INDIAN ASSOCIATION OF PHYSICS TEACHERS

IAPT

PHYSICS EDUCATION

COVID-19 Outbreak

Prevention
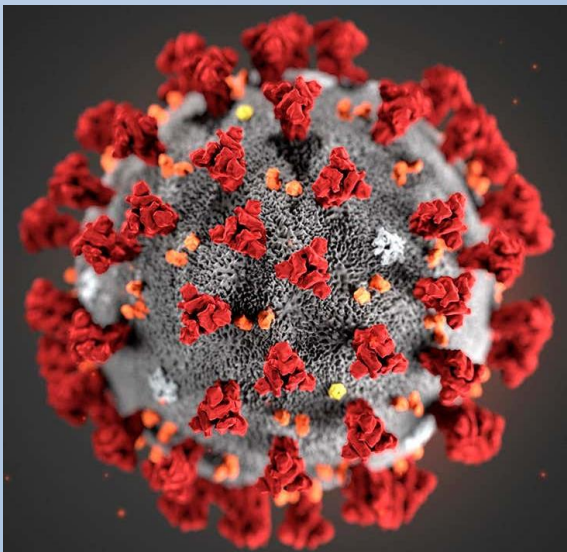
✓ Wash you hands with soap
✓ Keep Safe distance
✓ Don't touch your face
✓ Cover your mouth and nose

www.physedu.in

Volume 36, Number 1

In this Issue

Kaprekar Numbers in Binary Form

M.N.Murty[1] and Binayak Padhy[2]

[1]Visiting Faculty, Department of Physics,

Sri Sathya Sai Institute of Higher Learning,

Prasanthi Nilayam -515134, Andhra Pradesh, India.

[2]Department of Physics, Khallikote Autonomous College,

Berhampur -760001, Odisha, India.

mnarayanamurty@rediffmail.com

Abstract
The Kaprekar numbers were introduced by an Indian recreational mathematician D.R. Kaprekar. A Kaprekar number is such that its square can be split into two parts so that the sum of these two parts is equal to the number. By extending the definition of Kaprekar number to binary number, the Kaprekar number can be expressed in binary form. The multiplication of Kaprekar numbers in binary form is discussed by taking two algorithms used in digital signal processing. Multiplier circuits for squaring and multiplying kaprekar numbers in binary form are given.
Tags: Kaprekar numbers, Binary numbers, Algorithm, Multiplier.

1. Introduction
The Kaprekar numbers were introduced by Dattathreya Ramachandra Kaprekar(1905-1986). D.R.kaprekar was an Indian recreational mathematician.
A Kaprekar number for a given base is a non-negative integer, whose square can be split up into two parts, which add up to the original number again.

Let X be a non-negative integer and n is a positive integer. X is known as a n- Kaprekar number for the base b, if there exist a non-negative integer A and positive integer B satisfying the following pair of equations.

$$X^2 = Ab^n + B ,$$

Where, $0 < B < b^n$,

$$A \geq 0 \ \& \ X \geq 1 \qquad (1)$$
$$X = A + B \qquad (2)$$

The second part B in RHS of (2) must be non-zero.

$X = 1$ is taken as a Kaprekar number for all $n \geq 1$, because $1^2 = 0 \times b^n + 1 ; 1 = 0 + 1$ .Here $A = 0$ and $B = 1$.

The following is the sequence of Kaprekar numbers in base 10.

1, 9, 45, 55, 99, 297, 703, 999, 2223, 2728, 4879, 4950, 5050, 5292, 7272, 7777, 9999, 17344, 22222,...............

The following expansions show that the Kaprekar numbers satisfy the equations (1) and (2).
$$1^2 = 1 = 0 \times 10^2 + 1; 0 + 1 = 01$$

$9^2 = 81 = 8 \times 10^1 + 1; 8 + 1 = 9$

$45^2 = 2025 = 20 \times 10^2 + 25; 20 + 25 = 45$

$55^2 = 3025 = 30 \times 10^2 + 25; 30 + 25 = 55$

$99^2 = 9801 = 98 \times 10^2 + 01; 98 + 01 = 99$

$297^2 = 88209 \,; 88 + 209 = 297$

$703^2 = 494209 \,; 494 + 209 = 703$

By extending the definition of Kaprekar numbers (1) and (2), the Kaprekar numbers can be expressed in binary form. The square of a kaprekar number in binary form can be found out with multiplier. The multiplication of binary numbers is used in the field of digital signal processing for the design of digital multipliers. The squaring operation also forms the backbone in cryptography.

The rest of the paper is organized as follows. Kaprekar numbers in binary form are mentioned in Section 2. In Section 3, a multiplier circuit for squaring 2-bit kaprekar number is given. Algorithm for multiplying two kaprekar numbers in binary form is given in Section 4. Serial-parallel multiplier for multiplying two kaprekar numbers is mentioned in Section 5. Baugh-Wooley algorithm for multiplying kaprekar numbers is given in Section 6.

2. Kaprekar Numbers in Binary form

0 and 1 are called binary numbers. Computer circuits process binary numbers and they do not process decimal numbers. The basic rules of binary addition are $0 + 0 = 0$; $0 + 1 = 1$; $1 + 0 = 1$; $1 + 1 = 10$; $10 + 1 = 11$; $11 + 1 = 100$. The rules for binary multiplication are $1 \times 0 = 0$, $0 \times 1 = 0$ and $1 \times 1 = 1$. The following calculations using the rules of binary multiplication and addition for the 7 binary numbers show that the square of a binary number can be split up into two parts, which add up to give the original binary number. Hence these binary numbers represent Kaprekar numbers with binary base 2.

(a) $1^2 = 1 \times 1 = 1 = 01; 0 + 1 = 1$

(b) $11^2 = 11 \times 11 = 1001; 10 + 01 = 11$

(c) $110^2 = 110 \times 110 = 100100; 10 + 0100 = 0110 = 110$

(d) $111^2 = 111 \times 111 = 10001; 110 + 001 = 111$

(e) $1010^2 = 1010 \times 1010 = 100100; 110 + 0100 = 1010$

(f) $1111^2 = 1111 \times 1111 = 11100001; 1110 + 0001 = 1111$

(g) $11100^2 = 11100 \times 11100 = 1100010000; 1100 + 010000 = 011100 = 11100$

The first 26 Kaprekar numbers in binary form are

1, 11, 110, 111, 1010, 1111, 11100, 11111, 100100, 110011, 111111, 1010101, 1011011, 1111000, 1111111, 10001000, 10010011, 10101011, 10111011, 11001101, 11111111, 101010110, 101011111, 101101101, 111110000, 111111111, ...

The following are three important aspects of kaprekar numbers in binary form.

- Every even perfect number is a Kaprekar number in the binary base $b = 2$.

Example: An even perfect number in the binary base $b = 2$ has the form $2^{n-1}(2^n - 1)$, where $(2^n - 1)$ is prime. For $n = 3$, we have $2^2(2^3 - 1) = 28$. The binary equivalent of 28 is $(11100)_2$, because

$$2^4 \times 1 + 2^3 \times 1 + 2^2 \times 1 + 2^1 \times 0 + 2^0 \times 0 = 28$$

$(11100)_2$ is a kaprekar number in binary base as shown against the point (g) above.

- $b^{n-1}(b^n - 1)$ and $b^{n-1}(b^n + 1)$ are complementary Kaprekar numbers in the binary base $b = 2$.

Example: For $b = 2$ and $n = 3$, we get

$$b^{n-1}(b^n - 1) = 2^2(2^3 - 1) = 28$$

and

$$b^{n-1}(b^n + 1) = 2^2(2^3 + 1) = 36.$$

Already it has been shown that the number 28 in binary form is a Kaprekar number. The binary equivalent of the number 36 is $(100100)_2$ because

$2^5 \times 1 + 2^4 \times 0 + 2^3 \times 0 + 2^2 \times 1 + 2^1 \times 0 + 2^0 \times 0 = 36$

$X = 100100$ is a Kaprekar number in binary form, because X^2 can be split up into two parts A and B, whose sum is equal to X as mentioned below

$X^2 = 100100^2 = 1010001000$; $10100(A) + 010000(B) = 100100(X)$

Hence, $(11100)_2$ and $(100100)_2$ are complementary Kaprekar numbers in binary form.

- For any Kaprekar number $X_1 = b^{n-1}(b^n - 1)$ with base 2, there exists another Kaprekar number $X_2 = b^{n-1}(b^n + 1)$ with base 2 such that $X_1 + X_2 = 2^{2n}$.

Example: let $X_1 = b^{n-1}(b^n - 1)$ =28 for $b = 2$ and $n = 3$. or $X_1 = (28)_{10} = (11100)_2$

Similarly, let $X_2 = b^{n-1}(b^n + 1)$ = 36 for $b = 2$ and $n = 3$. or $X_2 = (36)_{10} = (100100)_2$

$X_1 + X_2 = (11100)_2 + (100100)_2 =$ $(1000000)_2 \Rightarrow (28)_{10} + (36)_{10} = (64)_{10} = 2^{2n}$ for $n = 3$.

3. Multiplier for squaring 2-bit Kaprekar Number

Let the product of two 2-bit binary numbers $a_1 a_o$ and $b_1 b_o$ is $P_3 P_2 P_1 P_o$ as shown in Fig.1. The hardware implementation for squaring 2-bit Kaprekar number is shown in Fig.2. It consists of four AND gates (1, 2, 3 & 4) and two half adders (HA-1 & HA-2). A half adder is a logic circuit that adds two binary digits and produces a 2-bit data, i.e., sum and carry. The 1st half adder is used to add $a_1 b_o$ and $a_o b_1$, the outputs of the 2nd and 3rd AND gates respectively. The 2nd half adder is used to add the carry c_1 generated from 1st half adder and the output $a_1 b_1$ of the 4th AND gate. The output of 1st AND gate is P_o. The sum of 1st half adder is P_1 and the sum of 2nd half adder is P_2. The carry of 2nd half adder gives P_3.

For example, while squaring the Kaprekr number $(11)_2$, we have to take $a_0 = 1, a_1 = 1, b_0 = 1$ & $b_1 = 1$ in the circuit shown in Fig.2. Then, we get $p_3 = 1, p_2 = 0, p_1 = 0$ & $p_0 = 1$ as $11^2 = 11 \times 11 = 1001$. Similarly, for squaring 3-bit Kaprekar number $(110)_2$ or $(111)_2$, a logic circuit containing 9 AND gates, 3 Full Adders and 2 Half adders is required.

$$\frac{\begin{array}{cccc} b_1 & b_0 & \times\ a_1 & a_0 \\ & & a_0 b_1 & a_0 b_0 \end{array}}{}$$

$$a_1 b_1 \quad a_1 b_0$$

$$p_3 \quad p_2 \quad p_1 \quad p_0$$

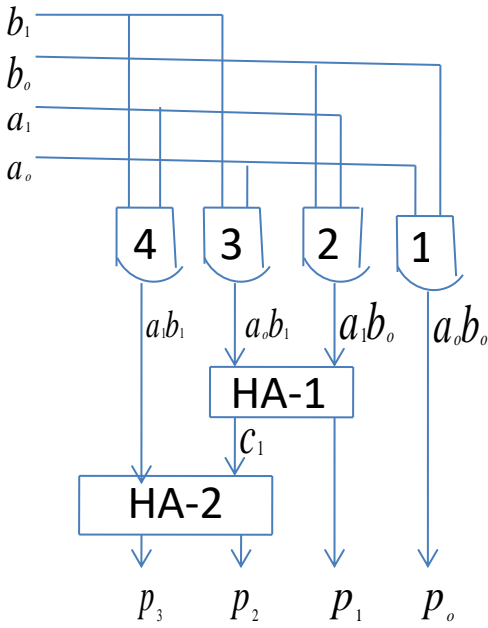Figure 1: Multiplication of two 2-bit binary numbers.

Figure 2. Block diagram of the multiplier for squaring 2-bit Kaprekar number

4. Algorithm for multiplying two Kaprekar Numbers in Binary form

The following algorithm [5] can be used for multiplying two positive or negative decimal numbers in binary form. Consider two numbers $A(m) = a_{m-1}, \ldots, a_0$ and $B(n) = b_{n-1}, \ldots, b_0$ represented by m and n bits, respectively. The product of these numbers is given by.

$$P = A(m)\, B(n)$$

In two's complement, $A(m)$ and $B(n)$ can be represented as

$$A(m) = -a_{m-1}2^{m-1} + \sum_{i=0}^{m-2} a_i 2^i \qquad (3)$$

and

$$B(n) = -b_{n-1}2^{n-1} + \sum_{j=0}^{n-2} b_j 2^j \qquad (4)$$

Then the product of these two's complement operands can be written as

$$P = \left[-a_{m-1}2^{m-1} + \sum_{i=0}^{m-2} a_i 2^i\right] \times$$

$$\left[-b_{n-1}2^{n-1} + \sum_{j=0}^{n-2} b_j 2^j\right]$$

$$= a_{m-1}b_{n-1}2^{m+n-2} + 2^{n-1}\sum_{i=0}^{m-2} 2^i (-a_i b_{n-1}) +$$

$$2^{m-1}\sum_{j=0}^{n-2} 2^j (-a_{m-1}b_j) + \sum_{i=0}^{m-2}\sum_{j=0}^{n-2} 2^{i+j} a_i b_j$$

Using the relation $C + \bar{C} = 1$ from Boolean algebra, we have

$$P = (-\overline{a_{m-1}b_{n-1}} + 1)\, 2^{m+n-2} + 2^{n-1}\sum_{i=0}^{m-2} 2^i (\overline{a_i b_{n-1}} - 1)$$

$$+ 2^{m-1}\sum_{j=0}^{n-2} 2^j (\overline{a_{m-1}b_j} - 1) + \sum_{i=0}^{m-2}\sum_{j=0}^{n-2} 2^{i+j} a_i b_j =$$

$$-\overline{a_{m-1}b_{n-1}}2^{m+n-2} + 2^{m+n-2} +$$

$$2^{n-1}\sum_{i=0}^{m-2} 2^i\, \overline{a_i b_{n-1}} - 2^{n-1}\sum_{i=0}^{m-2} 2^i$$

$$+ 2^{m-1}\sum_{j=0}^{n-2} 2^j\, \overline{a_{m-1}b_j} - 2^{m-1}\sum_{j=0}^{n-2} 2^j +$$

$$\sum_{i=0}^{m-2}\sum_{j=0}^{n-2} 2^{i+j} a_i b_j \qquad (5)$$

The sum of n terms in geometric progression is given by the formula

$$S_n = 1 + r + r^2 + r^3 + \cdots + r^{n-1} =$$

$$\sum_{i=0}^{n-1} r^i = \frac{1-r^n}{1-r} \qquad (6)$$

Using (6), we have

$$\sum_{i=0}^{m-2} 2^i = \frac{1-2^{m-1}}{1-2} = 2^{m-1} - 1 \qquad (7)$$

Similarly, $\sum_{j=0}^{n-2} 2^j = 2^{n-1} - 1 \qquad (8)$

Substituting (7) & (8) in 4th and 6th terms of (5), we obtain

$$-2^{n-1}\sum_{i=0}^{m-2}2^i - 2^{m-1}\sum_{j=0}^{m-2}2^j =$$

$$-2^{n-1}(2^{m-1}-1) - 2^{m-1}(2^{n-1}-1)$$

$$= -2^{m+n-2} + 2^{n-1} - 2^{m+n-2} + 2^{m-1} \quad (9)$$

Using (9) in (5), we get

$$P = \sum_{i=0}^{m-2}\sum_{j=0}^{n-2}2^{i+j}a_i b_j +$$
$$2^{n-1}\sum_{i=0}^{m-2}2^i\overline{a_i b_{n-1}} +$$
$$2^{m-1}\sum_{j=0}^{n-2}2^j\overline{a_{m-1}b_j} -$$
$$2^{m+n-2}\overline{a_{m-1}b_{n-1}} + 2^{m-1} +$$
$$2^{n-1} - 2^{m+n-2} \quad (10)$$

5. Serial - Parallel Multiplier for multiplying two Kaprekar Numbers in Binary form

For $m = n = 4$, (10) can be written as

$$P = 2^0 a_0 b_0 + 2^1(a_1 b_0 + a_0 b_1)$$
$$+2^2(a_2 b_0 + a_1 b_1 + a_0 b_2) +$$
$$2^3(\overline{a_3 b_0} + a_2 b_1 + a_1 b_2 + \overline{a_0 b_3})$$
$$+2^4(\overline{a_3 b_1} + a_2 b_2 + \overline{a_1 b_3} + 1) +$$
$$2^5(\overline{a_3 b_2} + \overline{a_2 b_3}) - 2^6(\overline{a_3 b_3} + 1) \quad (11)$$

The multiplication algorithm (11) resulting from (10) for a word length 4 is illustrated by the multiplication table shown in Table 1. The partial product terms are formed by ANDing each multiplicand bit with each multiplier bit. For two's complement multiplication, one NAND and three AND operations are performed in each of the first three rows of the bit products, and four NAND operations in the fourth row. The final product is computed by adding '1' to the fifth and the seventh columns along with all partial product terms. The excessive carry bit out of $(2n - 1) = 7$ bit product $P_6 P_5 P_4 P_3 P_2 P_1 P_0$ is ignored.

Fig.3 shows the steps of multiplication procedure for multiplying two Kaprekar numbers (110) and (111) in binary form using the multiplication algorithm (11). The decimal equivalent of (110) is 6 and that of (111) is 7. 0's are added to the left of these numbers to express them in 2's complement form because

$$-2^3 \times 0 + 2^2 \times 1 + 2^1 \times 1 + 2^0 \times 0 = 6$$
$$-2^3 \times 0 + 2^2 \times 1 + 2^1 \times 1 + 2^0 \times 1 = 7$$

The product is (0101010) and the excessive carry bits 10 out of this 7 bit product are ignored. The decimal equivalent of the product is 42 because

$$-2^6 \times 0 + 2^5 \times 1 + 2^4 \times 0 + 2^3 \times 1 + 2^2 \times 0 + 2^1 \times 1 + 2^0 \times 0 = 42$$

The 2's complement serial - parallel multiplier [5] comprising of a logic unit and an adder unit is shown in Fig. 4. The logic unit consist of three AND gates, one NAND gate, and three XOR gates. In this multiplier $A(m)$ is stored in parallel and $B(n)$ is entered serially from the least-significant bit to the most-significant bit. A flip-flop (FF) saves the carry bit and a full adder adds the result of the partial product. The output of the multiplier is obtained from the adder unit using a carry-save and add-shift technique (the CSAS technique) together with one control signal Q. In the first three and last two clock cycles $Q = 0$, but in the fourth, fifth and sixth clock cycles $Q = 1$. The extra '1' for the fifth and seventh columns are automatically provided by the control signal Q. Four zeros are appended to the left of the MSB of $B(n)$ for input / output synchronisation. The excessive carry bit out of $(2n\text{-}1)$ bit product is ignored.

Table 1: 4 × 4 Bit two's complement multiplication

				b_3	b_2	b_1	b_0
Multiplicand				b_3	b_2	b_1	b_0
Multiplier				a_3	a_2	a_1	a_0
			$\overline{a_3b_0}$	a_2b_0	a_1b_0	a_0b_0
		$\overline{a_3b_1}$	a_2b_1	a_1b_1	a_0b_1	
	$\overline{a_3b_2}$	a_2b_2	a_1b_2	a_0b_2		
$\overline{a_3b_3}$	$\overline{a_2b_3}$	$\overline{a_1b_3}$	$\overline{a_0b_3}$			
1		1				
P_6	P_5	P_4	P_3	P_2	P_1	P_0

```
        0  1  1  0   (6)
        0  1  1  1   (7)
       ─────────────
        1  0  0  0
     1  1  1  1
  1  1  1  1
1  1  1  1
1     1
─────────────────────
0  1  0  1  0  1  0   (42)
─────────────────────
```

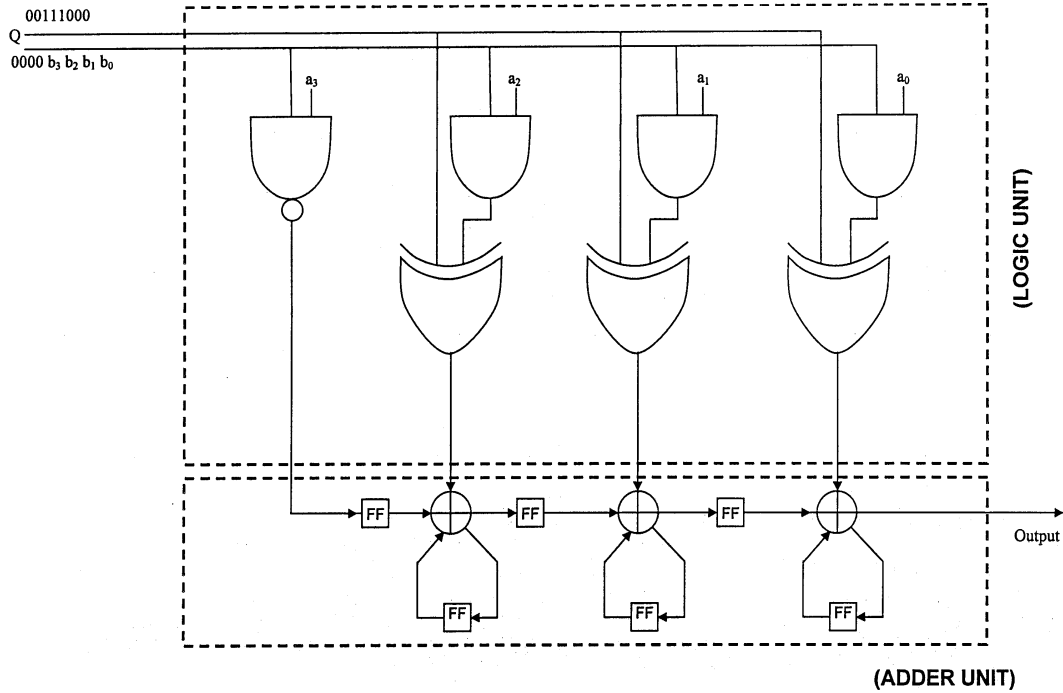Figure 3: Steps for multiplication of Kaprekar numbers 110 & 111.

Figure 4: Serial-parallel 4-bit multiplier

6. Baugh- Wooley algorithm for multiplying two Kaprekar numbers in binary form

The following is the derivation of Baugh-Wooley algorithm [1]. This algorithm can be used for multiplying two positive or negative decimal numbers in binary form.

The product P of two numbers $A = a_{n-1} \ldots a_1 a_0$ and $B = b_{n-1} \ldots b_1 b_0$ expressed in 2's complement form is given by

$$P = AB = \left(-2^{n-1}a_{n-1} + \sum_{i=0}^{n-2} 2^i a_i\right)\left(-2^{n-1}b_{n-1} + \sum_{j=0}^{n-2} 2^j b_j\right)$$

$$= \sum_{i=0}^{n-2}\sum_{j=0}^{n-2} 2^{i+j} a_i b_j + 2^{2n-2} a_{n-1} b_{n-1}$$

$$- 2^{n-1}a_{n-1}\sum_{j=0}^{n-2} 2^j b_j - 2^{n-1}b_{n-1}\sum_{i=0}^{n-2} 2^i a_i \qquad (12)$$

Where n is the number of bits in either of the numbers. Using the relation $C + \bar{C} = 1$ from

Boolean algebra in 3rd and 4th terms of (12), we get

$$P = \sum_{i=0}^{n-2}\sum_{j=0}^{n-2} 2^{i+j} a_i b_j +$$

$$2^{2n-2}a_{n-1}b_{n-1} + 2^{n-1}\left[\sum_{j=0}^{n-2} 2^j \left(\overline{a_{n-1}b_j} - 1\right) + \sum_{i=0}^{n-2} 2^i \left(\overline{a_i b_{n-1}} - 1\right)\right] \qquad (13)$$

From (6), we obtain

$$\sum_{j=0}^{n-2} 2^j = 2^{n-1} - 1 \qquad (14)$$

$$\sum_{i=0}^{n-2} 2^i = 2^{n-1} - 1 \qquad (15)$$

Substituting (14) and (15) in (13), we have

$$P = \sum_{i=0}^{n-2}\sum_{j=0}^{n-2} 2^{i+j} a_i b_j + 2^{2n-2} a_{n-1} b_{n-1}$$

$$+ 2^{n-1}\left[\sum_{j=0}^{n-2} 2^j \overline{a_{n-1}b_j}\right.$$

$$\left. + \sum_{i=0}^{n-2} 2^i \overline{a_i b_{n-1}}\right]$$

$$+ 2^n - 2^{2n-1} \qquad (16)$$

Solving the Baugh- Wooley algorithm (16) for $n = 4$, we obtain

$$2^2(a_0b_2 + a_1b_1 + a_2b_0) +$$
$$2^1(a_0b_1 + a_1b_0) + 2^0a_0b_0 \qquad (17)$$

$$P = -2^7 + 2^6a_3b_3 + 2^5(\overline{a_2b_3} + \overline{a_3b_2}) +$$
$$2^4(\overline{a_1b_3} + a_2b_2 + \overline{a_3b_1} + 1) +$$
$$2^3(\overline{a_0b_3} + a_1b_2 + a_2b_1 + \overline{a_3b_0}) +$$

Bit-level Baugh-Wooley multiplication algorithm (17) for $n = 4$ is shown Table 2.

