Big Bang, yet it assumes the constancy of the Universe density; so it concerns to a continuous creation of matter. After the general acceptation of the Cosmic Background Radiation and the Universe expansion, the Steady State theory was discarded.

4) The Λ theory, intended to justify a static Universe, was finally rejected by Einstein about 1931, after knowing the Hubble discovery. However, the Λ sign has been taken again by new cosmologists to show a theory that would result opposed to Einstein's original aim: it is that the Universe expansion has been originated by a mysterious dark energy [24].

5) A probably more realistic theory has been proposed by the author, in references [1] [3], based on the measurable Hubble parameter and a Hubble scalar field.

6) The favorite trending theory to combine general relativity with quantum mechanics is the super strings theory, whose results are not clear enough until now [25] [26].

5. Additional Comments on the Scientific Theories

In previous paragraphs they were mentioned the ancient beliefs (religious, philosophical and scientific) about the Universe creation. Additional comments on the scientific theories will be intended in the following section. In this respect, just two views thrive: the eternal Universe and the *creation ex-nihilo*.

1) The eternal existence of the Universe, as mentioned in paragraph (IV-2), has been sustained by some cosmologists. However, this model has been rejected because some facts are not accountable within the theory, *i.e.*, the Cosmic Background Radiation and the already confirmed Universe expansion since, if it had an origin, it would correspond to a singularity.

2) There is the Conformal Cyclic Cosmology by R. Penrose, stating that the Big Bang singularity may occur after every aeon ($\backsim 10^{100}$ years). He applies a model based on the Second law of Thermodynamics and the Phase Space, to show that the last step of the Universe expansion should be formed by photons [27].

3) P. Teilhard de Chardin (1881-1955), a French priest. He was the founder of the Cosmic Theology and did argue that the Omega point stands for the maximum level of knowledge that human kind could reach at the end of the Universe. This theory was later developed by F. Tipler [28]. Both Chardin and Tipler kept concerned with the Universe destiny.

4) In respect to the *creation ex-nihilo,* there are still two options: the Big Bang and the Quantum Theory.

Regarding the later, reference [29] states that vacuum is not empty; its energy seems to be zero because it fluctuates between positive and negative energy; so, "the Universe emerged spontaneously from a random quantum fluctuation in a primordial vacuum"; M. Gleisser adds: "a scientific explanation to the origin of the Universe cannot be self-contained. Sometimes we must have the humility to accept that our modes of explanation have limits". The same reference mentions a S. Hawking conclusion: "The Universe came out of quantum nothingness, though the absolute nothingness is not scientific. That is the best that Science can do"[2].

5) The opinion of J. Polkinghorne [31] about the Quantum Theory acceptance seems very interesting: "The discovery of Quantum Theory had produced the

[2]The final paragraph of Hawking's book expresses, in the Spanish translation [30]… "si descubrimos una teoría completa… comprensible para todos…entonces seremos capaces de tomar parte en la discusión de por qué existe el Universo y por qué existimos nosotros… sería el triunfo definitivo de la razón humana porque entonces conoceríamos el pensamiento de Dios".

deepest modification that Newtonian Physics had ever suffered… It was much more radical than that of the time relativity… Einstein was the last one of the ancients... his resistance to the modern quantum theory was a real loyalty to the older thinking forms". However, in spite of its success in Nuclear Physics, the Quantum Theory has not obtained acceptation equal to that of the Big Bang Theory respecting the cosmological origin of the Universe. Nevertheless, the extensive use of Quantum Theory in Nuclear Physics, in Chemistry and other sciences, has motivated big changes in the concept of the inherent incertitude of the micro and macroscopic measurements, as it were a generalization of the Heisenberg Principle. So, the original quantum singularity theory would still have something more to say.

6) Though the Big Bang has been universally accepted as the fittest cosmological theory (so, being accepted by the Catholic Church nowadays [32]), the cosmologist F. Hoyle had proposed some alternate models, probably originated in his own definition of religion as "escapism". He frequently uses the Democritus sentence: *ex nihilo, nihil fit*.

7) As it was said above respect to the Big Bang theory, G. Lemaitre proposed the first model of the Primeval Atom (1927), later named the Big Bang by G. Gamow (1942). Lemaitre's theory was soon confirmed by E. Hubble's discovery of the Universe expansion in 1929. It was in 1998 that reference [33] demonstrated the acceleration of this expansion. Reference [3] proposed in 2016 the constant value of such acceleration. These tests of Universe expansion, beside the already mentioned CRB and the matching chronology of nucleo-synthesis with the stars formation, make the Big Bang the most probable origin of the Universe and time. Besides, it works well with the Standard Model of Nuclear Physics. So, it may be assumed that, in the *creation ex nihilo*, the Universe and time started in a singularity (the Big Bang) *i.e.*, the Universe (space, matter and time) had a common beginning; and space is expanding itself at a constant acceleration.

8) Associated to the Big Bang, there is the theory of *creation continua*, which assumes that Creation was not complete at the Big Bang and it continues forever. Many years before, Aquinas prevented on such idea when he wrote in (XIII a. C.) [7]: "The World process is dynamical, *i.e.* it implies changes though it does not *create* new things from nothing. Any created being may *create* nothing, absolutely. Only God can *create* and He did it already, at the beginning". The theory of *creation continua* coincides, partially, to that of paragraph (IV-3) of 'a steady state', which assumes a continuous creation of matter in order to maintain a constant density in the Universe. It should be, rather, a continuous change in the original matter.

6. Metaphysical Aspects

1) In his book on "The Spirit of God", P. Davies [34] wrote[3]:

[3]A metaphysical interpretation is necessary to understand the Universe origin.

After some concepts on the relevance of the Physical laws[4], he suggested that the "Cosmic Code" *i.e.* the set of all of the physical laws would have been created simultaneously to the Universe and, also, the idea that all of the physical laws could be unified in a unique mathematical order, *i.e.* the Theory of Everything.

2) Davies continues by asking: why has God decided to create the Universe?[5].

In the French translation Davies states [34]: "Il me semble que, si l'on persévère avec le principe de raison suffisante… nous n'avons d'autre choix que de chercher une explication qui transcende le monde physique-dans quelque chose de métaphysique-parce que un Univers physique contingente ne peut renfermer en soi sa propre explication".

The frequent question about the instant when the Universe was created has been refuted by the assumption that time began along with singularity. Conjectures about a time before the Big Bang seem to be non-sense nowadays. Anyway, the imputed fate of God creating the Universe has not fully been explained to date. A cautious and preliminary attempt will be made in the following paragraphs.

3) Those assuming God as the creator of the Universe (that contents matter, space and time) have defined Him as an eternal, all-mighty, omnipresent and omniscient being. It seems that God happily existed in his own reign, free from any apparent need. Suddenly, He decided to create the Universe. Matter was so created in the form of very hot radiation (in the new space), so fulfilling a divine order to expand, cool and eventually condense itself in masses which must combine during indefinite time and under a pre-set of laws.

These laws induced matter to partially appear as radiations and masses such as the first elementary particles that evolved to form nuclei that reacted and added up to make big bodies, as stars, planets and galaxies. Later, by means of His own laws, God breathed life on Earth and also, by evolution processes, He printed some type of intelligence on his first_creatures. Accordingly to reference [7] the word "creation" has two meanings: the *creation ex-nihilo* and some transformations of matter at given times; (the best example would be the human concept *pro-creation*).

4) From Genesis [35], God said "that created was good" and also "Let us make the man at our image and similarity". So, by the above mentioned evolution, man was created along with a soul that includes the highest degree of intelligence which mankind has developed in its thriving to survive and understand his own origin as well as the God's will. In such a task, man has no other means than his own culture, religion and imagination. Such attempts have been always based on the already referred similarity between the man soul and God.

5) Why has God done so? The similarity between man and God is not referred, obviously, to the qualities of eternity or infinite wisdom, rather to the

[4]L'existence de régularités dans la Nature est un fait mathématique objectif. En revanche, les formulations dites lois, rencontrés dans les manuels scientifiques, sont des inventions humaines destinées à refléter des propriétés réelles de la Nature.

[5]S'il était satisfait d'être de toute éternité sans Universe, pourquoi a-t-il soudain décidé d'en créer un?

so-called spirit that includes intelligence, conscience and free-will. Therefore, the man may dare to explore the possible answers to the big questions of the Universe origin and God's will. Such boldness is supposedly intended in this paper. The first question, that concerning to physical theories, has been answered enough in the twenty century by the Big Bang theorists. However, it lacks a hypothesis about the possible God's intentions in both subjects (transcendence and ethics) a hypothesis that would necessarily be similar to others generated by the human mind. So, a logical first reason would be a God's "programmed need" to apply, in a totally new way, the infinite magnitude of His own power. That may sound rare or absurd; however, it would be difficult to imagine a different cause for such a decision.

Genesis [35] says that, at the 6th day, God had already created the Universe, the Earth and the vegetal and animal lives and He decided (rather, He had programmed), at the 7th day, to create the human beings "at His own image and similitude", *i.e.* including immortal soul or spirit. So, in addition to life, two transcendent faculties He imprinted in humans (free-will and an ethical code), failed with the first couple. However, He decided to maintain such faculties for the future human kind, who is now applied to look for answers to the above mentioned questions.

6) God's intention in creating both the Universe and rational beings has been fairly discussed in previous paragraphs; pending is the other one and perhaps most difficult question in thinking through God's will. In addressing that, it is maintained the assumption that God's mind and that (infinitely lower) of humans are similar. So, it would be necessary to add some assumptions: after the free-will failure or sin of the first couple, God decided to print the code of ethics in the human kind, initially in the human's conscience and later reinforced by the Ten Commandments of the ancient Jewish religion (later adopted by tributary religions). So, both the spiritual transcendence and the ethics of the Creation would be self-evident for God's masterpiece: the human being. Yet, what could have been God's reasons for providing such endowments to humans (transcendence and ethics)? The concept of transcendence has implications for both: the future time and the human beings aware of such endorsement. The future time, that is, the promise, was anyway contained in God's redemptive plan: the human capacity to grasp God's message; however, ethics must be printed in human souls through the conscience and explicitly listed in the Ten Commandments or in the core beliefs of most of the ancient religions. As it is well-known, human history attests that God's will has been frequently ignored or misunderstood, rather than obeyed.

7. The Time Limit

Since the Big Bang, time had been passing, accordingly to Newton, with an apparent normal course during almost 15 billion years. It was in the Special Theory of Relativity that the time magnitude appeared as a function of the relative ve-

locities of a mass and its observer. Then, a new term was coined: the proper time τ, the time value recorded by one observer when he travels with a mass in a given trajectory, while another observer, at rest, records a "personal" time t for that trajectory. As mentioned in reference [3], the relation between both times is: $\tau = t/\gamma$, being γ a function of the mass velocity ($\mathbf{v}$) in such a way that, if $\mathbf{v} < \mathbf{c}$ (the velocity of light), $\tau > 0$; if $\mathbf{v} = \mathbf{c}$, $\tau = 0$; if $\mathbf{v} > \mathbf{c}$, τ becomes an imaginary number. In general, τ is assumed as the time that corresponds to the own and instantaneous situation of the mass. So, when $\mathbf{v} = \mathbf{c}$, the proper time τ results equal to 0, which means a time freezing, *i.e.* time does not pass for a mass that would move at the velocity of light; it would mean that matter could exist in an eternity. Otherwise, If $\mathbf{v} > \mathbf{c}$, the proper time τ (the time measured at the mass position) results to be an imaginary time, *i.e.* it goes outside the four dimensional space of Minkowski; however, it could be assumed to exist in the Euclidic space (**Figure 3**) that includes an imaginary time coordinate [36].

Regarding the arrow of time, whose common image is an arrow traveling outward the space at a constant speed, it could be thank similar to the constant acceleration vector $\boldsymbol{\Gamma}_H$ of the Universe expansion [3]. In fact, out of the three types of arrow of time mentioned by reference [37], only those related to the entropy limit and the Universe expansion may coincide with such similitude.

Now, it will be intended to consider an obvious question: if light velocity ($\mathbf{c}$) is, according to the Special Theory of Relativity, the maximum speed that a mass may reach, does it exist a place or circumstance in the Universe where such a limit could really been reached out? Answering this question implies some conditions. In the first place, the maximum speed $\mathbf{c}$ has been feasible, stated for matter that moves through inertial frames of reference. However, according to the proposed Hubble field, space-time is not a reference frame (rather it contains all of reference frames) and it could be assumed as expanding respect to the Big Bang coordinates ($\mathbf{r} = t = 0$), at a constant acceleration ($\boldsymbol{\Gamma}_H$) that causes, at long term, a higher than $\mathbf{c}$ expansion velocity.

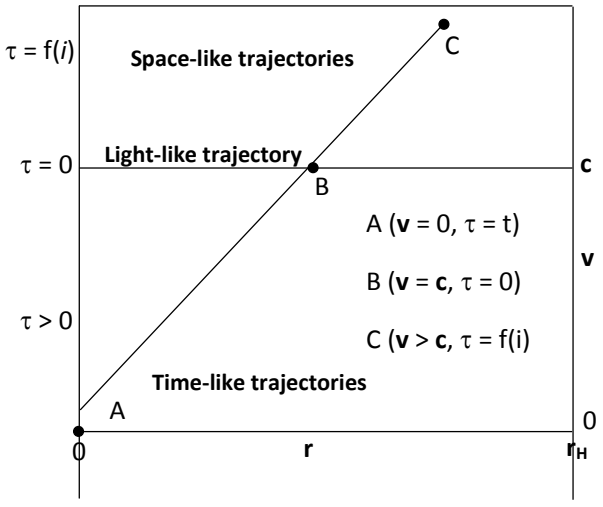


Figure 3. A trajectory in the Euclidic space diagram.

Throughout this expansion the total masses of the Universe have been delayed due to the gravitational acceleration ($\boldsymbol{\Gamma}_G$) that diminishes their net velocities, so driving to the equilibrium now called as the Cosmological Principle. The space-time expansion pulls up all matter either in frames, in co-moving coordinates and, even, if it would have a peculiar velocity. However, beyond the Hubble radius, where speeds would surpass the **c** value, the situation should change because of the time factor. It has been shown [21] that, at the Hubble radius ($\mathbf{r}_H$), the expansion velocity of the space would be equal to **c**. This means that, at a distance ($\mathbf{r}_H$) the proper time of a mass would be 0 and, beyond, the proper time would become an imaginary function. Moreover, at high velocities, masses would have been converted in radiation by annihilation processes; so all matter near the Hubble length, in a tiny gravitational field, it should acquire the acceleration corresponding to that of the space, $\boldsymbol{\Gamma}_H$. Then, by physical laws, how a space being coupled with imaginary time could it to be represented? **Figure 3** shows a simple representation of the Euclidic space, including an assumed matter world-line from a resting state up to a velocity higher than **c**.

Still pending is a meta-physical inquiry. It could be assumed that, since human soul persists after death, it must remain eternally in a space without time, i. e. similar to the actually known as the expanding Hubble sphere surface. Respect to the most outer space, with time assumed as an imaginary function, outlining such a place seems not feasible. The only argument would be that of the Religion that states, for the ethic souls, the ascent to Heaven, a space that remains imaginary for scientific minds.

8. Conclusion

They have been revised some of the ancient cosmological theories, from prehistory to the present time. Mention was made of those founded on ancient religions as well as those of the Middle Age and the Renaissance, developed from astronomical observations. Besides, various derivatives of the theories of relativity and quantum physics are described. An analysis is intended from the meta-physical point of view. A main conclusion is that the most probable origin of the Universe (space, time and matter) was the Big Bang. As well, they have been explored the consequences of overcoming the light velocity, beyond the Hubble sphere, in an imaginary proper time.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Lartigue, J.G. (2016) *Journal of Modern Physics*, **7**, 1607-1615. https://doi.org/10.4236/jmp.2016.712145

[2] Lartigue, J.G. (2017) *Journal of Modern Physics*, **8**, 2159-2166.

https://doi.org/10.4236/jmp.2017.813131

[3] Lartigue, J.G. (2018) *Journal of Modern Physics*, **9**, 1317-1325. https://doi.org/10.4236/jmp.2018.96080

[4] Davies, P. (1995) L'Esprit de Dieu. Éditions du Rocher, Paris, 175.

[5] Nun-Egyptian God. www.britannica.cpm/topic/nun

[6] Marin, G. (2014) La Religión y el Dios Verdadero. Ed. Weatherblink, Mexico.

[7] Ryan, J.K. (1961) *Encyclopedia Americana*, **2**, 114.

[8] Marshall, R. (1961) *Encyclopedia Americana*, **19**, 272.

[9] Ancona, J.A. (1940) Cosmografía. Universidad de Yucatán, México, 53.

[10] Encyclopedia Americana (1961) Americana Corporation, New York, 752.

[11] (2018) Agustín de Hipona y la Filosofía. www.corazones.org

[12] Ancona, J.A. (1940) Cosmografía. Universidad de Yucatán, México, 68.

