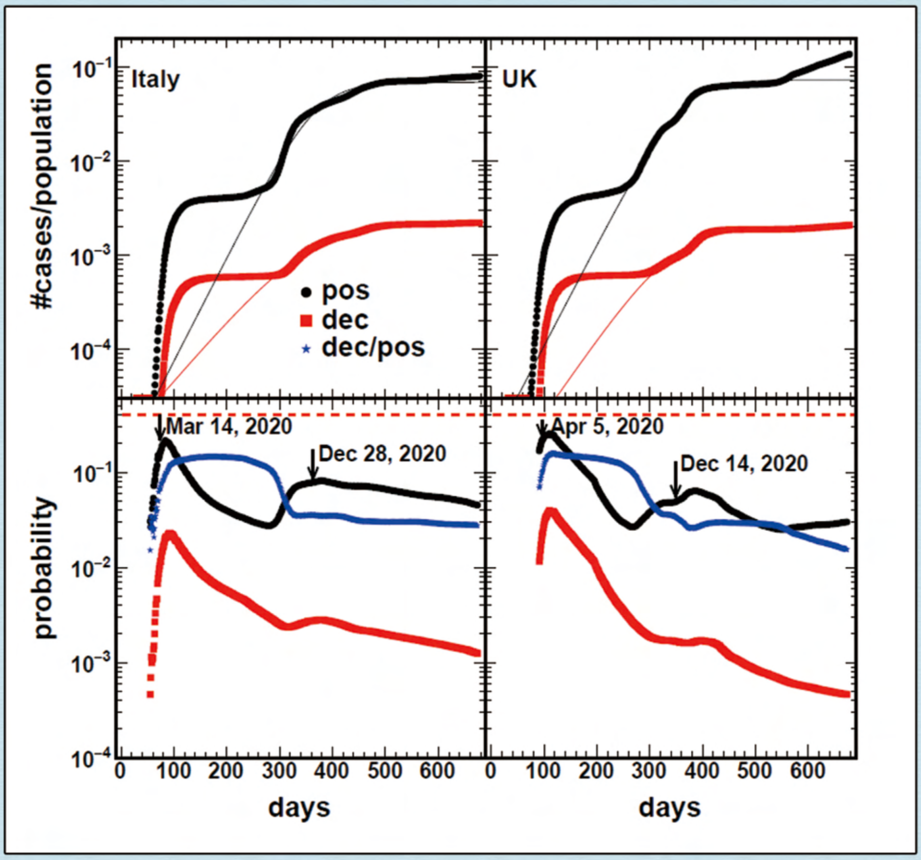
ISSN: 2153-1196 Volume 13, Number 1, January 2022

Scientific Research Publishing

Journal of Modern Physics



ISSN: 2153-1196

https://www.scirp.org/journal/jmp

Journal Editorial Board

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

Scientific Research Publishing

Journal of Modern Physics, 2022, 13, 1-88
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Table of Contents

Volume 13 Number 1 January 2022

The figure on the front cover is from the article published in Journal of Modern Physics, 2022, Vol. 13, No. 1, pp. 1-15 by Zheng, H. and Bonasera, A.

Journal of Modern Physics (JMP)
Journal Information

SUBSCRIPTIONS

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

Subscription rates:
Print: $89 per issue.
To subscribe, please contact Journals Subscriptions Department, E-mail: sub@scirp.org

SERVICES

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

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

COPYRIGHT

PRODUCTION INFORMATION

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

Journal of Modern Physics, 2022, 13, 1-15
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Controlling the Worldwide Chaotic Spreading of COVID-19 through Vaccinations

Hua Zheng[1*], Aldo Bonasera[2,3]

[1]School of Physics and Information Technology, Shaanxi Normal University, Xi'an, China
[2]Cyclotron Institute, Texas A&M University, College Station, USA
[3]Laboratori Nazionali del Sud, INFN, Catania, Italy
Email: *zhengh@snnu.edu.cn

How to cite this paper: Zheng, H. and Bonasera, A. (2022) Controlling the Worldwide Chaotic Spreading of COVID-19 through Vaccinations. *Journal of Modern Physics*, **13**, 1-15.
https://doi.org/10.4236/jmp.2022.131001

Received: December 15, 2021
Accepted: January 15, 2022
Published: January 18, 2022

Abstract

The striking differences and similarities between the "Spanish-flu" of 1918 and the Coronavirus disease of 2019 (COVID-19) are analyzed. Progress in medicine and technology and in particular the availability of vaccines has decreased the death probability from about 2% for the Spanish-flu, to about 10^{-4} in the UK and 10^{-3} in Italy, USA, Canada, San Marino and other countries for COVID-19. The logistic map reproduces most features of the disease and may be of guidance for predictions and future steps to be taken in order to contrast the virus. We estimate 6.4×10^7 deceases worldwide without the vaccines, this value decreases to 1.6×10^7 with the current vaccination rate. In November 2021, the number of deceased worldwide was 5.1×10^6. To reduce the fatalities further, it is imperative to increase the vaccination rate worldwide to at least 120 millions/day and the AstraZeneca vaccine due to its efficacy and cost is a possible route to accomplish this.