Table 2: Baugh- Wooley multiplication algorithm for $n = 4$

----

			a_3	a_2	a_1	a_0
			b_3	b_2	b_1	b_0

				$\overline{a_0b_3}$	a_0b_2	a_0b_1	a_0b_0
			$\overline{a_1b_3}$	a_1b_2	a_1b_1	a_1b_0	
		$\overline{a_2b_3}$	a_2b_2	a_2b_1	a_2b_0		
	a_3b_3	$\overline{a_3b_2}$	$\overline{a_3b_1}$	$\overline{a_3b_0}$			
1			1				

P_7	P_6	P_5	P_4	P_3	P_2	P_1	P_0

The multiplication of two Kaprekar numbers (110) and (111) as per the algorithm (17) is given in Fig.5. This multiplication is done by making the substitutions $a_3 = 0, a_2 = 1, a_1 = 1, a_0 = 0$ and $b_3 = 0, b_2 = 1, b_1 = 1, b_0 = 1$ in Table 2. The excessive carry bits (10) out of the 7-bit product (0101010) are ignored.

				0	1	1	0	(6)
				0	1	1	1	(7)
				1	0	0	0	
			1	1	1	1		
			1	1	1	1		
		0	1	1	1			
	1		1					

| | 0 | 1 | 0 | 1 | 0 | 1 | 0 | (42) |

Figure 5: Steps for multiplication of Kaprekar numbers 110 &111 using Baugh-Wooley algorithm.

The Baugh-Wooley algorithm is a most popular algorithm in digital signal processing applications. The multipliers based on this algorithm are used for implementation of discrete orthogonal transforms like discrete Fourier transform (DFT), discrete Hartley transform (DHT) and discrete cosine transform (DCT). The algorithm discussed in Section-4 can also be used for realization of discrete orthogonal transforms.

References

[1] C. R. Baugh and B. A. Wooley, A two's complement parallel array multiplication algorithm, IEEE Trans. Comput., C-22, (1973) 1045 -1047.

[2] D. E. Iannucci and B. B. John, The kaprekar numbers, Journal of Integer Sequences, 3 , (2000), 1-5.

[3] G. B. Colin, Some properties of the Kaprekar numbers and a means of generation, Science Asia, 27,(2001) 133-136.

[4] K. M. Patil and N. P. Shah, Kaprekar numbers and its analogue equations, International Journal of Mathematical Archive, 7(6),(2016) 126-134.

[5] M. N. Murty, S. S. Nayak, B. Padhy and S. N. Panda, 4-Bit serial-parallel multiplier and bit- level systolic architecture for implementation of discrete orthogonal transforms, Communications in Computer and Information Science, Springer, 269, Part-1, (2012) 91 – 99.

[6] P. Sumathi and V. Preethy, A new approach in assignment problem, International Journal of Pure and Applied Mathematics, 113 (13) , (2017) 122-131.

Which of arithmetic averaging and least squares fitting is best for estimating time periods?

N. Inui

Graduate School of Engineering, University of Hyogo,

2167, Shosha, Himeji, Hyogo, 671-2280, Japan.

`inui@eng.u-hyogo.ac.jp`

Abstract

Arithmetic averaging and the least squares method are often employed in experimental physics. However, it is not fully understood that the least squares method may provide inferior estimates compared with the arithmetic mean. We present a simple example for the estimation of time periods, in which the least squares method always provides a larger standard deviation than arithmetic averaging.

1   Introduction

The measurement of time periods is an important subject in experimental physics, and various experiments have been conducted regarding this [1]. As an example, we consider the measurement of the time period of a stroboscope [2] using a stop watch. When the strobe light glows, we start the watch, and we stop the watch when the light extinguishes. We obtain the time period for one cycle through this procedure. By re-peating this procedure and calculating the arithmetic mean, we obtain the best estimate for the time period [3]. More simply, if the times at which the light glows are recorded, then the difference between successive times gives the time period. When we plot the time as a function of the number of measurements, we will observe a linearly increasing point sequence. Then, we apply the least squares method to the measured data, [4, 5, 6] and estimate the slope of the linear fitting function. Here, a natural question arises: which of the arithmetic mean and the least squares method is better for estimating the time period? The answer is clear, the arithmetic mean is superior. However, it is not easy to explain why the least squares method is inferior, and to what extent it is. In order to answer this question, we consider the dependence of the standard deviation of the obtained slope on the amount of data. Before going into a detailed discussion, we emphasize that the least squares method is very useful for analyzing experi-

ments [7]. As example, we can estimate a speed of a car moving at a constant speed by applying the least squares methods to the positions observed at regular time intervals. A least squares fit provides a better estimate than a simple arithmetic averaging in this case. However, we must carefully use the least squares method.

2   Mathematical model

We consider a stroboscope whose time period τ obeys a probability density function $f(\tau|\mu,\sigma)$, with the mean time period μ and standard deviation σ. The measurement begins at time $t = 0$, and the time at which the light illuminates is recorded n times. We denote the i-th time period by τ_i, and assume that the sequence of time periods is statistically independent. The time at which the light is illuminated for the i-th time is expressed as

$$T_i = \sum_{j=1}^{i} \tau_j. \tag{1}$$

Figure 1 shows an example of the time intervals with $n = 4$, where the dashed line shows the linear fitting function, $T = ai$.

For simplicity, we consider a linear function through the origin. By applying the least squares method to the measurement result, the slope a is given by

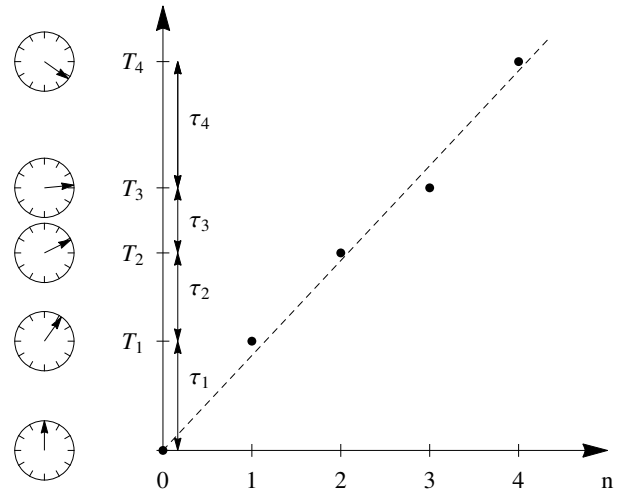$$a(\{\tau_i\}) = \frac{\displaystyle\sum_{i=1}^{n} iT_i}{\displaystyle\sum_{i=1}^{n} i^2}. \tag{2}$$

Figure 1: Plot of time when a strobe light glows. The dashed straight line shows a linearly fitting function.

Substituting (1) into (2), the slope a is expressed as

$$a(\{\tau_i\}) = \sum_{i=1}^{n} c_{i,n}\tau_i, \tag{3}$$

where

$$c_{i,n} = \frac{3(n+i)(n+1-i)}{n(1+n)(1+2n)}. \tag{4}$$

Accordingly, the ensemble average of the slope $\langle a \rangle$ is given by

$$\langle a(\{\tau_i\}) \rangle = \sum_{i=1}^{n} c_{i,n}\langle \tau_i \rangle. \tag{5}$$

Using $\langle \tau_i \rangle = \mu$ and the equation

$$\sum_{i=1}^{n} c_{n,i} = 1, \tag{6}$$

we find that $\langle a \rangle$ is equal to μ. Thus, there is no difference between arithmetic mean and the slope obtained by the least squares method for the estimate of the time period.
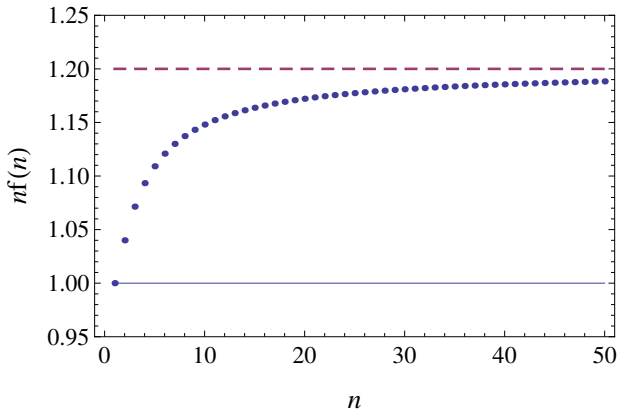
Figure 2: Plot of $nf(n)$ as function of the number of sample n, which exists between 1 and 6/5 = 1.2.

3   Variance and standard deviation of the time period

Because we assumed that τ_i is independent, the variance of the time period estimated using the arithmetic mean is given by

$$\left\langle \left(\mu - \frac{1}{n} \sum_{i=1}^{n} \tau_i \right)^2 \right\rangle = \frac{\sigma^2}{n}. \qquad (7)$$

Similarly, the variance of the time period estimated using the least squares method is defined by

$$V_L \equiv \langle (\mu - a(\{\tau_i\}))^2 \rangle. \qquad (8)$$

Substituting (2) into (8), we obtain

$$V_L = f(n)\sigma^2, \qquad (9)$$

where

$$f(n) \equiv \frac{6(1 + 2n(n+1))}{5n(n+1)(2n+1)}. \qquad (10)$$

The asymptotic behaviour of $f(n)$ in the limit of $n \to \infty$ can be expressed as

$$f(n) \approx \frac{6}{5n} + \mathcal{O}(n^{-2}). \qquad (11)$$

Figure 2 shows $nf(n)$, which converges to 6/5 in the limit $n \to \infty$. We find that the variance of the time period estimated using the least squares method becomes gradually worse as the amount of data increases. However, it is bounded below $6/5n$. For large n, the standard deviation for the time period obtained using the least squares method is

$$\sigma_L \approx \sqrt{\frac{6}{5}} \frac{\sigma}{\sqrt{n}}. \qquad (12)$$

As a result, the standard deviation becomes about 10% worse for large n.

4   Linear fitting with non-zero intercept

In the Section 2, we used a linear function intercepting the origin. We change the fitting function. The slope A obtained by fitting the data T_i $(i = 1, \ldots, n \geq 2)$ to the function $T = Ai + B$ can be expressed as

$$A(\{\tau_i\}) = \frac{n \sum_{i=1}^{n} iT_i - \sum_{i=1}^{n} i \sum_{i=1}^{n} T_i}{n \sum_{i=1}^{n} i^2 - \left(\sum_{i=1}^{n} i \right)^2}. \qquad (13)$$

By substituting (1) into (13) and calculating the ensemble average, we find that $\langle A \rangle = \mu$. The variance calculated using the procedures described in the Section 3 is given by

$$V_G \quad \equiv \quad \langle (\mu - A(\{\tau_i\}))^2 \rangle \qquad (14)$$

$$= \quad g(n), \qquad (15)$$

where

$$g(n) \equiv \frac{6(n^2 + 1)}{5n(n^2 - 1)}. \qquad (16)$$

A comparison between $ng(n)$ and $nf(n)$ is shown in figure 3. In contrast to $nf(n)$, the function $ng(n)$ decreases as n increases. However, $ng(n)$ is greater than $6/5$ for any n. In particular, a large deviation is observed for small n. In conclusion, the estimation of the time period is not improved by generalising the fitting function.
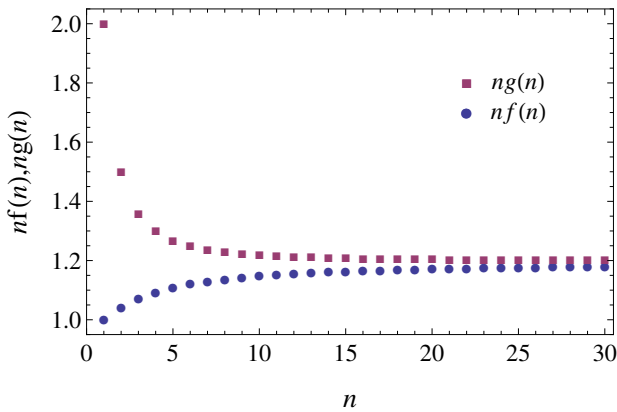

Figure 3: Comparison between $nf(n)$ (solid circle) and $ng(n)$ (solid square).

5 Estimate of a time-dependent period

We assumed that the time period is statistically independent. If a periodically changing signal is imposed on the time period, then the above results will not necessarily hold. For example, we replace the probability density function $f(\tau|\mu,\sigma)$ with the following function depending on the parity of the number of flashes:

$$F(\tau_i|\mu,\sigma) = f(\tau_i|\mu,\sigma) + (-1)^i \epsilon, \quad (17)$$

where $\epsilon > 0$ is a constant and n is assumed to be an even number. By using arithmetic averaging, we have that

$$\frac{1}{n}\sum_{i=1}^{n} \tau_i F(\tau_i|\mu,\sigma) = \mu. \quad (18)$$

On the other hand, the slope obtained by fitting to the linear function $T = \alpha i$ is given by

$$\langle \alpha \rangle = \mu - \frac{3n}{2 + 6n + 4n^2}\epsilon. \quad (19)$$

Therefore, we have different estimates. Because the time period cannot be characterised by a single value, we cannot question superiority of one method over the methods. However, it should be noted that not only the variance, but also the ensemble average may differ.

6 Conclusion

Using spreadsheet applications makes it easy to create graphs [8], and even the least squares method can be easily applied. The least squares method seems sophisticated in comparison with the arithmetic mean in some cases. However, as shown in this study, the least square method may yield inferior estimates. The reason for this can be explained using equation (3). The left-hand side of equation (3) indicates that the slope is given by the weighted sum of τ_i. Because each measured value is independent and obeys the same probability distribution function, measured values should be treated equally. However, the weight $c_{i,n}$ is a monotonically decreasing function of i, so that the measured values for small i have a large influence on the estimation of a. Although the

least squares method is useful, we keep in mind that it is not a tool that reduces uncertainty.

References

[1] V. Oliveira (2014) Phys. Educ. 49, 387.

[2] J. Birriel (2016) Phys. Teach. 54, 54.

[3] I. Duerdoth (2009) Phys. Educ. 44, 138.

[4] P. G. Hoel (1947) *Elementary statistics* (Wiley series in probability and mathematical statistics, Applied probability and statistics), p. 79.

[5] E. J. Burge (1966) Phys. Educ. 29, 59.

[6] C. Cooke (1974) Phys. Educ. 9, 321.

[7] R. J. E. Merry, M. J. G. can de Molengraft, and M. Steinbuch(2010) Mechatronics 20, 20.

[8] B. C. Reed (2010) Phys. Educ. 45, 93.

The influence of the second degree of freedom on the peculiarities of thread pendulum oscillations

V. Ivchenko

Department of Natural Sciences Training, Kherson State Maritime Academy, Kherson 73000, Ukraine.

reterty@gmail.com

Abstract

We investigate the influence of the second degree of freedom of a thread pendulum on the deviations from simple pendulum model. It is found that this factor has less but comparable effect than the non-linearity of the oscillations, especially at the low angular amplitudes for which the experiments are usually performed. Therefore, this correction should be added to the set of corrections to the oscillation period of simple pendulum while determining the precise value of the local gravity acceleration.

1 Introduction

The scientific ideal physics models are one of the most common classes of concepts, which are considered in the high school course of physics [1]. Among them the limit transition abstractions play a key role, i.e. the models constructed by limit transitions for the selected number of characteristics to their maximum, minimum or constant val-

ues [2]. For example, in the model of a rigid body one approaches its hardness to infinity. Within the model of an ideal gas the molecule sizes and the values of interaction forces between them are tended to zero. For the model of uniform vector field its vector-valued function is assumed to be independent from the spatial coordinates, i. e. it is constant in direction and modulus. Such idealizations describe the behaviour of real physical objects not only qualitatively but also (most importantly) quantitatively.

The correctness of the applicability of a particular model depends on the specific physical situation itself and, strictly speaking, should be determined by means of a physical experiment. However, for education purposes the accuracy of a model, i.e. the errors arising as a consequence of neglecting several factors considered to be insignificant in the construction of the model can be estimated theoretically within the extended model taking into account these neglected factors.

Let us consider as an example the model of simple gravity pendulum. The characteristic simplifications of such an idealization are the following:

1) the amplitude of pendulum oscillations is small;

2) the buoyancy force acting on the bob from a medium and its added mass [3] are neglected;

3) the rod or cord on which the bob swings is massless;

4) the rod or cord is inextensible;

5) the rod or cord always remains taut;

6) the bob is a point mass;

7) the support does not move;

8) the motion does not lose energy to friction or air resistance;

9) motion occurs only in two dimensions, i.e. in vertical plane.

It is well known that the period of oscillations of this pendulum is defined by the following simple relation:

$$T_0 = 2\pi\sqrt{\frac{l}{g}},\qquad(1)$$

where l is the rod length, g is the gravity acceleration. The comprehensive analysis of the corrections to Eq. (1) in the cases of violation of the conditions (1-8) was provided in article [3]. However, no attempt was made to investigate the influence of the second degree of freedom on the period of thread pendulum oscillations.

The condition (9) is important in the practical use of a simple pendulum to determine the gravity acceleration. The point is that in the process of performing a laboratory experiment the thread pendulum oscillations are not "perfectly vertically flat" due to the influence of random factors. Such a motion of a bob is described by the model of a spherical pendulum [4], which has two degrees of freedom. Consequently, the model of a simple pendulum is the limiting case of the model of a spherical pendulum. The aim of the present article is the quantitative analysis of the influence of the second degree of freedom on the deviation of the period of the spherical pendulum oscillations from the period of the simple pendulum oscillations.

2 The spherical pendulum

The theory of a spherical pendulum is well-developed ([5], [6]) and contains analytical expressions for all quantities describing the motion of such a system. Let $l = OM = $ const be the radius of the sphere on which the point mass moves (the length of the suspension). The OZ axis is directed vertically downward from the center of the sphere O. First, we assume that the pendulum is deflected from this axis through the initial angle θ_0 (angular amplitude) and released without the initial velocity in the vertical plane. Without loss of generality we also assume that at the instant of its release the point mass was given some velocity $\vec{v}_0$ in the horizontal direction normal to the plane passing through the OZ axis and the suspension. Then the trajectory of the point mass

in the horizontal plane is the unclosed curve (Fig. 1) close to the ellipse, which in its turn rotates about the OZ axis in the direction determined by the direction of the initial velocity $\vec{v}_0$ [7].
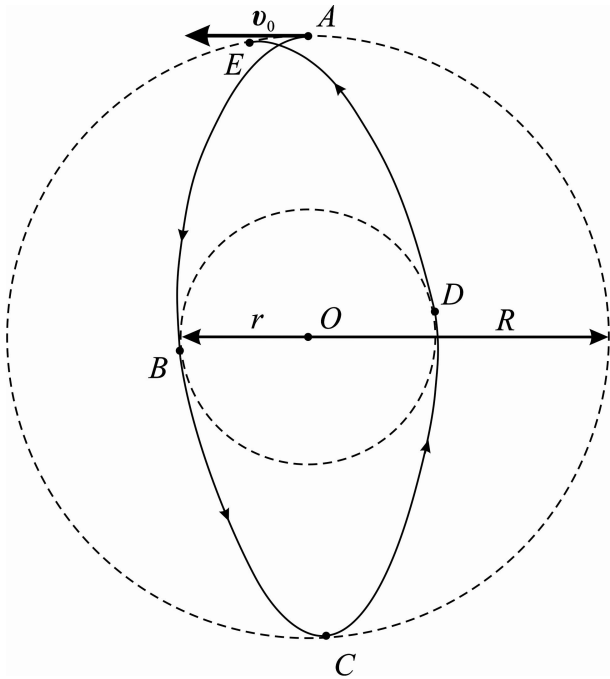

Figure 1: The trajectory of the point mass in the horizontal plane for the case of the spherical pendulum.

Such a rotation of the plane of the spherical pendulum oscillations of is much faster than the Foucault precession and can be easily observed in the experiment. The trajectory of the motion of this pendulum should alternately touch two circles with radii $R = l \sin \theta_0$ and $r = l \sin \theta_1$ ($R > r$) in points A, B, C, D, E.... . Here $\theta_1 < \theta_0$ is the maximum

angle of the pendulum deflection from the initial vertical plane.

In the theory of the spherical pendulum [7] it is proved that the time interval between two consecutive contacts of the point mass with these circles is constant and equal to

$$\int_{u_0}^{u_1} \frac{du}{\sqrt{F(u)}}, \qquad (2)$$

where $u = \cos \theta$ ($u_1 = \cos \theta_1$, $u_0 = \cos \theta_0$); $F(u)$ is the following cubic function of u (see Eq. (16) in Ref. [8]):

$$F(u) = \frac{8\pi^2}{T_0^2}(u_0 - u)[u^2 + k(u + u_0) - 1].$$
$$(3)$$

Here

$$k = 2 \left(\frac{\pi v_0}{g T_0} \right)^2.$$

Equation $F(u) = 0$ has three following roots: u_0 and

$$u_{1,2} = \frac{-k \pm \sqrt{k^2 + 4(1 - ku_0)}}{2}, \qquad (4)$$

which are found by solving the quadratic equation corresponding to the square bracket in Eq. (3) being zero. Roots u_0 and u_1 are the turning points [8], whereas $u_2 < -1$ has only formal sense (spurious solution). Thereby, we can factor:

$$F(u) = \frac{8\pi^2}{T_0^2}(u - u_0)(u_1 - u)(u - u_2).$$

Then, using Eq. (2), for the quasi-period of the spherical pendulum oscillations we get:

$$T = 4 \int_{u_0}^{u_1} \frac{du}{\sqrt{F(u)}} = \frac{\sqrt{2}}{\pi} T_0 \int_{u_0}^{u_1} \frac{du}{\sqrt{(u - u_0)(u_1 - u)(u - u_2)}}.$$

Let us introduce the change of variable in the last integral: $u = u_0 + (u_1 - u_0)\cos^2\phi$ ($0 \leq \phi \leq \pi/2$). Then, finally we derive:

$$T = \frac{2\sqrt{2}}{\pi\sqrt{u_1 - u_2}} T_0 \int_0^{\frac{\pi}{2}} \frac{d\phi}{\sqrt{1 - m^2 \sin^2\phi}} = \frac{2\sqrt{2}\mathbf{K}(m)}{\pi\sqrt{u_1 - u_2}} T_0, \tag{5}$$

where $\mathbf{K}(m)$ is the complete elliptic integral of the first kind [9] with elliptic modulus

$$m = \sqrt{\frac{u_1 - u_0}{u_1 - u_2}}.$$

Since the experimentally observed values are angles θ_0 and θ_1 rather than k, we express root u_2 through the other two roots of cubic equation $F(u) = 0$. Using Eq. (4) we have:

$$k = \frac{1 - u_1^2}{u_0 + u_1}$$

and

$$u_2 = \frac{1}{2}\left(-\frac{1 - u_1^2}{u_0 + u_1} - \sqrt{\left(\frac{1 - u_1^2}{u_0 + u_1}\right)^2 + 4\frac{u_1 + u_1^2 u_0}{u_0 + u_1}}\right). \tag{6}$$

For the very small angles θ_0 and θ_1 ($u_{0,1} \to 1$, $u_2 \to -1$) we have: $m \to 0$. In this case we can use the approximate expression [9]

$$\mathbf{K}(m) \approx \frac{\pi}{2}\left(1 + \frac{m^2}{4}\right)$$

and obtain

$$T \approx T_0\left(1 + \frac{\sin^2\theta_0 + \sin^2\theta_1}{16}\right). \tag{7}$$

This expression is in accordance with Eq. (26) of Ref. [6].

3   The numerical results and discussion

Fig. 2 shows the results of numerical calculations of relative error $\varepsilon_T = (T - T_0)/T$, depending on angle θ_0, based on Eqs. (5), (6) at different values of θ_1. It is seen that taking into account non-zero value of θ_1 will increase the period of pendulum oscillations (this result follows even from Eq. (7)). Since at $\theta_1 = 0$ Eq. (5) gives ones the value of the period of non-linear oscillations of simple pendulum, we are able to compare the deviations from simple pendulum model due
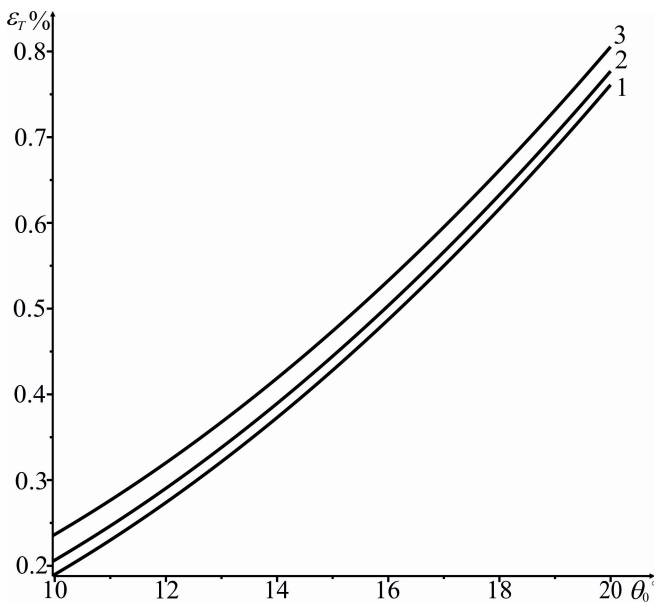
Figure 2: Dependence $\varepsilon_T(T)$ at the angles: 1) $\theta_1 = 0$; 2) $\theta_1 = 3°$; 3) $\theta_1 = 5°$.

to the non-linearity and the presence of the second degree of freedom. We come to the conclusion that the second factor has less but comparable effect than the first one, especially at low angles of θ_0 for which the experiments are usually performed. Therefore, this correction should be added to the set of corrections [3] to the oscillation period of simple pendulum while determining the precise value of the local gravity acceleration.