[13] Knickerbocker, W.S. (1940) Classics of Modern Science. F. S. Crofts, New York, 29.

[14] Maggie, W.F. (1935) A Source Book in Physics. Mc Graw Hill Book Co., New York, 92.

[15] Maggie, W. F. (1935) A Source Book in Physics. Mc Graw Hill Book Co., New York, 50.

[16] Knickerbocker, W.S. (1940) Classics of Modern Science. F. S. Crofts, New York, 138.

[17] Taylor, E.F. and Wheeler, J.A. (2000) Exploring Black Holes. Addison Wesley, New York.

[18] Bergstron, L. and Goobar, A. (2004) Cosmology and Particle Astrophysics. 2nd Edition, Springer, Berlin, Heidelberg, 59-60.

[19] Friedmann, A. and Lawden, D.F. (1982) Introduction to Tensor Calculus, Relativity and Cosmology. Dover Publications, New York, 189.

[20] Lemaitre, G. and Tolman, R.C. (1958) Relativity, Thermodynamics and Cosmology. Clarendon Press, Oxford, 362.

[21] Sartory, L. (1996) Understanding Relativity. University of California Press, Los Angeles, 299.

[22] Milne, E.A. and Tolman, R.C. (op. cit.) 364.

[23] Hawking, S.W. (1998) Historia del Tiempo. Editorial Grijalbo, México, 74.

[24] Dark Energy Task Force Report (2006).

[25] Isham, C.J. (2002) La creación del Universo como un proceso cósmico. In: *Física, Filosofía y Religión*, Observatorio Vaticano, Roma, 457.

[26] Hawking, S.W. (1998) Historia del Tiempo. Editorial Grijalbo, México, 182.

[27] Penrose, R. (2010) Cycles of Time. Bodley Head, London, 137. http://www.bodleyhead.co.uk

[28] Tipler, F. (1997) The Physics of Immortality. Random House, New York, 353.

[29] Gleisser, M. (2013) The Origin of the Universe. https://www.npr.org/sections/13.7

[30] Hawking, S. (1998) Historia del Tiempo. Editorial Grijalbo, México, 224.

[31] Polkinhorne, J. (2002) El Mundo Cuántico. In: *Física, Filosofía y Teología*, El Observatorio Vaticano, 377.

[32] El Papa Francisco apoya la Teoría del Big Bang. Diario El Clarín, Buenos Aires.

[33] Riess, A., *et al.* (1998) *The Astronomical Journal*, **116**, 1009-1038.

[34] Davies, P. (1995) L'Esprit de Dieu. Éditions du Rocher, Paris, 175.

[35] Sagrada Biblia-Génesis, trad. F. Torres (1980), Servagroup Ed., Madrid, ch. II, v. 7, 17.

[36] Hawking, S. (1998) Historia del Tiempo. Editorial Grijalbo, México, 179.

[37] Hawking, S. (1998) Historia del Tiempo. Editorial Grijalbo, México, 191.

Journal of Modern Physics, 2018, 9, 1827-1837
http://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

A Possible Solution to the Disagreement about the Hubble Constant

Frank R. Tangherlini

San Diego, CA, USA
Email: frtan96@gmail.com

How to cite this paper: Tangherlini, F.R. (2018) A Possible Solution to the Disagreement about the Hubble Constant. *Journal of Modern Physics*, **9**, 1827-1837.
https://doi.org/10.4236/jmp.2018.99116

Received: July 19, 2018
Accepted: August 13, 2018
Published: August 16, 2018

Abstract

The current disagreement about the Hubble constant H_0 was described as a "Crisis in Cosmology", at the April (2018) Meeting of the American Physical Society, and hence its resolution is of utmost importance. This work proposes that the solution to the disagreement between the Planck Collaboration cosmic microwave background (CMB) value of H_0, together with the very close BOSS Collaboration baryon acoustic oscillation (BAO) value, and the significantly higher value of H_0 found by the SHOES Collaboration cosmic distance ladder (CDL) work, is due to the fact that the CMB and BAO values of H_0 are not for an accelerating universe, as generally believed, but are actually the values for a decelerating universe. In contrast, the CDL value of H_0 is indeed that for an accelerating universe. It is shown that by replacing the negative deceleration parameter in the expression for $\log H_0$ in the CDL work by a positive deceleration parameter, the value of H_0 can be brought down to agree with the CMB and BAO lower values. There is a brief review of the author's decelerating model based on the Einstein de Sitter universe, augmented by a model of dark energy that does not have a negative pressure, but instead has a non-dispersive index of refraction n, causing the speed of light through the dark energy of intergalactic space to be reduced to c/n. As reported earlier, this assumption is sufficient to accommodate the increase in apparent magnitude of the Type Ia supernovae (SNe Ia). Additional support for the model is presented, together with a proposal for astronomical falsification.

Keywords

Hubble Constant Disagreement, CMB & BAO Determinations, Decelerating Interpretation, CDL Determination & Modification

1. Introduction

The current disagreement about the Hubble constant H_0 is widely recognized as

a serious problem in cosmology, indeed, as indicated in the Abstract, it has been recently described as a "Crisis in Cosmology." The purpose of this work is to show that there is a possible solution to the problem based on the author's proposal in [1], and further developed in [2] [3] [4] [5], that the universe is not accelerating, but is actually decelerating. In Section 2, for the reader's convenience, this alternative model that involves a reduction in the speed of light through the dark energy of intergalactic space (IGS) will be briefly reviewed. In Section 3, it will be proposed that the disagreement between the Planck collaboration [6] CMB value $H_0 = 67.90 \pm 0.55$ km·s^{-1}·Mp·c^{-1}, that utilizes the ΛCDM model, and the SHOES collaboration [7] cosmic distance ladder (CDL) value, $H_0 = 73.24 \pm 1.74$ km·s^{-1}·Mp·c^{-1}, is due to the CMB value for H_0 actually being that for a *decelerating* universe, whereas the CDL value for H_0 is indeed, as presented, that for an *accelerating* universe. The BOSS collaboration [8] BAO value, $H_0 = 67.6 \pm 0.5$ km·s^{-1}·Mp·c^{-1}, is so close to the CMB value, and because it is simpler to deal with analytically, it will be used to justify the decelerating interpretation of the lower CMB and BAO values. It will also be shown that when the CDL expression for $\log H_0$, given in [7], that uses a negative deceleration parameter in its kinematic term, $q_0 = -0.55$, is changed to a positive deceleration parameter, $q_0 = 0.50$, which is appropriate for the decelerating Einstein de Sitter (EdS) universe that is taken as reference, the ratio of the Hubble constant for the accelerating universe to the Hubble constant for the decelerating universe is reduced. In fact, it is shown that for reasonable values of redshifts, the ratio of the two Hubble constants is just the ratio of two kinematic terms that in turn become equal to $H_0(CDL)/H_0(BAO)$, or $H_0(CDL)/H_0(CMB)$. Although the EdS universe was used as a reference decelerating universe, it was shown in [3] that, for Ω_{tot} sufficiently close to unity, a closed universe might alternatively be used. It is also pointed out there that the closed universe yields an upper bound on Λ that is too small to accommodate the accelerating universe, and if the EdS universe can be seen as the limit of a closed universe with infinite radius, then the upper bound on Λ goes to zero.

Since the above results do not necessarily prove that the universe is decelerating, but only make it reasonable, in Section 4, two tests of the decelerating model, that have been discussed earlier, particularly in [4] [5], are briefly reviewed. The first test is based on the search for neutrinos correlated with gamma ray bursts (GRBs), and the second test is based on the possibility of alternative galactic sources for discordant redshifts. In Section 5, there are conclusions.

2. Brief Review of the Decelerating Model

In the decelerating model, it is assumed that the dark energy of intergalactic space (IGS), instead of having a negative pressure, that causes the expansion of the universe to be presently accelerating, has rather an index of refraction n, so that the speed of light through IGS is reduced to c/n. A least squares fit of the EdS universe to the accelerating ΛCDM universe, that is presently favored to ex-

plain the increased apparent magnitude of the Type Ia supernovae (SNe Ia), found by Riess *et al.* [9], Schmidt *et al.* [10], and Pearlmutter *et al.* [11], yielded $n \approx 1.5$. It is further assumed that there is no dispersion across the entire electromagnetic spectrum, based on the original work in the optical range [9] [10] [11], and later investigations covering a wide range of the spectrum by Costa *et al.* [12], Metzger *et al.* [13], and Frail *et al.* [14]. However, within the galaxies themselves, the speed of light remains the vacuum speed *c*. Since Riess *et al.* [15] found that at $z = 1.65 \pm 0.15$ there was no acceleration, it is assumed in this model that before this value of redshift there is negligible dark energy, and there is only dark matter in IGS as well as in the galaxies. However, as the universe expanded and reached the above value of redshift, it is further assumed that because of expansion cooling, the dark matter in IGS underwent a phase transition into dark energy, characterized by the development of an index of refraction that eventually became the above value, $n \approx 1.5$. Only the dark matter in IGS undergoes the phase transition, because it is only IGS that expands, not the galaxies themselves. At the ill-defined boundaries of the galaxies there will be a transition region for n which will be ignored in this preliminary, highly-simplified model. It is outside of the scope of this work to attempt to hypothesize on the nature of the particle or particles that constitute dark matter, since it is reasonable that one should first determine whether this phenomenological model is valid. The least squares fit for n in the author's first three papers [1] [2] [3] assumed the phase transition was completed by $z = 1.0$, so that n was taken as constant from $z = 1.0$ to $z = 0$. However, subsequent work [4] [5] suggested that n became constant somewhat later, possibly at $z \approx 0.6$. The reduced speed of light through IGS allows the EdS universe the additional time to expand to its size at the present epoch, so as to give rise to the increase in distance that is needed to explain the increased apparent magnitude of the SNe Ia. It was shown in [1] that the increase in apparent magnitude, δm, under the assumption n is a constant over the range of redshifts under consideration, is given by

$$\delta m = 5\log\left(1+\left(n-1\right)\ln\left(1+z\right)\right) \tag{1}$$

while the corresponding logarithmic fractional increase in distance d is given by

$$d = \log\left(1+\left(n-1\right)\ln\left(1+z\right)\right) \tag{2}$$

The above value for d *is* used when comparing the logarithmic distance to the standard ruler of the BAO that is fitted by the ΛCDM model, with that fitted by the EdS model that has been augmented by d; see, e.g., Equation (16) in [1], and also (3) below, after putting it into logarithmic form. Since $\delta m = 5d$, the percentage of disagreement for comparison of the model with the ΛCDM fit to the SNe Ia should be the same as the percentage of disagreement for the comparison of the model with the ΛCDM fit to the BAO. This was shown in [1] to be the case by using the tables of comparison given there.

3. Application to the Hubble Constant Disagreement

The resolution of the disagreement about the Hubble constant H_0 that is pro-

posed here begins with the observation that the CMB value not only disagrees with the CDL value, *but is significantly lower than the CDL value*. This would be the case if the CMB value of H_0 were not for an accelerating universe, but were actually for a decelerating universe. In contrast, the CDL value of H_0 is certainly for an accelerating universe, because it explicitly involves the deceleration parameter q_0, to which is assigned the value $q_0 = -0.55$ [7]. The proposed alternative interpretation of the CMB value for H_0 arises from the observation that since $H_0 \equiv \dot{a}_0 / a_0$, where a(t) is the expansion parameter for the flat Friedmann LeMaître Robertson Walker (FLRW) line element $\mathrm{d}s^2 = c^2 \mathrm{d}t^2 - a(t)^2 \mathrm{d}r^2 + r^2 \left(\mathrm{d}\theta^2 + \sin^2 \theta \mathrm{d}\phi^2 \right)$, and since a_0 is necessarily the same for both models, and also since $\dot{a}_0 (accel.) > \dot{a}_0 (decel.)$, then $H_0 (accel.) > H_0 (decel.)$. The inequality for the expansion parameter velocities arises in the following way: In the very early universe, before the accelerating term became significant, both models would be expanding at essentially the same rate, but as the effect of the accelerating term became larger, the ΛCDM universe's deceleration became increasingly less than that of the EdS universe, and eventually the accelerating term became dominant, and caused the universe to accelerate, so that by the times (which are different, less for the ΛCDM universe) each universe has expanded to a_0 the above inequality for the Hubble constants resulted. To be sure, the above qualitative analysis does not guarantee that $H_0 (accel.) > H_0 (decel.)$ by the observed amount, but it is suggestive, and it serves to motivate the quantitative analysis given further below. It will be shown there that by replacing in the expression for $H_0 (CDL)$ a value for the deceleration parameter that is the value for a decelerating universe, so that $q_0 > 0$ instead of $q_0 < 0$ as was used in [7], one can obtain a value of H_0 that agrees with $H_0 (BAO)$ which, as noted above, is essentially the same as $H_0 (CMB)$.

But before carrying out this analysis, it is necessary to show how it could be possible that the reported CMB and BAO values of H_0 could be that for a decelerating universe, since they are presented as being the Hubble constants for the accelerating ΛCDM universe. The key is to note a well-known ambiguity in the term "dark energy." On the one hand, such as in the author's work, one of the interpretations leads to a density parameter for dark energy $\Omega_{de} \approx 0.7$ that is needed to arrive at $\Omega_{tot} \approx 1$, since, as is well-known, dark matter and baryonic matter together only account for about 30% of Ω_{tot}. This was shown independently of the CMB determination, by Verde *et al.* [16], and Hawkins *et al.* [17], based on studies that utilized the 2 dF Galaxy redshift survey. On the other hand, the more frequent use of the term assumes that dark energy is associated with a negative pressure *P*, and in the case of the ΛCDM model, with $P = wU$, in which $w = -1$, and where U is the positive energy density associated with the cosmological term, which necessarily leads to a density parameter of ~0.7 as well. Thus one has the relation $\Omega_\Lambda = \Omega_{de}$, even though the corresponding energy-stress source tensors used in the Einstein field equations are different. Since the way the CMB workers extract H_0 from their measurements is rather complicated, and well beyond the scope of this work, it will be shown for the comparable BAO

determination of the Hubble constant, that $H_0(BAO)$ can be alternatively interpreted as being for a decelerating universe that has the same density of dark energy as the accelerating ΛCDM universe, but does not have the negative pressure associated with the cosmological term.

As was shown in [1], one can obtain the same distances that are found in the ΛCDM universe to the "standard ruler" of the BAO by using instead the decelerating EdS universe, augmented by the fractional distance increase provided by the reduced speed of light, the logarithm of which is given in (2). From the analysis in Equation (9) to Equation (18) of [1], one obtains an Equation for the effective distance, $a_0 r(z)$, that involves the Hubble constant, and that relates the two models, that is given by

$$cH_0^{-1}\mathrm{X}_\Lambda(z) = cH_0^{-1}\mathrm{X}_m(z)\left(1+(n-1)\ln(1+z)\right) \tag{3}$$

where $\mathrm{X}_\Lambda(z)$ and $\mathrm{X}_m(z)$ are dimensionless integrals derived from the field equations for energy densities for the two different universes, and in which each side of the above equation is equal to the effective distance to the standard ruler of the BAO, while z is its redshift. Actually there are two distances that are determined observationally: The angular distance $D_A(z) = a_0 r(z)/(1+z)$, and the longitudinal distance $D_L(z) = a_0 r(z)(1+z)$, so that the effective distance $a_0 r(z)$ is given by $a_0 r(z) = \left(D_A(z) D_L(z)\right)^{1/2}$. For the determination of these distances, that are presented as supportive of the flat ΛCDM universe, see Anderson *et al.* [18], and Anderson *et al.* [19]. The determination of H_0 from the left hand side of (3) by setting $a_0 r(z) = cH_0^{-1}\mathrm{X}_\Lambda(z)$ would seem to yield the Hubble constant for an accelerating universe given by $H_0 = c\mathrm{X}_\Lambda / a_0 r(z)$, which is the current interpretation. However, one can alternatively regard $a_0 r(z)$ as having been set equal to the right hand side of (3) , and hence it would yield the Hubble constant for a decelerating universe with reduced speed of light, which is the interpretation proposed here. To resolve the conflict in the two interpretations, it will be shown next that when the CDL expression for H_0 is taken to be that for a decelerating universe, it can be lowered in value to equal the BAO determination.