Keywords

COVID-19, Vaccination, Logistic Map

1. Introduction

The origins of the "Spanish-flu" of 1918 are unknown and so are the reasons why it stopped its deadly action in 1920, *i.e.*, after a period of three years. Data from such disease are not complete as we would like, since science, medicine and technology were not sufficiently advanced. At that time, we were scientifically in better shape than in previous epidemics reported in history, but not enough to be able now to draw strict parallels with the Coronavirus disease first reported in 2019 (COVID-19). In 1918, we were at the end of World War I (WWI) and this probably contributed to the virus spreading in different continents so quickly. Furthermore, countries at war were not willing to share information about the disease in order to avoid advantages for the enemy. Spain, which did not participate in the conflict, openly reported about the high mortality rate of the disease in the country and this is the origin of the name Spanish-flu (S1918). Presently, there are no major wars but still some countries prefer not to make public all the relevant data. We know that there were many "waves" in the period 1918-20, each more deadly than the previous one and it affected mostly young people. An explanation for this was that older people built their immune defenses from the pandemic of 1889-90 [1] but there is no definite proof about this. In any case it is a factor to be taken into account since "what does not kill you makes you stronger". The 1889 pandemic was as serious as the S1918 but lasted a shorter time maybe due to the smaller flow of people travelling from different countries and continents. The many successive waves observed for the S1918 may be connected to seasonal variations or an evolution of the disease, *i.e.*, different variants. The striking difference of the S1918 to the COVID-19 is the different age groups involved, most deaths in the latter case are older people even though new variants may change this unless we strengthen our immune defense through vaccines or survive to previous COVID-19 variants as in the 1889 case.

One hundred years ago we had little understanding of viruses, their spreading and how to contrast their deadly action. Masks and social distancing were adopted and some "homemade" remedies with little if no efficacy. Thus, we can assume that the disease followed its course almost undisturbed. The method of transmission is through the vicinity of affected people to unaffected ones (thus the mask utility when no other defenses are available), where one person may affect n-persons with $n \ge 1$ depending on the virus transmissibility. This can be expressed mathematically using chaotic maps and in particular we will discuss the logistic map:

$$x_{n+1} = r \times x_n \left(1 - x_n\right). \tag{1}$$

It is a simple iteration which, given an initial value $0 < x_0 < 1$, produces successive new values depending on the control parameter *r*. For $r \le 4$ the map is confined in the interval $[0,1]$, this is similar to the available phase-space in dynamical systems. For $r < 3$ the map gives two fixed points x_f which can be found assuming $x_{n+1} = x_n$, they are $x_f = 0$ and $x_f = 1 - 1/r$ [2]. For larger values of *r* a flip bifurcation occurs, *i.e.*, the map oscillates between two fixed values and this remains true for $r < 3.449479$, these values can be analytically recovered [2]. For larger values of *r*, the map oscillates among four fixed points ($r < 3.544090$) and so on, increasing the number of fixed points up to a critical value $r_c = 3.569946$ above which there are an infinite number of bifurcations and the map is fully chaotic [2] [3]. A powerful approach to quantify the "degree of chaoticity" is to calculate an ensemble of *N* pairs of trajectories, separated initially by a very small distance d_0. We introduce the mean distance between them at the iteration *n* as:

$$d_n = \frac{1}{N}\sum_{i=1}^{N}\left|x_n^{(i)} - y_n^{(i)}\right|, \tag{2}$$

where

$$x_n^{(i)} = f^n\left(x_0^{(i)}\right); \quad y_n^{(i)} = f^n\left(x_0^{(i)} + d_0\right). \tag{3}$$

The initial starting points $x_0^{(i)}$ are chosen from a uniform distribution spanning the defining interval of the map. For fully chaotic maps the average distance d_n after *n*-iteration may be expressed by the relation

$$d_n = \frac{d_0 d_\infty}{d_0 + d_\infty e^{-\lambda n}}, \tag{4}$$

where the Lyapunov exponent $\lambda > 0$ indicates that nearby trajectories diverge to a finite (since the phase space is finite) value $d_\infty\left(<1\right)$. In the ergodic limit $r = 4, \lambda = 2$ which can be considered as the "highest chaoticity" which can be reached. In the same limit [4] $d_\infty = \frac{4}{\pi^2}$, the largest average distance between nearby trajectories. Notice that even in the chaotic regime there are values of the control parameter where $\lambda < 0$, *i.e.*, the map is not chaotic. We may argue that for the S1918 pandemic one of such values was accidentally hit and the epidemic disappeared suddenly. Notice that in such cases Equation (4) gives $d_\infty = d_0 \to 0$ [4].