References

[1] E. Etkina, A. Warren and M. Gentile (2006). Phys. Teach. 44(1), 34

[2] M. Weisberg (2007). J. Philos. 104(12), 639

[3] R. A. Nelson, M. G. Olsson (1986). Am. J. Phys. 54(2), 112

[4] F. Ghorbel and J. B. Dabney (1999). IEEE Trans. Educ. 42(4), 359

[5] M. G. Olsson (1978). Am. J. Phys. 46(11), 1118

[6] M. G. Olsson (1981). Am. J. Phys. 49(6), 531

[7] L. D. Landau and E. M. Lifshitz Mechanics (Oxford: Pergamon, 3rd edn, 1976)

[8] H. Essen and N. Apazidis (2009). Eur. J. Phys. 30(2), 427

[9] M. Abramowitz and I. A. Stegun Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables (U. S. Government Printing Office, Washington, D. C., 1964)

The fallacy of Schott energy-momentum

Ashok K. Singal

Astronomy and Astrophysics Division, Physical Research Laboratory
Navrangpura, Ahmedabad 380 009, India.

ashokkumar.singal@gmail.com

Submitted on 02-11-2019

Abstract

The incompatibility between Larmor's formula for radiation losses (at a rate proportional to square of the acceleration of the electric charge) and the radiation reaction (the rate of loss of momentum of the accelerated charge proportional to its rate of change of acceleration) has recently been shown to arise because a proper distinction is not kept between radiation losses calculated in terms of a retarded time and those expressed in terms of a "real time". However, the occurrence of this disparity between two formulations is usually reconciled in literature by proposing an acceleration-dependent term, called Schott energy, lying somewhere in the nearby electromagnetic fields of an accelerated charge. But nobody has yet unambiguously demonstrated where the Schott energy, if any, actually lies in the fields. By scrutinizing electromagnetic fields of a uniformly accelerated charge, a mathematically tractable case, we show that contrary to the ideas prevalent in the literature, there is no evidence of any acceleration-dependent Schott energy-momentum in the electromagnetic fields, anywhere in the near vicinity of the charge or elsewhere. Accordingly, we expose the fallacy of this elusive Schott energy-momentum term, which should henceforth be abandoned, in the electromagnetic radiation formulation.

1 Introduction

According to Larmor's formula (or its relativistic generalization, Liénard's formula), electromagnetic power is radiated from an accelerated charge at a rate proportional to square of its acceleration [1, 2, 3, 4]. From that one can also infer the rate of momentum carried by the electromagnetic radiation which turns out to be directly proportional to the velocity vector of the charge multiplied by square of its acceleration [5, 6]. The picture here does not seem to be complete however, because if one attempts to compute the consequential rate of energy-momentum loss from the radiating charge, one encounters not only a direct violation of the energy-momentum conservation law but also sees a conflict with the special the-

ory of relativity [7].

Abraham [8, 9] and Lorentz [10, 11] derived for an accelerated charge, a formula for the self-force, widely known as radiation reaction, which gives the rate of loss of momentum of the accelerated charge proportional to its rate of change of acceleration [12, 13, 14, 15]. The same formula is also obtained, independently, from momentum conservation law by using the Maxwell stress tensor to calculate the rate of electromagnetic momentum flow across a spherical surface surrounding the neighbourhood of a point charge [16]. A scalar product of the radiation reaction with the velocity of the charge yields the rate of power loss of the accelerated charge. The radiative power loss can also be obtained directly from the Poynting flux in the neighbourhood of a point charge in arbitrary motion, leading exactly to the same formula [17].

The disparity between the two power loss formulas (one proportional to the square of acceleration and the other proportional to the scalar product of the velocity and the rate of change of acceleration of the charge), has remained a nagging puzzle for almost a century. However, recently it has been explicitly shown that this well-known incompatibility in the two formulas is succinctly resolved when a proper distinction is made between radiation losses expressed in terms of a retarded time and those expressed in terms of a "real time" [18]. Further, from an examination of the electromagnetic fields of a uniformly accelerated charge, it was

demonstrated why no radiation occurs in the case of such a charge [19]. It so happens in this case, *at all distances r* from the charge, the acceleration fields ($\propto a/cr$) strangely get cancelled by a transverse term of the velocity fields ($\propto v/r^2$). This is because in the case of a uniform acceleration, the time-retarded value of velocity, computed for a time interval r/c earlier, has an acceleration-dependent component ($\propto -ar/c$); the resulting term in velocity fields cancels the acceleration fields for all r. Consequently, no electromagnetic radiation will be detected by any observer from a uniformly accelerated charge, even in the far-off zone. This of course, contradicts Larmor's formula, according to which a uniformly accelerated charge would radiate power at a constant rate, proportional to the square of its acceleration. On the other hand, the absence of radiation from such a charge is in concurrence with the absence of radiation reaction in this case since the rate of change of acceleration of the charge here is nil.

Nevertheless, according to the conventional wisdom, the radiative power loss is believed to be given correctly by Larmor's formula, while at the same time, the rate of loss of momentum is thought to be described correctly by the radiation reaction formula, and these two apparently conflicting formulations are reconciled by proposing the presence of an extra term, called Schott term [12], in the fields of an accelerated charge. However, a physical meaning of this acceleration-dependent Schott

term is still eluding [6, 20, 21, 22, 23, 24, 25] and one does not encounter such an acceleration-dependent energy-momentum term elsewhere in physics.

The Schott energy-momentum is thought to be present in the electromagnetic fields in the vicinity of an accelerated charge. From the differences in the two conflicting formulas, we shall first arrive at the expression for the Schott term in a 4-vector form. Thereafter we shall examine the case of a uniformly accelerated charge, where the expression for the electromagnetic fields is relatively simple and the Schott term, if present therein, should be tractable mathematically in an exact manner. From a careful scrutiny of the electromagnetic fields of such an accelerated charge, we explore whether there really is some evidence of the Schott energy-momentum term in the vicinity of the charge as postulated in the literature, or is it merely an assertion based on fallacious arguments.

2   Larmor's/Liénard's radiation formula

According to Larmor's radiation formula, the radiative power loss for an accelerated charge in an inertial frame $\mathcal{S}'$, in which the charge is moving with a non-relativistic velocity ($v' \ll c$), is [2, 3, 4]

$$\mathcal{P}'_1 = \frac{2e^2\dot{\mathbf{v}}'^2}{3c^3}\,. \tag{1}$$

Throughout, a prime (') over a quantity indicates its value in frame $\mathcal{S}'$, where the mo-

tion of the charge is non-relativistic, in order to distinguish from its value in a relativistic formulation, elsewhere.

However, the consequential net rate of momentum loss to Larmor's radiation by such a charge is nil

$$\mathbf{F}'_1 = 0\,. \tag{2}$$

This is because the radiation pattern possesses an azimuth symmetry ($\propto \sin^2\phi$) in the case of a non-relativistic motion [2, 3, 4].

For a relativistic motion, the rate of energy-momentum loss of an accelerated charge due to *radiation damping* can be expressed in a 4-vector form as

$$\mathcal{F}^\mu_1 = \frac{2e^2}{3c^5}\,\dot{v}^\alpha\dot{v}_\alpha v^\mu\,, \tag{3}$$

where in all covariant equations, dot represents a proper time derivative [6].

The time part, $\mathcal{F}^0_1$, is $\gamma\mathcal{P}_1/c$, where

$$\mathcal{P}_1 = \frac{2e^2}{3c^3}\,\dot{v}^\alpha\dot{v}_\alpha = \frac{2e^2\gamma^4}{3c^3}\left[\dot{\mathbf{v}}\cdot\dot{\mathbf{v}} + \frac{\gamma^2(\dot{\mathbf{v}}\cdot\mathbf{v})^2}{c^2}\right] \tag{4}$$

is Liénard's formula (in cgs units) for power being lost by a radiating charge [2, 3, 4] in say, frame $\mathcal{S}$, where the velocity v of the charge may be relativistic, while the space part, $\mathcal{F}^i_1$, is γ times the ith component, for $i = 1, 2, 3$, of the rate of momentum being lost in frame $\mathcal{S}$ into radiation [5, 6]

$$\mathbf{F}_1 = \frac{\mathcal{P}_1}{c^2}\mathbf{v}\,. \tag{5}$$

3 Abraham-Lorentz radiation reaction

Larmor's formula purportedly uses Poynting's theorem of energy conservation to relate Poynting flux through a spherical surface of radius r at a time t, to the rate of loss of kinetic energy of the radiating charge at a retarded time $t - r/c$. However, in Poynting's theorem *all quantities* are supposed to be calculated for the *same instant of time* [2, 3, 4]. A correct application of the Poynting's theorem, using real-time values of the charge motion, gives instantaneous power loss of the charge in frame $\mathcal{S}'$, where the motion of the charge is non-relativistic, as [17]

$$\mathcal{P}_2' = -\frac{2e^2}{3c^3}\ddot{\mathbf{v}}' \cdot \mathbf{v}' . \qquad (6)$$

Similarly the electromagnetic momentum flow across a surface surrounding the vicinity of a point charge, computed employing the Maxwell's stress tensor, yields a rate of *loss of mechanical momentum* of the charge in frame $\mathcal{S}'$ as [16]

$$\mathbf{F}_2' = -\frac{2e^2}{3c^3}\ddot{\mathbf{v}}' . \qquad (7)$$

Equation (7) is the famous Abraham-Lorentz radiation reaction formula, derived usually in a quite involved way by computing the net self-force on the accelerated charge [8, 9, 10, 11, 12, 13, 14, 15]. For this one considers the charge to be in the shape of a small spherical shell and the force on every tiny bit of the shell, due to the time-retarded fields of the rest of the shell charge

distribution, is calculated and then the total force is calculated by summing over the whole spherical shell. But the same result now has been obtained in an independent manner from the momentum conservation theorem [16].

A relativistic generalization of Eq. (6) yields the instantaneous power loss of the charge in frame $\mathcal{S}$ [15, 16]

$$\mathcal{P}_2 = -\frac{2e^2\gamma^4}{3c^3}\left[\ddot{\mathbf{v}} \cdot \mathbf{v} + \frac{3\gamma^2(\dot{\mathbf{v}} \cdot \mathbf{v})^2}{c^2}\right] , \qquad (8)$$

while a relativistic generalization of Eq. (7) gives the rate of loss of mechanical momentum of the charge in frame $\mathcal{S}$ as [14, 15, 16]

$$\begin{aligned}\mathbf{F}_2 &= -\frac{2e^2\gamma^2}{3c^3}\left[\ddot{\mathbf{v}} + \frac{\gamma^2(\ddot{\mathbf{v}} \cdot \mathbf{v})\mathbf{v}}{c^2}\right.\\ &\left.+ \frac{3\gamma^2(\dot{\mathbf{v}} \cdot \mathbf{v})\dot{\mathbf{v}}}{c^2} + \frac{3\gamma^4(\dot{\mathbf{v}} \cdot \mathbf{v})^2\mathbf{v}}{c^4}\right].(9)\end{aligned}$$

We can express Eqs. (8) and (9) in a 4-vector form $\mathcal{F}_2^\mu$, where

$$\mathcal{F}_2^0 = \frac{\gamma\mathcal{P}_2}{c}, \qquad (10)$$
$$\mathcal{F}_2^i = \gamma\mathbf{F}_2^i, \quad i = 1, 2, 3 . \qquad (11)$$

4 Schott energy-momentum term

Power loss by the charge due to radiation reaction (Eq. (6)), is related to the radiated power by Larmor's formula (Eq. (1)) in frame $\mathcal{S}'$, i.e., for a non-relativistic motion, as

$$\mathcal{P}_2' = \mathcal{P}_1' - \frac{2e^2}{3c^3}\frac{\mathrm{d}(\dot{\mathbf{v}}' \cdot \mathbf{v}')}{\mathrm{d}t} . \qquad (12)$$

The last term on the right hand side in Eq. (12) is known as the Schott term, after Schott [12] who first brought it to attention. Schott term is a total derivative

and is thought in literature to arise from an acceleration-dependent energy, $-2e^2(\dot{\mathbf{v}}' \cdot \mathbf{v}')/3c^3$, in electromagnetic fields [20, 21, 22, 23, 24, 25].

We can also express $\mathbf{F}_2'$ in terms of $\mathbf{F}_1'$ (Eqs. (7) and (2)), again in frame $\mathcal{S}'$ for a non-relativistic case, as

$$\mathbf{F}_2' = \mathbf{F}_1' - \frac{2e^2}{3c^3}\ddot{\mathbf{v}}' . \tag{13}$$

The last term on the right hand side, again, is a total derivative, assumedly arising from an acceleration-dependent momentum, $-2e^2\dot{\mathbf{v}}'/3c^3$, apparently in electromagnetic fields.

Radiation reaction in the covariant form [26] yields a 4-vector, $\mathcal{F}_s^\mu$, for the Schott term

$$\mathcal{F}_2^\mu = \mathcal{F}_1^\mu + \mathcal{F}_s^\mu = \mathcal{F}_1^\mu - \frac{2e^2}{3c^3}\dddot{v}^\mu . \tag{14}$$

$\mathcal{F}_s^\mu$ is a proper-time derivative of the Schott energy-momentum, $\mathcal{E}_s^\mu = -2e^2\dot{v}^\mu/3c^3$. The 4-acceleration $\dot{v}^\mu$ is obtained from the 4-velocity $(\gamma c, \gamma \mathbf{v})$ by a differentiation with proper time and the 4-vector $\mathcal{E}_s^\mu$ then is

$$\mathcal{E}_s^0 = -\frac{2e^2\gamma^4}{3c^4}\dot{\mathbf{v}} \cdot \mathbf{v}, \tag{15}$$

$$\mathcal{E}_s^i = -\frac{2e^2\gamma^2}{3c^3}\left[\dot{\mathbf{v}}^i + \frac{\gamma^2(\mathbf{v} \cdot \dot{\mathbf{v}})\mathbf{v}^i}{c^2}\right]. \tag{16}$$

Equation (14) can now be explicitly verified by a proper-time differentiation of $\mathcal{E}_s^\mu$, in conjunction with Eqs. (4), (5), (8) and (9), to give

$$\mathcal{F}_s^0 = -\frac{2e^2}{3c^3}\dddot{v}^0 = \frac{\gamma(\mathcal{P}_2 - \mathcal{P}_1)}{c}, \tag{17}$$

$$\mathcal{F}_s^i = -\frac{2e^2}{3c^3}\dddot{v}^i = \gamma(\mathbf{F}_2^i - \mathbf{F}_1^i) . \tag{18}$$

It may be noted that in case of radiation reaction, power and force (Eqs. (8) and (9)) are related by $\mathcal{P}_2 = \mathbf{F}_2.\mathbf{v}$, implying

$$\mathcal{F}_2^0 = \frac{\mathcal{F}_2^i v_i}{\gamma c}, \tag{19}$$

where v_i ($i = 1, 2, 3$) stands for ith component of the 4-velocity. In contrast, the relation between force and power in case of Larmor's radiation formula, $\mathbf{F}_1 = \mathcal{P}_1\mathbf{v}/c^2$ (Eq. (5)), implies

$$\mathcal{F}_1^0 \neq \frac{\mathcal{F}_1^i v_i}{\gamma c} . \tag{20}$$

Also

$$\mathcal{F}_s^0 \neq \frac{\mathcal{F}_s^i v_i}{\gamma c}, \tag{21}$$

because

$$\dddot{v}^0 = \frac{(\dddot{v}^i v_i + \dot{v}^\mu \dot{v}_\mu)}{\gamma c} . \tag{22}$$

An uncomfortable question in Larmor's radiation loss formula arises in case of an accelerated charge in its instantaneous rest frame. Due to zero velocity of the charge, it could not lose any kinetic energy into radiation. However, Larmor's formula, according to which the radiated power is proportional to square of acceleration of the charge, yields a finite power loss. Even if the external force causing the acceleration of the charge, were considered to be responsible for the radiative power as well, it could not have done so, because at that instant the rate of work being done by the external force would also be zero as the system has a zero velocity. This embarrassing question

is purportedly resolved by proposing that there is an equivalent decrease in the Schott energy (an acceleration-dependent internal energy!) stored within the electromagnetic fields in the close vicinity of the accelerated charge. According to this argument, even if the Schott energy term may be zero in the instantaneous rest frame (Eq. (15)), its temporal derivative (Eq. (17)) yields a finite power loss for the instantly stationary charge equal to that expected from Larmor's formula (Eq. (1)).

However, even if this might seem to resolve the particular energy conservation problem, it gives rise to another awkward question about the presence of momentum for an instantly stationary charge. From Eq. (16) we infer that there is a finite momentum, $-2e^2\dot{\mathbf{v}}'/3c^3$, in electromagnetic fields in the vicinity of the charge, even at the instant when the charge is stationary ($\mathbf{v}' = 0$). Now this apparent momentum, which is directly proportional to the acceleration of the charge, and is strangely independent of the velocity of the charge, at least in the non-relativistic case, raises a vexing question – How come there is supposedly a finite momentum in the fields of a stationary charge when there is *no motion of any kind* at that instant?

We shall here endeavour to verify this, apparently an ad hoc, and to an extent even paradoxical but nonetheless widely-believed, assumption of the presence of acceleration-dependent Schott terms in electromagnetic fields of an accelerated charge.

For this, we will carefully scrutinize the case of a uniformly accelerated charge, where it may be possible to tract the question in exact mathematical details whether the electromagnetic fields really harbour the Schott energy-momentum, somewhere in the close vicinity of such a charge (or even elsewhere), as opined in the literature [22, 23].

5 Electromagnetic fields around the "present" position of a uniformly accelerated charge – no trace whatsoever of Schott energy-momentum anywhere

A uniformly accelerated motion is understood to imply a motion with a constant proper acceleration, say, $\mathbf{g}$. For simplicity, we may assume a one-dimensional motion, $\mathbf{v} \parallel \mathbf{g}$, since, using a Lorentz transformation, we can always switch to an appropriate inertial frame in which the velocity component perpendicular to the acceleration vector is zero. In fact, at any given event, transforming to the instantaneous rest frame guarantees that for a constant proper acceleration the motion will be along one dimension.

In the case of a uniform proper acceleration $\mathbf{g}$, Schott term, $\mathcal{F}_s^\mu$ is the 4-vector

$$\mathcal{F}_s^0 = -\frac{2e^2}{3c^4}\gamma\mathbf{g}^2, \qquad (23)$$

$$\mathcal{F}_s^i = -\frac{2e^2}{3c^5}\gamma\mathbf{g}^2\mathbf{v}^i, \quad i = 1,2,3, \quad (24)$$

while from Eqs. (15) and (16), the Schott

energy-momentum 4-vector $\mathcal{E}_s^\mu$ is given by

$$\mathcal{E}_s^0 = -\frac{2e^2}{3c^4}\gamma \mathbf{g} \cdot \mathbf{v}, \qquad (25)$$

$$\mathcal{E}_s^i = -\frac{2e^2}{3c^3}\gamma \, \mathbf{g}^i, \quad i = 1, 2, 3. \qquad (26)$$

As it was mentioned earlier, this Schott energy-momentum is thought to be present in the electromagnetic fields in the near vicinity of an accelerated charge.

In order to explore the electromagnetic fields around the charge in its neighbourhood, we could attempt to express fields at a time t with respect to the position and motion of the charge also at the same instant t. It may not be quite feasible to do so for an arbitrary motion of the charge. However, for a uniformly accelerated charge, it is possible to solve the expression for electromagnetic fields not necessarily in terms of motion of the charge at retarded time, instead wholly in terms of the "real-time" motion of the charge [27].

We assume that the constant proper acceleration vector, $\mathbf{g}$ $(= \gamma^3 \dot{\mathbf{v}})$, is along the $+z$ axis and that the charge, coming from $z = \infty$ at time $t = -\infty$, initially moves along $-z$ direction, getting constantly decelerated till it comes to rest momentarily at a point $z = \alpha$ at time $t = 0$, and then onwards moves with an increasing speed along the $+z$ direction. Without any loss of generality, we can choose the origin of the coordinate system so that $\alpha = c^2/g$, then the position and velocity of the charge at a time t are given by $z_c = (\alpha^2 + c^2 t^2)^{1/2}$ and $\beta = v/c = ct/z_c$. The charge happens to be at the same point on the z-axis at times $-t$ and t, but with velocity in opposite directions, i.e., $\boldsymbol{\beta}(t) = -\boldsymbol{\beta}(-t)$

The expression for electromagnetic fields of such a charge [28], in a spherical coordinate system (r, θ, ϕ) with origin at the instantaneous charge position [29], is

$$E_r = \frac{e(1 + \eta \cos\theta)}{r^2 \gamma^2 (1 + 2\eta \cos\theta + \eta^2 - \beta^2 \sin^2\theta)^{3/2}}$$

$$E_\theta = \frac{e\eta \sin\theta}{r^2 \gamma^2 (1 + 2\eta \cos\theta + \eta^2 - \beta^2 \sin^2\theta)^{3/2}}$$

$$B_\phi = \frac{e\beta \sin\theta}{r^2 \gamma^2 (1 + 2\eta \cos\theta + \eta^2 - \beta^2 \sin^2\theta)^{3/2}}, \qquad (27)$$

with $\eta = gr/2\gamma c^2$. All the remaining field components are zero.

Since the magnetic field B_ϕ has linear dependence on β (Eq. (27)), therefore at any given location (r, θ, ϕ), $B_\phi(t) = -B_\phi(-t)$. On the other hand, the electric field components E_r, E_θ do not have such linear dependence on β and $\mathbf{E}(t) = \mathbf{E}(-t)$.

From Eq. (27) we can infer the following:

(i) For $g = 0$, $\eta = 0$ and in that case the fields reduce to that of a charge moving with a uniform velocity $\boldsymbol{\beta}$, with the electric field everywhere in a radial direction from the present position of the charge with $\mathbf{B} = \boldsymbol{\beta} \times \mathbf{E}$ [2, 3, 4].

(ii) For a finite g ($\eta \neq 0$), the radial component of the Poynting vector, $\mathcal{S}_r = c(E_\theta B_\phi)/4\pi$, at any time $t < 0$, when the charge is getting decelerated, is everywhere (i.e., at any field point r, θ, ϕ) *pointing inward*, toward the present position of the charge.

(iii) At $t = 0$, $\beta = 0$, implying $\mathbf{B} = 0$ everywhere. Thus there is no Poynting vector seen anywhere at $t = 0$.

(iv) At time $t > 0$, when the charge is accelerating, the radial component of the Poynting vector is everywhere pointing away from the present position of the charge. In fact, everywhere the Poynting vector at time t_1 is equal and opposite to that at time $-t_1$, for all t_1 values.

The electromagnetic field energy in a volume $\mathcal{V}$ is given by the volume integral

$$\frac{1}{8\pi} \int_{\mathcal{V}} dV \left(E^2 + B^2 \right). \qquad (28)$$

The field energy density, $(E^2 + B^2)/8\pi$, being equal at times t_1 and $-t_1$, its volume integral over *any chosen* $\mathcal{V}$, whether in the vicinity of the charge or in some far-off zone, is also equal at times t_1 and $-t_1$. Now, the acceleration-dependent Schott energy term, according to Eq. (25), is equal but opposite at t_1 and $-t_1$ (because $v = -c^2 t_1/z_c$ at $-t_1$). Thus the Schott energy should be making a positive contribution at $-t_1$ and a negative contribution at t_1, which is not consistent with the fact that the actual field energy, computed from Eq. (28), is identical at t_1 and $-t_1$.

One can also compute the electromagnetic field momentum contained within a volume $\mathcal{V}$ from

$$\frac{1}{4\pi c} \int_{\mathcal{V}} dV \left(\mathbf{E} \times \mathbf{B} \right). \qquad (29)$$

Since $\mathbf{B} = \mathbf{0}$ at $t = 0$ (Eq. (27)), there is no momentum in the electromagnetic fields anywhere, whether in the vicinity of the charge or in the far-off regions, in the instantaneous rest frame. Therefore Eq. (26)

is clearly violated where the Schott momentum is proportional to $-\mathbf{g}$ even at $t = 0$, when the charge is instantly stationary.

Further, for $t \neq 0$, from Eq. (29) in conjunction with Eq. (27), the electromagnetic field momentum at times t_1 and $-t_1$ is not only equal but in opposite directions; it being directly proportional to the instantaneous velocity β. Now, this again is not in agreement with Eq. (26), where at times t_1 and $-t_1$, not just the magnitude but also the direction of the Schott momentum vector should remain the same, being directly proportional to the acceleration g but opposite in direction to $\mathbf{g}$, unlike the electromagnetic field momentum which is not only directly proportional to the instantaneous velocity β, but also in same direction as β. In fact, a finite Schott momentum for an accelerated charge in its instantaneous rest frame, as inferred from Eq. (26), would from the strong principle of equivalence [30] imply a finite momentum $2e^2 g/3c^3$ associated with a charge that is continually at rest in a gravitational field of strength g, an unpalatable inference for an otherwise completely static system.