In [7], from their Equation (9), one has the following expression for $\log H_0$

$$\log H_0 = \frac{M_x^0 + 5a_x + 25}{5} \tag{4}$$

where M_x^0 is a distance modulus, and from Equation (5) in [7] one has that

$$a_x = \log\left(cz\left\{1+\frac{1}{2}\left[1-q_0\right]z - \frac{1}{6}\left[1-q_0-3q_0^2+j_0\right]z^2 + O\left(z^3\right)\right\}\right) - 0.2m_x^0 \tag{5}$$

where $j_0 = 1$ from prior deceleration, and m_x^0 is another distance modulus. It will be convenient to combine the two distance moduli and the additional constant term into the combined distance modulus M defined by

$$M \equiv \frac{M_x^0 - m_x^0 + 25}{5} \tag{6}$$

and upon introducing the kinematic term $k(z, q_0, j_0)$ defined by

$$k(z,q_0,j_0)\equiv\left\{1+\frac{1}{2}[1-q_0]z-\frac{1}{6}\left[1-q_0-3q_0^2+j_0\right]z^2\right\} \tag{7}$$

in which the correction term $O(z^3)$ has been omitted, one can rewrite $\log H_0$ as

$$\log H_0 = M+\log(cz)+\log k(z,q_0,j_0) \tag{8}$$

In this form, the distance modulus M has an intuitive interpretation. The original Hubble relation, based on the first order Doppler effect, takes the form $H_0D=cz$ where D is the proper distance to the galaxy with recessional speed *cz.* For larger values of redshift, and when acceleration is taken into account, the revised Hubble law takes the form $H_0D=czk(z,q_0,j_0)$, upon neglect of higher order terms, but including a period of deceleration represented by j_0. Hence it follows, upon taking the logarithm of the preceding equation and utilizing (8), that $M=-\log D$. However, the following argument does not make use of the value of *M*, since one is going to subtract from $\log H_0$ for an accelerating universe, $\log H_0$ for a decelerating universe. To simplify the notation, primes will be used to denote the latter, so that upon subtracting, and suitably rewriting the result, one has

$$\log\left(\frac{H_0}{H_0'}\right)=M-M'+\log\left(\frac{k(z,q_0,j_0)}{k(z,q_0',j_0)}\right) \tag{9}$$

After setting $q_0=-0.55$ for the accelerating universe as given in [7], and $q_0=0.5$ for the decelerating EdS universe, equation (9) takes the form

$$\log\left(\frac{H_0(-0.55)}{H_0(0.5)}\right)=M-M'+\log\left(\frac{1+0.775z-0.274z^2}{1+0.25z-0.125z^2}\right) \tag{10}$$

One can determine M from (8) as a function of redshift since the value of $H_0(-0.55)$ is known, or alternatively from its definition in (6). However, in order to obtain M' it will be necessary for the astronomers to re-interpret their data under the assumption that the universe is decelerating, so that one can check whether $H_0(0.5)$ is equal to the BAO value of H_0 after allowing for the measurement uncertainties. In the present case, given the lack of such empirical information for M', one can proceed as follows: Set the ratio of the two Hubble constants, which will be denoted by η, equal to $H_0(CDL)/H_0(BAO)$; in addition, for simplicity, take their mean values as fiducial, so that one has $\eta=73.24/67.6=1.083$. Next, find the behavior of $M-M'$ as function of redshift by using (10) with $\eta=1.083$. Since it is mentioned in [7] that the primary fit was based on the redshift range, $0.0233<z<0.15$, it is of interest to determine $M-M'$ for, say, $z=0.03$ and $z=0.15$. One finds for these two redshifts, $M-M'=0.0323$, and $M-M'=0.0084$, respectively. Thus $M-M'$ is approaching zero with increasing redshift, for this range of redshifts. The redshift z^* for which $M-M'=0$ is of special interest, since under these circumstances $\log\eta=k(-0.55,z^*,j_0)/k(0.5,z^*,j_0)$. Because of the lack of empirical values for M' one cannot predict $H_0(0.5)$, so that a reasonable alternative is

to predict the values of z^* for which $M - M' = 0$. From (10), when $M - M' = 0$, one obtains for z^*, after removing the logarithms, the quadratic equation

$$\left(z^*\right)^2 - \frac{0.775 - \eta 0.250}{0.274 - \eta 0.125} z^* + \frac{\eta - 1}{0.274 - \eta 0.125} = 0 \tag{11}$$

For $\eta = 1.083$, one has the two roots $z^* = 0.17$, and $z^* = 3.47$. Because terms of $O\left(z^3\right)$ have been omitted, the higher redshift value is possibly less reliable than the lower value, and in any case this higher value is unacceptable. Since, as indicated in [7], more than 600 SNe Ia were used in a Hubble plot, with redshifts ranging, $0.01 < z < 0.4$, it follows that the higher value of z^* lies well outside the range of redshifts considered in the CDL work, unlike the lower value. Also, if one takes the fiducial value of η to be that for $H_0(CDL)/H_0(CMB)$ so that $\eta = 73.24/67.90 = 1.079$, one finds that the two roots are $z^* = 0.16$, and $z^* = 3.47$. Hence the smaller root is clearly more sensitive to the value of η than the larger root, which, as indicated above, is to be excluded.

Although the above analysis makes it reasonable that the origin of the disagreement about the Hubble constant is due to the fact that the CMB and the BAO values are for a decelerating universe, in contrast with the CDL value, which is clearly for an accelerating universe, the analysis does not provide a rigorous proof that the universe is decelerating, since the above lower values of z^*, when uncertainties are allowed for, have not yet been confirmed empirically, and consequently there is always the possibility, until it is shown otherwise, that the CMB and BAO values for H_0 could be brought up to be in agreement with the CDL value. Therefore, other possible proofs that the universe is decelerating, and that the speed of light in IGS is ~2*c*/3, are required. In the next section, two further predictions of the model that support this proposed reduction of the speed of light in IGS are described. They have been mentioned earlier in [2] [4] [5], but for completeness, they are briefly presented here, together with some additional comments.

4. Additional Predictions of the Model

A prediction that fully supports the model, although it does not confirm it, was first discussed in [2], and in greater detail in [5]. It is based on attempts to find correlation of neutrinos with gamma ray bursts (GRBs). Significantly, to date, none have been found. The most impressive search, in terms of the number of null events, is that of the Boreximo collaboration, as given in Agostino *et al.* [20]. They searched from 2007-2015 for neutrinos and anti-neutrinos in the energy range, 1.5 MeV - 17 MeV, correlated with 2350 observations of GRBs, and found no statistically significant excess above background. Since in the model the speed of light through the dark energy of IGS is ~2*c*/3, and since for these energies neutrinos travel very close to *c*, if *D* is a typical distance to the cosmological sources of the GRBs, the time difference for the arrival of the neutrinos and the

GRBs would be ~$D/2c$, after neglecting the time the GRBs spent traveling through the Milky Way with speed c. Since the GRBs come from cosmologically distant sources, $D/2c$ is of the order of millions of years, so that even if there were a sufficient fluence of neutrinos to be detectable, the neutrinos would have arrived millions of years earlier than the GRBs. Hence the model predicts that there cannot be any correlation of neutrinos with GRBs, which is what is observed. This has also been shown to be true with numerous other searches as well [21]-[30]. To be sure, it could be the case that none of the GRB sources produce a sufficient fluence of neutrinos to be detectable above the background, which would therefore give rise to the same null result. However, there is a possible counter-example to this alternative: although Artsen *et al.* [31] reported that the IceCube collaboration had detected two PeV neutrinos without any correlation with GRBs, it is nevertheless possible that a GRB was produced along with these very high energy neutrinos, but the GRB will not arrive until millions of years in the future. Furthermore, as emphasized in [5], if one assumes that none of the GRB sources produced a measureable fluence of neutrinos, that amounts to thousands of assumptions: one assumption for each case of detected GRB unaccompanied by neutrinos. Hence, according to Occam's Razor, this large number of assumptions makes this explanation less preferable than that of the proposed model that uses only one assumption about the reduced speed of light through IGS.

Although the above absence of correlations supports the model, as well as the proposed resolution of the disagreement over the Hubble constant, neither fully confirms it. On the other hand, the following tests based on alternative sources for discordant redshift galaxies would confirm the model, and can be carried out quite readily. It was shown by Bahcall [32] that of the 64 cases of discordant redshifts, compiled and edited by Arp [33], ~40 of them could be explained as due to accidental superposition, *i.e.*, the higher redshift galaxy (HRG) being behind the lower redshift galaxy (LRG), at a cosmological distance, and shining through. He concluded that these ~40 cases were sufficiently close in number to the 64 observed cases that all the cases could be explained as due to accidental superposition. However, upon re-examining Bahcall's estimate, I found that he had rounded-up two of the numbers he had used in the product that led to the ~40 cases, and that when I did not round up these numbers, the product yielded only 30 cases that could be explained as due to accidental superposition. This suggested that about half the remaining cases are not due to accidental superposition, but are due to something else. It turns out the proposed model can explain the remaining cases, because it predicts that the dark energy of IGS has an index of refraction of ~1.5, while galaxies have an index of refraction of unity. Consequently, a light ray from an HRG that is not behind the LRG, but off to the side, angle-wise, upon being incident upon a LRG, at a suitable angle relative to the local normal, could be refracted in such way that it would pass through the LRG, and after being further refracted upon exiting the LRG, it could travel along a path that would bring it to the astronomer who would see the HRG as a

discordant redshift galaxy, while at the same time by suitably redirecting the telescope, depending on the location of the discordant redshift image, the astronomer would see the HRG directly. (Please see Figure 1 in [4].) Since galaxies do not have sharp edges, a light ray will encounter a variable behavior of n upon entering and leaving the LRG, as well as entering the Milky Way, so that the above scenario is highly simplified, but the qualitative picture should be valid. Hence it is a prediction of the model that the proposed HRGs will be found as a supplementary source of discordant redshifts. However, if they should not be found, this would not necessarily falsify the model, since it could turn out that the numbers Bahcall proposed did not accurately describe the situation. On the other hand, the model would definitely be falsified if it were found that an HRG was suitably located off to the side of a LRG in such a way that it should have given rise to a discordant redshift but did not, since that would mean the ray from the HRG had most likely gone through the LRG with negligible refraction, and would thereby demonstrate that IGS does not have an observably higher index of refraction than the LRG, in conflict with the model.

5. Conclusion

The above work has shown that one can resolve the current disagreement about the Hubble constant by assuming that the CMB and BAO determinations of H_0 are not for the accelerating ΛCDM universe, but for a decelerating EdS universe, in which the density parameter for the dark energy in the EdS universe satisfies the relation $\Omega_{de} = \Omega_\Lambda$. However, unlike the cosmological term, the dark energy in this model does not have a negative pressure associated with it, but instead it has an index of refraction n. As a test of this proposal, it was shown that the CDL's higher value for H_0 could be lowered to agree with the CMB and BAO determinations by re-evaluating the CDL value of H_0 for a decelerating universe, rather than for an accelerating universe. In order to obtain the additional distance in the EdS decelerating universe that is needed to explain the increased apparent magnitude of the SNe Ia that led astronomers [9] [10] [11] to infer that the universe is accelerating, it is necessary to assume the speed of light through the dark energy of IGS is reduced to c/n, where $n \approx 1.5$. This assumption gives rise to a challenging problem: How can such a very low density substance as the dark energy have an index of refraction comparable to some types of glass, and moreover do it without introducing any dispersion? Furthermore, since it has been shown recently from a binary neutron star merger that the resulting gravitational waves (GWs) Abbott *et al.* [34] [35] arrive at essentially the same time (~1.7s) as the GRBs Goldstein *et al.* [36] and Savchenko *et al.* [37], this index of refraction for electromagnetic radiation would have to hold for gravitational waves as well! Consistent with this, the Borexino collaboration, as reported in d'Agostino *et al.* [38], did not find any neutrinos above background correlated with the arrival of GWs. The only explanation that seems to embrace all these findings, as remarked in [5], is that one is possibly encountering consequences of that long-sought unified theory of electromagnetism and gravitation that

might predict how dark energy could influence the propagation of EM radiation and GWs in this way. However, before one can conclude that dark energy does indeed have such extraordinary properties, it is essential to ascertain whether the above predictions concerning the Hubble constant and discordant redshifts actually hold. Hopefully this work, as well as those that preceded it, will encourage astronomers to undertake such investigations.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Tangherlini, F.R. (2015) *Journal Modern Physics*, **6**, 78-87. https://doi.org/10.4236/jmp.2015.61010

[2] Tangherlini, F.R. (2015) *Journal Modern Physics*, **6**, 1360-1370. https://doi.org/10.4236/jmp.2015.69141

[3] Tangherlini, F.R. (2016) *Journal Modern* Physics, **7**, 1829-1844. https://doi.org/10.4236/jmp.2016.713163

[4] Tangherlini, F.R. (2017) *Journal Modern* Physics, **8**, 622-635. https://doi.org/10.4236/jmp.2017.84042

[5] Tangherlini, F.R. (2018) *Journal Modern Physics,* **9**, 573-583. https://doi.org/10.4236/jmp.2018.94039

[6] Ade, P.A.R., *et al.* (2016) *Planck Collaboration*: *Astronomy and Astrophysics*, **594**, A13.

[7] Riess, A.G., *et al.* (2016) *SHOES Collaboation*: *Astrophysical Journal*, **826**, 56. https://doi.org/10.3847/0004-637X/826/1/56

[8] Shadab, A., *et al.* (2016) *BOSS Collaboration*: *Monthly Notices Royal Astronomical Society*, **000**, 1-38.

[9] Riess, A.G., *et al.* (1998) *Astronomical Journal*, **116**, 1009-1038. https://doi.org/10.1086/300499

[10] Schmidt, B., *et al.* (1998) *Astrophysical Journal*, **507**, 45-63. https://doi.org/10.1086/306308

[11] Perlmutter, S., *et al.* (1999) *Astrophysical Journal*, **517**, 565-586. https://doi.org/10.1086/307221

[12] Costa, E., *et al.* (1997) *Nature* (*London*), **387**, 783-785. https://doi.org/10.1038/42885

[13] Metzger, M.R., *et al.* (1997) *Nature* (*London*), **387**, 878-880. https://doi.org/10.1038/43132

[14] Frail, D.A., *et al.* (1997) *Nature* (*London*), **389**, 261-263. https://doi.org/10.1038/38451

[15] Riess, A.G., *et al.* (2001) *Astrophysical Journal*, **560**, 49-71. https://doi.org/10.1086/322348

[16] Verde, L., *et al.* (2002) *Monthly Notices Royal Astronomical Society*, **335**, 432-440. https://doi.org/10.1046/j.1365-8711.2002.05620.x

[17] Hawkins, E., *et al.* (2003) *Monthly Notices Royal Astronomical Society*, **346**, 78-96.

https://doi.org/10.1046/j.1365-2966.2003.07063.x

[18] Anderson, L., *et al.* (2012) *Monthly Notices Royal Astronomical Society*, **427**, 3435-3457. https://doi.org/10.1111/j.1365-2966.2012.22066.x

[19] Anderson, L., *et al.* (2014) *Monthly Notices Royal Astronomical Society*, **439**, 83-101. https://doi.org/10.1093/mnras/stt2206

[20] Agostini, M., *et al.* (2017) *Borexino Collaboration*: *Astroparticle Physics*, **86**, 1-17.

[21] Achteberg, A., *et al.* (2007) *Astrophysical Journal*, **664**, 397-410. https://doi.org/10.1086/518596

[22] Vieregg, A.G., *et al.* (2011) *Astrophysical Journal*, **736**, 50. https://doi.org/10.1088/0004-637X/736/1/50

[23] Adrian-Martinez, S., *et al.* (2013) *Astronomy & Astrophysics*, **559**, A9. https://doi.org/10.1051/0004-6361/201322169

[24] Adrian-Martinez, S., *et al.* (2017) *European Physical Journal C*, **77**, 20. https://doi.org/10.1140/epjc/s10052-016-4496-8

[25] Avrorin, A., *et al.* (2011) *Physics of Particles and Nuclei Letters*, **8**, 704-716. https://doi.org/10.1134/S154747711107003X

[26] Fukuda, S., *et al.* (2002) *Astrophysical Journal*, **578**, 317-324. https://doi.org/10.1086/342405

[27] Aharmim, B., *et al.* (2014) *Astroparticle Physics*, **55**, 1-76. https://doi.org/10.1016/j.astropartphys.2013.12.004

[28] Asakura, K., *et al.* (2015) *Astrophysical Journal*, **806**, 87. https://doi.org/10.1088/0004-637X/806/1/87

[29] Kochkarov, M.M., *et al.* (2015*) Physics of Particles and Nuclei*, **46**, 197-200. https://doi.org/10.1134/S1063779615020124

[30] Abassi, R., *et al.* (2012) *Nature* (*London*), **484**, 351-354. https://doi.org/10.1038/nature11068

[31] Aartsen, M.G., *et al.* (2013) *IceCube Collaboration*: *Physical Review Letters*, **111**, Article ID: 021103.

[32] Bahcall, J.N. (1976) Redshifts as Distance Indicators. In: Field, G.B., Bahcall, J.N. and Arp, H., Eds., *The Redshift Controversy*, W. A. Benjamin, Reading, 61-121.

[33] Arp, H. (1966) *Astrophysical Journal Supplement*, **14**, 1-20.