In refs. [5] [6] [7] [8] we discussed the fact that the cumulative probability to be infected by COVID-19 (number of cases divided by number of tests) follows the same Equation (4) with the iteration *n* substituted by time *t*, thus if we know the Lyapunov exponents for the data and the map we can easily make a connection between time and iteration *i.e.*,

$$t = \frac{\lambda}{\Lambda} n. \tag{5}$$

Under this assumption it follows that the maximum probability to become infected is $\frac{4}{\pi^2}$. In 1918 the world population was about 2 billion which means that about 0.8 billions got infected by the S1918 in the period 1918-20. As we will discuss below about 2% of the infected died which gives 16 millions deaths, a number comparable to the casualties of WWI. Extending these estimates to 2020, the world population is about a factor 4 higher, which increases the estimates above by the same factor. However, in November 2021, 7.3 billion vaccine doses have been administrated which means that about 3.65 billion people have been fully vaccinated. Since the mortality rate is about 2% of the positive cases, then we expect the mortality rate (mostly non vaccinated people) to be given by $(8 - 3.65) \times 2\% \times 4/\pi^2 = 35$ millions. Also by the end of the year, people who received the first vaccine dose, will get the second thus reducing the maximum estimate to 18 millions. Optimistically 75% of the world population may be vaccinated (excluding children age less than 12 for whom vaccination is not possible yet, but more recently the use of some vaccines have been extended to children aged 5 or older). This gives a mortality rate $0.25\times8\times10^9\times2\%\times4/\pi^2=1.6\times10^7$, currently, the number of recorded deaths is 5.1 million. Our estimate above assumes that the spreading is fully chaotic and as we will show below this is not the case. If we take Italy as an example, the average positive probability is about a factor 4 below the ergodic limit of the logistic map. These are in any case very important numbers thus the urgency to increase the number of vaccines worldwide, the faster we do the more we protect children and people with medical conditions and that cannot be vaccinated. We would like to stress that the extension of vaccinations to the youngest population may be necessary since there is some part of the population refusing to get vaccinated denying the clear success vaccines have obtained so far. But this is not enough especially if new variants will appear.

In Figure 1 (top panel) we plot the cumulative number of cases (positive and deceased) as function of time starting from January 1, 2020 for Italy (top-left panel) and the UK (top-right panel) [9] [10]. In order to directly compare the two countries we have divided the number of cases by their population (60,177,000 in

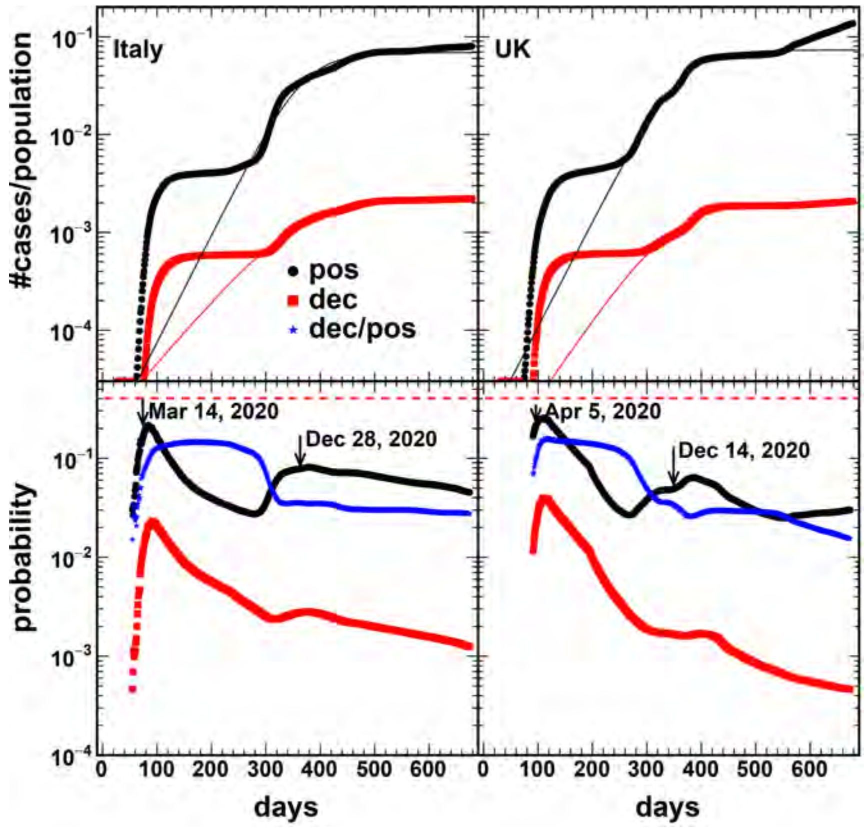


Figure 1. (Color online) Updated to November, 2021. (Top) The cumulative number of cases (positive and deceased) over population versus time for Italy and UK respectively. January 1, 2020 is the day 1; (Bottom) the corresponding probabilities as well as the ratio deceased over positive versus time. The arrows for March 14, 2020 and December 28, 2020 correspond to complete lockdown and start of vaccinations for Italy. Similarly, April 5, 2020 and December 14, 2020 correspond to the UK prime minister hospitalization for COVID-19 and start of vaccinations respectively.