Thus we find no signature of the acceleration-dependent Schott energy-momentum terms that were in accordance with Eqs. (25) and (26). We may add here that the introduction of the Schott energy term to account for the power loss into radiation but without any equivalent rate of decrease of kinetic energy of the radiating charge, say, in the instantaneous rest

frame, is akin to the proposal of the loss of internal (rest mass!) energy [31] without a loss of momentum (c.f. Eqs. (1) and (2)). However, in the case of Scott energy it is thought to be an acceleration-dependent extraneous kind of energy (neither the rest mass energy nor the kinetic energy, not even some kind of potential energy that may depend upon location in an external field) present in the electromagnetic fields and which does not seem to make an appearance elsewhere in physics. In any case, we see no evidence of the presence of such an energy term in the fields of a uniformly accelerated charge. Actually it has recently been shown that the Schott term is merely a difference in rate of change of energy in self-fields of the charge between retarded and real times [18, 32, 33]. and contrary to the ideas that have been proposed in the literature [6, 20, 21, 22, 23, 24, 25], there is no acceleration-dependent extra energy term lurking somewhere in the electromagnetic fields whether in the near vicinity of the charge or elsewhere.

6 Conclusions

From the difference between Larmor's/Liénard's radiation formula and Abraham-Lorentz radiation reaction formula, we arrived at the expression for Schott energy-momentum for an accelerated charge. We demonstrated that in the electromagnetic fields of a uniformly accelerated charge there is no evidence, what-

soever, of the Schott energy-momentum terms, whether in the near vicinity of the charge or elsewhere. The presence of such terms would have been, even otherwise, in conflict with the strong principle of equivalence as one would then infer from them, among other things, a finite momentum for a charge continually at rest in a gravitational field, an unpalatable inference for an otherwise completely static system. Since the difference between two formulations is resolved when a proper distinction is made between radiation losses calculated in terms of a retarded time and those expressed in terms of a "real time", the proposition of the elusive Schott energy-momentum terms in fields is superfluous and needs to be abandoned.

References

[1] J. Larmor "On the theory of the magnetic influence on spectra; and on the radiation from moving ions," Phil. Mag. **44** 503-512 (1897)

[2] J. D. Jackson, *Classical electrodynamics*, 2nd ed. (Wiley, New York, 1975).

[3] W. K.H . Panofsky and M. Phillips, *Classical electricity and magnetism*, 2nd ed. (Addison-Wesley, Massachusetts, 1962).

[4] D. J. Griffith, *Introduction to electrodynamics*, 3rd ed. (Prentice, New Jersey, 1999).

[5] F. V. Hartemann and N. C., Luhmann Jr., "Classical electrodynamical derivation of

the radiation damping force," Phy. Rev. Lett. **74**, 1107-1110 (1995).

[6] F. Rohrlich, "The dynamics of a charged sphere and the electron," Am. J. Phys. **65**, 1051-1056 (1997).

[7] A. K. Singal, "Disparities of Larmor's/Liénard's formulations with special relativity and energy-momentum conservation," J. Phys. Commun. **2**, 031002 (2018).

[8] M. Abraham, "Zur Theorie der Strahlung," Ann. Phys. **14** 236-287 (1904)

[9] M. Abraham, *Theorie der elektrizitat, Vol II: Elektromagnetische theorie der strahlung* (Teubner, Leipzig, 1905).

[10] H. A. Lorentz, "Weiterbildung der Maxwellschen Theorie," Encykl. Mathe. Wiss. **2**, 145-280 (1904).

[11] H. A. Lorentz, *The theory of electron* (Teubner, Leipzig, 1909); 2nd ed. (Dover, New York, 1952).

[12] G. A. Schott, *Electromagnetic radiation* (Univ. Press, Cambridge, 1912).

[13] L. Page and N. I. Adams, Jr., *Electrodynamics* (D. Van Nostrand, New York, 1940).

[14] W. Heitler, *The quantum theory of radiation* (Oxford, Clarendon, 1954).

[15] A. D. Yaghjian, *Relativistic Dynamics of a charged sphere*, 2nd ed. (Springer, New York, 2006).

[16] A. K. Singal, "Radiation reaction from electromagnetic fields in the neighbourhood of a point charge," Am. J. Phys. **85**, 202-206 (2017).

[17] A. K. Singal, "Poynting flux in the neighbourhood of a point charge in arbitrary motion and radiative power losses," Eur. J. Phys. **37**, 045210 (2016).

[18] A. K. Singal, "Compatibility of Larmor's formula with radiation reaction for an accelerated charge," Found. Phys. **46**, 554-574 (2016).

[19] A. K. Singal, "Why no radiation occurs in the case of a uniformly accelerated charge," Phys. Ed. **35**, 01, 02 (2019).

[20] C. Teitelboim, "Splitting of the Maxwell tensor: radiation reaction without advanced fields," Phys. Rev. D **1**, 1572-1582 (1970).

[21] J. A. Heras and R. F. O'Connell "Generalization of the Schott energy in electrodynamic radiation theory," Am. J. Phys. **74**, 150-153 (2006).

[22] E. Eriksen and Ø. Grøn, "The significance of the Schott energy in the electrodynamics of charged particles and their fields," Ind. J. of Phys., **82**, 1113-1137 (2008).

[23] Ø. Grøn, "The significance of the Schott energy for energy-momentum conservation of a radiating charge obeying the Lorentz–Abraham–Dirac equation," Am. J. Phys. **79**, 115-122 (2011).

[24] R. T. Hammond, "Relativistic particle motion and radiation reaction in electrodynamics," El. J. Theor. Phys. **23**, 221-258 (2010).

[25] D. R. Rowland, "Physical interpretation of the Schott energy of an accelerating point charge and the question of whether a uniformly accelerated charge radiates" Eur. J. Phys. **31** 1037-1051 (2010).

[26] M. Von Laue, "Die wellenstrahlung einer bewegten punktladung nach dem relativitätsprinzip," Ann. Phys. **28**, 436-442 (1909).

[27] M. Born, "Die Theorie des starren Elektrons in der Kinematik des Relativitätsprinzips," *Ann. Physik* **30**, 1-56 (1909).

[28] T. Fulton and F. Rohrlich "Classical radiation from a uniformly accelerated charge," Ann. Phys. **9** 499-517 (1960)

[29] A. K. Singal, "The equivalence principle and an electric charge in a gravitational field II. A uniformly accelerated charge does not radiate," Gen. Rel. Grav. **29**, 1371-1390 (1997).

[30] C. W. Misner, K. S. Thorne, J. A. Wheeler, *Gravitation* (Freeman, San Fransisco, 1973)

[31] W. B. Bonnor, "A new equation of motion for a radiating charged particle," Proc. R. Soc. London A **337**, 591-598 (1974).

[32] A. K. Singal, "Reply to comment on 'Poynting flux in the neighbourhood of a point charge in arbitrary motion and the radiative power losses' Eur. J. Phys. **39**, 018002 (2018).

[33] A. K. Singal, "Erratum: Reply to comment on 'Poynting flux in the neighbourhood of a point charge in arbitrary motion and the radiative power losses' Eur. J. Phys. **39**, 039601 (2018).

New Possibilities of Efficiency Improvement in Multi-Junction Solar Cells

Shalini Dubey, Arpit Swarup Mathur, Nidhi, BP Singh

Department of Physics, Institute of Basic Sciences, Dr. B. R. Ambedkar University, Agra 282002, India

dubey.shalini12@gmail.com

Abstract

The efficiency is the most important feature in solar cell performance. In recent years III-V multi-junction solar are much popular among all solar cells due to their high energy conversion ratio or efficiency. There are some major ways to improve efficiency in multi- junction solar cells. The solar cells have different combination of sub cell band gaps. The design optimizations of each sub cells for spectral radiance of a particular region can result in improve solar cell performance. In multi-junction solar cells, using nanotechnology to carry out band gap engineering provides better matching to the solar spectrum and enables improved the current matching for better efficiency. The multi-junction solar cells are applicable in space applications and they are very expensive in cost. The new technology named concentrator photovoltaic overcomes these issues in multi-junction solar cells which is also economically viable. The concentration ratio improves the performance of photovoltaic cell even more. Semiconductor materials properties allow solar cells to operate more efficiently in concentrated light, as long as the cell junction temperature is kept cool by suitable heat sinks.

1. Introduction

Sufficient and secure energy sources are the main need of social and economic development of society. Energy sources may classify in two major categories one is conventional energy sources and other is renewable energy sources. The conventional energy sources are mainly depended on cow dung cakes, wood, fossil fuels such as oil, natural gases and coal. These sources have been utilized for centuries to power generation and industries. These all energy sources are very popular among all but unfortunately these all sources of energy are limited. In next two decades, the conventional sources of energy are subject to depletion that's why the whole world looking forward for renewable energy sources. Renewable energy sources are also called unconventional energy sources. They are clean, eco-friendly and inexhaustible. There are so many forms of renewable sources of energy like wind energy, hydro energy, ocean energy, tidal energy, geothermal energy and solar energy. In the last few decades, we have been seen unpredictable changes in the consumption of energy resources. When we talk about India, the power sector is mainly dominated by fossil fuel especially by coal. The recent study shows that in 2014, India is the 4th

largest energy consumer (7,124.002 TWh) and 5th largest energy producer (4547.946 TWh) in the world[1]. Indian government is pushing an increased investment in renewable energy.

Renewable energies are playing key role in new global energy strategy. It gives the independence from conventional energy source among all renewable energy hydro power, wind power, biomass and solar energy appear the most promising. Despite all of these approaches to renewable energies still have deficiency of that limit their ability to stabilize global climate. Utilization of the power of sun light is certain one of the most viable way to solve the foreseeable world's energy crisis [2]. Solar energy is the cleanest, echo friendly, non-toxic source of energy. Solar energy, mostly classified in two techniques: passive solar energy technique and active solar technique. A passive solar technique used in building design orientation to provide electricity, mechanical power, heat and lighting through the sun and active solar technique used photovoltaic system, concentrating solar system [3]. In all these techniques photovoltaic system is the most useful and promising technique for the future.

As the name referred photovoltaic is the combination of two words Photo means light and voltaic means voltage. So we can say that light can be used for creating voltage or electricity. The photovoltaic system is the arrangement or the combination of different components, including the solar panels.

A solar panel or solar module is the most important component of the photovoltaic system. The solar panel is the device which converts light into electricity. The solar panel is the bunch or collection of solar cells. The large number of small solar cells spread over a large sheet is called a solar panel. Solar panel used photons (light energy) from the sun to generate electricity. Solar cells are connected one to another electrically in series or in parallel pattern, externally. The parallel connected panel gives the desired current while in series connected solar panel provide the desired voltage. Some special photovoltaic module used concentrator that means light focused on small photovoltaic cell through lenses or mirrors. If we talk about its maintenance or wear-and-tear, solar panels are very hardy [4]. There are two types of approaches used in solar panel configuration: stand alone and grid connect interface.

2. Background or History of Solar Cells-

The concept of Photovoltaic cell firstly came in light in 1839 when A. E. Becquerel at the age of 19 discovered the Photovoltaic effect in his father's lab. He observed while experimenting with solid electrodes in an in an electrolyte solution, a voltage develops when light fell upon the electrodes [5].The first genuine solar cell was developed by Charles E. Fritts, who used to junction in the coating of selenium with the thin layer of gold [6]. The first solar cell demonstration and press release is shown in FIG (1). This cell achieved the conversion efficiency of 1-2%. After that selenium is the subject of discovery in the field of solar cells development, but in 1946, an American physicist Russel Ohl from Bell Laboratory, developed a p-n junction semiconductor silicon solar cell [7]. After that silicon is being the most promising and effective material in solar cell formation. In 1954, the three researcher Gordon Pearson, Calvin Fuller and Darryl Chapin from the same laboratory developed first silicon based solar cell with achieving 6% of conversion efficiency [8]. They created an array of several strips of silicon based solar cells [9-10]. At that time the word solar panel came in trend.

After this revolutionary discovery researcher were keenly interested in solar cell development with

various materials and junctions. There are three main generations of solar cells mentioned in the TABLE (1).

As per given table it is clear that third generation solar cells having more possibilities in the field of energy improvement. Traditional solar cells are having limitations like key ingredient Si has become more expensive, which can't reduce the solar cell cost. Also, Si has other physical limitations that's why to overcome these issues physicist, developed other optical active materials which makes solar cells more effective, but they are less popular in comparison to conventional Si solar cells. Third generation Photovoltaic cells are designed to combine the advantages of both first and second generation solar cells. This paper mainly focuses on the attempts to improve efficiency of multi-junction solar cells above the Shockley-Queisser efficiency limit and combination of semiconductor material to more efficiently capture a target range of photon. Multi-junction solar cells have the highest theoretical limit of conversion efficiency as compared to other photovoltaic technologies [11-13]. Before we talk about the fundamentals of multi- junction solar cell we should know about some basics of solar cells.

Features	First generation	Second generation	Third generation
Technology	Silicon wafer based	Thin film based	Multi-layer or tandem
Stability	Stable	Less Stable	Less Stable
Efficiency	Efficient	Less Efficient	High Efficient
Toxicity	Non-Toxic	Highly Toxic	Toxic
Production Cost	Costly	Cheaper than previous generation	Cheaper than previous one
Market Share (Commercially)	Very big share	Small	Very small
Examples	Mono crystalline, Polly crystalline Si based Solar cells	CdTe, CGIS, CIS, CZTS, a-H Si based solar cells	Organic, Quantum Dots, Perovskite, dye-sensitizer and multi-junction

Table (1) Featuring all three generations merits and demerits

The behavior of a solar cell is represented by the current-voltage curve (I-V curve). The solar cell performance depends upon some basic parameters of solar cells. Short circuit current (ISC) is the current through the solar cell when the voltage

across the solar cell is zero (i.e. solar cell is short circuited) this is the largest current drawn from the solar cell. It depends upon the area of the solar cell, the number of incident photons or intensity of light, light spectrum and optical properties of the solar cell (i.e. absorption and reflection) the relation of short circuit current and I-V curve diagram (FIG. 4) are given below

$$J_{SC}=I_{SC}/A$$

Short circuit current density (JSC) is the amount of short circuit current per unit area of the solar cell14. Open circuit voltage (VOC) is the maximum voltage from a solar cell and occurs when the net current through the device is zero that means when the circuit is open circuited, no load connected to the circuit so the current across the cell will be minimized and voltage is maximum. It is also the measurement of the recombination of the device. It depends upon the saturated current and light generated current of the cell. The normally open circuit voltage of solar cell measured around 0.5-0.6 volts [15]. Another most important term in solar cell behavior or performance is Fill Factor (FF), a ratio that describes how close the I-V curve of a solar cell resemblance a perfect rectangle. The fill factor is the parameter which, in conjunction with ISC and VOC , determine the maximum power from a solar cell. The equation is given below

$$FF=P_m/I_{SC}\times V_{OC}$$

The main goal of the any solar cell design is having maximum power. For each point on the I-V curve, to calculate the power of the cell current and voltage should be multiplied. Maximum power point (MPP) is the point on the I-V curve of the solar cell corresponding to the maximum output electric power. It denotes as

$$P_m=V_{max}\times I_{max}$$

In multi-junction solar cell compared with single junction solar cell, reducing current because the fixed number of photons distributed over the increasing number of layers in the solar cell. At the same time excited electrons having high electric potential and energies, so the reduction of current is compensated for by an increase in voltage, and overall power of the cell is increased so that the design of multi-junction is advantageous [16]. Quantum Efficiency (QE) is the value refers the amount of current when irradiation of photons of particular wavelength or it indicates the percentage of the absorbed photon that produces an electron-hole pair. While Conversion Efficiency (η) has refers the percentage of the incident solar radiation on the solar cell that converted into the useable electric power, when the solar cell connected with an electric circuit [17]. There are so many factors that affecting the efficiency of solar cell like wavelength of the incident light, recombination, temperature, reflection of the light from the cell surface and surface area of the panel etc .

3. Fundamentals of Multi-junction Solar Cells-

These photovoltaic cells are photovoltaic solar cells with multiple p-n junctions of different semiconductor materials. In other words, we can say that several single junction solar cells stacked upon each other. Each material has different energy band-gap which produces an electric current with their respective wave length of light. The band gap of each layer decreases from top to bottom so that it can be passes more energy photons through the top layer and reach the absorber bottom layer that means collector. The multi-junction solar cell has

been subject of research since 1960 [18]. But the very first multi-junction solar was developed and display in the early 1980's with conversion efficiency of 16% [19].

Multi-junction solar cells having their own limitations, relating to material availability with optimal bang-gap which allow higher efficiency through low defect densities simultaneously. Multi-junction used p-n junction always in forward bias. They are designed in such manner to capture photons from the light spectrum. Photons hit the top layer of cells are either reflected or absorb by the layer. Absorb photon have potential to give their energy hυ to an electron. If hυ≥Eg, it's generating electron-hole pair. There are three basic characteristics are playing major role in multi-junction solar cell performance and affected the conversion efficiency. The multi-junction solar cells are tested under different numbers of suns because they are often used in concentrator photovoltaic systems which allow us to reduce the size or number of cells needed.

There are three major things which keep in mind when multi-junction solar cell is fabricated.

3.1 Band-gap
To be much efficient solar cell the cell should absorb radiation as much as possible. It only happens when band-gap of material should cover the wide range of the spectrum of light. Band-gaps of an adjacent layer should differ by as small amount as possible because of the amount of excess energy from light converted to heat is equal to the difference between the photon energy and the band-gap of the

absorbing material. Triple junction solar cells are currently using GaInP(1.86eV)/GaAs(1.4eV)/Ge(0.65eV).
For the quad junction AlGaInP (1.9eV)GaAs (1.4eV)/GaInAsP (1.1eV)/GaInAs(0.7eV) [20].

In 1990 J M Olson et al. [21] using 1.7eV-1.1eV energy band gap for Ga0.5In0.5P/GaAs and received the conversion efficiency of 27.3%. Paul R Sharp et al. [22] gives the general idea for theoretical band gap-efficiency ratio as they suggest 1.9eV/1.33eV/0.92eV = 34.9%-38.8% for triple junction, 2eV/1.46eV/1.08eV/0.77eV = 37.6%-41.8% for quad junction, 2.13eV/ 1.46eV/1.08eV/0.77eV = 39.5%-43.9% for five junctions   and 2.22eV/1.76eV/1.42eV/1.15eV/0.92eV/0.71eV = 40.8%-54.3% for six junctions solar cell.

3.2 Lattice constant
For the maximum current conductivity in multi-junction solar cells, all layers of the different semiconductor materials which grown directly on the top of the layers, having similar crystal structure. The lattice constant describes the spacing of the atom location in the crystal structure. Crystal lattice constant mismatching in different layers creates dislocation in the lattice of the cell layers, which affect the efficiency of the cell. NREL (National Renewable Energy Laboratory USA) showed the lattice mismatching of very small amount around 0.01% decreasing the current production in solar cell [23].

3.3 Current Matching
In the case of multi-junction solar cell, where the sub-cells are connected in series, the resulting current of such a tandem cell is the minimum of the two contributing sub-cells i.e. the weaker sub-cells are limiting the overall solar cell current. If the both sub-cell generate the same current, the device is currently-matched [24]. The current matching due to the design of sub-cell (thickness) and illuminating spectrum quality (if the illuminating spectrum quite differs from AM1.5D). In multi-junction solar cell, the cell current can't be adjusted

by a simple multiplicative correction factor because of further impact such as fill factor of the stacked cell [25].

After material is selected with desired band gaps and lattice constant, the thickness of each layer must be determined based on the material's absorption constant and number of incident photon so that each layer will generate the photo current.

4. Possible ways of efficiency improvement in multi-junction solar cells

4.1 Increase in the number of Junctions

The most possible way to efficiency improvement in multi-junction solar cells, is to device more junctions. Five and six junctions solar cell design partition the solar spectrum into narrower wave length ranges than triple junction cells that allows all the sub cells to be better current matched to the low current producing sub cell. Frank Dimroth et al. conclude that four or more junctions will be necessary to reach conversion efficiency above 50% under concentrated sunlight [26]. Theoretical efficiency limits for multi-junction devices based on thermodynamic fundamentals are 37, 50 and 56% for 1, 2 and 3 band- gap correspondingly [27].

4.2 Design Optimization of each sub-cells

The second thing we can do with efficiency improvement is design optimization of each sub cells. The solar cell have different combination of sub cell band gaps. An alternative to GaInP top layer, AlGaAsS or AlGaInP with the band gap of 1.98 eV can be used, which has the similar lattice constant and band gap energy. In the past, the major drawback of this approach was the high sensitivity of these materials to oxygen and water contamination, but recent results in this field is promising and hopeful [28-30]. As we know that the energy loss due to the series resistance is the cause of decrement in solar cell efficiency. The components of series resistance such as electrode resistance, tunnel junction resistance and lateral electrodes between electrodes will be estimated separately.

4.3 Inclusion of nano-structures

In recent years, it has been proposed and experimentally verifies that the use of nanostructure such as quantum wells, quantum dots, quantum wires, nano rodes, nanotubes offers the high photovoltaic efficiency potential. Theoretically, a single intermediate electronic band created by quantum dots would offer a 63.2% efficiency of an ordinary solar cell, which greatly exceeds the maximum conversion efficiency of 31% for even a single junction device [31]. A system with an infinite number of sizes of quantum dots has the same theoretical efficiency as an infinite number of band gaps [32].

4.4 Concentration Photovoltaic

As we know that the multi-junction solar cells are applicable in space applications and they are very expensive in cost. The new technology named concentrator photovoltaics overcomes these issues in multi-junction solar cells. This technology is not only help in efficiency improvement but also economically viable [33]. Concentrator operation well suited for multi-junction photovoltaic cells because increasing the concentration ratio improve the performance of photovoltaic cell even more.

Concentrator photovoltaic opens the new door of opportunity in the efficiency improvement and cost reduction. Multi-junction solar cells have not found yet general application as compared to Si based flat PV module. Several researchers have investigated the possibilities of employing optical devices such as mirrors, Fresnel lenses, diachronic films and light guides in multi-junction solar cells [34- 36].

By this way it is possible to collect solar light and concentrate its energy in a single small area solar cell. This reduces the total cell are an amount equal to the concentration ratio and thus decreases the cost of Photovoltaic system [37,38]. Concentrators collects direct and diffused both solar radiations simultaneously [39].

5. Conclusion

The global population and economic growth will increase the rate of energy consumption more than double by the middle of 21th century. So we need to use another source of energy besides the conventional source of energy. At the present time the most efficient solar cell is multi-junction solar cells. But unfortunately, these cells are much famous not as comparable to Si based solar cells because these cells are costly and used in space applications. The most exciting aspect of multi-junction solar cell is that there are still many possibilities to explore it. In this present study it was shown that the current design of multi-junction device can be improved by design optimization of each layer, by increasing the number of junctions in the photovoltaic cell structure or including the nanomaterials (such as quantum dots, nanorodesetc) which offers high potential for efficiency improvement. The gap between the ideal (theoretical) and real value of conversion efficiency in expected to decrease due to fundamental advances in understanding of material behavior. Concentrator photovoltaic technology in multi-junction solar cells makes it applicable on earth and economically worthy. Thus, it expected that

the use of highly efficient multi-junction solar cell with innovative concept of concentrator photovoltaic has the potential to establish a new milestone in history of photovoltaic.

References

1. Ellabban O., Abu-Rub H. ,F. Blaabjerg . Renewable energy resources: Current status. future prospects and their enabling technology. Renew Sustain Energy Review,39,748 (2014). doi:10.1016/j.rser.2014.07.113.

2. INDIA RENEWABLE ENERGY STATUS REPORT 2014 (2014).

3. F. Manzano-Agugliaro, A. Alcayde, F. G. Montoya, A. Zapata-Sierra ,C. Gil Scientific production of renewable energies worldwide: An overview. Renew Sustain Energy Review.18,134,(2013). doi:10.1016/j.rser.2012.10.020.

4. P. Enzymes. United States Patent (19) 54, 96:62 (1980). doi:US005485919A.

5. Y. Y. Gurevich, Y. V. Pleskov. Photoelectrochemistry of Semiconductors. 19, (1983). doi:10.1016/S0080-8784(08)60277-X.

6. C. E. Fritts. On a new form of selenium cell, and some electrical discoveries made by its use. American J Science 18833-26,465 (1983). doi:10.2475/ajs.s3-26.156.465.

7. X. Zhang, Z. Tang, D. Hu, D. Meng,S. Jia. Nanoscale p-n junctions based on p-type ZnSe nanowires and their optoelectronic applications. Mater. Lett.,168, 121 (2016). doi:10.1016/j.matlet.2016.01.044.

8. D. M. Chapin ,C. S. Fuller,G. L. Pearson. A new silicon p-n junction photocell for

converting solar radiation into electrical power. J. Appl. Phys., 25,676 (1954). doi:10.1063/1.1721711.

9. J. J. Loferski. The First Forty Years: a Brief History of the Modern Photovoltaic Age. prog. photovol. : Res. Appl, 1, 67 (1993).

10. G. Jones,L. Bouamane." Power from Sunshine ": A Business History of Solar Energy, http://nrs.harvard.edu/urn-3:HUL.InstRepos:9056763. Accessed 15 july 2018.

11. W. Shockley, H. J. Queisser. Detailed balance limit of efficiency of p-n junction solar cells. J Appl Phys, 32,51 (1961). doi:10.1063/1.1736034.

12. C. B. Honsberg, R.C. Corkish, Bremner SP. A new generalized detailed balance formulation to calculate solar cell efficiency limits. Georgia Institute of Technology;1,3 (2001).