[34] Abbott, B.P., *et al.* (2017) *LIGO Scientific Collaboration and Virgo Collaboration*, *Astrophysical Journal Letters*, **848**, L12. https://doi.org/10.3847/2041-8213/aa91c9

[35] Abbott, B.P., *et al.* (2017) *LIGO Scientific Collaboration and Virgo Collaboration*, *Physical Review Letters*, **119**, Article ID: 161101. https://doi.org/10.1103/PhysRevLett.119.161101

[36] Goldstein, A., *et al.* (2017) *Fermi-GBM*, *Astrophysical Journal Letters*, **848**, L14. https://doi.org/10.3847/2041-8213/aa8f41

[37] Savchenko, V., *et al.* (2017) *INTEGRAL Collaboration*, *Astrophysical Journal Letters*, **848**, L15. https://doi.org/10.3847/2041-8213/aa8f94

[38] Agostini, M., *et al.* (2017) *Borexino Collaboration*, *Astrophysical Journal*, **850**, 21. https://doi.org/10.3847/1538-4357/aa9521

Scientific Research Publishing

Journal of Modern Physics, 2018, 9, 1838-1865
http://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

The Principles and Laws of Black Hole Universe

T. X. Zhang

Department of Physics, Alabama A & M University, Normal, AL, USA
Email: tianxi.zhang@aamu.edu

How to cite this paper: Zhang, T.X. (2018) The Principles and Laws of Black Hole Universe. *Journal of Modern Physics*, **9**, 1838-1865.
https://doi.org/10.4236/jmp.2018.99117

Received: July 15, 2018
Accepted: August 14, 2018
Published: August 17, 2018

Abstract

A new alternative cosmological model called black hole universe was recently developed by the author on the basis of the following three fundamentals: 1) the principle of spacetime black hole equivalence, 2) the cosmological principle of spacetime isotropy and homogeneity, and 3) the Einsteinian general theory of relativity that describes the effect of matter on spacetime. According to this black hole model of the universe, the author up-to-dately has self-consistently described the origin, structure, evolution, expansion, and acceleration of the universe, quantitatively explained the measurements of cosmic microwave background radiation, type Ia supernovae's luminosity distance and redshift, and dynamic properties of star-like, massive, and supermassive black holes such as gamma-ray bursts, X-rays flares from galactic centers, and quasars, and fully overcome the difficulties of the conventional model of the universe such as the problems of horizon, flatness, monopole, inflation, dark matter, dark energy, and so on. In this paper, the author will examine and overview thoroughly this new cosmological model and completely describe its development from the three fundamentals and its creative explanations to the existing observations of the universe. From this comprehensive investigation of the new cosmological model, the author will further reveal the fundamental regularities and laws of the black hole universe with respect to the spacetime mass and radius, spacetime equilibrium, spacetime expansion and acceleration, spacetime radiation energy, and spacetime entropy variation. These efforts will help us to uncover various regularities and mysteries of the universe.

Keywords

Cosmology, Black Hole, Spacetime, Universe

1. Introduction

Brewing the idea—the universe as a black hole—that was unexpectedly emerged

from his mind in 2004 at reading the famous paper entitled "Mach's principle and a relativistic theory of gravitation [1]" in the library of the University of Alabama in Huntsville to develop his new electric redshift mechanism [2] and his new gravitational shielding theory [3] for three years, Dr. T. X. Zhang formally proposed a new cosmological model called black hole universe in the 211th American Astronomical Society (AAS) meeting held at Austin, Texas (USA), on January 7-11, 2008 [4]. This new multiverse cosmological model is consistent with Mach's principle, governed by Einstein's general theory of relativity, and able to explain all the existing observations of the universe according to the well-developed physics without encountering any fates of the universe or difficulties [5]. Seven years later, Zhang [6] further successfully convened and chaired a special session entitled "Observation and Theory for Multiverse" for the 224th AAS meeting held at Boston, Massachusetts (USA), on June 1-5, 2014. In this special session of the 224th AAS meeting, many famous cosmologists from the world were brought together to discuss cosmological issues of multiverse [7]. Zhang [8] also gave a talk on his hierarchically multilayered black hole universe model and greatly impacted the communities of astronomy, astrophysics, and cosmology with this new cosmological model.

The early studies done by the author up to the date have 1) self-consistently described the origin of the universe from black holes, the structure of the entire space with infinite layers, the iterative and cyclic evolution in an endless and beginningless style, and the physical expansion outward in one-way along the time or entropy increasing direction, 2) quantitatively explained the measurement of 2.725 K cosmic microwave background radiation with the blackbody spectrum, the observation of type Ia supernovae's luminosity distance and redshift for the acceleration of the universe, and the dynamic properties of star-like, massive, and supermassive black holes such as gamma-ray bursts, X-rays flares from galactic centers, and quasars, and 3) fully overcome the difficulties of the conventional standard big bang model of the universe such as the problems of the horizon, flatness, monopole, inflation, dark matter, dark energy, and so on. The results obtained from these early studies have been presented in a sequence of AAS meetings [4] [6]-[17] and published in a number of refereed journals and conference proceedings [5] [18]-[25].

The big bang model of the universe, though being the currently accepted standard cosmological model, critically relies on a growing number of hypothetical entities [26]. Without making these supplementary assumptions, the big bang theory severely contradicts with the observations of the universe. Theoretically, the big bang universe stands only on the basis of two fundamentals, which are the cosmological principle of spacetime isotropy and homogeneity and Einstein's general theory of relativity, but needs to face lots of uncertainties and difficulties that are emerged in explaining the observations of the universe ([13] [25], see also left panel of **Figure 1**). To overcome these difficulties, cosmologists had to make a number of supplementary assumptions. For instances, the inflation epoch was assumed in order for the big bang model to solve the initial value

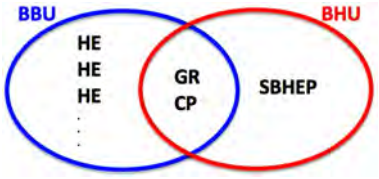


Figure 1. The big bang universe (BBU) vs. the black hole universe (BHU) in fundamentals [13] [25]. BBU was developed on the basis of Einstein's general relativity (GR) and the cosmological principle (CP) with innumerable hypothetical entities (HEs) to explain observations and overcome cosmic problems and difficulties. BHU was developed on the basis of GR, CP and the principle of spactime black hole equivalence (SBHEP). With this new principle (*i.e.* SBHEP), BHU can also explain all the existing observations of the universe and meantime overcome the cosmic problems and difficulties in terms of the well-developed physics without needing any other HEs.

problems such as the horizon, flatness, and monopole problems [27]. The dark energy component is introduced in order for the big bang theory to address the acceleration of the universe as shown by the measurement of type Ia supernovae's redshift and luminosity distance [28] [29]. These hypothetical entities are growing in number as we explore the universe more deeply and more completely.

In the black hole model of the universe [4] [5], Zhang proposed a new postulate or principle: *the Principle of Spacetime Black Hole Equivalence (SBHEP)*, in addition to the cosmological principle of spacetime isotropy and homogeneity and Einstein's general theory of relativity [13] [25]. By upstanding on these three fundamentals, this new black hole universe model not only self-consistently explains all the existing observations of the universe, but also naturally overcomes the difficulties of the universe without requiring any extra hypotheses (see the right panel of **Figure 1**). Therefore, one principle can remove all hypothetical entities from the big bang model of the universe. In this paper, we will thoroughly examine this new cosmological model and fully describes its development from the three fundamentals and its explanations to the existing observations of the universe. From this complete investigation of the new cosmological model, we will further reveal the fundamental regularities and laws of the black hole universe with respect to the spacetime mass and radius, spacetime equilibrium, spacetime expansion and acceleration, spacetime radiation conservation, and spacetime entropy variation. These efforts will help us to develop an elegant, perfect, and complete cosmological model that reveals various aspects and mysteries of the universe.

2. The Principles of Black Hole Universe

2.1. The Principle of Spacetime Black Hole Equivalence

The first principle or fundamental of black hole universe is the principle of spacetime black hole equivalence. More specifically, a black hole constructs an individual spacetime and a spacetime encloses a black hole. Black holes and spacetimes are equivalent (**Figure 2**). This principle comes from the fact or

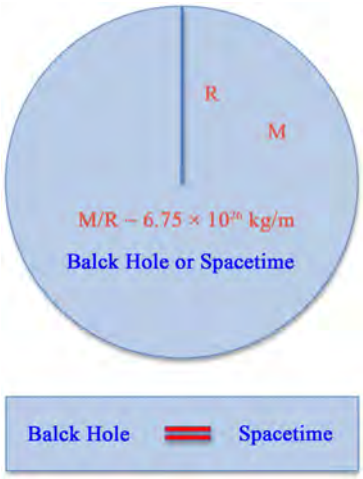


Figure 2. Equivalence between spacetime and black hole. A region, where matter accumulates up to a critical point such that $M/R \sim 6.75\times10^{26}$ kg/m, forms a black hole and meantime constructs its own space to be singular to the outside and its own time to be noncausal to the outside. Black holes and spacetimes are equivalent. This principle of spacetime black hole equavalence is one of the three fundamentals of the black hole universe model.

result of Einstein's general theory of relativity and Schwarzschild's solution of the field that a region, where matter accumulates up to a critical point such that the mass-radius ratio is equal to $M/R = c^2/2G \sim 6.75 \sim 10^{26}$ kg/m (or ~0.3375 solar masses/km), forms a black hole and meantime builds or constructs its own space being singular to the outside and its own time being noncausal to the outside. When a black hole is formed, an individual spacetime is immediately wrapped. Here c is the speed of light traveling in the free space and G is the gravitational constant.

According to this newly proposed principle, our four-dimensional (4D) spacetime universe is a black hole, extremely supermassive and fully expanded. All the inside, currently observed, star-like, massive, and/or supermassive black holes are subspacetimes of our black hole universe (**Figure 3**). The bottom panel of **Figure 3** represents our universe spacetime as a line, which breaks with or contains a number of holes as subspacetimes. Here, we, for simple, only drew four subspacetimes, corresponding to four star-like or supermassive black holes.

The spatial boundary of a spacetime or black hole including our black hole universe is the Schwarzschild absolute event horizon determined by [30]

$$\frac{2GM}{c^2R}=1 \tag{1}$$

where M and R are the mass and radius of the spacetime or black hole. This mass-radius relation is also the relation of the effective mass and radius of the universe from the Mach principle [31] [32] or the relation of the observable mass and radius of the universe. For convenience, here we name this relation as the Mach-Schwarzschild mass-radius (M-R) relation or as the Mach-Schwarzschild law of spacetime mass and radius, or the 0^{th} law of the black hole universe.

Figure 3. Our spacetime universe is an extremely supermassive and fully expanded black hole. The inside star-like, massve, and/or supermassive black holes observed are subspacetimes of our black hole universe.

Therefore, the black hole model of the universe does not exist the horizon problem. Within a black hole or a spacetime, an event occurred at one point is observable to all other points and events occurred at different points are all causally related one another.

Equation (1) indicates that the radius of a spacetime or black hole including the black hole universe is proportional to its mass ($R \propto M$), *i.e.* the ratio remains a constant or is conserved. For a star-like black hole with 3 solar masses, its radius is about 9 km. For a supermassive black hole with 3 billion solar masses, its radius is about 9×10^9 km. For the present black hole universe with hundred billion trillion solar masses, its radius is about 10^{23} km. The black hole model of the universe is supported by the Mach principle, the observations of the universe, and the Einsteinian general theory of relativity.

A black hole curves the spacetime maximally, so that it becomes singular and noncausal to the outside. It is highly curved, wrapped and closed. The 3D space curvature constant of a closed spacetime or a black hole including our black hole universe is

$$k = 1 \tag{2}$$

Therefore, the black hole model of the universe does not exist the flatness problem. Having neither the horizon problem nor the flatness problem, the black hole model of the universe does not need an inflation epoch, a nonphysical process that the current widely accepted big bang model of the universe is relying on [27], at all. Most inflation models predict the gravitational waves to be stronger and stronger, but observations showed opposite results [33] [34].

It is reasonable to suggest that our black hole universe was originated from a star-like black hole with several solar masses, which was formed from a massive star when the inside thermonuclear fusion has completed, and grew up through a supermassive black hole with billions of solar masses to the present state with

hundred-billion-trillions of solar masses, by accreting the ambient matter or merging with other black holes (Figure 4). It is not important which black hole is the first one to form the present universe because its contribution is waned out and negligible. This is analogy to that a big river is originated or formed by merging a great number of streams, but which stream is the first one is not essential. The black hole model of the universe is not sensitive to its past (no initial value or fine-tuning problem) or to the properties of star-like or supermassive black holes such as their angular momenta, temperatures, and sizes. The recent detections of gravitational waves by the Laser Interferometer Gravitational-Wave Observatory (LIGO) have confirmed the existence of stellar black holes and their mergers in the universe [35].

2.2. The Cosmological Principle of Spacetime Isotropy and Homogeneity

The second principle or fundamental of black hole universe is the cosmological principle of spacetime isotropy and homogeneity. This is also one of the two fundamentals in the big bang theory. According to this principle, a spacetime or black hole including our black hole universe, if it is viewed on a scale that is sufficient large or comparable to the scale of the spacetime, is homogeneous and isotropic. This principle implies that there is no special location and direction in a spacetime or black hole including the black hole universe. The properties of the universe are the same for all observers in the universe. More strongly, physical laws are all universal. If a physical law is applicable to the earth, then it can be applied to everywhere.

The metric of an isotropic and homogeneous spacetime or black hole including our black hole universe is given by the Friedmann-Lemaitre-Robertson-Walker (FLRW) metric [36] [37] [38] [39].

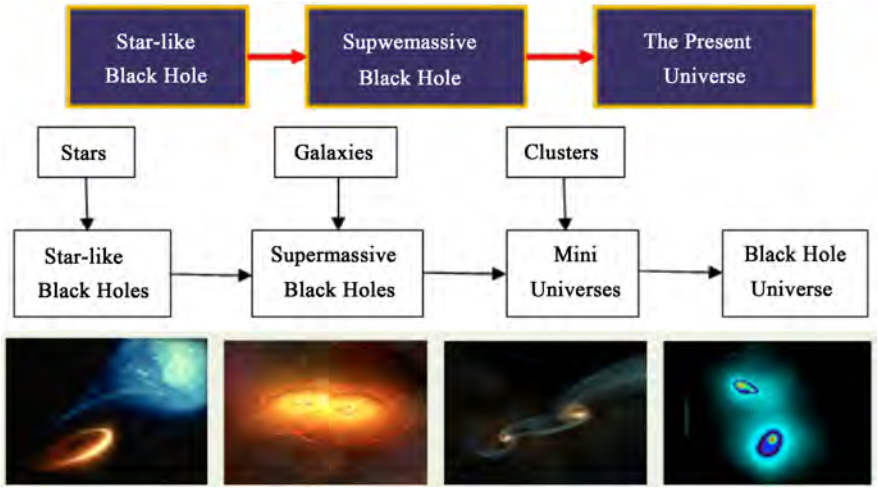


Figure 4. Our present black hole universe was originated from a star-like black hole with several solar masses, which grew through a supermassive black hole with billions of solar masses and further to the present state with hudred-billion-trillions of solar masses by accreting its ambient matter and merging with other black holes.

$$ds^2 = c^2dt^2 - R^2(t)\left[\frac{dr^2}{1-kr^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right] \quad (3)$$

where ds is the line element of 4D spacetime, r, θ, ϕ are co-moving space-coordinates, t is the time, $R(t)$ is the radius of curvature of the space, and k is the curvature constant of the space. In the black hole model of the universe, $R(t)$ is the radius of the black hole or spacetime and $k = 1$.