Italy and 68,207,116 in the UK). The two cases are strikingly similar and we have observed no major differences to other countries, see also supplemental material. After the first rapid increase there is a plateau, which roughly starts at the end of the lockdowns and the beginning of the summer season. In the fall of 2020 the cases increase rapidly again forcing new lockdowns to slow them down and fortunately in December 2020 vaccinations started. The final increase in the UK is due to the Delta variant of the disease. The number of deceased is roughly a factor 50 smaller than the number of positives and no increase is seen at later times thanks to vaccinations. If we treat these data as representative of cases worldwide we can estimate a total number of positive to the virus in 790 millions (10% of the current world population) and a total number of deceased in 16 millions (0.2% of the current world population), in November 2021 those values are 253 millions and 5.1 millions deaths respectively. If this trend is followed, *i.e.*, no increase in the number of vaccines and no other more deadly variant, we can estimate the maximum total duration of the disease in 5 years, two more than the S1918. These values are of course not acceptable and a worldwide effort is needed, especially vaccinations. We notice in passing that some countries, notably S. Korea and others, have adopted strict distancing, tracking and wearing mask measures, which successfully stopped the spreading, see supplemental material. Other countries have not been able to follow these examples because of unrest from part of the population thus leaving vaccinations and new medicines to contrast the spreading. Unfortunately, in the same countries there has been some resistance to vaccination as well but this may be overcome through compulsory vaccinations or the natural spreading of the virus from vaccinated to unvaccinated ones. In the latter case, the weak part of the population, which cannot be vaccinated, may pay a high price unwillingly. A fit to the top part of **Figure 1** using Equation (4) gives about 7% and 0.2% total number of positives and deceased of the population respectively. These values are confirmed by many other countries, see supplements, with the notable exception of S. Korea reducing to 0.6% and 0.004% respectively. The approach and methods used by S. Korea and other Asiatic countries should be part of the civic study curricula of students starting in elementary schools worldwide to avoid the drama for the next pandemic, which could occur in the next 10 or 100 years.

The estimates above could be misleading because, especially at the very beginning of the disease, only part of the population was tested. Since the number of tests each day are known we can define probabilities as the number of cases divided the number of tests [5] [6] [7] [8]. These probabilities are reported in the bottom panels of **Figure 1**. The quantities reported in **Figure 1**, top and bottom, would be the same if the total number of tests equals the countries' population. The dashed lines in the bottom panels are given by the logistic map value $\frac{4}{\pi^2}$ and in no cases it is reached apart Norway, see **Figure S1**, at the very beginning of the pandemic. In particular the rapid increase at the beginning reaches a maximum value on March 14, 2020 when the lockdown was imposed in Italy and April 5, 2020 when the UK prime minister was admitted at the hospital in serious conditions having contracted COVID-19. As we noticed before the deceased probability follows the same trend and it is about a factor 50 lower than the positives, shifted in time of about one week. Notice that the "population" for the deceased is given by the positives, thus we can define another probability as the number of deceased divided by the number of positives. This quantity is plotted in the bottom part of Figure 1 (blue symbols) as function of time and it is very close to the probability to be positive to COVID-19 as expected. In particular, after a first rapid increase at the beginning it plateaus at about 10% and jumps down to 2% - 3% again in December 2020 when the vaccination campaigns started, clearly demonstrating their efficacity. We stress that in the S1918 case, if the disease was completely out of control, then the number of deaths may be a factor of 5 higher than estimated above, *i.e.*, of the order of 80 millions.

In order to compare to the logistic map, we plot in Figure 2 the positive probabilities in two different periods for Italy (top panel) and UK (bottom panel). The logistic map follows rather well (red symbols) the data and the values of the Lyapunov and d_∞ can be extracted for all cases. From ref. [4] we know that d_∞ can be parametrized as

$$d_\infty = c_1 \left(r - r_\infty\right)^{\nu/2}, \tag{6}$$

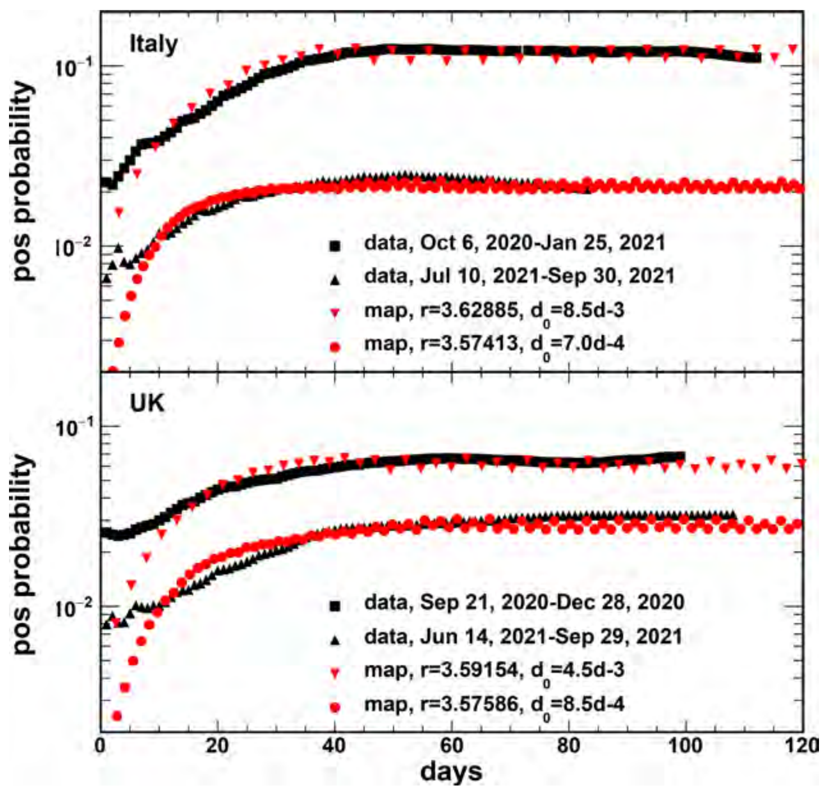


Figure 2. (Color online) The comparison between the logistic map results with the data in Italy (top panel) and UK (bottom panel) for the second wave in 2020 and the third wave in 2021.

as function of r with $\nu/2 = \frac{\ln\alpha}{\ln\delta}$, $\alpha = 2.502807$, $\delta = 4.6692016$ and $r_\infty = 3.569946$. And similarly for the Lyapunov exponent

$$\lambda = c_2\left(r - r_\infty\right)^\beta \tag{7}$$

where $\beta = \frac{\ln 2}{\ln\delta}$. The values c_1 and c_2 can be fixed to their analytical values in the ergodic limit *i.e.*, $r = 4$.