13. M. A. Green, Y. Hishikawa,E. D. Dunlop, D. H. Levi,J. Hohl-Ebinger, a W Y Ho-Baillie. Solar cell efficiency tables (version 51). Prog Photovolt: Res Appl, 26, 3 (2018). doi:10.1002/pip.2978.

14. Q. M. Aish. Temperature Effect on Photovoltaic Modules Power Drop,11,62 (2015).

15. L.X Wang, Z Q Zhou,T N Zhang, X Chen,M. Lu., High Fill Factors of Si Solar Cells Achieved by Using an Inverse Connection Between MOS and PN Junctions. Nano Res Lett ,11,3 (2016). doi:10.1186/s11671-016-1678-0.

16. M. A. Green. Photovoltaic principles. Physica E: Low-Dimensional Systems and Nanostructures,14,11(2002). doi:10.1016/S1386-9477(02)00354-5.

17. S E Shaheen , C J Brabec, S J Sariciftci,F Padinger,T Fromherz, Hummelen JC. 2.5% Efficient Organic Plastic Solar Cells. Appl Phys Lett,78,841, (2002). doi:10.1063/1.1345834.

18. B V Mathiesen.,H Lund, D Connolly,H Wenzel,P A Ostergaard,B Möller, et al. Smart Energy Systems for coherent 100% renewable energy and transport solutions. Appl Energy .145,139 (2015). doi:10.1016/j.apenergy.2015.01.075.

19. M Babar, E A Al-Ammar. Simulation and Modeling of Multi-junction Solar Cell for Concentrated Photovoltaics using MATLAB/Simulink. Recent Advances in Energy, Environment and Development (ISBN: 978-1-61804- 157-9)1,36 (2013).

20. F Dimroth, T N D Tibbits, Niemeyer M, Predan F, Beutel P, Karcher C, et al. Four-junction wafer-bonded concentrator solar cells. IEEE J Photovolt ,6,343 (2016). doi:10.1109/JPHOTOV.2015.2501729.

21. J M Olson,S R Kurtz,A E Kibbler, P A Faine. 27.3% efficient Ga0.5In0.5P/GaAs tandem solar cell. Appl Phys Lett. 56,623,(1990). doi:10.1063/1.102717.

22. P R Sharps, A Cornfeld,M Stan,A Korostyshevsky,V Ley,B Cho, et al. The future of high efficiency, multi-junction space solar cells. 2008 33rd IEEE Photovolatic Specialists Conference ,1–6 (2008). doi:10.1109/PVSC.2008.4922854.

23. P C Choubey,A Oudhia, Dewangan,R Autonomous VYTPG. A review : Solar cell current scenario and future trends. Recent Res Sci Technol 2012 4,99 (2012).

24. Feitknecht L. Evaluations on current-matching of multi-junction photovoltaics under outdoor conditions. Energy Procedia. 15,396 (2012). doi:10.1016/j.egypro.2012.02.048.

25. Guter W, Schöne J, Philipps SP, Steiner M, Siefer G, Wekkeli A, et al. Current-matched triple-junction solar cell reaching 41.1% conversion efficiency under concentrated sunlight. Applied Physics Letters,94,94.(2009). doi:10.1063/1.3148341.

26. F Dimroth,C Baur, A W Bett,M Meusel, Strobl G. 3-6 Junction Photovoltaic Cells for Space and Terrestrial Concentrator Applications. Conference Record of the Thirty-First IEEE Photovoltaic Specialists Conference,2005,525(2005). doi:10.1109/PVSC.2005.1488185.

27. G S Bonett, R M Gerdes. Progress in Next-Generation Photovoltaic Devices, Iowa State University,Ames, IA, April 15, (2006).

28. T Takamoto,T Agui,K Kamimura,M Kaneiwa. Multijunction solar cell technologies - high efficiency,radiation resistance, and concentrator applications, 3rd World Conference on Photovoltaic Energy Conversion, Osaka Japan; May 11-18,(2011).

29. R R King, C M Fetzer, P C Colter, K M Edmondson, J H Ener, H L Cotal,H Yoon , A P Stavrides, Kinsey G, D D Krut, N H Karam. High-Efficiency Space and Terrestrial Multijunction Solar Cells Through Bandgap Control in Cell Structures . Conference record on twenty ninth IEEE photovoltaic specialists conference; 2002; May 19-24,(2002). dio:10.1109/PVSC.2002.1190685.

30. C M Fetzer,H Yoon, R R King, D C Law, T D Isshiki, N H Karam. 1.6/1.1 eV metamorphic GaInP/GaInAs solar cellsgrown by MOVPE on Ge. J Crystal Growth. 276,48 (2005).

31. S L Castro, S G Bailey, R P, K K Raffaelle Banger , A F Hepp. Nanocrystalline Chalcopyrite Materials (CuInS2 and CuInSe2) via Low- Temperature Pyrolysis of Molecular Single-Source Precursors. Chem Mater 15(16), 3142. doi:10.1021/cm034161o.

32. N J Ekins-Daukesa, D B Bushnella, J P Connollya, K W J Barnhama, M Mazzerb, J S Roberts,G Hill, R Airey. Strain-balanced quantum well solar cells. Physica E ,14, 132 (2002). PII: S 1386-9477(02)00378-8.

33. A W Bett,F Dimroth,M Hein,G Lange,M Meusel,U Schubert,G Siefer. Development of ill-v-based concentrator solar cells and their application in pv-modules. Conference record

of the twenty ninth IEEE photovoltaic specialists conference 2002, May 19-24 (2002); doi: 10.1109/PVSC.1190711.

34. Richard M. Swanson. The Promise of Concentrators. Prog Photovolt: Res Appl. 8,93 (2000).

35. M Khamooshi,H Salati,F, Egelioglu, A H Faghiri,J Tarabishi, S. A Babadi Review of Solar Photovoltaic Concentrators. International J Photoenergy 2014;(2014). doi:10.1155/2014/958521.

36. M Bosi, C Pelosi. The Potential of III-V Semiconductors as Terrestrial Photovoltaic Devices. Prog Photovolt: Res Appl. 15,51(2007). doi: 10.1002/pip.715.

37. P Pérez-Higueras, E Munoz ,G Almonacid,P G Vidal. High Concentrator PhotoVoltaics

efficiencies: Present status and forecast. Renew Sustain Energy Rev.15,1810(2011).doi: 0.1016/j.rser.2010.11.046.

38. L El Chaar, L.A. lamont, N. El Zein. Review of photovoltaic technologies. Renew Sustain Energy Rev.15,2165(2011). doi:10.1016/j.rser.2011.01.004.

39. N X Tien, S. Shin A Novel Concentrator Photovoltaic (CPV) System with the Improvement of Irradiance Uniformity and the Capturing of Diffuse Solar Radiation. Applied Sciences. 6,251(2016). doi:10.3390/app6090251.

The Dual-Spring-Cannon-Propelled Cart

Carl E. Mungan

Physics Department, United States Naval Academy

Annapolis, MD 21402-1363 USA

mungan@usna.edu

Abstract

Two identical spring-loaded cannons pointing backward are mounted on a frictionless railcart. The cannons are to be fired to propel the cart forward as rapidly as possible. Does it matter whether the two cannonballs are launched simultaneously or sequentially? The velocities and kinetic energies of the two balls and of the cart in various reference frames are analyzed in detail in order to explain the answer.

Figure 1. A pair of spring cannons bolted together and mounted on wheels. The cannons launch two balls, each of mass m, to propel the cart of mass M (exclusive of the two balls) forward. There is no axle friction, rolling friction, or air drag. The launcher travels on a pair of horizontal rails to ensure that its motion is one-dimensional. The springs are initially compressed such that they each have elastic potential energy E.

1. Introduction

A dual-cannon "rocket" cart is sketched in Fig. 1. A similarly powered trolley has been considered by Brun [1] with the assumption that each ball has the same final velocity relative to the initial velocity of the cart, regardless of when it is launched. Pinheiro [2], on the other hand, assumes that a ball's final velocity is to be measured relative to the final velocity of the cart. Either way, the assumption that the relative exhaust speed of the balls will always be the same is questionable, at least for nonchemically powered cannons [3] such as the spring-loaded ones treated here. Edmonds [4] has specifically shown that the cart ends up with a different final velocity if the two balls are launched simultaneously than if they are launched one after
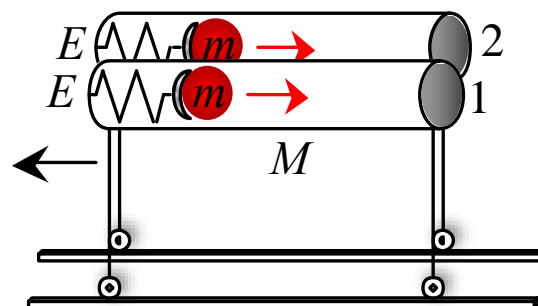
the other. He first assumes that each ball has the same exhaust speed relative to the final velocity of the cart regardless of when it is launched. Next he considers the possibility that the total kinetic energy gained by the two balls and the cart is the same, whether the balls are released sequentially or simultaneously, but he does not spell out how that can be accomplished. In fact, in a follow-up letter [5] he says that the exhaust speeds must be carefully calculated and the launch mechanism

adjusted (in some unspecified fashion) to ensure it happens. His follow-up was in part a response to comments by Scaife [6] and by Hinson [7]. Hinson showed that the cart will end up with the same final velocity when the two balls are released simultaneously as when they are launched sequentially in the limit that the mass of the balls is much less than that of the cart, $m \ll M$. The standard rocket equation is obtained in that limit when a continuous stream of exhaust is emitted, such as in the form of burning gas [8]. Specifically, during an infinitesimal emission time dt an infinitesimal amount of mass dm is launched out the tail, rather than the impulsive non-infinitesimal amount m in the present problem. Scaife has further commented that it is more efficient (in the sense of leading to a higher final rocket velocity for a fixed initial amount of fuel) to emit the gas in chunks (or clouds) for which all internal portions have the same exhaust velocity (and which thereby are dynamically equivalent to discrete balls). That claim anticipates Gowdy's "perfect" rocket [3] and also verifies Edmonds's hypothesis in his second letter [5] that for sequential launching the second ball should have the same kinetic energy (and hence velocity) as the first ball, in order to maximize the final velocity of the rocket.

In the present article, the situation is reconsidered, but this time with a specific launch mechanism (namely spring-loaded cannons). In that way, no ad hoc assumption about the exhaust speeds of the balls is necessary. Instead, an explicit calculation of those velocities is carried out, by combining the conservation laws of mechanical energy and of linear momentum for the isolated system of balls plus cart. The result is a well-defined problem (free of questionable assertions), that is solvable at the introductory physics level (using only elementary classical mechanics), and with three possible answers (namely simultaneous

firing leads to the greatest final cart speed, or sequential firing does, or it is a tie) which all appear plausible and thus find proponents when posed as a ConcepTest clicker question [9] to students.

Prior to revealing the correct answer, anticipation will be heightened if a class discussion includes consideration of the following two points [10, 11]. On the one hand, the gain in kinetic energy of the system of two cannonballs and cart in Fig. 1 must be equal to the elastic potential energy $2E$ lost by the two cannons. This equality holds in any inertial frame of reference, which might suggest that the order of firing is irrelevant. On the other hand, consider the center-of-mass reference frame of the ground (in which the cart and balls are initially at rest). If the two cannons are fired sequentially, then the second ball will have an exhaust velocity relative to an initially forward moving cart. Its velocity relative to the ground will then normally be smaller than that relative velocity. Hence, if the two cannonballs are fired with nearly (or exactly) the same relative exhaust velocities, the second ball will acquire less kinetic energy relative to the ground than the first ball. That leaves more kinetic energy to be gained by the cart, which is similar to the Oberth effect for rocket propulsion [12]. This idea suggests that sequential firing might have an advantage over simultaneous firing.

2. Analysis using the nonrelativistic conservation laws of classical mechanics

This problem is an excellent example of a situation where a straightforward mathematical analysis is more convincing than conceptual arguments. Begin by considering the general problem of a cart of mass M carrying a single cannonball of mass m. (Let M be the mass of the cart without the ball.) Suppose the cart is initially traveling forward at speed U before it fires the ball. The cannon is taken to be ideal in the sense that 100% of its elastic

potential energy E gets converted into translational kinetic energy of the cart and ball. Adopt a reference frame moving with the initial speed U of the cart. After firing the cannon, denote the final speed of the ball and of the cart in this frame as υ and V, respectively. Conservation of linear momentum implies that

$$m\upsilon = MV \tag{1}$$

while conservation of mechanical energy leads to

$$E = \tfrac{1}{2}m\upsilon^2 + \tfrac{1}{2}MV^2. \tag{2}$$

Simultaneous solution of Eqs. (1) and (2) gives

$$\upsilon = \sqrt{\frac{2EM}{(M+m)m}} \quad \text{and} \quad V = \sqrt{\frac{2Em}{(M+m)M}}. \tag{3}$$

Armed with this result, the final cart speed can be determined when it initially carries two balls as in Fig. 1. First, if the two cannons are fired simultaneously with the cart initially at rest relative to the ground, then replace $m \to 2m$ and $E \to 2E$ in the expression for V to get

$$V_{\text{simultaneous}} = \sqrt{\frac{8Em}{(M+2m)M}}. \tag{4}$$

Next, for the sequential firing case, after the first cannon has been fired, the final speed V of the cart becomes the initial speed U for the second firing, except that it is necessary to replace $M \to M+m$ because one cannonball is still left onboard. Thus, the speed of the cart after the first cannon firing but before the second firing is

$$U = \sqrt{\frac{2Em}{(M+2m)(M+m)}}. \tag{5}$$

After the second firing, the cart has a speed given by precisely V in Eq. (3) relative to speed U. Thus

the final speed of the cart relative to the ground becomes

$$V_{\text{sequential}} = V + U = \sqrt{\frac{2Em}{(M+m)M}} + \sqrt{\frac{2Em}{(M+2m)(M+m)}}. \tag{6}$$

The ratio of Eq. (6) to (4) is

$$\frac{V_{\text{sequential}}}{V_{\text{simultaneous}}} = \frac{\sqrt{M} + \sqrt{M+2m}}{2\sqrt{M+m}}. \tag{7}$$

This expression reduces to unity if $m \ll M$, so that one gets a tie in that limit, as Hinson [7] remarked. However, it is less than 1 otherwise. For example, it is equal to 96.6% if $m = M$, which admittedly requires heavy cannonballs. (Interestingly, the ratio is independent of E, however, and so it does not matter how fast the balls are launched.)

Therefore the two cannons should be fired simultaneously rather than sequentially, in order to maximize the final speed of the cart. To better understand this result, consider the speeds and kinetic energies of the launched cannonballs. In the simultaneous case, the speed and kinetic energy of the two balls relative to the ground (or equivalently, relative to the initial velocity of the cart) are respectively

$$\upsilon_{\text{simultaneous}} = \sqrt{\frac{2EM}{(M+2m)m}} \quad \text{and}$$

$$K^{\text{balls}}_{\text{simultaneous}} = \frac{1}{2}(2m)\upsilon^2_{\text{simultaneous}} = \frac{2EM}{M+2m} \tag{8}$$

where $\upsilon_{\text{simultaneous}}$ was obtained by again replacing $m \to 2m$ and $E \to 2E$ in the expression for υ in Eq. (3), just as was done to get Eq. (4). As a check, the kinetic energy of the cart relative to the ground is

$$K^{\text{cart}}_{\text{simultaneous}} = \frac{1}{2} MV^2_{\text{simultaneous}} = \frac{4Em}{M+2m} \qquad (9)$$

so that one correctly obtains

$$K^{\text{balls}}_{\text{simultaneous}} + K^{\text{cart}}_{\text{simultaneous}} = 2E . \qquad (10)$$

Alternatively, the speed of the simultaneously fired cannonballs relative to the final velocity of the cart is

$$\upsilon'_{\text{simultaneous}} = \upsilon_{\text{simultaneous}} + V_{\text{simultaneous}} = \sqrt{\frac{2E(M+2m)}{Mm}} \qquad (11)$$

with corresponding kinetic energy of

$$K'^{\text{balls}}_{\text{simultaneous}} = \frac{1}{2}(2m)\upsilon'^2_{\text{simultaneous}} = \frac{2E(M+2m)}{M} . \qquad (12)$$

The cart has no final kinetic energy in this frame of reference. However, the system is initially moving backward at speed $V_{\text{sequential}}$ in this frame, so that the initial mechanical energy of the system is

$$2E + \frac{1}{2}(M+2m)V^2_{\text{simultaneous}} = \frac{2E(M+2m)}{M} \qquad (13)$$

in agreement with Eq. (12). This result verifies the first point suggested for class discussion at the end of Sec. 1. It demonstrates that, unlike kinetic energy, potential energy is frame independent [13]. Each cannon has initial elastic energy E stored in its compressed spring regardless of the frame of reference adopted, whereas the kinetic energies of the balls and cart have different values depending on the motion of the frame.

In the sequential case, the speed and kinetic energy of the first ball relative to the ground are respectively

$$\upsilon_{\text{first}} = \sqrt{\frac{2E(M+m)}{(M+2m)m}} \qquad \text{and}$$

$$K^{\text{first ball}}_{\text{sequential}} = \frac{1}{2} m\upsilon^2_{\text{first}} = \frac{E(M+m)}{M+2m} \qquad (14)$$

where υ_{first} was obtained by replacing $M \to M+m$ in the expression for υ in Eq. (3), just as was done to get Eq. (5). In analogy to Eq. (6), the speed and kinetic energy of the second ball relative to the ground are respectively

$$\upsilon_{\text{second}} = |\upsilon - U| = \left| \sqrt{\frac{2EM}{(M+m)m}} - \sqrt{\frac{2Em}{(M+2m)(M+m)}} \right|$$

and $\quad K^{\text{second ball}}_{\text{sequential}} = \frac{1}{2} m\upsilon^2_{\text{second}} . \qquad (15)$

As discussed at the end of Sec. 1, the second ball's backward ground velocity is diminished by the forward velocity of the cart; the absolute value bars are needed because either square root term can be larger than the other, depending on the relative sizes of m and M. For example, $\upsilon = U$ when $m/M = 1 + \sqrt{2}$ in which case the second ball has a final speed of zero relative to the ground! In general, since the kinetic energy of the cart relative to the ground is

$$K^{\text{cart}}_{\text{sequential}} = \frac{1}{2} MV^2_{\text{sequential}} \qquad (16)$$

where $V_{\text{sequential}}$ is given by Eq. (6), one again correctly finds that

$$K^{\text{first ball}}_{\text{sequential}} + K^{\text{second ball}}_{\text{sequential}} + K^{\text{cart}}_{\text{sequential}} = 2E \qquad (17)$$

after some algebra. Although υ_{second} is smaller than $\upsilon_{\text{simultaneous}}$ when $m/M < 3 + 2\sqrt{3}$, υ_{first} is sufficiently larger than $\upsilon_{\text{simultaneous}}$ that $K^{\text{cart}}_{\text{sequential}}$ always ends up being smaller than $K^{\text{cart}}_{\text{simultaneous}}$ in contrast to the suggestion implied by the Oberth effect above.

3. Closing comments

One can use these results to show that the exhaust speeds of the cannonballs (relative to either the initial or final velocities of the cart at the instant of their firings) are all different from each other, as

Table 1. Speeds of the balls relative to either the initial or final velocities of the cart at the instant of firing. Here $v_{\text{simultaneous}}$ is from Eq. (8), $v'_{\text{simultaneous}}$ is from Eq. (11), v_{first} is from Eq. (14), $v'_{\text{first}} = v_{\text{first}} + U$ from Eqs. (5) and (14), $v''_{\text{second}} = v$ from Eq. (3), and $v'_{\text{second}} = v + V$ from Eq. (3).

ball	exhaust speed relative to initial cart speed	exhaust speed relative to final cart speed
simultaneous	$v_{\text{simultaneous}} = \sqrt{\dfrac{2EM}{(M+2m)m}}$	$v'_{\text{simultaneous}} = \sqrt{\dfrac{2E(M+2m)}{Mm}}$
first sequential	$v_{\text{first}} = \sqrt{\dfrac{2E(M+m)}{(M+2m)m}}$	$v'_{\text{first}} = \sqrt{\dfrac{2E(M+2m)}{(M+m)m}}$
second sequential	$v''_{\text{second}} = \sqrt{\dfrac{2EM}{(M+m)m}}$	$v'_{\text{second}} = \sqrt{\dfrac{2E(M+m)}{Mm}}$

listed in Table 1. Only in the limit $m \ll M$ do they all converge to the same value of $\sqrt{2EM/m}$. Otherwise, the exhaust speeds relative to the initial cart velocity are in the order $v_{\text{first}} > v''_{\text{second}} > v_{\text{simultaneous}}$ while those relative to the final cart velocity are in the opposite order $v'_{\text{simultaneous}} > v'_{\text{second}} > v'_{\text{first}}$. A simple example of these inequalities can be verified by putting $m = M$ into the six formulas in Table 1.

The solution that simultaneous ejection gives a larger final railcart speed than sequential ejection of two balls becomes more intuitive in the opposite limit where the final payload is much lighter than the exhausted fuel. In the sequential case, after the first half-mass m is ejected, both it and the second half-mass m (plus its tiny attached M) each gain $E/2$ amount of kinetic energy by symmetry; then after the second spring is fired, there is almost no change in the kinetic energy of the second (now detached) half-mass because the payload M is so light compared to it. (The speed of a moving car does not noticeably change if the driver flicks a cigarette butt out the window!) Thus each fuel half-mass m ends up with $E/2$ amount of kinetic energy, and so the cart must end up with the remaining E amount of released kinetic energy (all relative to the ground). On the other hand, in the simultaneous case, almost all $2E$ goes into the payload and little into the heavy full-fuel-mass $2m$

ejected as a unit. The ratio of final payload kinetic energies in the two cases is therefore

$$\frac{\frac{1}{2}MV_{\text{sequential}}^2}{\frac{1}{2}MV_{\text{simultaneous}}^2} = \frac{E}{2E} = \frac{1}{2} \qquad (18)$$

which agrees with Eq. (7) evaluated in the limit $m \gg M$.

     In conclusion, the efficiency or utilisation [14] of the fuel to accelerate the payload is maximized by launching all of the fuel out the tail end at once, rather than dribbling it out in sequential bits. However, that is only true if the rocket can maintain the same launch energy per unit mass of fuel either way! Specifically, to ensure that is true in the present setup, two cannons are used to fire the balls simultaneously. The rocket would be lighter (and less costly to build) if it instead had only one cannon which was used to fire the two balls sequentially. For a real rocket, there are trade-offs in the weight and power of the engine that favor sequential operation.

     Finally, it would be interesting to explore what happens if the second spring cannon were fired after the first cannon is fired but before the first spring has fully decompressed and released its ball. Presumably the results would smoothly interpolate between the simultaneous and sequential cases analyzed in this article. That avoids the disconcerting "discontinuity" which otherwise occurs as the time interval t between the ejection of the two balls is reduced with a sudden change in the final speed of the rocket just as $t \to 0$.

Acknowledgment

References

[1] J.L. Brun, "Motion of a trolley powered by ejecting balls," *Eur. J. Phys.* **27**, 1357–1362 (Nov. 2006).

[2] M.J. Pinheiro, "Some remarks about variable mass systems," *Eur. J. Phys.* **25**, L5–L7 (Jan. 2004).

[3] R.H. Gowdy, "The physics of perfect rockets," *Am. J. Phys.* **63**, 229–232 (Mar. 1995).

[4] J.D. Edmonds, "The put put rocket," *Phys. Educ.* **25**, 304–305 (Nov. 1990).

[5] J.D. Edmonds, "Put put rocket revisited," *Phys. Educ.* **27**, 8–10 (Jan. 1992).

[6] J. Scaife, "Generalized rockets," *Phys. Educ.* **26**, 143–144 (May 1991).

[7] D. Hinson, "Putting the put put rocket right," *Phys. Educ.* **26**, 144 (May 1991).

[8] A.R. Marlow, "The pedagogy of the rocket," *Phys. Teach.* **33**, 124–125 (Feb. 1995).

[9] E. Mazur, *Peer Instruction: A User's Manual* (Upper Saddle River NJ, Prentice Hall, 1997).

[10] B. Faulkner and H.D. Young, "Why the difference in energy?" *Phys. Teach.* **16**, 397 (Sep. 1978).

[11] M. Iona, "Efficiency in rocket propulsion," *Phys. Teach.* **16**, 600 (Dec. 1978).

[12] P.R. Blanco and C.E. Mungan, "Rocket propulsion, classical relativity, and the Oberth effect," *Phys. Teach.* **57**, 439–441 (Oct. 2019).

[13] R.C. Hilborn, "Galilean transformations of kinetic energy, work, and potential energy," *Phys. Teach.* **57**, 40–43 (Jan. 2019).

[14] P.R. Blanco, "A discrete energetic approach to rocket propulsion," *Phys. Educ.* **54**, 065001 (Nov. 2019).