The density of matter (ρ, defined as mass M divided by the volume V) in an isotropic and homogeneous spacetime or black hole including our black hole universe can be determined as

$$\rho \equiv \frac{M}{V} = \frac{3c^6}{32\pi G^3 M^2} = \frac{3c^2}{8\pi G R^2} \quad (4)$$

i.e. $\rho R^2 = \text{constant}$ and $\rho M^2 = \text{constant}$. Here, we have applied the Mach-Schwarzschild M-R relation (Equation (1)) and $V = 4\pi R^3/3$. It is seen that the density of matter in a spacetime is inversely proportional to the square of the mass ($\rho \propto M^{-2}$) or the square of the radius ($\rho \propto R^{-2}$). Both the mass and radius of the universe can be determined from the density of the universe. According to the current observations of the universe, the density of the present universe ρ_0 is about the critical density $\rho_0 \sim \rho_C \equiv 3H_0^2/(8\pi G) \sim 9.2\times10^{-27}$ kg/m^3. Then, from Equation (4), we can determine the radius and mass of the present universe as $R_0 = \sqrt{3c^2/(8\pi G\rho_0)} \sim 1.32\times10^{26}$ m (or ~14 billion light years) and $M_0 = R_0c^2/(2G) \sim 8.8\times10^{52}$ kg, respectively. Figure 5 plots the density of a black hole (solid line) as a function of the radius (or mass by 1 km = 0.3375 solar masses). The dotted line refers to the density of the black hole universe at the present time ρ_0. Its intersection with the solid line determines the radius and mass of the present black hole universe. The big bang model is not able to know

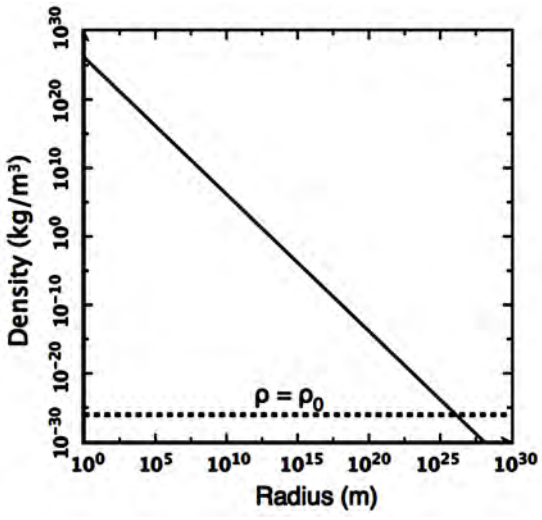


Figure 5. The density of a spacetime or black hole including our black hole universe versus the radius. The dotted line refers to $\rho = \rho_0$, so that its intersection with the solid line determines the radius and thus mass of the present black hole universe R_0 and M_0 to be in the orders of ~10^{26} m and ~10^{53} kg, respectibvely.

the mass of the universe and thus cannot determine the radius of the present universe according to the measurement of the density. Equation (4) can be considered as another form of the 0^{th} law of black hole universe.

3. The Theory of Black Hole Universe

3.1. The Theory of Spacetime

The theory or third fundamental of black hole universe is the Einsteinian general theory of relativity, which is a theory of spacetime or a geometric theory of gravitation that describes the effect of matter on spacetime [40]. The direct relation of the curvature of spacetime to the energy and momentum of matter in the spacetime is given by the Einsteinian field equation,

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \tag{5}$$

where $G_{\mu\nu}$ is the Einsteinian curvature tensor of spacetime and $\underline{T_{\mu\nu}}$ is the energy-momentum tensor of matter in spacetime. Einstein's another theory called special relativity describes the effect of motion on spacetime [41].

The theoretical predictions of general relativity such as the gravitational redshift of light from the Sun, the deflection of starlight by the Sun, the precessions of planetary perihelia, and the time delays of radar echoes have been confirmed in all observations and experiments to date. The Schwarzschild metric solution of the Einsteinian field equation of general relativity implies the existence of black holes, regions of space in which space and time are distorted in such a way that nothing, not even light, can escape from [30]. The observed gamma-ray bursts, X-rays flares from galactic centers, and quasars are believed to result from the activities of dynamic star-like, massive, and supermassive black holes, respectively. Not event light, except for the Hawking radiation [42], being able to escape from a black hole is only true if the black hole is in static state, in which the black hole has an enclosed event horizon. It should be noted that the Hawking radiation is a quantum effect obtained when the quantum mechanics (rather than a quantum gravity which has not yet been developed) is inappropriately applied to the surface of a black hole. As pointed out by Zhang in [18], a black hole, if it accretes its ambient matter or merges with other black holes, becomes dynamic. The event horizon of a dynamic black hole is temporarily broken and inside hot blackbody radiation leaks out and produces a gamma ray burst, X-ray flare, or quasar if the black hole is a star-like, massive, or supermassive black hole, respectively [12] [16] [19] [22] [23] [24].

3.2. The Theory of Cosmology

Substituting the FLRW metric of spacetime (Equation (3)) into the Einsteinian field equation of general relativity (Equation (5)), we have the Friedmann equation [43],

$$H^2(t) \equiv \frac{\dot{R}^2(t)}{R^2(t)} = \frac{8\pi G\rho(t)}{3} - \frac{c^2}{R^2(t)} \tag{6}$$

where $H(t)$ is the Hubble parameter or the expansion rate of the universe. The dot sign refers to the derivative with respect to time, $\dot{R}(t) \equiv \mathrm{d}R(t)/\mathrm{d}t$. The Hubble constant H_0 is the expansion rate of the universe at the present time and was measured as $H_0 \sim 70$ km/s/Mpc [44] [45] [46]. In addition, to derive Equation (6), we have considered Equation (2) for $k = 1$. All spacetimes are governed by the Friedmann equation that is derived from the Einsteinian general theory of relativity with the FLRW metric, the Mach-Schwarzschild M-R relation, and the positive space curvature.

4. The Laws of Black Hole Universe

4.1. The Law of Spacetime Equilibrium

Substituting the density equation (Equation (4)) into the Friedmann equation (Equation (6)), we have

$$\dot{R}(t) = 0 \text{ or } H(t) = 0 \tag{7}$$

This implies that an individual spacetime or black hole including our black hole universe, if there is not any external influence, is in the static state and equilibrium. A black hole in the static or equilibrium state is quite or does not emit. We here call this regularity as the law of spacetime equilibrium (or the 1st law of black hole universe). To have a static universe, Einstein [47] introduced a cosmological constant Λ into the field equation (Equation (5)) of his general relativity. In the black hole model of the universe, we do not need a cosmological constant in order for the universe to be static. A spacetime or black hole including our black hole universe remains in the static state until an external influence such as matter entering is applied. Any internal event may also disturb or ripple the spacetime [35] [48]. A static spacetime does not expand and shrink due to equilibrium. In the spacetime equilibrium, the black hole remains its radius $R(t)$ or mass $M(t)$ as a constant, *i.e.* independent of time.

This law of spacetime equilibrium has overturned the traditional idea that all the matter inside a black hole is piled up in a singular point at the center, while other parts are empty, though nobody understands how this central singularity works. According to this law, the matter inside a black hole is homogeneously and isotropically distributed in the spacetime that the black hole forms. The spacetime formed by a black hole is highly curved and power enough to hold or sustain exactly all the matter of the black hole that forms the spacetime. The gravity of the matter inside a black hole is balanced by the curvature of the spacetime formed by the black hole. This does not exclude any possible anisotropy and inhomogeneity in a small scale, a scale that is significantly smaller than the scale of the spacetime formed by the black hole. The balance force produced by a highly curved spacetime (here we name it as the spacetime tension) prevents the matter of black hole from further crunching to the center to form a singular point. This is analogy to that a spring or string, when stretched by an object that is hung on, produces a tension that prevents the object from falling to the ground.

4.2. The Law of Spacetime Expansion

When a spacetime or black hole including our black hole universe accretes matter from the outside or merges with other black holes, it becomes dynamic and expands [4] [5]. Considering that a spacetime or black hole including our black hole universe with mass $M(t)$ and radius $R(t)$ accretes matter of amount $\mathrm{d}M(t)$ from the outside, we have

$$\frac{2G\left(M(t)+\mathrm{d}M(t)\right)}{c^2R(t)}>1 \tag{8}$$

In this case, the spacetime breaks its equilibrium state and expands its size or radius from $R(t)$ to $R(t)+\mathrm{d}R(t)$, where the radius increment $\mathrm{d}R(t)$ can be determined by

$$\frac{2G\left(M(t)+\mathrm{d}M(t)\right)}{c^2\left(R(t)+\mathrm{d}R(t)\right)}=1 \tag{9}$$

Substituting the Mach-Schwarzschild M-R relation Equation (1) into Equation (9), we have

$$\frac{2G}{c^2}\frac{\mathrm{d}M(t)}{\mathrm{d}R(t)}=1 \tag{10}$$

Therefore, a spacetime or black hole including our black hole universe expands when it inhales matter from the outside. The expansion rate (or the rate of change in radius) is obtained as

$$\dot{R}(t)=\frac{2G}{c^2}\dot{M}(t) \tag{11}$$

and the Hubble parameter is given by

$$H(t)\equiv\frac{\dot{R}(t)}{R(t)}=\frac{\dot{M}(t)}{M(t)} \tag{12}$$

Equation (11) or Equation (12) indicates that the rate, at which a spacetime or black hole including the black hole universe expands, is proportional to the rate, at which it inhales matter from the outside. We here define this relation as the law of spacetime expansion, $\dot{M}(t)=H(t)M(t)$ or $\dot{R}(t)=H(t)R(t)$. A spacetime or black hole becomes dynamic when it accretes matter. A dynamic spacetime expands due to non-equilibrium.

Considering that a black hole with three solar masses accretes matter at a rate of 10^{-5} solar masses per year from its outside [49], we have $\dot{R}(t)=10$ cm/years and $H(t)\sim 10^7$ km/s/Mpc. Considering that a supermassive black hole with one billion solar masses swallows matter at a rate of one thousand solar masses in one year to run a quasar, we have $\dot{R}(t)\sim 3\times 10^3$ km/years and $H(t)\sim 10^6$ km/s/Mpc. When the black hole merges with other black holes, the growth rate should vary and be much larger in average than that when the black hole accretes its ambient matter. For our universe at the present state, the value of the Hubble parameter is measured as $H(t_0)\sim 70$ km/s/Mpc [44] [45] [46]. If the

radius of the universe is chosen as 13.7 billion light years, we have $\dot{R}(t_0) \sim c$, which implies that our universe is expanding in about the light speed at present. To have such fast expansion, the universe must inhale $\dot{M}(t_0) = \left[c^2/(2G)\right]\dot{R}(t_0) \sim c^3/(2G) \sim 2\times 10^{35}$ kg/s, about 10^5 solar masses in one second, or swallows a supermassive black hole in about a few hours. If Λ is introduced to the Einsteinian field equation, the black hole universe model determines it as $\Lambda = 3H_0^2 = 3\left(\dot{M}_0/M_0\right)^2 \sim 1.5\times 10^{-35}$ s^{-2}.

To expand, a spacetime or black hole including our black hole universe must have an outside, where matter is available for accretion and other black holes or parallel universes for mergers. Therefore, the black hole model of the universe suggests that the entire space is structured with layers, hierarchically and family-likely built [4] [5]. To see the hierarchical structure of the space, we sketch, in Figure 6, the innermost three layers of the black hole universe including the mother or parent universe, our universe itself, and child universes (*i.e.* stellar or supermassive black holes). For the sketch to be simple, we have only drawn three child universes and did not draw sister universes, which are universes parallel to our black hole universe. There should have a number of child universes and may also have many sister universes. A child universe grows by accreting materials from this universe or by merging with other child universes. This universe grows up by accreting materials from the mother universe or by merging with sister universes. The mother universe will also grow up if it has an outside. Otherwise, it is static.

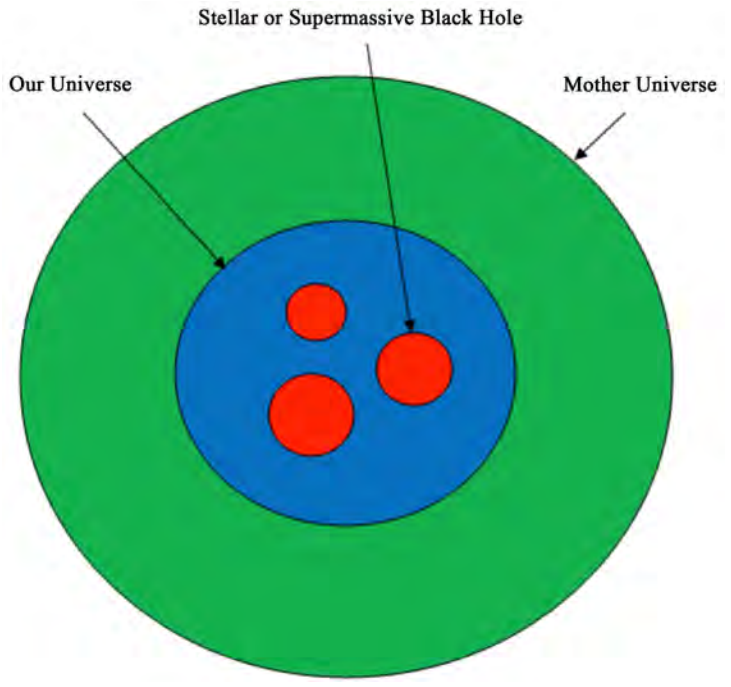


Figure 6. The innermost three layers of the entire space that is structured hierarchically up to mother universe (coded as green) [4] [5]. Here within our black hole universe (coded as blue), we only drew three child universes (coded as red), corresponding to the star-like or supermassive black holes.

To see the multilayer structure of the space in a larger (or more complicate) view, we plot, in **Figure 7**, the innermost four layers of the black hole universe [4] [5]. It includes the grandmother universe, aunt universes, mother universe, sister universes, cousin universes, our universe itself, child universes, and niece universes. Here again for the sketch to be simple, we only drew a few of universes for each layer. If the whole space is finite, then the matter in the whole space is finite and thus the number of layers is finite. Otherwise, it has infinite layers and the outermost layer corresponds to the limit of zero degree for the absolute temperature, zero for the density, and infinity for the radius and mass. These four generations of the universe family can also be represented by a universe family tree (see **Figure 8**) [4] [5]. The mother and aunt universes are children of the grandmother universe. The cousin universes are children of the aunt universes. Our universe, the sister universes, and the cousin universes have their own children, which are the star-like or supermassive black holes.

The evolution of the space structure is iterative [4] [5]. In each of iterations, the matter reconfigures and the universe is renewed rather than a simple repeat or bouncing back. **Figure 9** shows a series of sketches for the cartoon of the universe evolution in a single iteration from the present universe to the next similar one (from the top left to bottom right). This whole spacetime evolution process does not have the end and the beginning. As our universe expands, the child

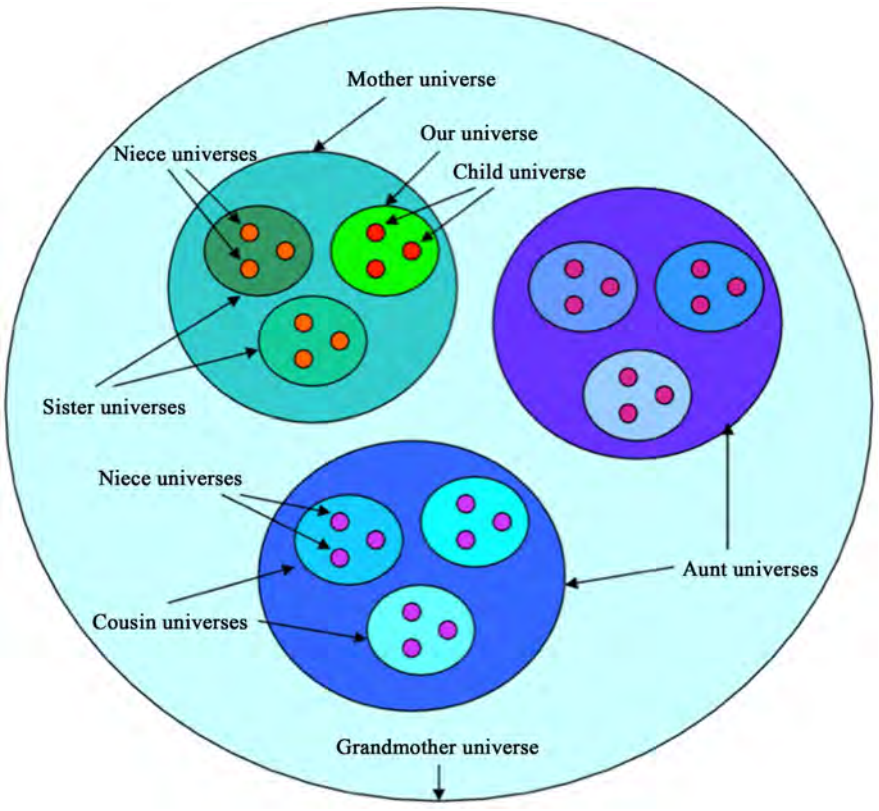


Figure 7. The innermost four layers of the entire space that is structured hierarchically up to grandmother universe [4] [5]. Here we also drew two sister universes that are parallel to our black hole universe and two aunt universes that are parallel to our mother universe.

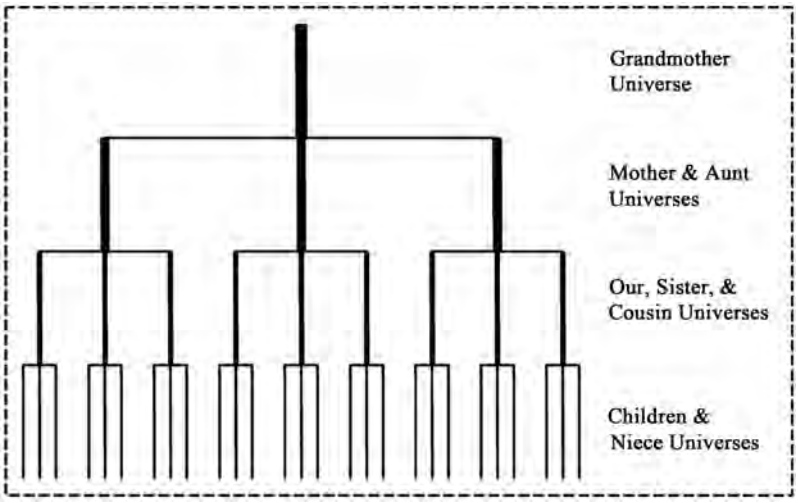


Figure 8. A family tree for the youngest four generations of the universe family [4] [5]. The generation one includes the child and niece universes; the generation two includes our universe itself and the sister universes; the generation three includes the mother and aunt universes; and the generation four includes the grandmother universe.