From the number of positives we know the value of the asymptotic probability which plays the role of d_∞ and it is the same as for the logistic map, thus we can extract the corresponding value of r from Equation (6) and, knowing r we obtain the Lyapunov exponent for the map Equation (7). From the ratio of the Lyapunov exponent for the map and the data we can make the transformation from the number of iterations to time, Equation (5). In **Table 1** we have collected

Table 1. The r and λ calculated from Equations (6) and (7) for **Figure 1** and **Figure S1**.

Country	pos/dec	d_∞	r	λ	$r - r_\infty$
Italy	pos	0.0689	3.592	0.182	0.0219
	dec	0.00225	3.570	0.0137	0.000074
UK	pos	0.0731	3.594	0.190	0.0242
	dec	0.00198	3.57	0.0125	0.000054
Sweden	pos	0.108	3.617	0.256	0.0468
	dec	0.00148	3.570	0.00100	0.000034
Norway	pos	0.0271	3.575	0.0899	0.00458
	dec	0.000166	3.570	0.00191	4×10^{-6}
Denmark	pos	0.0493	3.582	0.141	0.0125
	dec	0.000435	3.570	0.00396	4×10^{-6}
South Korea	pos	0.00634	3.570	0.0300	0.000404
	dec	4.138×10^{-5}	3.570	0.000669	4×10^{-6}
Mongolia	pos	0.0974	3.609	0.236	0.0392
	dec	0.000270	3.570	0.00276	4×10^{-6}
USA	pos	0.118	3.624	0.273	0.0543
	dec	0.00237	3.570	0.0142	0.0000764
Canada	pos	0.0413	3.579	0.124	0.00929
	dec	0.000725	3.570	0.00582	0.0000104
Russia	pos	0.0470	3.581	0.136	0.0115
	dec	0.00143	3.570	0.00971	0.0000326
San Marino	pos	0.158	3.658	0.340	0.0880
	dec	0.00271	3.570	0.0158	0.0000955

some results for different countries. Large values for both, Lyapunov exponent and d_∞, are obtained by Sweden which opted for herd immunization. Fortunately, the low population density and good health organization decreased the possible catastrophic effect of such a choice as we saw in the UK which were following the same strategy at the beginning of the pandemic [5] [6] [7] [8]. In the last column of **Table 1** we report the value of $r - r_\infty$. If this quantiy becomes zero or negative then we have reached "herd-immunization". As we see from the table, no country gets such values but it gets very close to it especially for the deaths probabilities which are the most important quantity. The combined action of vaccine [11] and new effective medicines which reduces mortality may finally take us to herd-immunization. People may still get COVID-19 but will not die because of it!

2. Controlling Chaos

We have discussed in the introduction the connection between the control parameter r of the logistic map with λ and d_∞. In order to have $r < r_c$ we can decrease the value of d_∞ through lockdowns, masks, tracing and vaccinations. In particular we can write:

$$\text{Ratio} = d_\infty^1 / d_\infty^0 = a_0 - a_1\chi + a_2/2\,\chi^2 + \cdots, \quad (8)$$

where, d_∞^1 is the current value of the probability while d_∞^0 is the same quantity but exactly in the same period of the year before. χ is the controlling factor which we assume equal to

$$\chi = \left(N^v + N^p\right)/\text{population}. \quad (9)$$

N^v and N^p are the number of vaccinated and positive to COVID-19 respectively, since $N^p \ll N^v$ it can be neglected for all practical estimations. The expansion in Equation (8) may need higher order terms because of lost of efficacity of the vaccine, of previous infections and/or new variants. Unvaccinated people contribute to Equation (9) if they get infected, thus another reason to get vaccinated as soon as possible. Also vaccinated people may be justified not wearing masks in presence of unvaccinated by choice, this may spread the virus faster when it is harmless to vaccinated and there are medicines which can save the life of the unvaccinated, keeping in mind that 2% - 3% of the positives will die, see **Figure 1**. Future, more deadly variants may develop thus the importance to stop the epidemic as quickly as possible. In **Figure 3** we plot the probabilities as function of the vaccination ratio χ. The decrease is very encouraging but it maybe misleading since could be due to lockdowns or simply to a major awareness from the population. A more suitable quantity may be given by the ratio of probabilities or cases in two different periods, see Equation (8).