Numerical Solution of Square Well Potential With Matrix Method Using Worksheets

O.S.K.S Sastri[1], Aditi Sharma, Jyoti Bhardwaj, Swapna Gora, Vandana Sharda and Jithin B.P

Department of Physical and Astronomical Sciences
Central University of Himachal Pradesh, HP 176215, Bharat(India)
[1]sastri.osks@gmail.com

Submitted on 08-11-2019

Abstract

We solve the time-independent Schrödinger equation for a square well potential, using matrix methods based numerical technique. Implementation of numerical method in a computer requires limiting the region of potential from infinity to some finite value, which is equivalent to embedding it within an infinite well potential. The eigen functions of the infinite well are utilised as the basis to build the hamiltonian matrix for the system. The matrix eigen equation is solved using Free Open Source Software (FOSS) Gnumeric which is a worksheet based environment that is easy for Implementation at the UG level.

1   Introduction

Marsiglio et.al., have introduced matrix methods for solving various quantum mechanical potentials [1, 2, 3] by embedding them within an infinite potential. This tech-

nique has the advantage of obtaining the energy eigen values and the corresponding eigen functions in a single iteration. With many FOSS offering eigen value solvers, this technique is easy to implement as part of simulation activities for quantum mechanics lab course at the UG level. Simulations have become an important approach for learning the concepts of physics[4, 5, 6]. There are many good simulations available at PhET [7, 8] and Compadre websites [9] written in Java and available for running both online and offline in order to understand the effect of various system parameters and their inter-relationships. Understanding and implementing the numerical methods involved to simulate a physical system is an important skill that needs to be developed while training to be a good physicist. Here, we focus on the pedagogical aspects involved in learning to perform simulations at the UG level, by taking the example of square well potential, which students encounter in introductory course of quan-

tum mechanics.

2 Methodology

To solve any physics problem using a computer simulation, we take up the following approach:

Step 1: Modelling the Physics problem
The primary step is to have clear understanding of *Mathematical model* that describes the problem at hand. Here, one could use the modeling methodology suggested by Hestenes [10] to *describe* the system w.r.t. its interactions and process involved in it. After giving full description of the system, one can proceed to *formulate* the model by introducing the force or potential obtained from the *interaction law* into the *dynamical law* that underlies the process and introducing the appropriate *initial* and *boundary conditions* that need to be complied with.   Once, the mathematical model is ready, one can solve the problem using analytical technique, if available or alternatively one can resort to numerical methods.

Step 2: Preparing the system to be solved using numerical approach

- **Choice of Numerical Technique**
  By looking at all the available numerical methods for solving the current problem, one has to arrive at the technique that is *stable*, *accurate* and *reasonably*

fast among the various methods available.

- **Rephrasing the Physics problem in appropriate units**
  This is important so as to avoid very large or very small numbers in computation that could lead to round off errors. Here, either the variables involved are made dimensionless or have to be expressed in an appropriate set of units. For example, for astronomical systems, one can use astronomical units such as specifying the masses in terms of mass of earth and distances in terms of distance between Sun and Earth, time in light years and so on.

- **Discretizing the continuous variables**
  As the computer can handle only discrete data, it is inevitable to sample the continuous variable with appropriate step size and also limit the variables to a finite range in accordance with the region of interest for solving the problem. This is a must to translate the problem into one that could be programmable.

Step 3: Implementation of the numerical method in computer

- **Algorithm**
  The inputs required, the equations to be iterated, the functions that need to be called in the main processing block and outputs generated (to be displayed as tabular data or plots) needs to be clearly defined in a step by step procedure so

that it is easy to program in a chosen software. Then, the code is written.

- **Running the program**

  This is initially to be taken up for a problem for which we know the analytical solution. This helps to troubleshoot the code for any mistakes and ensures its proper working.

- **Varying the parameters related to the numerical algorithm**

  We have to run the code by changing the various inputs parameters that effect the algorithm to minimize the various errors associated with them. Typical parameters that need to be optimized would be step-size, number of points representing the variable, tolerance limits on the iterated-quantities, values that need to remain constant with increasing number of iterations, etc.

Step 4: Simulation

- **Varying the physical properties associated with the physical system**

  Once the code is optimized, the variables associated with the physics of the problem can be varied. Plots & tabular data could be generated for understanding the relationships that would lead to analysis, interpretation & discussion of the results.

- **Modifications to the Physical system at hand**

It is possible to vary the physics problem, in the sense, that the functions, initial conditions, boundary conditions, etc., involved are changed but yet the same numerical technique could be employed to study the system.

3 Implementation

3.1 Modelling the Physics Problem:

Consider a microscopic particle in 1-Dimensional system (*reference system*)such as an electron which is of mass m and energy E (*object variables*) is interacting with a finite attractive potential well of width 'b' and depth V_0 (*interaction variables*). The interaction potential V(x) can be written in equation form as

$$V(x) = \begin{cases} V_0, & -\infty < x < -b/2 \\ 0, & -b/2 \leqslant x \leqslant b/2 \\ V_0, & b/2 < x < \infty \end{cases} \quad (1)$$

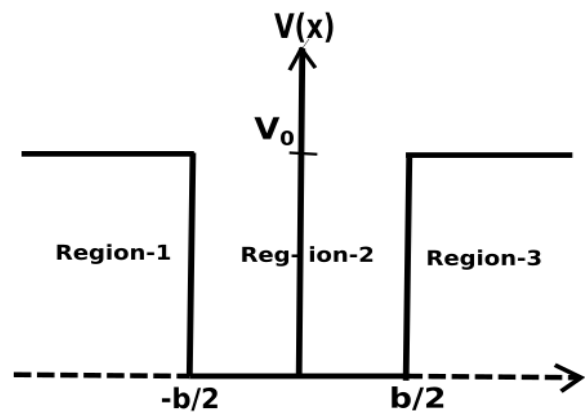and is plotted below in Figure 1.

Figure 1: Finite square well in 1-D

In the microscopic domain, considering that the wave nature associated with a particle plays significant role in understanding the probability of it being found at various locations (dependent on **state function**, $\psi(x)$), we need to solve the 'Time Independent Schrödinger Equation(TISE)' (*dynamical equation*) in the regions 1, 2 and 3, to obtain the wave functions for the particle with energy E i.e

$$H\psi(x) = E\psi(x) \qquad (2)$$

where H is the hamiltonian of the system given by

$$H = \frac{-\hbar^2}{2m}\frac{d^2}{dx^2} + V(x) \qquad (3)$$

Considering the case $E < V_0$, the analytical bound state solutions of TISE involving graphical approach have been discussed in various standard texts of quantum mechanics[11, 12, 13].

Taking specific values for parameters V_0 and b as: $V_0 = 14eV$ and $b = 2\mathring{A}$, we obtained the energy values for bound states by solving the eigen value conditions [11], utilising **Newton-Raphson method**[14, 15] to 3 decimal places as

$$E1 = 1.467eV$$
$$E2 = 5.711eV$$
$$E3 = 11.918eV \qquad (4)$$

Now, we focus on getting numerical solutions by using matrix diagonalization method suggested by Marsiglio [3, 16].

3.1.1 Marsiglio's matrix method

The central idea in matrix matching method is to limit the potential of interest to a finite region, which is equivalent to embedding it within an infinite square well potential of width 'a'. The TISE eq.(2) gets modified as

$$\left[H_\infty + V_w(x)\right]\psi(x) = E\psi(x) \qquad (5)$$

where H_∞ is the hamiltonian of infinite square well potential and $V_w(x)$ is the potential limited to finite region [0,a].

The energy eigen values and normalized eigen functions of H_∞ are given by

$$E_n^\infty = \frac{n^2\pi^2\hbar^2}{2ma^2} \qquad (6)$$

and

$$\Phi_n(x) = \sqrt{\frac{2}{a}}sin\left(\frac{n\pi x}{a}\right) \qquad (7)$$

This suggests the solution $\psi(x)$ to be written as a linear combination of sine functions as basis. So, $\psi(x)$ is expressed as

$$\psi(x) = \sqrt{\frac{2}{a}}\sum_{n=1}^{\infty} C_n sin\left(\frac{n\pi x}{a}\right) \qquad (8)$$

where the coefficients C_n's need to be determined. This is akin to performing Fourier analysis.

Substituting eq.(8) in eq.(2), we get

$$H\sum_{n=1}^{\infty} C_n\Phi_n(x) = E\sum_{n=1}^{\infty} C_n\Phi_n(x) \qquad (9)$$

Pre multiplying eq.(9) by $\Phi_m^*(x)$ on both sides and integrating over all space,

$$\int_{-\infty}^{\infty} \left(\Phi_m^*(x) \left[H_\infty + V_w(x) \right] \sum_{n=1}^{\infty} C_n \Phi_n(x) \right) dx$$
$$= E \int_{-\infty}^{\infty} \left(\Phi_m^*(x) \sum_{n=1}^{\infty} C_n \Phi_n(x) \right) dx \tag{10}$$

Since integration and hamiltonian are linear operators, eq.(10) can be written as

$$\sum_{n=1}^{\infty} C_n \int_{-\infty}^{\infty} \Phi_m^*(x) H_\infty \Phi_n(x) dx +$$
$$\sum_{n=1}^{\infty} C_n \int_{-\infty}^{\infty} \Phi_m^*(x) V_w(x) \Phi_n(x)$$
$$= E \sum_{n=1}^{\infty} C_n \int_{-\infty}^{\infty} \Phi_m^*(x) \Phi_n(x) \tag{11}$$

Introducing the notation for

$$\int_{\infty}^{\infty} \Phi_m^*(x) V_w(x) \Phi_n(x) = V_{mn} \tag{12}$$

using

$$H_\infty \Phi_n(x) = E_n^\infty \Phi_n(x) \tag{13}$$

and

$$\int_{-\infty}^{\infty} \Phi_m^*(x) \Phi_n(x) = \delta_{mn} \tag{14}$$

we get

$$E_m^\infty C_m + \sum_{n=1}^{\infty} V_{mn} C_n = E C_m \tag{15}$$

for m=1, 2, 3,...

These are any array of m equations that can be expressed as

$$E_m^\infty C_{m \times 1} + V_{m \times n} C_{n \times 1} = E C_{m \times 1} \tag{16}$$

Because m takes same values as n, we have

$$HC = EC \tag{17}$$

where

$$H = \begin{bmatrix} E_1^\infty + V_{11} & V_{12} & . & . & . & V_{1n} \\ V_{21} & E_2^\infty + V_{22} & . & . & . & V_{2n} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ V_{n1} & V_{n2} & . & . & . & E_n^\infty + V_{nn} \end{bmatrix}$$

and

$$C = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ . \\ . \\ . \\ C_n \end{bmatrix}$$

or we can express H as

$$H_{n \times n} = E_n^\infty I_{n \times n} + V_{n \times n} \tag{18}$$

3.2 Application to finite square well potential

The hamiltonian for the finite square well potential is given by

$$H = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) = H_\infty + V_w(x) \tag{19}$$

where, H_∞ is the hamiltonian of infinite square well potential and $V_w(x)$ is restricted to a finite region [0,a] defined as

$$V_w(x) = \begin{cases} V_0, & \text{if } b1 \leqslant x \leqslant b2. \\ 0, & \text{if } 0 < x < b1 \text{ and } b2 < x < a \\ \infty, & \text{x=0 and x=a} \end{cases}$$
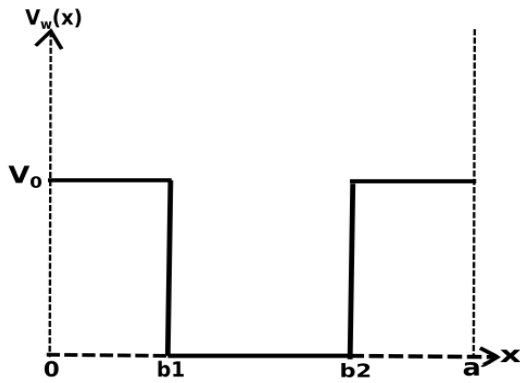$$\tag{20}$$

and is shown in Figure 2.

Figure 2: Square well embedded inside an infinite well

Now, V-matrix consists of diagonal elements, V_{nn} for which m=n and can be determined as

$$V_{nn} = \frac{n^2\pi^2}{2a^2} + \frac{V_0}{a}\left[b1 + a - b2 + \right.$$

$$\left.\frac{a}{2n\pi}(sin(\frac{2n\pi b2}{a}) - sin(\frac{2n\pi b1}{a}))\right] \quad (21)$$

and non-diagonal elements, V_{nm} which are to be determined as

$$V_{nm} = V_0\left[\frac{sin\frac{(n-m)\pi b1}{a}}{(n-m)\pi} - \frac{sin\frac{(n-m)\pi b2}{a}}{(n-m)\pi} \right.$$

$$\left.+\frac{sin\frac{(n+m)\pi b2}{a}}{(n+m)\pi} - \frac{sin\frac{(n+m)\pi b1}{a}}{(n+m)\pi}\right] \quad (22)$$

Theoretically, H-matrix is infinite in dimensions, which needs to be limited to a finite dimension, say N for sake of computation.

3.3 Choice of units: Atomic units

We prefer to work with atomic units where $\hbar$, m_e, c and Coulomb factor, $k_e = \frac{1}{4\pi\epsilon_0}$ all are equal to unity. This would mean that distances are choosen in Bohr and energies are

in hartree[17]. That is,

$$L = 1 Bohr = 0.52917725 Å$$

$$E = 1\ hartree = 27.211396\ eV$$

3.4 Discretizing position variable

We can only represent the position variable at certain discrete points within the finite region[0,a]. Typically, it is chosen as equally spaced points over this region with step size of say, h. That is, the values of x are limited to [0,h,2h,3h,....nh,...a].

3.5 Implementation in Gnumeric Spreadsheet

We enlist the various steps involved in implementing matrix method formalism for obtaining the energy spectrum and wave functions of 1-dimensional square well problem using free open source software, **'Gnumeric'** which is a spreadsheet environment.

Step 1: **Initializing the parameters**

	A	B	C
1	I-D Square well potential		
2	Initializing the parameters		
3			
4	$V_0(x)=$	14.000	0.514
5	a=	8.000	15.118
6	b=	4.000	7.559
7	b1=	(a-b)/2	3.779
8	b2=	(a+b)/2	11.338

Figure 3: Initializing the various input parameters in gnumeric worksheet

The initial values to the various parameters, $V_0 = 14eV$, $a = 8\text{Å}$ and $b = 4\text{Å}$ involved in square well potential are entered in worksheet (as shown in Figure 3):
• In cells A4, A5, A6, we mention the parameters as labels.
• In cells B4, B5, B6 we enter their respective values as 14, 8 and 4.
• In cell C4, we have given formula $" = \$B\$4/27.211396"$ so as to convert the potential $V_0(x)$ into hartree units and in cells C5 and C6 we have given the formula $" = \$B\$5/0.52917725"$ and $" = \$B\$6/0.52917725"$ respectively so as to convert the distances ('a' and 'b') into Bohr units.
• In cells A7 and A8 we label the parameters b1 and b2 respectively and show their respective formulae in B7 and B8. In cells C7 and C8, we enter the formulae as $" = (\$C\$5 - \$C\$6)/2"$ and $" = (\$C\$5 + \$C\$6)/2"$, respectively.

Step 2: Generating the hamiltonian matrix

Initially, let's choose 5 basis functions, N = 5, so we need to generate a 5×5 matrix. Now, we give the values of m & n from 1 to 5 in cells A17-A21(rows) and B16-F16(columns) as shown in Figure 4. Next, enter the formula for H-matrix in cell B17 as shown using **if condition expression** which is $" = if(condition, [truevalue], [falsevalue])"$
In our case, the if condition is given as: $\$A17 = B\$16$. True value is for diagonal elements given by eq. (21), which is to be
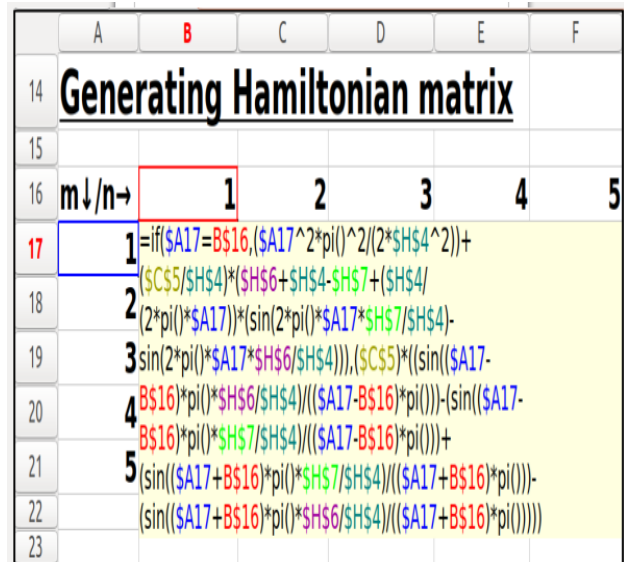
typed as

Figure 4: Generating the matrix in gnumeric sheet

```
($A17^2*pi()^2/(2*$C$5^2))+
($C$4/$C$5)*($C$7+$C$5-$C$8+
($C$5/(2*pi()*$A17))*
(sin(2*pi()*$A17*$C$8/$C$5)-
sin(2*pi()*$A17*$C$7/$C$5)))
```

whereas false value is for non-diagonal elements, given by eq.(22), and is typed as:

```
($C$4)*((sin(($A17-B$16)*pi()*
$C$7/$C$5)/(($A17-B$16)*pi()))-
(sin(($A17-B$16)*pi()*$C$8/$C$5)/
(($A17-B$16)*pi()))+(sin(($A17+B$16)*
pi()*$C$8/$C$5)/(($A17+B$16)*pi()))-
(sin(($A17+B$16)*pi()*$C$7/$C$5)/
(($A17+B$16)*pi())))
```

After entering desired formula, press Enter key, first matrix element will appear in cell

B17. To generate remaining values select cell B17 and drag the cursor upto cell F17 horizontally. After this, select entire row from B17:F17 and drag the cursor vertically and drop at F21. The complete H matrix having 25 elements is shown in Figure 5. **Step**

Figure 5: 5×5 H matrix in gnumeric sheet

3: Calculating the eigenvalues and corresponding eigenvectors

Gnumeric has an eigen value solver, which is not available in other worksheet environments such as MS-excel or open Office Calc. Here, we have generated 5×5 matrix in Step 2 which has 5 eigen values and 5×1 eigen vector for each of them. Now, select 6×5 matrix from B28 to F33 by dragging the mouse and then we type in formula bar:
$" = eigen(B17 : F21)"$
and press Ctrl+Shift+Enter keys together. With this, all 5 eigen values will appear in cells B28:F28 in decreasing order along with their corresponding 5 column vectors below each of them as shown in Figure 6.
Since all energy values come out to be in

hartree, we need to convert them into original units i.e in eV by multiplying with a factor of 27.211396 so as to compare with analytical values and this is done by typing formula in cell B35 as: $" = B\$28 * 27.211396"$ and drag cursor upto F35.

Figure 6: Calculating eigen values and eigen functions

Step 4: Generating basis functions
In order to determine the wavefunctions in eq.(8), we need to generate the basis functions, $\phi_n(x)$. In Figure 7, we generate x-values first from A43:A58 at an interval of 1 i.e 16 values and the values of 'n' are to be fed in cells B41:F41 as 1, 2, 3, 4, 5. The cells B42:F42 are labelled for clarity as $phi_1(x)$, $phi_2(x)$, $phi_3(x)$, $phi_4(x)$ and $phi_5(x)$ respectively. Now, select the entire array i.e. range of cells from $B43:$B58 and then enter in formula bar:

```
=sqrt(2/$C$5)$sin(B$41*pi()*
($A43:$A58)/$C$5)
```

	A	B	C	D	E	F
39	**Generating basis function(phi(x))**					
40						
41	x↓/n→	1	2	3	4	5
42		phi$_1$(x)	phi$_2$(x)	phi$_3$(x)	phi$_4$(x)	phi$_5$(x)
43	0	=sqrt(2/$C$5)*sin(B$41*pi()*($A43:$A58)/$C$5)				
44	1	0.075	0.147	0.212	0.269	0.313
45	2	0.147	0.269	0.345	0.362	0.318
46	3	0.212	0.345	0.348	0.220	0.009
47	4	0.269	0.362	0.220	−0.066	−0.309
48	5	0.313	0.318	0.009	−0.309	−0.322
49	6	0.345	0.220	−0.205	−0.350	−0.018
50	7	0.361	0.084	−0.342	−0.163	0.304
51	8	0.362	−0.066	−0.350	0.130	0.326
52	9	0.348	−0.205	−0.227	0.339	0.027
53	10	0.318	−0.309	−0.018	0.326	−0.299
54	11	0.275	−0.360	0.198	0.101	−0.330
55	12	0.220	−0.350	0.339	−0.190	−0.036
56	13	0.155	−0.280	0.352	−0.357	0.294
57	14	0.084	−0.163	0.233	−0.291	0.334
58	15	0.009	−0.018	0.027	−0.036	0.044

Figure 7: The five basis functions of infinite square well potential

After giving the formula, press Ctrl+Enter. The remaining phi vectors will be generated by cursor drag from $B43:$B58 and drop at $F43:$F58. We can observe that all these phi functions are nothing but the 5 sine functions of 1-D infinite square well.

Step 5: **Obtaining the wave functions**

For obtaining the wave functions, x values are generated first ranging from 0-15 in cells A64:A79 just as we have done in Step 4. Now, we need to multiply eigen vectors with basis functions to get wave functions(psi) as per eq. (7) By looking into Figure 6 and Figure 7 one can generate fifth wavefunction as

$psi_5(x) = C51 * phi_1(x) + C52 * phi_2(x) + C53 * phi_3(x) + C54 * phi_4(x) + C55 * phi_5(x)$

where C51, C52, C53, C54, C55 are the components of 5th eigen vector(coefficients

of fifth wave function) from B29:B33. To determine this, first select the range of cells from B64:B79 and then enter the following expression in the formula bar as:

$= \$B\$29 * \$B\$43 + \$B\$30 * \$C43 + \$B\$31 * \$D43 + \$B\$32 * \$E43 + \$B\$33 * \$F43$

and press Ctrl+Enter. This will generate only fifth wavefunction. In order to generate all ψ functions, one needs to give the formula in such a way that all ψ values will appear simultaneously. To achieve this, first select the range of cells from B64:F79 and then enter the following in the formula bar.

$= B\$29 * \$B\$43 + B\$30 * \$C43 + B\$31 * \$D43 + B\$32 * \$E43 + B\$33 * \$F43$

and press Ctrl+Enter, which will generate all the ψ values as shown in Figure 8. It is to be observed that ψ functions that are generated are in descending order.

	A	B	C	D	E	F
61	**Wave functions**					
63	x↓/psi(x)→	psi$_5$(x)	psi$_4$(x)	psi$_3$(x)	psi$_2$(x)	psi$_1$(x)
64	0	0.000	0.000	0.000	0.000	0.000
65	1	0.384	0.306	0.038	−0.005	0.000
66	2	0.462	0.448	0.165	0.054	0.022
67	3	0.207	0.361	0.342	0.191	0.080
68	4	−0.119	0.122	0.414	0.348	0.175
69	5	−0.223	−0.111	0.263	0.429	0.289
70	6	−0.055	−0.196	−0.054	0.364	0.394
71	7	0.162	−0.100	−0.321	0.153	0.459
72	8	0.177	0.081	−0.338	−0.122	0.463
73	9	−0.026	0.193	−0.093	−0.346	0.404
74	10	−0.215	0.131	0.231	−0.430	0.303
75	11	−0.147	−0.090	0.408	−0.363	0.188
76	12	0.166	−0.338	0.359	−0.210	0.089
77	13	0.447	−0.449	0.186	−0.067	0.026
78	14	0.413	−0.334	0.048	0.003	0.001
79	15	0.053	−0.040	0.002	0.002	0.000

Figure 8: The wave functions of square well potential

4   Results and Discussions

4.1   Visualising wavefunctions and probability densities

Plotting the wave functions '$\psi_n(x)$':
We plot the first three wavefunctions as a superposition of the basis functions with their corresponding weightages (coefficients). The plots are shown in Figure 9.
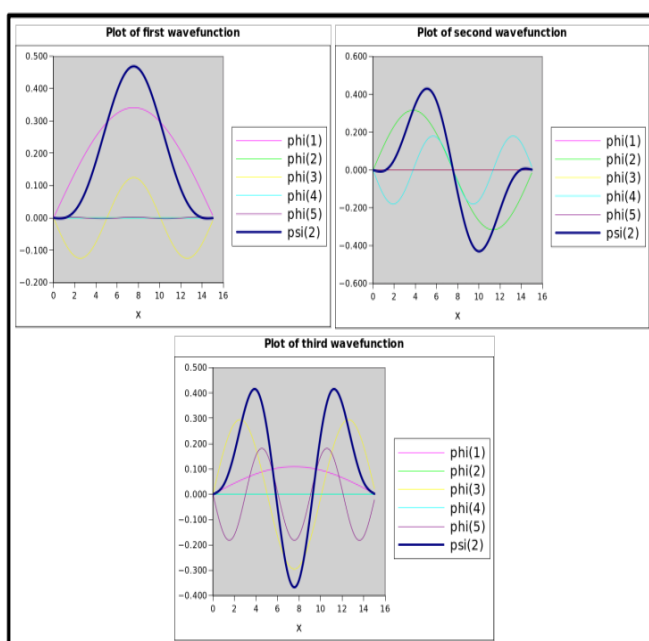

Figure 9: Graph of $\psi_1(x)$, $\psi_2(x)$ and $\psi_3(x)$ eigen functions

Probability density functions :
Once we obtain the wavefunctions, we can determine probability densities i.e probability of finding a particle(electron) in state $\psi(x)$ at position x by multiplying conjugate of ψ with ψ i.e $||\psi|| = \psi^*\psi$. The probability densities corresponding to first three wavefunctions which are of our interest are shown in Figure 10.
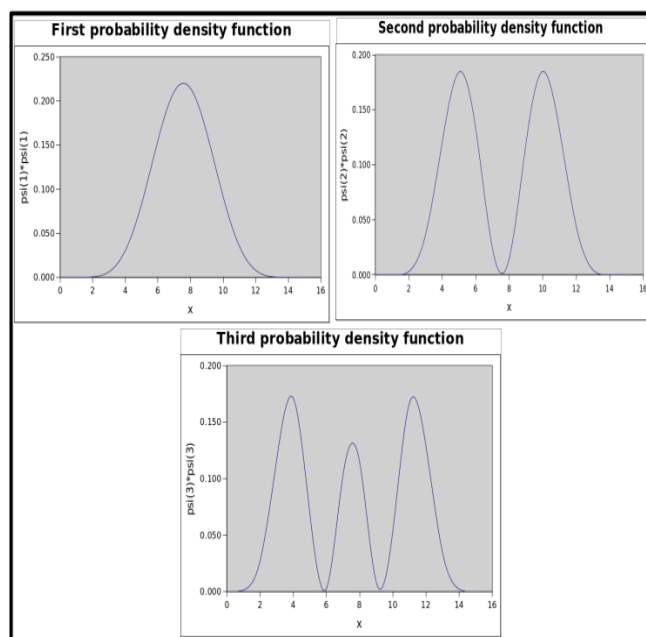
Figure 10: Getting the probability densities for square well potential

4.2   Study of variation of algorithm parameters: N and a

Effect of number of basis functions 'N':
During the implemetation stage, we have generated the results for N = 5 but these do not give the expected energy values to a good accuracy. So, we increased N in steps of 5 upto N=30, where the accuracy is found to be less than 2%. This effect could be clearly seen in Table 1.