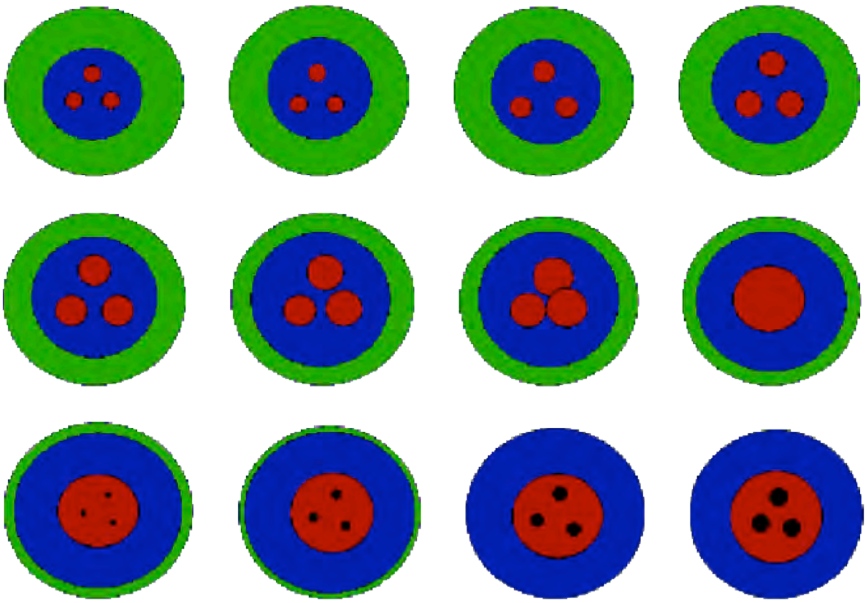


Figure 9. A series of sketches (or a cartoon, from left to right and then top row to bottom row) for the black hole universe to evolve in a single iteration from the present universe to the next similar one [4] [5]. This is an irreversible process, in which matter and spacetime reconfigure rather than a simple repeat or bouncing back. One universe is expanded to die out and a new universe is born from inside.

universes (*i.e.* the inside star-like and supermassive black holes) grow and merge each other into a new universe. Therefore, when one universe expands out, a new similar universe is born from inside. Like the living things in the nature, the universe passes through its own birth, growth, and death process and iterates this process endlessly. Its structure evolves iteratively forever without the beginning and end.

There is nothing more natural than to consider the entire space to be infinite large without an edge and have infinite number of layers [4] [5]. For the outermost layer, the radius tends to infinity, while the density and absolute temperature both tend to zero. We call this outermost layer as the whole/grand space universe because it contains all universes. To represent this infinite layer structure of the whole space, we use the mathematical set concept (see **Figure 10**). We let the whole/grand space universe be the set (denoted by U) of all universes; the child universes (also the niece universes) are null sets ($C=\{\ \}$ or $N=\{\ \}$); our universe is a set of the child universes ($O=\{C,C,C,\cdots,C\}$); the sister universes are sets of the niece universes ($S=\{N,N,N,\cdots,N\}$); the mother universe is a set of our universe and the sister universes ($M=\{S,S,S,\cdots,O\}$); the aunt universes are sets of the cousin universes ($A=\{K,K,K,\cdots,K\}$); the grandmother universe is a set of the aunt universes and the mother universe ($G=\{A,A,A,\cdots,M\}$); and so on.

The black hole universe model gives a fantastic structure of the whole space. All universes are self-similar and governed by the same physics, the Einsteinian general theory of relativity with the FLRW metric, Mach-Schwarzschild M-R relation, and positive curvature constant. The expansion of a black hole universe increases its mass and radius and decreases its density and temperature, but it does not alter the laws of physics. This infinite hierarchically layered space can also be represented as **Figure 11**. The top layer is the entire space, *i.e.* the grand universe. The bottom layer is a child universe. The second layer from the bottom is our universe. A child universe is a subspacetime of our universe; our universe is a subspacetime of the mother universe; the mother universe is a subspacetime of the grandmother universe; and so on. The total number of universes in the whole space is given by $n=\sum_{i=1}^{L} n_i$, where the subscript i is the layer number,

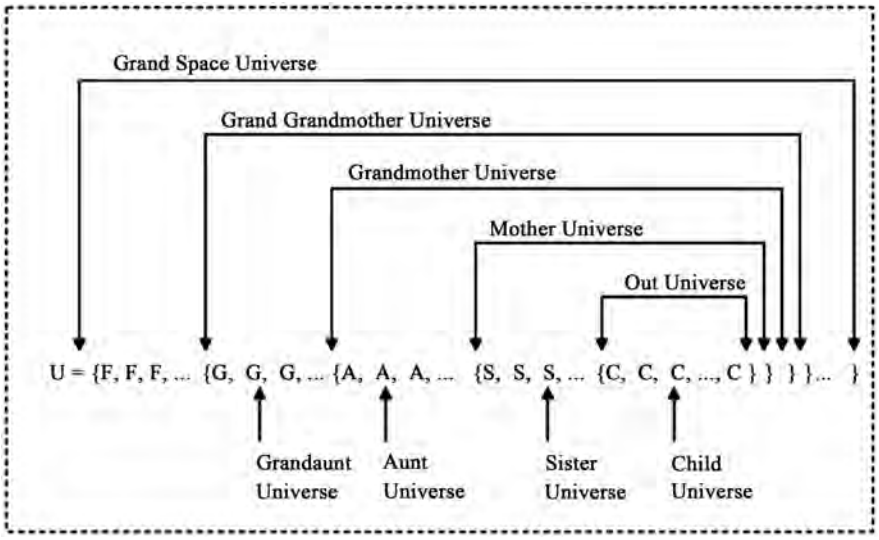


Figure 10. A mathematical representation of sets of universes for an infinite large and layered space [4] [5]. An inner layer universe set is a subset of the outer layer universe set. The niece and child universes are null sets because they do not contain any sub-spacetime.

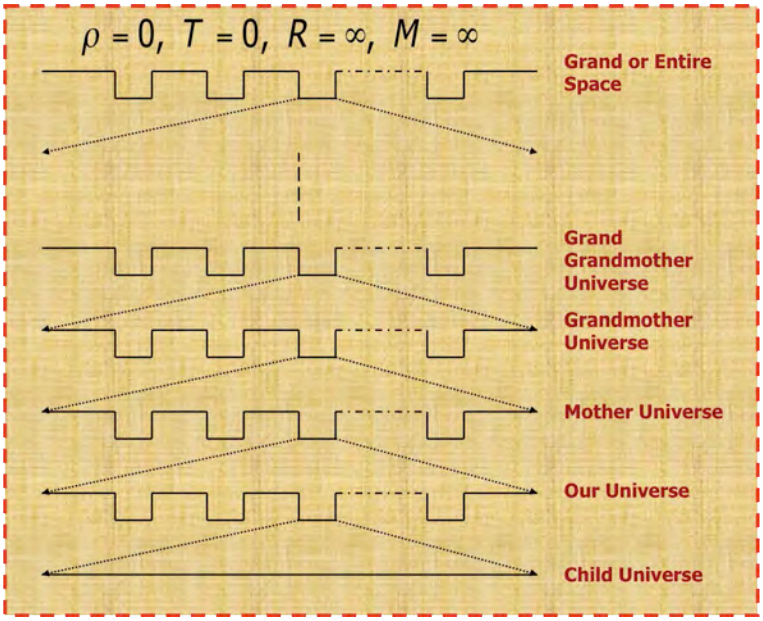


Figure 11. The entire space with infinite layers or subspacetimes [9] [18]. The bottom layer is a child universe or an empty spacetime. The child universe is a subspacetime of the universe in which we live. Similarly, our universe is a subspacetime of the mother universe, and so on. The top layer is the entire space of all subspacetimes.

n_i is the number of universes in the i^{th} layer, and L refers to the number of layers in the whole space. For the four layer (or generation) black hole universes sketched in **Figure 7** or **Figure 8**, we have $L = 4$ and $n = 27 + 9 + 3 + 1 = 40$. If the whole space includes infinite number of layers (*i.e.* $L = \infty$) and then the total number of universes is infinity (*i.e.* $n = \infty$). To explain the observation of cosmic microwave background radiation, the black hole model of the universe also needs the entire space to have infinite layers [9] [18]. The map of cosmic background radiation measured by the Planck Satellite has provided the first hard evidence for multiverse. The black hole universe model is a multiverse model and thus supported by the recent Planck CMB measurements.

4.3. The Law of Spacetime Acceleration

When a spacetime or black hole including our black hole universe accretes matter in an increasing rate, *i.e.* $\ddot{M}(t) > 0$, it accelerates its expansion [13] [21]. This regularity of spacetime is called the law of spacetime acceleration. It is combined with the law of spacetime expansion as the 2nd law of black hole universe. The dimensionless deceleration parameter is usually defined as

$$q(t) \equiv -\frac{R(t)\ddot{R}(t)}{\dot{R}^2(t)} = -\frac{M(t)\ddot{M}(t)}{\dot{M}^2(t)} \tag{13}$$

where the double dot sign refers to the second order derivative with respect to time. From Equation (13), we see that whether the black hole universe accelerates or not depends on whether the double dot of mass is positive or not. For a positive $\ddot{M}(t) > 0$, we have a negative deceleration parameter $q(t) < 0$, *i.e.* ac-

celeration of the universe.

According to the type Ia supernova measurements and their empirical distance-redshift relation, Daly *et al.* [50] obtained the deceleration parameter of the present universe as $q_0 \sim -0.48$ for the flat FLRW metric of spacetime with $k = 0$ (note that q_0 is smaller for $k = 1$). To explain the recent acceleration of the universe, the big bang universe model suggests that the universe is dominated by dark energy up to about 75%. In the black hole universe model, however, the universe accelerates because it inhales the outside matter in an increasing rate, *i.e.* a positive $\ddot{M}(t) > 0$. To accelerate its expansion with $q_0 \sim -0.48$, the present black hole universe only needs to inhale the outside matter in an increasing rate given by $\ddot{M}(t_0) = -q_0 M_0 H_0^2 = -c^2 q_0 R_0 H_0^2 / (2G) \sim 2.2 \times 10^{17}$ kg/s^2 (or about 110 solar masses per year square).

To explain the type Ia supernova measurements in accordance with the black hole universe model, we first solve Equation (13) to find the mass and Hubble parameter as functions of time. For a constant acceleration expansion universe, the time-dependent mass can be analytically solved from Equation (13) as,

$$M(t) = M_0 \left[(q+1) H_0 t + 1 \right]^{\frac{1}{q+1}} \tag{14}$$

from which the Hubble parameter is derived as

$$H(t) = \frac{\dot{M}(t)}{M(t)} = \frac{H_0}{(q+1) H_0 t + 1} \tag{15}$$

The time t can be replaced by the redshift z, because

$$1 + z \equiv \frac{R_0}{R(t)} = \frac{M_0}{M(t)} = \left[(q+1) H_0 t + 1 \right]^{\frac{1}{q+1}} \tag{16}$$

This mass-redshift relation, $M(t) = M_0 / (1+z)$, does not depend on the deceleration parameter q. Using Equation (16) to replace the time of Equation (15), we obtain

$$H(z) = H_0 (1+z)^{q+1} \tag{17}$$

To explain the type Ia supernova measurements according to the black hole universe model, we then determine the luminosity distance-redshift relation from the mass-time relation Equation (14) as

$$d_L = (1+z) R_0 \sin\left[\int_t^0 \frac{c \mathrm{d}t}{R(t)} \right] = (1+z) R_0 \sin\left[\int_t^0 \frac{c^3 \mathrm{d}t}{2GM(t)} \right] \tag{18}$$

Substituting Equation (14) into Equation (18), completing the integration and then using Equation (16), we have

$$d_L = (1+z) R_0 \sin\left[\frac{c^3}{2GM_0 H_0} \frac{1 - (1+z)^{-q}}{q} \right] \tag{19}$$

This redshift and luminosity distance relationship (Equation (19)) reduces to the Hubble law $z = d_L H_0 / c$ at $z \ll 1$.

Figure 12 plots the luminosity distance-redshift relation (red line) along with the type Ia supernova measurements (blue dots. Credit: Union 2.1 compilation of 580 SNIA data from Supernova Cosmology Project [51] [52]). In this plot the Hubble constant is chosen to be $H_0 = 70$ km/s/Mpc and the deceleration parameter is chosen to be $q_0 = -0.6$. In the upper panel of Figure 12, the distance modulus, which is defined by $\mu = 5\log_{10} d_L - 5$ with d_L in parsecs, is plotted as a function of redshift; while in the lower panel of Figure 12, the distance modulus difference between the measured SNIA data and analytical results derived from Equation (19). The results have shown that the black hole universe model can perfectly explain the measurements of type Ia supernovae.

Recently, the author pointed out that the luminosity distance-redshift relation that was usually applied to analyze the measurement of distant type Ia supernovae is an approximate expression that is only valid for nearby objects with

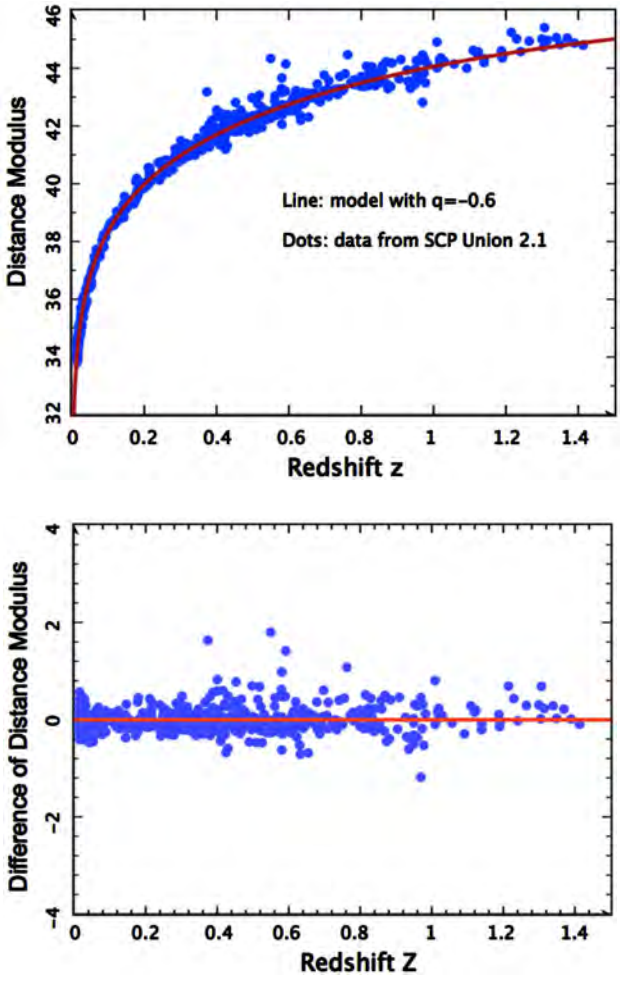


Figure 12. Luminosity distance-redshift relation of type Ia supernovae [13] [21]. Blue dots are measurements credited by the Union2 compilation of 580 SNeIa data from Supernova Cosmology Project [51] [52]. Red lines are analytical results from this study with $q = -0.6$. The upper panel plots the distance modulus as a function of redshift, while the lower panel plots the distance modulus difference between the measurement data and theoretical results.

$z \ll 1$ [14] [15] [20]. The luminosity distances of all distant type Ia supernovae with $z \geq 1$ had been underestimated. Zhang [14] [20] has corrected the luminosity distance expression with an extra redshift factor $\sqrt{1+z}$, with which the type Ia supernova measurements do not lead to the acceleration of the universe and thus remove the need of dark energy.

In the black hole universe model, if we include this redshift factor, the luminosity distance expression (Equation (19)) becomes,

$$d_L = (1+z)^{3/2} R_0 \sin\left[\frac{c^3}{2GM_0H_0}\frac{1-(1+z)^{-q}}{q}\right] \tag{20}$$

Using this equation, we can also nicely fit the data of type Ia supernovae if the deceleration parameter is chosen about $q = 0.5$ (see **Figure 13**). This corrected luminosity distance-redshift relation excludes the universe acceleration, so that removes the need of dark energy for the big bang universe model and the need of increasing its mass in an increasing rate for the black hole universe model.