From Equation (8) we can define the probability ratios for the year 2021 divided the same period in 2020. Since, in Italy for instance, test data was taken starting Feb 24, 2020 that is the first day we can build the ratios (for the UK, US and Russia are Mar 31, 2020, Mar 1, 2020 and Mar 4, 2020 respectively). This ratio is plotted in Figure 4 as function of the parameter χ defined in Equation (9). The divergence for low vaccination rates is due to the low statistics at the very beginning of the pandemics when the virus was starting to diffuse among the

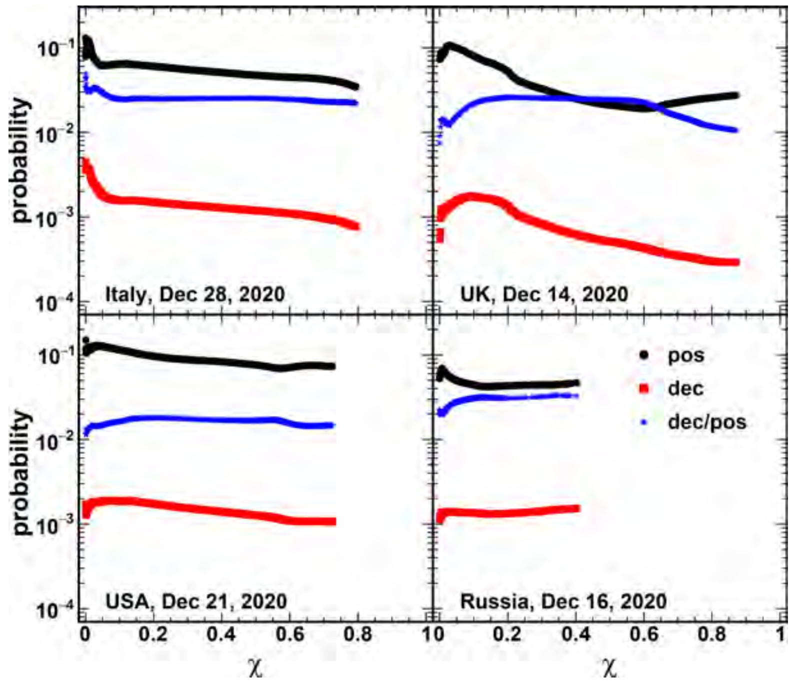


Figure 3. (Color online) The probabilities (positive, deceased and deceased/positive) versus the vaccination ratio χ for different countries as indicated. The date of the first day of vaccination has been indicated in the figure for each country.

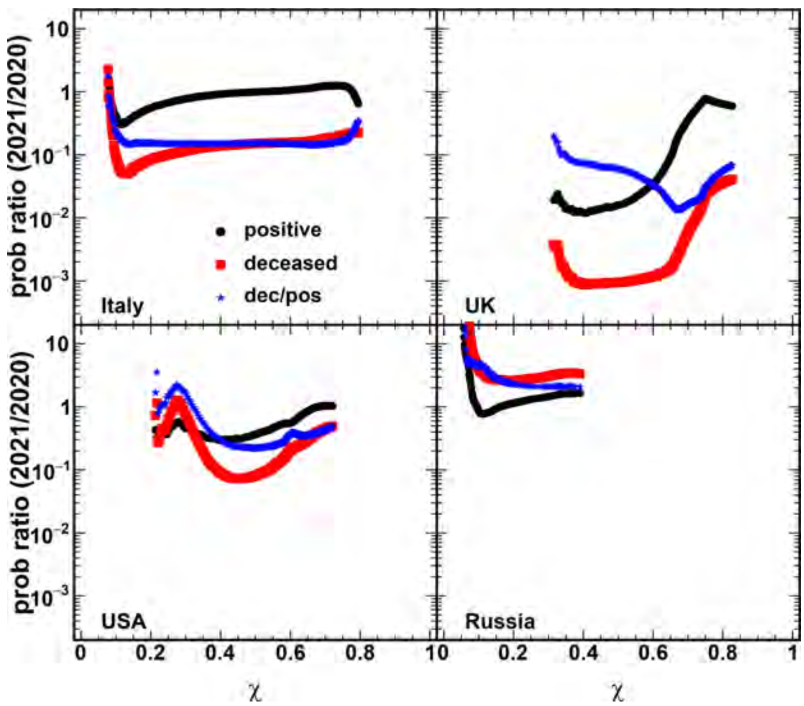


Figure 4. (Color online) The probability ratio for the same period in 2021 and 2020 versus the vaccination ratio χ for each country.

people. From the expansion, Equation (8), we should get an initial decrease followed by higher order corrections that could be positive or negative. If 75% of the population is vaccinated, we expect the number of cases to decrease of a factor at least 4 as respect to the year before. The figure displays a fast decrease at the beginning of the vaccinations followed by a later increase due most probably to the Delta variant and the release of strong restrictions and lockdowns. The fact that the ratio for positives is about 1 for large χ values indicates that herd immunity cannot be reached. On the other hand the ratio for deceased is about 20% for the US and Italy and 4% for the UK indicating that vaccines are working in preventing deaths but with different efficacities. As we noticed above, the number of deaths should be compared to the number of positives and not to the number of tests, thus in the same figure we include this ratio with blue symbols. Now the results are comparable to the deceased probabilities as expected with the exception of Russia. The important lesson to take from this plot is that the vaccines are working and in particular the approach used in the UK seems to be the most effective. A striking difference is given by Russia, which follows the same trend as other countries but the ratios are larger than one and the deceased are above the positives, **Figure 4**. This suggests that an important epidemic wave is underway in Russia more important than the previous year together with the resistance of people to get vaccinated. We had noticed in previous works [7] that in the year 2020 while the other countries were "under siege" from the virus, Russia was relatively free from it. The general behavior as function of vaccination confirms the quality of the vaccine used, in any case we can test this by analyzing cases from the Republic of San Marino (RSM), which adopted the same vaccine as Russia. **Figure 4** shows an increase in the ratio with increasing vaccination rate. We have attempted fits using Equation (8) adding terms in the expansion up to 9^{th} order. Those fits are quite bad showing strong oscillations thus confirming once more that we are nowhere near herd-immunity. This is not going in the direction of herd immunity but it could be an artifact due to the different number of tests in the different periods. To explore this feature we will discuss below the ratios of the number of cases in the two periods.