Table 1: Comparison of numerical and analytical energy values for a particle in finite square well for $V_0 = 14eV = 0.514$ hartree and $a = 8Å$=15.118 Bohr

	Energy eigen values(eV)			
	Analytical	Numerical		
		N=5	N=30	N⩾70
E1	1.467	1.511 (3.00%)	1.468 (0.07%)	1.468 (0.07%)
E2	5.712	5.967 (4.48%)	5.722 (0.19%)	5.721 (0.17%)
E3	11.919	12.678 (6.38%)	12.089 (1.43%)	12.087 (1.42%)

Effect of varying the infinite square well width 'a':

The effect of infinite square well of width(a) on bound state energy values is well presented in Table 2. It can be seen that the accuracy of given energy level decreases with the increase in 'a' for same value of N = 70. The explanation behind this is that as 'a' increases, spatial resolution gets decreased for particular basis state, ϕ_{30}(say) as compared to 30th basis state for smaller value of 'a'. Also, the spread of sine wave increases well beyond the region of interest. As a result, there will be poor representation of basis functions within the infinite square well region. Therefore, to achieve good convergence as 'a' is increased, we need to increase the number of basis functions, N and decrease the step-size, h for descretizing 'a' accordingly. Another important observation is that for well width b = $4A^o$, only for a ≥ 20 A^o, the numerical results converge to all the

three analytical values.

4.3 Simulation by varying the physical parameters of the system: V_0 and b

Effect of varying the finite square well width 'b':
Results of bound state energies for finite square well potential as a function of well-width 'b' are shown in Table3. It is clear that as 'b' increases, number of bound states also increases.

As $b \to 0$, the finite square well reduces to a delta potential where the probability of finding a particle becomes maximum at the centre of infinite square well. On other hand, when $b \to a$, numerical results began to resemble with those of the infinite square well. *Effect of varying the finite square well depth $'V_0'$*:

When infinite square well depth, V_0 is increased, the number of bound states also increased and such behaviour is also expected which can be seen clearly from data presented in Table 4. One can observe that the energy eigen values for $V_0 = 10^4$ are very close to those of the infinite well potential shown in last column.

Table 2: Variation of energy eigen values with infinite square well width 'a' for $b = 4\mathring{A} = 7.559$ Bohr and $V_0 = 14eV = 0.514$ hartree. N is varied till convergence is acheived.

	Analytical	Energy eigen values(eV)					
		Numerical					
		a=10Å		a=20Å		a=30Å	
		N=70	N=80	N=70	N=100	N=70	N=150
E1	1.467	1.467	1.467	1.467	1.467	1.468	1.467
E2	5.712	5.712	5.712	5.713	5.712	5.716	5.712
E3	11.919	11.954	11.954	11.921	11.919	11.926	11.919

5 Conclusions

We have introduced a simple worksheet based simulation methodology for solving the TISE for 1D finite square well for obtaining its bound state energy eigen values and corresponding eigen wavefunctions using the 1D infinite square well wavefunctions as basis. This approach could be easily extended to solve the Schrödinger equation for other potentials such as Harmonic, Anharmonic, Morse and Double well as well as N-square-well since it is possible to obtain the analytical expressions for the integrals involved in the determination of hamiltonian matrix elements for these. For central potentials such as Hydrogen atom, spherical square well, Yukawa and Woods-Saxon, we need to determine the integrals involved numerically and a FOSS Scilab is a good choice. Even then, the current suggested approach could be used as the first step to give clarity regarding the steps involved in implementation for the former potentials, so that it be-comes pedagogically easy to write the code in Scilab for the later ones.

Table 3: Variation of energy eigen values with finite square well width(b) for $N = 140$, $a = 8\text{Å} = 15.118\,\text{Bohr}$ and $V_0 = 14eV = 0.514$ hartree. The analytical value corresponding to energies are shown below in brackets.

	Numerical(Analytical) energy eigen values(eV)		
	b=2Å (a=20Å)	b=4Å (a=16Å)	b=6Å (a=16Å)
E1	3.915 (3.915)	1.467 (1.467)	0.756 (0.756)
E2	12.972 (12.972)	5.712 (5.712)	2.997 (2.997)
E3		11.919 (11.919)	6.619 (6.619)
E4			11.302 (11.302)

Table 4: Variation of energy eigen values with finite square well depth(V_0) for $N = 100$, $a = 20\text{Å} = 15.118\,\text{Bohr}$ and $b = 4\text{Å} = 7.559\,\text{Bohr}$. The values written inside brackets are analytical results corresponding to energies.

	Numerical(Analytical) energy eigen values(eV)			
	Finite well			Infinite well
	V_0=50eV	V_0=1000eV	V_0=10^4eV	
E1	1.812 (1.812)	2.212 (2.212)	2.346 (2.346)	2.350
E2	7.214 (7.214)	8.846 (8.846)	9.383 (9.382)	9.401
E3	16.084 (16.084)	19.900 (19.900)	21.112 (21.110)	21.152
E4	28.135 (28.135)	35.373 (35.372)	37.533 (37.530)	37.603
E5	42.510 (42.510)	55.259 (55.258)	58.646 (58.640)	

References

[1] Marsiglio F., "The Harmonic Oscillator in Quantum Mechanics:A third way", American Association of Physics Teachers 77, no. 3, (2009): 253-258.

[2] Jugdutt B.A. and Marsiglio F., "Solving for three- dimensional central potentials using numerical matrix methods", American Association of Physics Teachers 81, no. 5, (2013): 345-346.

[3] Jelic V and Marsiglio F., "The Double Well Potential in Quantum Mechanics:a simple, numerically exact formulation", European Journal of Physics, (2012):1-10.

[4] Sharda, V. , Sastri, O. S. K. S., Bhardwaj, J., & Jha, A. K. (2016), "A computer simulation using spreadsheets for learning concept of steady-state equilibrium", Physics Education 51, no. 2, (2016): 025007.

[5] Sharda, V. , Bhardwaj, J., Sastri, O. S. K. S., & Jha, A. "A. Learning to Construct Boltzmann Distribution Concept: Using Computer Simulation Activity", International Education Conference, Learning Technology in Education, (2015): 367-374.

[6] Aditya, Abhishek, Ravikanth and Sastri O.S.K.S, "Computer Simulation Approach to Teaching Fundamental Ideas in Statistical Physics", THE BEDE ATHENAEUM 3, no. 1, (2012): 1-5.

[7] McKagan et al., "Developing and researching PhET simulations for teachingquantum mechanics", American Journal of Physics 76, (2008): 406-417.

[8] Moore, E. B., & Lewis C., "Opportunity: Inclusive Design for Interactive Simulations", Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility, (2015): 395-396.

[9] `https://www.compadre.org/osp/` `search/\search.cfm?gs=231&b=1&qc=` `Compiled\%20Simulation.`

[10] David Hestenes, "Toward a modeling theory of physics instruction", American Journal of Physics 55, no. 5,(1987): 440-454.

[11] Goswami, Amit. The idealistic interpretation of quantum mechanics. 1989.

[12] Ghatak, Ajoy K., and S. Lokanathan. Quantum mechanics: theory and applications. Macmillan, 2004.

[13] Zettili, Nouredine. "Quantum mechanics: concepts and applications." (2003): 93-93.

[14] DeVries, Paul L., and Patrick Hamill. A first course in computational physics. (1995): 283-285.

[15] Chapra, Steven C., and Raymond P. Canale. Numerical methods for engineers. Boston: McGraw-Hill Higher Education, 2010.

[16] Dauphinee, and Marsiglio, "Asymmetric wave functions from tiny perturbations", American Journal of Physics 83, no. 10, (2015): 861-866 .

[17] `https://en.wikipedia.org/wiki/` `Hartree_atomic_units`

A Compendium on General Relativity for Undergraduate Students

Tushar Kanti Dey[1],† and Surajit Sen[1,2,§]

[1] Centre of Advanced Studies and Innovations

18/27 Kali Mohan Road, Tarapur, Silchar 788003, India

†Email: tkdey54@gmail.com

[2]Department of Physics

Gurucharan College,Silchar 788004, India.

§ Email: ssen55@yahoo.com

Abstract

We give a pedagogical introduction of the essential features of General Theory of Relativity (GTR) in the format of an undergraduate (UG) project. A set of simple MATHEMATICA® code is developed which enables the UG students to calculate the tensorial objects without prior knowledge of any package operation. The orbit equations of light and material particle in Minkowski and Schwarzschild spacetime are solved numerically to illustrate the crucial tests of GTR.

1   Introduction

In a reminiscence on General Theory of Relativity Einstein wrote, *"I was sitting in a chair in the patent office at Bern when all of a sudden a thought occurred to me. 'If a person falls freely, he will not feel his own weight'. I was startled. This simple thought made a deep impression on me. It impelled me to a theory of gravitation"* [1]. In contrast, Newton narrated the natural world as, *"...the whole burden of (natural) philosophy seems to consist of this - from the phenomena of motions to investigate the forces of nature, and then from these forces to demonstrate the other phenomena"* [2]. According to him, the attraction between two gravitating objects is due to the *'gravitational force'* between them, while Einstein *did not* endorse it as a 'force', but a manifestation of the curvature of *spacetime*, an entity which is formed by soldering space and time together. Einstein's GTR is regarded as one of

the foremost intellectual triumph of all time which always attracts the young undergraduate (UG) students. However, their enthusiasm is often impeded by two major obstacles: first, conceptually the tenet of GTR is completely different from Newtonian approach of gravitation, and second, the evaluation of the tensorial quantities of Riemannian geometry is very tedious and often leads to endless number of human errors. To address the first part of two major hindrances, the serious readers may go through books [3, 4] and large quantity of online materials available in Internet. However, for the second part, i.e., to compute the tensorial objects, they generally consult suitable software packages to minimize the human error and save their precious time. Today there exists a plenty of free as well as commercial packages based on MATHEMATICA® [5], MATLAB® [6] or other software to carry out tensor algebra on computer. However, due to short span a given semester allotted for the UG project, the students of the sophomore or advanced years often face great difficulty to learn the intricacies of these package operation in particular. To address this problem, in this paper we have provided a set of MATHEMATICA® code to calculate the tensorial objects like, Cristoffel symbol, Riemannian, Ricci, Einstein tensor and Geodesic equation which they require to learn GTR. Our approach is a easy alternative to popular MATHEMATICA® based packages like, MathGR® [7], GTRTensorII® [8] etc [9, 10], which a student can workout

easily without prior knowledge of package operation.

The remaining parts of the paper are organized as follows: In Section 2 we introduce the basic formula of Reimaniann geometry necessary to develop the subject. Section 3 gives the step by step usage of MATHEMATICA® software to calculate the tensorial objects from some simple metric. In Section 4, we discuss the orbit equation of light and the material particle by solving the geodesics equations in Minkowski spacetime. The Einstein field equation in Schwarzschild spacetime is solved in Section 5 and the corresponding orbit equations are studied. Section 6 illustrates the numerical solution of these equations using MATHEMATICA® code and discuss the bending of light and perihelion shift, two most coveted tests of GTR. Finally, we summarize our results and discuss outlook.

2 Mathematical preliminaries

2.1 Key formula and equations

The special theory of relativity ensures frame independence of the physical laws with respect to all inertial frame of reference, while the general theory of relativity extends the same idea to include the non-inertial frame also. Any object moving under the influence of the gravitational field is essentially a accelerating object and therefore fits under that paradigm of the non-inertial frame of reference. To introduce the key effects of gravitation within framework

of GTR, let us recall some basic formula of Riemannian geometry. For their detailed derivation, we refer the readers to consult any standard textbook on GTR [3, 4, 11].

Any vector quantity V^μ (a tensor of rank one) in four dimensional space-time can be expanded as [1]

$$\mathbf{V}(t,\mathbf{r}) = \hat{e}_\mu(t,\mathbf{r})V^\mu(t,\mathbf{r}) \tag{1a}$$

$$\mathbf{V}(t,\mathbf{r}) = \hat{e}^\mu(t,\mathbf{r})V_\mu(t,\mathbf{r}) \tag{1b}$$

where V_μ and V^μ are defined as the covariant and contra-covariant vectors with corresponding basis vectors $\hat{e}_\mu$ and $\hat{e}^\mu$, respectively. We can define the two-indexed covariant and contra-variant metric tensor of rank two,

$$g_{\mu\nu}(t,\mathbf{r}) = \hat{e}_\mu(t,\mathbf{r})\hat{e}_\nu(t,\mathbf{r}) \tag{2a}$$

$$g^{\mu\nu}(t,\mathbf{r}) = \hat{e}^\mu(t,\mathbf{r})\hat{e}^\nu(t,\mathbf{r}), \tag{2b}$$

which satisfies the orthogonality relation, namely,

$$g^{\mu\nu}g_{\mu\rho} = \delta^\nu_\rho. \tag{3}$$

The metric tensor have the ability to transform the contravariant vector into a covariant vector and vice versa and this operation is known as raising or lowering of indices,

$$V^\mu g_{\mu\nu} = V_\nu \qquad V_\mu g^{\mu\nu} = V^\nu, \tag{4a}$$

$$V^{\mu\nu}g_{\mu\sigma} = V^\nu_\sigma, \quad V^\mu_\nu g_{\mu\sigma} = V_{\nu\sigma},$$

$$V_{\mu\nu} = g_{\mu\rho}V^{\rho\sigma}g_{\sigma\nu}. \tag{4b}$$

In general, we may have a covariant, contravariant or a mixed tensor of rank n, m and (n,m), respectively,

$$V_{\mu_1\mu_2\mu_3...\mu_n}, \quad V^{\mu_1\mu_2\mu_3...\mu_m}, \quad V^{\mu_1\mu_2\mu_3...\mu_m}_{\nu_1\nu_2\nu_3...\nu_n}. \tag{5}$$

To develop a calculus on the spacetime continuum, it is customary to define Christoffel symbol which involves the derivative of the metric tensors with respect to spacetime coordinate $x^\mu(x^1, x^2, x^3, x^4)$,

$$\Gamma^\rho_{\alpha\beta} = \frac{1}{2}g^{\rho\gamma}\left(\frac{\partial g_{\gamma\alpha}}{\partial x^\beta} + \frac{\partial g_{\gamma\beta}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^\gamma}\right), \tag{6}$$

which is symmetric with respect to its lower indices. The covariant derivative of a second rank tensor can be defined in terms of the Cristoffel symbol, i.e.,

$$D_\rho V_{\mu\nu} = \partial_\rho V_{\mu\nu} - \Gamma^\alpha_{\mu\rho}V_{\alpha\nu} - \Gamma^\alpha_{\nu\rho}V_{\mu\alpha}, \tag{7a}$$

$$D_\rho V^{\mu\nu} = \partial_\rho V^{\mu\nu} + \Gamma^\mu_{\alpha\rho}V^{\alpha\nu} + \Gamma^\nu_{\beta\rho}V^{\mu\beta}. \tag{7b}$$

The covariant derivative of metric tensor has vanishing value, i.e.,

$$D_\alpha g_{\mu\nu} = 0 \qquad D_\alpha g^{\mu\nu} = 0. \tag{8}$$

The Riemann curvature tensor is defined as

$$(D_\mu D_\nu - D_\nu D_\mu)V_\rho = R^\lambda_{\rho\nu\mu}V_\lambda, \tag{9}$$

where,

$$R^\lambda_{\mu\nu\rho} = \partial_\nu \Gamma^\lambda_{\mu\rho} - \partial_\rho \Gamma^\lambda_{\mu\nu} + \Gamma^\eta_{\mu\rho}\Gamma^\lambda_{\eta\rho} - \Gamma^\eta_{\mu\nu}\Gamma^\lambda_{\eta\rho}, \tag{10}$$

and on contraction it gives the Ricci tensor and Ricci scalar, i.e.,

$$R_{\mu\nu} = g^{\lambda\rho}R_{\lambda\rho\mu\nu}, \qquad R = g^{\mu\nu}R_{\mu\nu}, \tag{11}$$

respectively. It is convenient to the define a second rank tensor called '*Einstein tensor*',

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}, \tag{12}$$

which is divergence less, namely,

$$D_\mu G^{\mu\nu} = 0, \tag{13}$$

1. Henceforth, the summation over the repeated indices is assumed following the Einstein convention.

Thus we can write celebrated *Einstein equation*,

$$G_{\mu\nu} = -\frac{8\pi G}{c^4} T_{\mu\nu}, \qquad (14)$$

where $T_{\mu\nu}$ be the divergence-free energy-momentum tensor of the gravitating matter which is responsible for producing the curvature in spacetime.

2.2 Geodesic equation in curved space-time

In Newtonian mechanics, the equation of motion of a free particle ($\mathbf{F} = 0$) in the Eucledean space is given by,

$$\frac{d^2\mathbf{r}}{dt^2} = 0, \qquad (15)$$

while in the special theory of relativity, which deals with the *inertial frame of reference*, the dynamics is governed by the equation of motion,

$$\frac{d^2 x^\mu}{d\tau^2} = 0. \qquad (16)$$

Here x^μ represents the spacetime coordinate in *Minkowski spacetime* with τ as the body-fixed proper time.

On the other hand in GTR, due to inhomogeneous character of the gravitational field, the inclusion of the acceleration becomes indispensable. Under such condition, the introduction of *non-inertial frame of reference* becomes inevitable and the equation of motion is given by so called, *Geodesic Equation*,

$$\frac{d^2 x^\mu}{d\tau^2} + \Gamma^\mu_{\nu\sigma} \frac{dx^\nu}{d\tau} \frac{dx^\sigma}{d\tau} = 0, \qquad (17)$$

which completely generalizes Eq.(16).

Thus from the calculation point of view, the GTR involves two steps:

- To solve the Einstein's equation Eq.(14) to find the metric tensor $g_{\mu\nu}$ which determines the geometry of spacetime continuum.

- To find the solution of the geodesic equation Eq.(17) to know the trajectory of the point mass or massless particle in that spacetime.

In this way, GTR is essentially a metric based geometric theory of gravitation which replaces the preeminent position of *force equation* advocated by Newton.

3 MATHEMATICA® Code for Tensorial Calculus

3.1 Basic flowchart

The calculation of the tensorial quantities in Riemannian geometry involves multiple derivative of the metric tensor with respect to four spacetime coordinates followed by a number of summation over the repeated indices. The purported nature of such operation often leads to endless number of errors. A suitable computer programme capable of doing symbolic calculation can perform such calculation in a error-free way within a very short span of time. In this section we present some MATHEMATICA® notebook code to calculate the tensorial objects for any arbitrary metric tensor (Grey Box online) and list them systematically (Red Box

online). The flowchart of their evaluation is given below (Download MATHEMATICA® NOTEBOOK file here or our Github repository):

Code I: Define list of four space-time coordinates (x1,x2,x3,x4**):**

```
x = List[x1,x2,x3,x4]/.{x1 → ♣,
x2 → ♣,x3 → ♣,x4 → ♣};
```

where ♣ be the unknown parameter which we need to be supplied externally.

Code II: Define the covariant metric tensor (gcv**) for a given line-element, find corresponding contra-variant (**gct**) metric tensor and check their orthogonality (**orthg**):**

```
        gcv = Table[0,{α,4}, {β,4}];
```
Now supply nonzero components of gcv, for example, gcv[[1,1]]$= -e^{\lambda[r]}$,
gcv[[2,2]]$= -r^2$, gcv[[3,3]]$= -r^2\sin^2\theta$, gcv[[4,4]]$= e^{\nu[r]}$.
```
        gct = Simplify[Inverse[gcv]]; MatrixForm[gct]
        orthg = FullSimplify[gcv.gct]; MatrixForm[orthg]
```

Code III: Calculation of Christoffel symbols (Γ) from Eq.(6):

```
Γa= FullSimplify[Module[{α,β,γ,δ}, Table[Sum[½gct[[ρ,γ]]
(D[gcv[[γ,α]],x[[β]]]+ D[gcv[[γ,β]], x[[α]]-D[gcv[[α,β]], x[[γ]])];
```

Code IV: List of components of the Christoffel Symbol:

```
listaffine := Table[If[UnsameQ[Γ[[α,β,γ]],0],
{ToString[TraditionalForm[Γ^{x[[α]]}_{x[[β]],x[[γ]]} ] ], "=",Γa[[α,β,γ]] }],
{α,4},{β,4},{γ,4}]
TableForm[DeleteCases[Flatten[listaffine,2],Null], TableSpacing → {1,1}]
```

Code V: Calculation of covariant Riemann curvature tensor (Reim**) from Eq.(10):**

```
Riem=Simplify[Module[{α,β,γ,δ,ρ},Table[ D[Γa[[ρ,α,γ]],x[[β]]]-
D[Γa[[ρ,α,β]],x[[γ]] ] +Sum[ Γa[[δ,α,γ]] Γa[[ρ,β,δ]] , {δ,4} ] -Sum[
Γa[[δ,α,β]] Γa[[ρ,β,δ]]  , {δ,4}], {α,4},{β,4},{γ,4},{ρ,4}]] ];
```

Code VI: Calculation of covariant Ricci tensor (`Ricicv`) from Eq.(11):

```
Ricicv =
Simplify[Module[{α,β,γ,δ},Table[Sum[D[Γa [[ρ,α,β]],x[[ρ]]],{ρ,4}]-
Sum[D[Γa [ρ,α,ρ]],x[[β]]],{ρ,4}] + Sum[Γa [[σ,α,β]] Γa [[ρ,ρ,σ]],
{σ,4},{ρ,4}]  - Sum[Γa [[σ,α,ρ]] Γa [[ρ,β,σ]],{σ,4},{ρ,4}] ]  ];
```

Code VII: List of components of covariant Ricci Tensor:

```
listRicicv := Table[If[UnsameQ[Ricicv[[α,β]]],0],
{ToString[TraditionalForm[R_{x[[α]],x[[β]]}] ] ,"=",Ricicv[[α,β]] }], {α,4},{β,4}]
TableForm[DeleteCases[Flatten[listRicicv,1], Null],TableSpacing → {1,1}]
```

Code VIII: Calculation of Ricci scalar (Rc) from Eq.(11)

```
Rc=
Simplify[Module[{α,β},Sum[gct[[α,β]] Ricicv[[α,β]],{α,4},{β,4}] ] ];
```

Code IX: Calculation of Einstein covariant tensor (Gmncv) from Eq.(12);

```
Gmncv=
Simplify[Module[{α,β},Table[(Ricicv[[α,β]]- 1/2
gcv[[α,β]] Rc),{α,4},{β,4}] ] ];
```

Code X: List of components of Covariant Einstein Tensor:

```
listGmncv := Table[If[UnsameQ[Gmncv[[α,β]]],0],
{ToString[TraditionalForm[G_{x[[α]],x[[β]]}] ] ,"=",Gmncv[[α,β]] }], {α,4},{β,4}]
TableForm[DeleteCases[Flatten[listGmncv,1], Null],TableSpacing → {1,1}]
```

Code XI: Calculation of mixed Einstein tensor (Gmnmx);

```
Gmnmx=
Simplify[Module[{α,β,γ},Table[Sum[gct[[α,γ]] Gmncv[[α,β]]]{γ,4},{α,4},{β,4}
] ]\, ];
```

Code XII: List of components of mixed Einstein Tensor:

```
listGmnmx := Table[If[UnsameQ[Gmnmx[[α,β]]],0,
{ToString[TraditionalForm[G_{x[[β]]}^{x[[α]]}] ] ,"=",Gmnmx[[α,β]] }], {α,4},{β,4}]
TableForm[DeleteCases[Flatten[listGmnmx,1], Null],TableSpacing → {1,1}]
```

Code XIII: Calculation of Geodesic (Orbit) equations (Geodesic) from Eq.(17):

```
Geodesic=
Module[{α,β,γ,s}, Table[Simplify[D[x[[α]][s],{s,2}]+
Sum[Γa[[α,β,γ]]*D[x[[β]][s],s]*D[x[[γ]][s],s], {β,4},{γ,4}]],{α,4}]];
```

Code XIV: List of components of geodesic equations

```
TableForm[Geodesic] // ExpandAll
```

In Code XIV, each expression of the list should be equated to zero to obtain required geodesic equations of point mass and light in a given spacetime.