4.4. The Law of Spacetime Radiation Conservation

According to Planck's law, the spectral energy density of the blackbody radiation within a black hole including the black hole universe can be written as

$$u(\nu,T) = \frac{8\pi h\nu^3}{c^3}\frac{1}{\exp\left(\frac{h\nu}{k_BT}\right)-1} \tag{21}$$

where ν is the radiation frequency, T is the temperature, h is the Planck constant, and k_B is the Boltzmann constant. In the SI unit system, the unit of $u(\nu,T)$ is J/m^3/Hz. **Figure 15** plots the spectral energy density as a function of frequency at temperature equal to 2.725, 10^4, 10^8, and 10^{12} K, respectively [18].

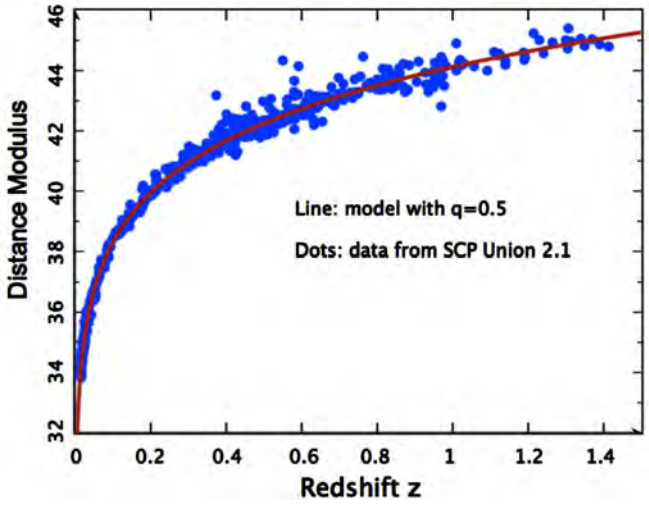


Figure 13. Distance modulus versus redshift, obtained from the corrected luminosity distance-redshift relation with the redshift factor $\sqrt{1+z}$ and $q = 0.5$ (red line) [14] [20]. Blue dots are again the measurements credited by the Union2 compilation of 580 SNeIa data from Supernova.

Integrating the spectral energy density (Equation (21)) with respect to the frequency of radiation in the entire range, we have the energy density of the blackbody radiation inside a black hole including the black hole universe,

$$\rho_\gamma \equiv \frac{U}{V} = \frac{8\pi^5 k_B^4}{15h^3c^3}T^4 = \beta T^4 \tag{22}$$

where U is the total radiation energy inside the black hole, V is the volume of the black hole, and the constant β is given by $\beta \equiv 8\pi^5 k_B^4 / \left(15h^3c^3\right) \sim 7.54\times10^{-16}$ J/m^3/K^4. Inside a black hole with the four temperatures given in Figure 15, the energy densities of radiation are ~4.15 × 10^{-14}, 7.54, 7.54 × 10^{16}, 7.54 × 10^{32} J/m^3, respectively.

As a spacetime or black hole including the black hole universe accretes its outside matter and radiation, it expands its volume from V to $V + \mathrm{d}V$ or radius from R to $R + \mathrm{d}R$. This type of expansion does not geometrically stretch the space of itself, instead it just takes the space as well as the matter and radiation from the outside. Considering that the increase of the Planck radiation energy within the black hole equals to the radiation energy inhaled from the outside space, we have

$$\beta\left(T+\mathrm{d}T\right)^4\left(V+\mathrm{d}V\right) - \beta T^4 V = \beta T_p^4 \mathrm{d}V \tag{23}$$

where T is the temperature of the black hole, $\mathrm{d}T$ is the change (or decrease because it is negative) of the temperature due to the expansion of the black hole, and T_p is the temperature outside the black hole, *i.e.* the temperature of the mother black hole or universe. Since $\mathrm{d}V/V = 3\mathrm{d}R/R$, we can reform Equation (23) by a first order development as

$$\frac{\mathrm{d}T}{\mathrm{d}R} = -\frac{3T}{4R}\left[1-\left(\frac{T_p}{T}\right)^4\right] \tag{24}$$

This equation governs the thermal history of the black hole universe that grew up from a star-like black hole through a supermassive black hole and a mini universe to the present state [9] [18]. The solution of Equation (24) for the dependence of T on R depends on T_p or on the relation between T and T_p. In the early studies cited above, Equation (24) was solved for the present black hole universe that grew up from a hot star-like black hole through a supermassive black hole.

For hot star-like or supermassive black holes, the temperatures inside should be much greater than the temperatures outside, *i.e.* $T \gg T_p$. In this case Equation (24) approximately reduces to

$$\frac{\mathrm{d}T}{\mathrm{d}R} = -\frac{3T}{4R} \tag{25}$$

which can be simply solved as

$$R^3T^4 = C \tag{26}$$

where C is an integral constant. Therefore, the temperature of a star-like or supermassive black hole decreases as it expands in size according to $T \propto R^{-3/4}$.

Substituting Equation (26) into Equation (22), we obtain the total radiation energy inside a star-like or supermassive black hole to be a constant and thus independent of its size,

$$U = V\beta T^4 = \frac{4}{3}\pi\beta R^3 T^4 = \frac{4}{3}\pi\beta C = \text{Constant} \tag{27}$$

This interesting result indicates that the blackbody radiation energy inside a star-like or supermassive black hole remains the same no matter it is static or dynamic. Attaining matter and growing its size do not increase its total radiation energy. Here, we call this regularity of constant radiation energy as the law of radiation energy conservation (or the 3rd law of black hole universe).

Figure 14 has shown that the spectral energy density of blackbody radiation inside a spacetime or black hole rapidly increases with the temperature and frequency. Inside a star-like black hole with temperature of 10^{12} K, the blackbody radiation dominates at the frequency of gamma rays. The radiation maximizes its spectral energy density up to ~10^{25} J/m^3/Hz at frequency of ~10^{23} Hz. Inside a massive or supermassive black hole with temperature of 10^8 K, the blackbody radiation dominates at the frequency of X-rays. The radiation maximizes its spectral energy density up to ~10^{12} J/m^3/Hz at frequency of ~10^{19} Hz.

A black hole, when it accretes its ambient matter or merges with another black hole, becomes dynamic. A dynamic black hole has a broken event horizon and thus cannot hold the inside hot (or high-frequency) blackbody radiation, which flows or leaks out of it and produces X-ray flares or gamma ray bursts. Dynamic star-like black holes with thousand billions of Kelvins (*i.e.* ~10^{12} K) radiate gamma rays, while dynamic massive or supermassive black holes with hundred millions of Kelvins (*i.e.* ~10^8 K) radiate X-rays such as X-ray emissions from

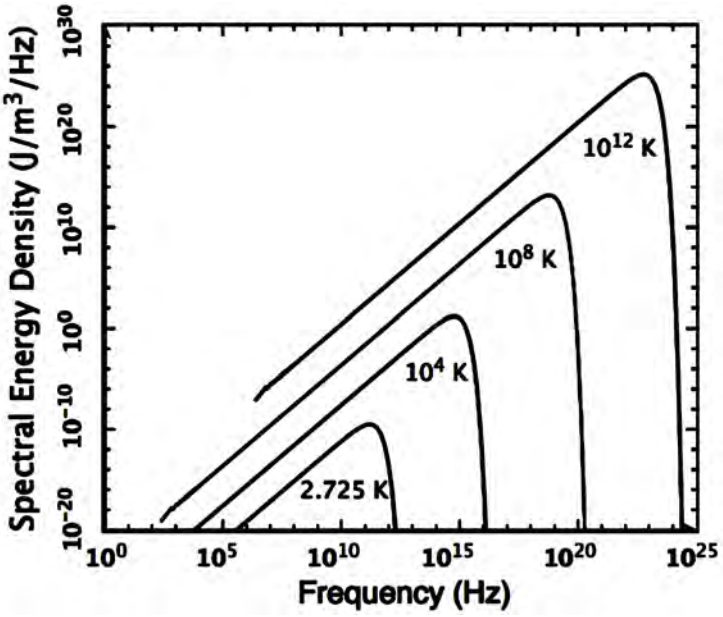


Figure 14. The spectral energy density of blackbody radiation as a function of radiation frequency at temperature equal to 2.725, 10^4, 10^8, and 10^{12} K, respectively.

quasars, supermasive black holes with billions of solar masses, and X-ray flares from Sagitarius A*, a massive black hole with millions of solar masses at the Milky Way center. This section describes how these dynamic properties of black holes are explained according to the black hole model of the universe [12] [16] [19] [22] [23] [24].

To reveal the possible thermal history of the black hole universe, we have considered a general case that the black hole universe decreases its relative temperature in a rate slightly greater than the mother universe as [18]

$$\frac{\mathrm{d}T_p}{T_p} = q\frac{\mathrm{d}T}{T} \tag{28}$$

which is equivalent to

$$T_p = \alpha T^q \quad \text{or} \quad \frac{T_p}{T} = \alpha T^{-\delta} \tag{29}$$

Here q is a constant slightly less than 1; $\delta \equiv 1-q$ is a small number; and α can be derived from q or δ according to the temperature and radius of the present universe. The other cases such as $q = 1$ or q to be significantly smaller than the unity could not explain the measurement of cosmic microwave background radiation as shown in [18].

In this case with q slightly smaller than the unity, the solution of Equation (24) can be analytically obtained as

$$T = R^{-3/4}\left(\alpha^4 R^{3\delta} + T_s^{4\delta} R_s^{3\delta}\right)^{1/4\delta} \tag{30}$$

where the constant α is given by

$$\alpha = \left[T_0^{4\delta} - \left(\frac{R_s}{R_0}\right)^{3\delta}\right]^{1/4} \tag{31}$$

Choosing q or δ appropriately, we can completely determine the thermal history of the black hole universe that evolved from a hot star-like black hole with temperature T_s and radius R_s to the present universe with temperature T_0 and radius R_0. In **Figure 15**, the radiation temperature of the black hole universe is plotted as a function of the universe radius with a fixed $q = 0.93$ or $\delta = 0.07$. Here we have chosen $T_0 = 2.725$ K, $R_0 = 13.7$ billion light years, $R_s = 9$ km, $T_s = 10^{12}$ K (solid line), $T_s = 10^{11}$ K (dotted-dashed line), and $T_s = 4\times 10^{12}$ K (dashed line).

It is seen that all temperature lines are concave downward and approach ~2.725 K as the black hole universe expands to the present size. Therefore, the initial temperature of the star-like black hole T_s is not critical to the present universe. The reason is because most matter and radiation are from the mother universe. This reason also explains why all other physical properties of the star-like black hole, including its size (or mass), angular momentum, and charge, and the evolution of the early universe are not critical to the present universe. Furthermore, the early process of material accretion and black hole mergers do not have significant leftover in the present universe.

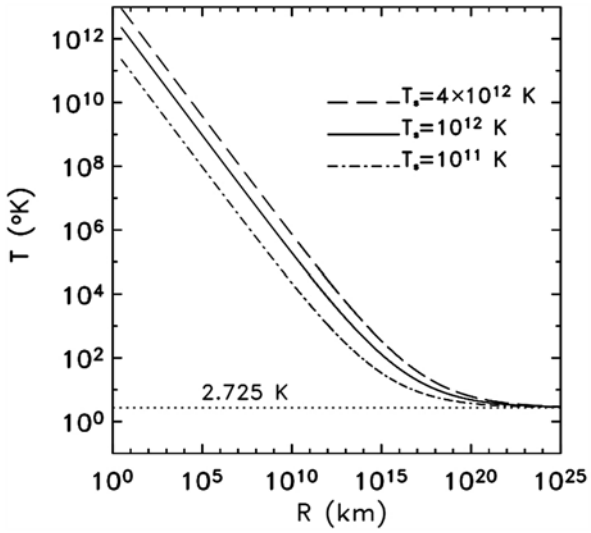


Figure 15. The possible thermal history of the black hole universe [18]. The solid line corresponds to $T_s = 10^{12}$ K, the dotted-dashed line corresponds to $T_s = 10^{11}$ K, and the dashed line corresponds to $T_s = 4\times10^{12}$ K. All temperature lines are curved by concaving downward and approach ~2.725 K at the present as the black hole universe expands to the present size.

4.5. The Law of Spacetime Entropy Increase

A spacetime or black hole including our black hole universe expands when it accretes matter and radiation from its outside. The expansion also alters its entropy given by [53]

$$\mathrm{d}S \equiv \frac{\mathrm{d}Q}{T} = \frac{c^2}{T}\mathrm{d}M + \frac{\beta T_b^4}{T}\mathrm{d}V = \left(\frac{c^4}{2GT} + \frac{4\pi\beta T_b^4 R^2}{T}\right)\mathrm{d}R \tag{32}$$

The first term of Equation (32) is the entropy increase due to the accretion of matter and the second term of Equation (32) is the entropy increase due to the accretion of radiation. It is seen from Equation (32) that the entropy of a spactime or black hole including our black hole universe does not decrease, $\mathrm{d}S \geq 0$, because $\mathrm{d}R \geq 0$. This regularity is called the law of spactime entropy increase (or the 4th law of black hole universe). **Figure 16** plots $\mathrm{d}S/\mathrm{d}R$ as a function of R. To plot this, we have used the temperature-radius (T-R) relations and constants associated, Equations (30) and (31).

Integrating Equation (32), we have the entropy of a spacetime or black hole with radius R or mass $M = c^2R/(2G)$,

$$S - S_s = \int_{R_s}^{R}\left(\frac{c^4}{2GT} + \frac{4\pi\beta T_b^4 R^2}{T}\right)\mathrm{d}R \tag{33}$$

where S_s is the entropy of the reference black hole with mass $M_s = 3$ solar masses (or radius $R_s = 8.89$ km) and temperature $T_s = 10^{12}$ K. Using Equations (30) and (31) to relate the temperatures T_b and T to the radius R, we can numerically solve Equation (33) and find the entropy of a spacetime or black hole as a function of its radius or mass.

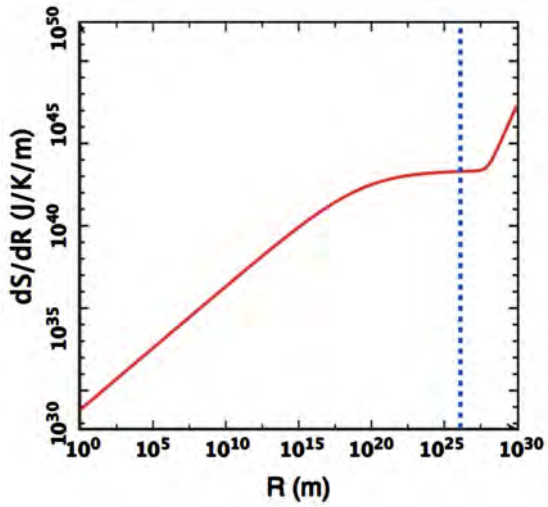


Figure 16. The slop for the increase of entropy of a spacetime or black hole including our black hole universe is plotted as a function of the radius.

For a star-like or supermassive black hole, we have $T_p \ll T$ and $T = T_s \left(R_s / R\right)^{3/4}$ as seen from Equation (26). Further using the Mach-Schwarzschild M-R relation Equation (1), we have $T = T_s \left[R_s c^2 / \left(2GM\right)\right]^{3/4}$. Then, the entropy of a star-like or supermassive black hole can be obtained by the following,

$$S_{BH} = \int_0^M \frac{c^2}{T} \mathrm{d}M = \frac{4c^2}{7T_s} \left(\frac{2G}{R_s c^2}\right)^{3/4} M^{7/4} \tag{34}$$

It is seen that the entropy of a star-like or supermassive black hole is proportional to the mass or radius according to $S_{BH} \propto M^{7/4}$ or $S_{BH} \propto R^{7/4}$. **Figure 17** plots the entropy of a star-like or supermassive black hole as a function of the black hole mass (solid red line). Let the mass equal one solar mass $M = M_{sun}$, then, from Equation (34), we obtain the entropy of 1-solar mass black hole as $S_{sun} = 3 \times 10^{35}$ J/K. For a supermassive black hole with one billion solar masses, its entropy is equal to about 10^{50} J/K. In **Figure 18**, we also plot the entropy of a star, S_{star}, which is usually estimated according to the Sackur-Tetrode equation, as a function of its mass,

$$S_{star} = k_B N_p \left\{ \ln \left[\frac{V_{star}}{N_p} \left(\frac{2\pi k_B m_p T_{star}}{h^2} \right)^{3/2} \right] + \frac{5}{2} \right\} \sim 18 k_B N_p \tag{35}$$

where N_p is the number density of proton in the star, m_p is the mass of proton, V_{star} is the volume of the star.