RSM does not provide the same data as for the cases in **Figure 4** but we can make the ratio of the number of deceased for the same period in the two years under study. In **Figure 5** (top panel) we plot the ratios as function of χ for different countries (top panel). We have added Canada as well since it has a large number of vaccinations. We also expect very many similarities between the UK and Canada, the weather, social organization etc., and we wanted to test if the different vaccines used in the two countries give different results. USA and Canada used the same vaccines (Moderna, Pfizer and Johnson&Johnson) and their behavior is very similar but slightly shifted. The RSM has lower statistics, thus higher fluctuations, but it plateaus at the lower values of the previous two countries, thus the Sputnik-V vaccine is as effective as the previous ones. The UK case, AstraZeneca vaccine used especially at the beginning of the vaccination

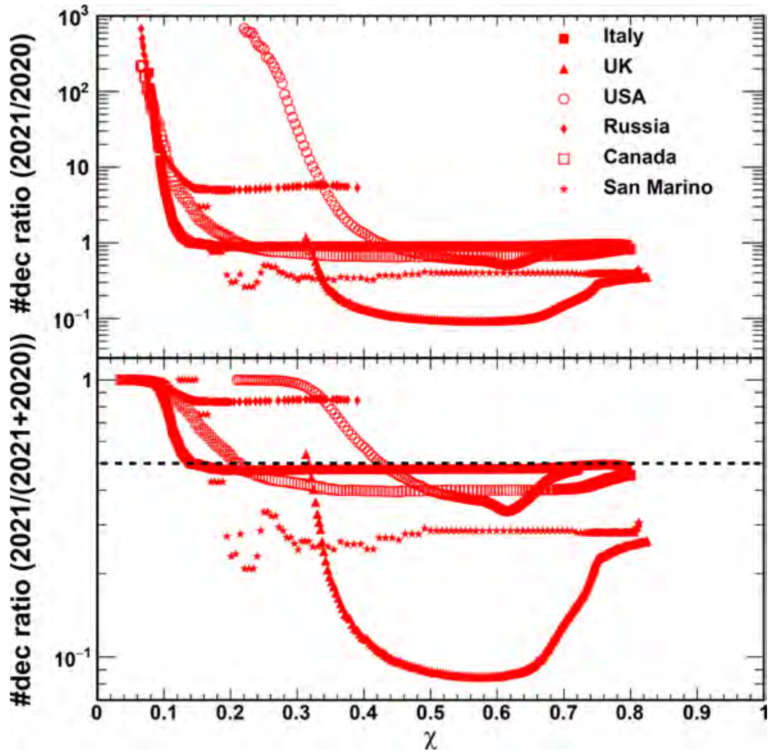


Figure 5. (Color online) (Top) Deceased case ratio (2021/2020) and (Bottom) Deceased case ratio (2021/(2020 + 2021)) of the same time period in 2021 and 2020 versus vaccination ratio χ for the countries indicated.

campaign, gives much lower ratios thus much more effective than the other vaccines. These cases confirm the results discussed for **Figure 4**, the different observables are consistent. Italy is slightly above the other countries but we believe it is due to the different ways of counting the deceased. A careful investigation should be performed. Other vaccines are used worldwide such as Sinovac, Novavax. However we have not found any country using these vaccines and providing complete data. In the supplement we discuss some partial results from the incomplete data available from Mongolia (not the ratios), which seem to confirm the quality of these vaccines as compared to the others.

As we discussed above, the divergence seen at low χ is due to low statistics data at the beginning of the pandemics, *i.e.*, early 2020. In order to compare different responses on the same scale we define the ratio of the period 2021/(2020 + 2021). In this case if the number of deaths in the period 2021 is much larger than the corresponding period in 2020, the ratio converges to 1. This is seen in **Figure 5** (bottom panel) for very low χ and it is due to statistics at the beginning of the pandemics. If the vaccines are not working, because of a new variant or loss of efficacity with time, then the ratio should converge to 0.5, dashed line in **Figure 5** (bottom panel). The latter is observed for Italy, USA and Russia (larger than 0.5). Canada seems to be approaching the 0.5 value while the UK and RSM are well below it meaning that the vaccines used in those countries are still effective. We notice the increase of the ratio for the UK at larger vaccination rates, corresponding to later times, which may be due to the choice of the UK government to use other vaccines instead of the AstraZeneca (AZ) for people below 40 years of age. This was due to the claim that the AZ vaccines may provoke blood clots. This claim was largely amplified by the press leading to many countries to halt or slow down the use of this particular vaccine. In the UK, 49 blood clots cases have been reported after 28.5 million AZ vaccines have been administrated, out of about 49 million vaccines administrated in the same period (Pfizer is the other one). This is about 2 cases per million people which is about double the value observed normally in the population, thus a negligible factor. Nevertheless many countries opposed it or slowed down its use (Germany for instance). Clearly the economical factor is huge, recall that the AZ vaccine was led by the university of Oxford-UK with the intent of creating a vaccines readily available to all for a low price, about $2, *i.e.*, an order of magnitude less than other more "popular" vaccines such as Pfizer and Moderna, see **Table 2**. Our results show a quite different scenario about the vaccines efficacity, thus while rich countries should continue their economical games, "third world" countries can confidently rely on the AZ vaccine. One question remains is why prices are so much different.