Using above set of codes, it is easy to evaluate various tensorial quantities for any arbitrary spacetime described by the metric tensor.

3.2 Application to some simple metrics:

The generic metric of a coordinate system is given by

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu \tag{18}$$

where $g_{\mu\nu} = g_{\mu\nu}(x^1, x^2, x^3, x^4) = (\mathbf{r}, ct)$ be the covariant metric tensor.

1. *Minkowski space in cartesian coordinate*:

   Step I: This spacetime is described by the line element,

   $$ds^2 = dt^2 - dx^2 - dy^2 - dz^2, \tag{19}$$

where the coordinates of the system are cartesian, i.e.,

$$x^1 \to x, \quad x^2 \to y, \quad x^3 \to z, \quad x^4 \to t. \tag{20}$$

Step II: The components of the metric tensor ($c = 1$) are,

$$g_{11} \to -1, \quad g_{22} \to -1, \quad g_{33} \to -1,$$
$$g_{44} \to 1, \tag{21}$$

which are constant. Thus in the Minkowski spacetime, the Cristoffel symbol vanishes and hence Riemann curvature, Ricci and Einstein tensors become trivial.

2. *Minkowski space in spherical polar coordinate*:

   Step I: The line element in this coordinate system is given by

   $$ds^2 = dt^2 - dr^2 - r^2 d\theta^2 - r^2 \sin^2\theta d\phi^2, \tag{22}$$

   where coordinates are,

   $$x^1 \to r, x^2 \to \theta, x^3 \to \phi, x^4 \to t. \tag{23}$$

   Step II: The components of metric tensor directly read off from the line element Eq.(22) are,

   $$g_{11} \to -1, \quad g_{22} \to -r^2,$$
   $$g_{33} \to -r^2 \sin^2\theta, \quad g_{44} \to 1. \tag{24}$$

   Step III: The non-vanishing Cristoffel symbols are found to be,

   $$\Gamma^\theta_{r\theta} = \frac{1}{r}, \quad \Gamma^r_{\theta\theta} = -r, \Gamma^r_{\phi\phi} = -r\sin^2\theta$$

   $$\Gamma^\theta_{\phi\phi} = -\cos\theta\sin\theta, \Gamma^\phi_{r\phi} = \frac{1}{r},$$

   $$\Gamma^\phi_{\theta\phi} = \cot\theta. \tag{25}$$

Using the programmes (Code-V to X), the Reimann curvature, Ricci and Einstein tensors are found to be zero which indicates that in the spherical polar coordinate the Minkowski spacetime is intrinsically flat.

3. *Schwarzschild spacetime (Static, spherically symmetric and non-rotating object)*:

   Step I: The line element of Schwarzschild spacetime is given by

   $$ds^2 = \gamma_M d\tau^2 - \frac{1}{\gamma_M} dr^2$$
   $$-r^2(d\theta^2 + \sin^2\theta d\phi^2), \tag{26}$$

   where, unlike previous cases, the metric tensor is now space-dependent, i.e., $\gamma_M = \gamma_M(r)$. The coordinates of such system is same as the spherical polar coordinate system, i.e.,

   $$x^1 \to r, \ x^2 \to \theta, \ x^3 \to \quad \phi, \ x^4 \to t. \tag{27}$$

   Step II: The components of metric tensor read off for such system are,

   $$g_{11} = -\frac{1}{\gamma_M}, \quad g_{22} = -r^2,$$
   $$g_{33} = -r^2\sin^2\theta, \quad g_{44} = \gamma_M, \tag{28}$$

   Step III: The non-vanishing Cristoffel symbols are given by

$$\Gamma_{tr}^{t} = \frac{\gamma_M'}{2\gamma_M}, \quad \Gamma_{tt}^{r} = \frac{1}{2}\gamma_M\gamma_M',$$

$$\Gamma_{rr}^{r} = -\frac{\gamma_M'}{2\gamma_M}, \quad \Gamma_{\theta\theta}^{r} = -r\gamma_M,$$

$$\Gamma_{\phi\phi}^{r} = -r\gamma_M\sin^2\theta, \quad \Gamma_{r\theta}^{\theta} = \frac{1}{r}$$

$$\Gamma_{\phi\phi}^{\theta} = -\cos\theta\sin\theta, \quad \Gamma_{r\phi}^{\phi} = \frac{1}{r},$$

$$\Gamma_{\theta\phi}^{\phi} = \cot\theta \tag{29}$$

where γ_M' be the derivative of γ_M with respect to r. The derivation of the factor γ_M and the geodesic equation in Schwarzschild spacetime is given in Section.5. Finally we note that our code can be easily extended for spacetime for the line element containing nonzero off-diagonal terms.

4 Geodesic equation in Minkowski spacetime

Before passing to GTR, let us ask 'what would be the trajectory of light or a free particle in the Minkowski space-time?'. In spherical polar coordinate, using Codes I to IV, we obtain the non-vanishing Cristoffel symbols given by Eq.(17), and then, by executing Code XIII-XIV, we get following

geodesic equations,

$$\frac{d^2t}{d\tau^2} = 0 \tag{30a}$$

$$\frac{d^2r}{d\tau^2} - r\left(\frac{d\theta}{d\tau}\right)^2 - r\sin^2\theta\left(\frac{d\phi}{d\tau}\right)^2 = 0 \tag{30b}$$

$$\frac{d^2\theta}{d\tau^2} + \frac{2}{r}\left(\frac{dr}{d\tau}\right)\left(\frac{d\theta}{d\tau}\right)$$
$$- \cos\theta\sin\theta\left(\frac{d\phi}{d\tau}\right)^2 = 0 \tag{30c}$$

$$\frac{d^2\phi}{d\tau^2} + \frac{2}{r}\left(\frac{dr}{d\tau}\right)\left(\frac{d\phi}{d\tau}\right)$$
$$+ 2\cot\theta\left(\frac{d\theta}{d\tau}\right)\left(\frac{d\phi}{d\tau}\right) = 0. \tag{30d}$$

Eq.(30a) readily gives $\frac{dt}{d\tau} = constant$ and assuming it to be Lorentzian factor γ, we obtain the well known time dilation law of special theory of relativity,

$$\frac{dt}{d\tau} = \gamma. \tag{31}$$

Now if we restrict ourselves to work on a plane by setting $\theta = \pi/2$, Eq.(30c) becomes trivial, while Eqs.(30b) and (30d) are simplified to,

$$\frac{d^2r}{d\tau^2} - r\left(\frac{d\phi}{d\tau}\right)^2 = 0, \tag{32a}$$

$$\frac{d^2\phi}{d\tau^2} + \frac{2}{r}\left(\frac{dr}{d\tau}\right)\left(\frac{d\phi}{d\tau}\right) = 0, \tag{32b}$$

respectively. From Eq.(32b) we obtain the angular momentum conservation law,

$$\frac{d}{d\tau}\left(r^2\frac{d\phi}{d\tau}\right) = 0 \quad \Rightarrow \quad r^2\left(\frac{d\phi}{d\tau}\right) = h, \tag{33}$$

where h the conserved angular momentum. Now plucking back Eqs.(33) into Eqs.(32a) we obtain,

$$\frac{d}{d\tau}\left(\left(\frac{dr}{d\tau}\right)^2 + \frac{h^2}{r^2}\right) = 0$$
$$\Rightarrow \quad \left(\frac{dr}{d\tau}\right)^2 + \frac{h^2}{r^2} = -\epsilon, \tag{34}$$

where ϵ be a constant. Finally noting the fact that,

$$\frac{dr}{d\tau} = \frac{dr}{d\varphi}\frac{d\varphi}{d\tau} = \frac{h}{r^2}\frac{dr}{d\varphi}, \qquad (35)$$

Eq.(34) gives the requisite 'orbit equation' in Minkowski spacetime in $r - \phi$ plane,

$$\frac{1}{r^4}\left(\frac{dr}{d\varphi}\right)^2 + \frac{1}{r^2} = -\frac{\epsilon}{h^2}. \qquad (36)$$

At closest distance of approach $r = r_0$, the derivative in Eq.(36) vanishes and we obtain $\epsilon = -\frac{h^2}{r_0^2}$. Plucking it back into Eq.(36), we obtained its solution given by the integral,

$$\varphi(r) = \pm \int \frac{dr}{r^2\sqrt{\frac{1}{r^2} - \frac{1}{r_0^2}}}. \qquad (37)$$

The angle of deflection of the starlight due to a typical star is measured by the formula [4]

$$\hat{\alpha} = 2\Delta\varphi - \pi, \qquad (38)$$

where $\Delta\varphi = \varphi(r_\infty) - \varphi(r_0)$. Integration of Eq.(37), in the limit $r \to \infty$, gives $\Delta\varphi = \frac{\pi}{2}$ and the angle of deflection of the starlight is found to be,

$$\hat{\alpha} = 0, \qquad (39)$$

which indicates that the bending of light is *impossible* in the Minknowski spacetime.

5   Schwarzschild solution

5.1   Solution of Einstein's field equation

In 1916, Karl Schwarzschild gave the exterior solution of the Einstein equation for a static, non-rotating and spherically symmetric object in vacuum ($T_{\mu\nu} = 0$). To find the

metric tensors for this spacetime, we consider the line element to be,

$$ds^2 = B(r)dt^2 - A(r)dr^2 \\ -r^2(d\theta^2 + \sin^2\theta d\phi^2). \qquad (40)$$

Here $A(r)$ and $B(r)$ are the unknown terms, which tend to unity in the asymptotic limit ($r \to \infty$), are to be determined. Using Code I to X, non-vanishing components of the Einstein tensors are given by,

$$r\frac{dA(r)}{dr} + A^2(r) - A(r) = 0, \qquad (41a)$$

$$r\frac{dB(r)}{dr} - A(r)B(r) + B^2(r) = 0. \qquad (41b)$$

The solution of Eq.(41) is given by,

$$A(r) = \frac{1}{1 - \frac{e^{C_1}}{r}}, \qquad (42a)$$

$$B(r) = (1 - \frac{e^{C_1}}{r})C_2, \qquad (42b)$$

where C_1 and C_2 are two constants. To get the Minkowski metric in the asymptotic limit we must choose $C_1 = ln(2M), C_2 = 1$ and thus Schwarzschild metric reads,

$$ds^2 = \gamma_M d\tau^2 - \frac{1}{\gamma_M}dr^2 \\ -r^2(d\theta^2 + \sin^2\theta d\phi^2), \qquad (43)$$

where the general relativistic correction factor γ_M is given by,

$$\gamma_M = (1 - \frac{2M}{r}), \qquad (44)$$

which, as expected, tends to unity for $\frac{2M}{r} << 1$.

5.2 Geodesic equation in Schwarzschild space-time

In this Schwarzschild spacetime, the non-vanishing Christoffel symbols are given in Eqs.(29), while the corresponding geodesic equations are obtained by Codes XI and XII, respectively. In particular, unlike the geodesic equations in the Minkowski space-time, we note that Eqs.(30a) and (30b) are generalized to

$$\frac{d^2 t}{d\tau^2} - \frac{\gamma'_M}{\gamma_M}\left(\frac{dr}{d\tau}\right)\left(\frac{dt}{d\tau}\right) = 0, \qquad (45a)$$

$$\frac{d^2 r}{d\tau^2} - \frac{\gamma'_M}{2\gamma_M}\left(\frac{dt}{d\tau}\right)^2 + \frac{\gamma_M \gamma'_M}{2}\left(\frac{dr}{d\tau}\right)^2$$
$$- r\gamma_M\left\{\left(\frac{d\theta}{d\tau}\right)^2 + \sin^2\theta\left(\frac{d\phi}{d\tau}\right)^2\right\} = 0 \quad (45b)$$

while, Eqs.(30c) and (30d) remain unchanged. Taking $\theta = \frac{\pi}{2}$, which corresponds to the particle moving in the equatorial plane, Eq.(45b) is simplified to,

$$\frac{d^2 r}{d\tau^2} - \frac{\gamma'_M}{2\gamma_M}\left(\frac{dr}{d\tau}\right)^2 + \frac{\gamma\gamma'_M}{2}\left(\frac{dt}{d\tau}\right)^2$$
$$- r\gamma_M\left(\frac{d\phi}{d\tau}\right)^2 = 0, \qquad (46)$$

while Eq.(45a) can be written as

$$\frac{d}{d\tau}\left(ln\left[\gamma_M\frac{dt}{d\tau}\right]\right) = 0 \Rightarrow \frac{dt}{d\tau} = \frac{E}{\gamma_M}, \quad (47)$$

with E be a constant to be evaluated from some boundary condition. Plucking it back into Eq.(46) yields,

$$\frac{d^2 r}{d\tau^2} - \frac{\gamma'_M}{2\gamma_M}\left(\frac{dr}{d\tau}\right)^2 + \frac{E^2\gamma'_M}{2\gamma_M} - \frac{h^2\gamma_M}{r^3} = 0, \qquad (48)$$

which can be further simplified to

$$\frac{d}{d\tau}\left[\frac{1}{\gamma_M}\left(\frac{dr}{d\tau}\right)^2 - \frac{E^2}{\gamma_M} + \frac{h^2}{r^2}\right] = 0$$
$$\rightarrow \quad \frac{1}{\gamma_M}\left(\frac{dr}{d\tau}\right)^2 - \frac{E^2}{\gamma_M} + \frac{h^2}{r^2} = -\epsilon. \quad (49)$$

Finally using Eq.(33) we can write Eq.(49) in terms of the azimuth angle ϕ which gives the desired '*Orbit Equation*' in Schwarzschild spacetime,

$$\frac{1}{\gamma_M r^4}\left(\frac{dr}{d\varphi}\right)^2 + \frac{1}{r^2} - \frac{E^2}{\gamma_M h^2} = -\frac{\epsilon}{h^2}. \quad (50)$$

At distance $r = r_0$, $\gamma_M = \gamma_{M_0}$, $(dr/d\varphi) = 0$ and Eq.(50) gives

$$E = \sqrt{\gamma_{M_0}\left(\epsilon + \frac{h^2}{r_0^2}\right)}, \qquad (51)$$

where γ_{M_0} be the value of γ_M at $r = r_0$ i.e., $\gamma_{M_0} = 1 - \frac{2M}{r_0}$. Finally substituting back the value of E in Eq.(45a), we obtain the 'time dilation' law for GTR,

$$\frac{dt}{d\tau} = \gamma_G(r_0, M, h), \qquad (52)$$

where,

$$\gamma_G(M, r_0, h) = \frac{1}{\gamma_M}\sqrt{\gamma_{M_0}\left(\epsilon + \frac{h^2}{r_0^2}\right)}. \quad (53)$$

Unlike Eq.(31) of the Minkowski spacetime, we note that the ratio of the incremental coordinate time and proper time is function of r, h, M_0, r_0, respectively. The solution of Eq.(50) is given by,

$$\varphi(r) = \pm \int \frac{dr}{r^2(\gamma_M)^{1/2}\sqrt{\frac{E^2}{h^2\gamma_M} - \frac{1}{r^2} - \frac{\epsilon}{h^2}}}, \qquad (54)$$

which is the Schwarzschild counterpart of Eq.(37) in Minkowski spacetime. In the next Section, we shall study the solution of Eq.(54) to find the trajectory of light and a point mass in Schwarzschild geometry, respectively.

5.3   Effective potential in Schwarzschild geometry

To find the effective potential in Schwarzchild geometry, from Eq.(49) we have [3],

$$\frac{1}{2}E^2 = \frac{1}{2}\left(\frac{dr}{d\tau}\right)^2 + U_{eff}^{Sch2}(r), \qquad (55)$$

where the effective potential is given by

$$U_{eff}^{Sch2}(r) = \frac{h^2}{2r^2} - \frac{h^2 M}{r^3} + \frac{\epsilon}{2}\left(1 - \frac{2M}{r}\right). \quad (56)$$

1. *Material particle ($\epsilon = 1$)*: Minimizing the effective potential $U_{eff}^{Sch}(r)$ at $r = r_M$ we obtain

$$r_M = \frac{h^2 \pm h\sqrt{h^2 - 12M^2}}{2M}. \qquad (57)$$

Setting $h = 2\sqrt{3}M$, we obtain the orbit with smallest possible radius, namely,

$$r_{mat}^{min} = 6M. \qquad (58)$$

for the material particle.

2. *Light ($\epsilon = 0$)*: Similarly minimizing the effective potential the minimum radius for light is found to be,

$$r_{ph} = 3M. \qquad (59)$$

This minimal sphere of this radius is often referred as '*Photon sphere*'.

Comparing the effective potentials of the point mass ($\epsilon \neq 0$) and massless particle ($\epsilon = 0$) in Schwarzschild spacetime, where we note that in GTR both of them can be trapped to form a bound state.

6   Experimental tests of GTR

In this section we study the trajectory of light and material particle from the orbit equation given by Eq.(48) and its solution Eq.(54), respectively. These orbits are generally classified as the unbound ($\epsilon = 0$) and bound orbit ($\epsilon \neq 0$) which explain the bending of starlight and the perihelion precession of mercury, respectively.

6.1   Motion of light in the unbound orbit - Bending of light ray:

To discuss the deflection of light ($\epsilon = 0$) grazing out from a star like sun at closest distance of approach ($r = r_0$), we find from Eq.(51),

$$E = \pm\frac{h\sqrt{\gamma_{M_0}}}{r_0}. \qquad (60)$$

Substituting Eq.(60) along with γ_M and γ_{M_0} in Eq.(54) we obtain,

$$\Delta\varphi_{ph} = \int_{r_0}^{\infty} dr f(r; r_0, M), \qquad (61)$$

where, $\Delta\varphi_{ph} = \varphi_{ph}(r_\infty) - \varphi_{ph}(r_0)$ and the integrand $f(r; r_0, M)$ is given by,

$$f(r; r_0, M) = \frac{\sqrt{\frac{r}{r-2M}}}{r^2\sqrt{\frac{r(r_0-2M)}{r_0^3(r-2M)} - \frac{1}{r^2}}}. \qquad (62)$$

The solution of above integral cannot be obtained exactly and there exists many approximate methods to solve it. Here we present a direct calculation of the deflection angle using `NIntegrate` program of MATHEMATICA®,

$$In[11] = \hat{\alpha}_{ph} = (2\,\text{NIntegrate}[\text{f, r, } \clubsuit, \text{ Infinity}] - \pi)\frac{180 \times 60 \times 60}{\pi}$$

Taking the spherical body as a prototype star like sun, i.e.,

$$
\begin{aligned}
M = M_\odot &= 1.989 \times 10^{30}\text{kg,}\\
r_0 = R_\odot &= 6.95 \times 10^5\text{km,}\\
\frac{M_\odot G}{c^2} &= 1.457\text{km,}
\end{aligned}
\tag{63}
$$

and taking template $\clubsuit \rightarrow R_\odot$, the deflection angle is found to be $\hat{\alpha}_{ph}^{M_\odot} = 1.75003$ arcsecond. In 1919, Sir Arthur Eddington and his team verified the bending of starlight during total solar eclipse [12]. It is worth mentioning here that, the radius of the photon sphere of the sun is $r_{ph} = 3M_\odot \equiv 4.425$ km obtained from Eq.(59) lies well within the sun, i.e., $r_{ph} << R_\odot$. On the contrary, for a ultra-compact object like a black hole, the photon sphere is quite large and falls outside that object. In consequence, the bending of light for such objects is quite considerable. Finally we mention here that the bending of light leads to a unique phenomenon known as 'Gravitational Lensing', which is another spectacular outcome of GTR.

6.2 Motion of particle in bound orbit - perihelion shift of planets

We finally consider the motion of a test particle ($\epsilon \neq 0$) orbiting around the sun in an elliptical orbit. At perihelion $r = r_P$ and at aphelion $r = r_A$, $\frac{dr}{d\phi}$ vanishes in Eq.(50) and we get two values of the constants,

$$\epsilon = h^2\frac{r_A^2\gamma_{M_P} - r_P^2\gamma_{M_A}}{r_P^2 r_A^2(\gamma_{M_A} - \gamma_{M_P})}, \tag{64a}$$

$$E = \pm h\sqrt{\frac{\gamma_{M_A}\gamma_{M_P}}{\gamma_{M_A} - \gamma_{M_P}}\frac{r_A^2 - r_P^2}{r_A^2 r_P^2}}, \tag{64b}$$

where, $\gamma_{M_A} = 1 - \frac{2M}{r_A}$ and $\gamma_{M_P} = 1 - \frac{2M}{r_P}$, respectively. Plucking back these values in Eq.(54) we obtain,

$$\Delta\varphi_P(r) = \int_{r_P}^{r_A} f(r; M, r_P, r_A)dr \tag{65}$$

where $\Delta\varphi_P = \varphi(r_P) - \varphi(r_A)$ is referred as the perihelion shift of the given planet and f is given by,

$$
\begin{aligned}
&f(r; M, r_P, r_A)\\
&= \frac{\sqrt{\frac{r}{r-2M}}}{r^2\sqrt{\frac{(r-r_A)(r-r_P)[rr_A r_P - 2M\{r_A r_P + r(r_A+r_P)\}]}{(2M-r)r^2 r_A^2 r_P^2}}},
\end{aligned}
\tag{66}
$$

Once again the integral can be done numerically using following program of MATHEMATICA®,

$$\text{In[5]} = \hat{\alpha}_P = \frac{365.25}{\clubsuit}\left(2\,\text{NIntegrate}[\text{f}, \{\text{r}, \text{r}_\text{P}, \text{r}_\text{A}\}] - \pi\right)\frac{180 \times 60 \times 60}{\pi}.$$

Substituting the values of r_A, r_P and ♣ (Period in days) into above programme, we obtain the value of perihelion shift of all planets per century including Mercury. In Table-I, we have compared the perihelion shift of different planets with corresponding observational values:

Table: Perihelion shift of planets (in arc-sec per century)

Planet	r_P (in km)	r_A (in km)	Period in days (♣)	$\hat{\alpha}_{th}$	$\hat{\alpha}_{obs}$ [13]
Mercury	4,60,01,200	6,98,16,900	87.97	42.9334	42.9
Venus	10,74,76,359	10,89,42,109	224.70	8.6233	8.6
Earth	14,70,98,074	15,20,97,701	365.35	3.8364	3.8
Mars	20,66,69,000	24,92,09,300	686.97	1.3438	1.3
Jupiter	74,05,73,600	81,65,20,800	4331.57	.0622	.06
Saturn	135,35,72,956	151,33,25,783	10,759.22	.0136	.014
Uranus	274,89,38,461	300,44,19,704	30,799.10	.0024	.002
Neptune	445,29,40,833	455,39,46,490	60,190.00	.0007	.0007

We note that the observational value of the perihelion shift of all planets coincides with the theoretical results predicted by GTR. This striking success of GTR has established it as a complete theoretical model of gravitation.

7 Conclusion and outlook

This paper gives a cursory overview of General Theory of Relativity perceivable to the UG students. We presuppose that, the UG students have a rudimentary knowledge of Reimaniann geometry from their regular course and familiar with the tenets of basic MATHEMATICA® code to undertake this study as a unsupervised (or as a

project with minimal supervision) review project. A set of easy-to-do command-line MATHEMATICA® code is developed to solve the cumbersome tensorial quantities and to solve the orbit equations numerically. We have explicitly calculated the magnitude of the bending of light and perihelion shift of all planets including Mercury which are precisely in agreement with observational result. Apart from the UG students, our code of obtaining the orbit equation in arbitrary spacetime may be helpful for the graduate freshmen who often need to deal with wide class of metric with several nontrivial attributes.

References

[1] J. Bernstein and P. M. Fishbane and S. Gasirowicz, Modern Physics, (Pearson Education, New Delhi, 2009)

[2] M. L. King, A Short History of the Renaissance in Europe, (University of Toronto Press, Toronto 2017)

[3] S. Carrol. Spacetime and Geometry: An Introduction to General Relativity, (Pearson Education, New Delhi 2013)

[6] Mathworks, MATLAB

[4] S. Weinberg, Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity, (John Wiley and Sons, New York 1972)

[5] Wolfram Inc., MATHEMATICA, version 11

[7] Yi Wang, MathGR: A tensor and GR computation package to keep it simple (arXiv:cs.MS/1306.1295v3)

[8] Kayal Lake, A Tool for General Relativity, arXiv:physics.ed-ph/0509108

[9] ChiDotPhi, Mathematica for Physics 2-General Relativity

[10] J. B. Hartle, Gravity : an introduction to Einstein's general relativity (Addison Wesley, San Fransisco, 2003)

[11] J. V. Narlikar, A text book on General Theory of Relativity (Cambridge,New Delhi 2017)

[12] F. W. Dyson and A. S. Eddington and C. R. Davidson (1920), Philosophical Transactions of the Royal Society, Series **A220** 291

[13] S. Carloni and D. Grumiller and F. Preis (2011), Phys. Rev. **D83** 124024