From **Figure 17**, it is seen that the entropy of a black hole with mass less than about 10 solar masses is lower than that of a star with the same mass. This implies that a star with mass less than 10 solar masses cannot directly produce or collapse into a black hole with the same mass at the end of its life. This is in agreement with Enistein's general theory of relativity, which suggests, to produce a star-like black hole, that the parent star must be 20 or more massive than the Sun. As a typical example, considering that a star with 20 solar masses forms a

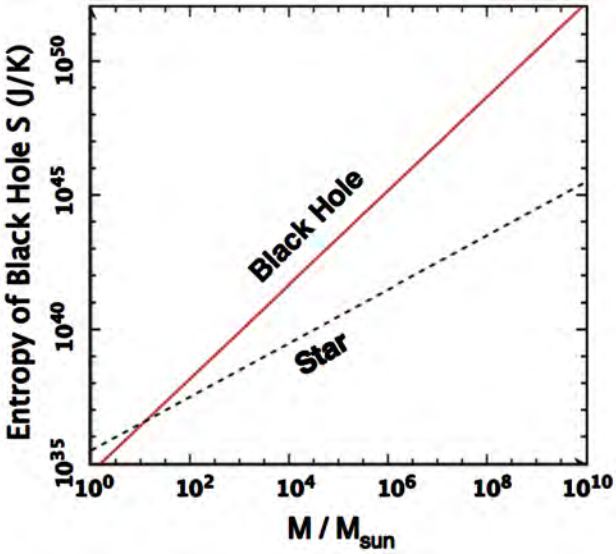


Figure 17. The entropies of a black hole and a star are plotted as functions of the mass. Results indicate that the entropy of black hole increases with mass faster than that of star does. But, the entropy of a black hole with the mass if less than 10 solar masses is lower than that of a star with the same mass.

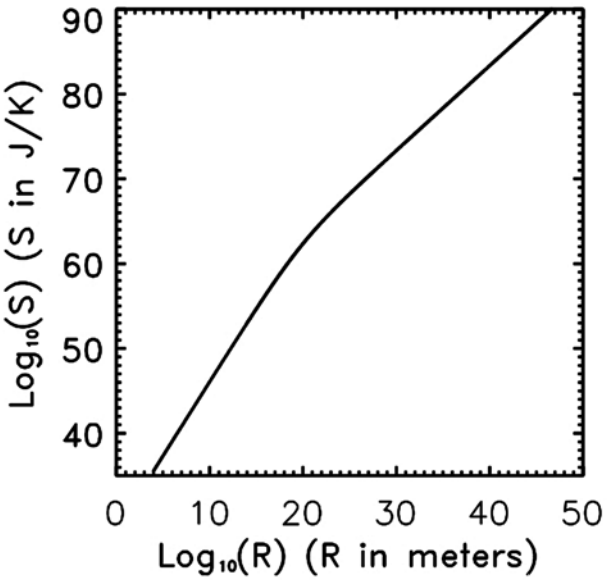


Figure 18. Entropy of a spacetime or black hole including our black hole universe. It is ploted as a function of the radius in logarithm scales. As starlike black hole grows up into a universe as big as the present universe, its entropy increases from about 10^{35} J/K to about 10^{70} J/K.

black hole with the same mass, the entropy increases to 10^{37} J/K from 10^{36} J/K. If the black hole formed has only 5 solar masses by a supernova explosion, the entropy decreases to 10^{35} J/K from 10^{36} J/K because the supernova explosion has brought a significant amount of entropy away. The entropy of a star with 50 solar masses has entropy of ~10^{37} J/K. When it forms a black hole with 10 solar masses, which has entropy of 7×10^{36} J/K, at the end of its life after supernova

explosion, the supernova explosion carries about 3×10^{36} J/K or more entropy away.

Conventionally, the entropy of a black hole is calculated according to the Hawking radiation. Considering the quantum effect near the event horizon, Hawking [42] theoretically showed that a black hole, even though in the static state (no accretion and no merger), can radiate blackbody emission called Hawking radiation with temperature of ~1.23×10^{23}/M K and power of ~3.56×10^{32}/M^2 W. Here M is the mass of black hole in kilograms. For a star-like or more massive black hole, both the temperature and power of the Hawking radiation are negligible small. In accordance with the temperature of Hawking radiation, the entropy of the black hole is obtained as 3×10^{-7} M^2 J/K, which is about 10^{20} times greater than that of star with the same mass as well as that of a dynamic black hole.

Now, we come back to Equation (33), to plot, in **Figure 18**, the numerically integrating result for the entropy of a black hole including our black hole universe as a function of radiation. It is seen that, for a star-like black hole to grow into a universe as big as the present universe, its entropy increases from about 10^{35} J/K to about 10^{70} J/K. The entropy of the present highly ordered universe is still keeping quickly growing. Since the black hole universe is not an isolated system, its entropy is mostly coming from the mother universe due to accretions of matter and radiation. This increase does not lose the order significantly. This explains why the present universe, although it fully grows with temperature less than 3 K, is still highly ordered.

5. Discussions and Conclusions

The black hole universe is not an isolated spacetime with $k = 1$ [4] [5]. Its mass must increase when it expands because $M(t)/R(t) = c^2/(2G) = \text{Constant}$. The density decreases by inversely proportional to the square of the radius (or the mass) of the black hole universe. This mass-radius or density-radius relation excludes the flatness and horizon problems. Both black hole and big bang universes are governed by the same physics—Einstein's general relativity with the FLRW metric of spacetime. The big bang universe model suggests that the observable and non-observable spaces are isotropically uniform, while the black hole universe model considers them to be hierarchically layered. Big bang universe is an isolated system with $\rho(t)R^3(t) = \text{Constant}$ when matter is dominant and $k = 0$, while black hole universe is a semi-open system with $\rho(t)R^2(t) = \text{Constant}$ and $k = 1$. Mathematically, there is only a small change from the big bang universe model to the black hole universe model, but physically the two universe models lead to completely different results.

As a summary, the author has examined and overviewed thoroughly the new black hole universe model and completely described its development from the three fundamentals and its creative explanations to the existing observations of the universe. From this comprehensive investigation of this new cosmological model, the author has further revealed the fundamental regularities and laws of

the black hole universe with respect to the spacetime origin, structure, evolution, equilibrium, expansion, acceleration, radiation energy, and entropy variation. These efforts will help us to uncover various regularities and mysteries of the universe. The results indicate that, standing on 1) the principle of spacetime black hole equivalence, 2) the cosmological principle of spacetime isotropy and homogeneity, and 3) the Einsteinian general theory of relativity, the black hole universe model can self-consistently describe the origin, structure, evolution, expansion, and acceleration of the universe, quantitatively explain the measurements of cosmic microwave background radiation, type Ia supernovae's luminosity distance and redshift, and dynamic properties of star-like, massive, and supermassive black holes such as gamma-ray bursts, X-rays flares from galactic centers, and quasars, and fully overcome the difficulties of the conventional model of the universe such as the problems of horizon, flatness, monopole, inflation, dark matter, dark energy, and so on.

The black hole universe is ruled by one theory of spacetime (*i.e.* the Einsteinian general theory of relativity), based on two principles of spacetime (*i.e.* the principle of spacetime black hole equivalence and the cosmological principle of spacetime isotropy and homogeneity), regulated with five laws of spacetime (*i.e.* the conversion of spactieme mass-radius ratio, the equilibrium of static spacetime, the expansion and acceleration of dynamic spacetime, the conservation of childish spacetime radiation energy, and the entropy increase of spacetime). It is governed by the Friedmann equation that can be derived from Einstein's general theory of relativity, with the FLRW metric that can be derived from the cosmological principle, the Mach-Schwarzschild mass-radius ratio, and the positive curvature of 3D space that can be obtained from the principle of spacetime black hole equivalence. Mathematically, the black hole universe is slightly different from the big bang universe (one has mass-radius ratio constant and $k = 1$, while another has mass constant and $k = 0$), but physically they are completely different.

Acknowledgements

The work that was done previously and involves partially with students was supported by the NSF/REU (Grant #: PHY-1263253 and PHY-1559870) at Alabama A & M University. The author is also thankful to the University's Title III Program for travel support of his attending conferences.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Brans, C.H. and Dicke, R. H. (1961) *Physical Review*, **124**, 925-935. https://doi.org/10.1103/PhysRev.124.925

[2] Zhang, T.X. (2006) *Astrophysical Journal Letters*, **636**, L61-L64. https://doi.org/10.1086/500255

[3] Zhang, T.X. (2010) *Astrophysical Journal Letters*, **725**, L117-L120. https://doi.org/10.1088/2041-8205/725/2/L117

[4] Zhang, T.X. (2007) *American Astronomical Society* 211th *Meeting*.

[5] Zhang, T.X. (2009) *Progress in Physics*, **5**, 3-11.

[6] Zhang, T.X. (2014) Observation and Theory for Multiverse. *Special Session* (*The Oral Session* #304) *of the* 224th *AAS Meeting*, Boston, 7-11 June 2014.

[7] Zhang, T.X. (2014) *American Astronomical Society* 224th *Meeting*.

[8] Zhang, T.X. (2009) *American Astronomical Society* 213rd *Meeting*.

[9] Zhang, T.X. (2009) *American Astronomical Society* 214th *Meeting*.

[10] Zhang, T.X. (2010) *American Astronomical Society* 215th *Meeting*.

[11] Zhang, T.X. (2011) *American Astronomical Society* 217th *Meeting*.

[12] Zhang, T.X. (2012) *American Astronomical Society* 219th *Meeting*.

[13] Zhang, T.X. (2012) *American Astronomical Society* 220th *Meeting*.

[14] Zhang, T.X. (2013) *American Astronomical Society* 221st *Meeting*.

[15] Zhang, T.X. (2013) *American Astronomical Society* 222nd *Meeting*.

[16] Zhang, T.X. (2014) *American Astronomical Society* 223rd *Meeting*.

[17] Zhang, T.X. (2016) *American Astronomical Society* 228th *Meeting*.

[18] Zhang, T.X. (2010) *Astrophysics and Space Science*, **330**, 157-165. https://doi.org/10.1007/s10509-010-0372-4

[19] Zhang, T.X. (2013) *Proceeding of GRB* 2013 *Symposium*, *SNSN*-323-63.

[20] Zhang, T.X. (2013) *Progress in Physics*, **9**, 33-38.

[21] Zhang, T.X. and Frederick, C. (2014) *Astrophysics and Space Science*, **349**, 567-573. https://doi.org/10.1007/s10509-013-1644-6

[22] Zhang, T.X. (2015) *Astrophysics and Space Science*, **358**. https://doi.org/10.1007/s10509-015-2409-1

[23] Zhang, T.X., Wilson, C. and Schamschula, M.P. (2016) *Progress in Physics*, **12**, 61-67.

[24] Zhang, T.X., Naka, P. and Guggilla, P. (2016) 8*th Huntsville Gamma-Ray Burst Symposium*, Huntsville, 24-28 October 2016, LPI Contribution No. 1962, Article ID: 4028.

[25] Zhang, T.X. (2016) *Progress in Physics*, **12**, 353-361.

[26] Arp, H., *et al.* (2004) An Open Letter to the Scientific Community. New Scientists.

[27] Guth, A.H. (1981) *Physical Review D*, **23**, 347-356. https://doi.org/10.1103/PhysRevD.23.347

[28] Riess, A.G., *et al.* (1998) *The Astronomical Journal*, **116**, 1009-1038. https://doi.org/10.1086/300499

[29] Perlmutter, S., *et al.* (1999) *Astrophysical Journal*, **517**, 565-586. https://doi.org/10.1086/307221

[30] Schwarzschild, K. (1916) *Mathematical Physics*, **1916**, 189-196.

[31] Sciama, D.W. (1953) *Monthly Notices of the Royal Astronomical Society*, **113**, 34-42. https://doi.org/10.1093/mnras/113.1.34

[32] Davidson, M.W. (1957) *Monthly Notices of the Royal Astronomical Society*, **117**, 212-224. https://doi.org/10.1093/mnras/117.2.212

[33] Slezak, M. (2014) Cosmic Inflation Is Dead, Long Live Cosmic Inflation. New

Scientists, September 25.

[34] Cortes, M., Liddle, A.R. and Parkinson, D. (2014) *Physical Review D*, **92**, Article ID: 063511. https://doi.org/10.1103/PhysRevD.92.063511

[35] Abbott, B.P., *et al.* (2016) *Physical Review Letters*, **116**, Article ID: 061102. https://doi.org/10.1103/PhysRevLett.116.061102

[36] Friedmann, A. (1922) *Zeitschrift fur Physik*, **10**, 377-386. https://doi.org/10.1007/BF01332580

[37] Lemaitre, G. (1931) *Monthly Notices of the Royal Astronomical Society*, **91**, 483-490. https://doi.org/10.1093/mnras/91.5.483

[38] Robertson, H.P. (1935) *Astrophysical Journal*, **82**, 284-301. https://doi.org/10.1086/143681

[39] Walker, A.G. (1937) *Proceedings of the London Mathematical Society*, **42**, 90-127. https://doi.org/10.1112/plms/s2-42.1.90

[40] Einstein, A. (1916) *Annalen der Physik*, **354**, 769-822. https://doi.org/10.1002/andp.19163540702

[41] Einstein, A. (1905) *Annalen der Physik*, **322**, 891-921. https://doi.org/10.1002/andp.19053221004

[42] Hawking, S.W. (1974) *Nature*, **248**, 30-31. https://doi.org/10.1038/248030a0

[43] Friedmann, A. (1924) *Zeitschrift fur Physik*, **21**, 326-332. https://doi.org/10.1007/BF01328280

[44] Hubble, E. (1929) *Proceedings of National Academy of Science of the United States of America*, **15**, 168-173. https://doi.org/10.1073/pnas.15.3.168

[45] Suyu, S.H., *et al.* (2010) *Astrophysical Journal*, **711**, 201-221. https://doi.org/10.1088/0004-637X/711/1/201

[46] Tytler, D., *et al.* (2004) *Astrophysical Journal*, **617**, 1-28. https://doi.org/10.1086/425226

[47] Einstein, A. (1917) *Sitzungsberichte der Koniglich Preussbischen Akademie der Wissenschaften* (*Berlin*), **8**, 142-152.

[48] Zhang, T.X. (2018) *Journal of Modern Physics*, **9**, 433-442.

[49] Shakura, N.I. and Syuanyaev, R.A. (1973) *Astronomy and Astrophysics*, **24**, 337-355.

[50] Daly, R.A., Djorgovski, S.G., Freeman, K.A., Matthew, P., O'Dea, C.P., Kharb, P. and Baum, S. (2008) *Astrophysical Journal*, **677**, 1-11. https://doi.org/10.1086/528837

[51] Amanullah, R., *et al.* (2010) *The Astrophysical Journal*, **716**, 712-738. https://doi.org/10.1088/0004-637X/716/1/712

[52] Suzuki, N., *et al.* (2012) *The Astrophysical Journal*, **746**, Article No. 85.

[53] Zhang, T.X. (2018) Entropy of Black Hole Universe, Astrophysics and Space Science. (To be Submitted)

Call for Papers

Journal of Modern Physics

ISSN: 2153-1196 (Print) ISSN: 2153-120X (Online)
http://www.scirp.org/journal/jmp

Journal of Modern Physics **(JMP)** is an international journal dedicated to the latest advancement of modern physics. 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 modern physics.

Editor-in-Chief

Prof. Yang-Hui He City University, UK

Executive Editor-in-Chief

Prof. Marko Markov Research International, Buffalo Office, USA

Subject Coverage

Journal of Modern Physics publishes original papers including but not limited to the following fields:

Biophysics and Medical Physics
Complex Systems Physics
Computational Physics
Condensed Matter Physics
Cosmology and Early Universe
Earth and Planetary Sciences
General Relativity
High Energy Astrophysics
High Energy/Accelerator Physics
Instrumentation and Measurement
Interdisciplinary Physics
Materials Sciences and Technology
Mathematical Physics
Mechanical Response of Solids and Structures
New Materials: Micro and Nano-Mechanics and Homogeneization
Non-Equilibrium Thermodynamics and Statistical Mechanics
Nuclear Science and Engineering
Optics
Physics of Nanostructures
Plasma Physics
Quantum Mechanical Developments
Quantum Theory
Relativistic Astrophysics
String Theory
Superconducting Physics
Theoretical High Energy Physics
Thermology

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 have been previously published nor be currently under consideration for publication elsewhere. Paper submission will be handled electronically through the website. All papers are refereed through a peer review process. For more details about the submissions, please access the website.

Website and E-Mail

http://www.scirp.org/journal/jmp **E-mail: jmp@scirp.org**

What is SCIRP?

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

What is Open Access?

All original research papers published by SCIRP are made freely and permanently accessible online immediately upon publication. To be able to provide open access journals, SCIRP defrays operation costs from authors and subscription charges only for its printed version. Open access publishing allows an immediate, worldwide, barrier-free, open access to the full text of research papers, which is in the best interests of the scientific community.

- High visibility for maximum global exposure with open access publishing model
- Rigorous peer review of research papers
- Prompt faster publication with less cost
- Guaranteed targeted, multidisciplinary audience

Scientific Research Publishing

Website: http://www.scirp.org

Subscription: sub@scirp.org

Advertisement: service@scirp.org