A comparison between **Figure 4** and **Figure 5** confirms that herd immunity may be not reached anytime soon with the notable exception of the ratio of the number of deceased for the UK and RSM. This may suggest further vaccinations for the other countries and maybe we should consider a wider usage of the AstraZeneca vaccine as confirmed by the different observables used in this work.

3. Conclusion

In this work we have explored the similarities and differences of COVID-19 with the Spanish flu of 1918 using the logistic map as guidance. We have made some predictions on the number of cases and estimated the progress reached through vaccinations. We have shown that herd immunity cannot be reached because of

Table 2. The prices of the vaccines, data are taken from ref. [12].

Vaccine name	Type	Price (dollar/dose)	total doses
Pfizer-BioNTech	mRNA	19.5	2
Moderna	mRNA	25 - 37	2
AstraZeneca	Adenovirus-based	2.15 (EU), 3 - 4 (UK), 5.25 (South Africa)	2
Johnson & Johnson	Adenovirus-based	10	1
Russia"s Sputnik V Vaccine	Adenovirus-based	10	2
Sinovac Biotech	Inactivated SARS-CoV-2 virus	29.75	2
Novavax	Protein-based	16	2
CanSino Biologics	Viral vector	unknown	1
Bharat Biotech	Inactivated SARS-CoV-2 virus	2	2

the Delta variant. However, the death probability has been largely reduced and it affects almost exclusively the unvaccinated. A comparison of different countries using different vaccine types and different policies to contrast the pandemic shows a larger efficacity for the UK, Norway, S. Korea and other countries. Cases are bouncing back (especially the positives) to the values of the previous year thus suggesting a loss of vaccines efficacity and the possibility for a third dose. Countries like the UK and the Republic of San Marino seem still doing well thus may be reasonable to offer to the population a third dose with the AstraZeneca or the Sputnik-V vaccines, given a green light from the doctors for each case. Also, suitable blood test investigations for the vaccinated people should be performed before administrating the third dose to exclude people who have enough antibodies. A statistics of cases for different vaccine types can help decide which route to follow to produce new vaccines able to confront possible new and more deadly variants if any. At this stage it seems to us not useful to implement further actions to force people to get vaccinated if they do not want to do that. The most important action in those cases is information. People that get positive to COVID have 1 probability out of 50 to die. Since about 70% - 80% of the positives are vaccinated and their probability to die is close to zero, it means that the unvaccinated positive to COVID has probability 1 out of 10 - 15 to die. Furthermore, given the high probabilities to be positive to the virus (about 20% - 30% of the tested), sooner or later we may become positive and it is better to be ready, *i.e.* vaccinated, to reduce the risk practically to zero to die.

Acknowledgments

This work was supported partly by the National Natural Science Foundation of China (Grant Nos. 11905120 and 11947416) and by the U.S. Department of Energy under Grant No. DE-FG03-93ER40773 and NNSA Grant No. DENA0003841 (CENTAUR).

"Note added in proof: extending the analysis of this paper to December 21, 2021 did not show any relevant changes due to the Omicron variant. Updated figures are available upon request from the authors."

Conflicts of Interest

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

References

[1] Sabbatini, S., *et al.* (2007) *Le Infezioni in Medicina*, **15**, 272-285.

[2] Turcotte, D.L. (1992) Fractals and Chaos in Geology and Geophysics. Cambridge University Press, Cambridge.

[3] Schuster, H.G. (1995) Deterministic Chaos (VCH). Wiley, New Jersey. https://www.amazon.com/Deterministic-Chaos-Heinz-Georg-Schuster/dp/3527293159

[4] Baran, V., Zus, M., Bonasera, A. and Paturca, A. (2015) *Romanian Journal of Phys-*

ics, **60**, 1263.

[5] Bonasera, A. and Zhang, S. (2020) *Frontiers of Physics*, **8**, 171. https://doi.org/10.3389/fphy.2020.00171

[6] Bonasera, A., Bonasera, G. and Zhang, S. (2020) *The European Physical Journal Plus*, **135**, 453. https://doi.org/10.1140/epjp/s13360-020-00494-6

[7] Zheng, H. and Bonasera, A. (2020) *The European Physical Journal Plus*, **135**, 799. https://doi.org/10.1140/epjp/s13360-020-00811-z

[8] Zheng, H. and Bonasera, A. (2020) Aggressive COVID-19 "Second Wave" in Italy. https://www.medrxiv.org/content/10.1101/2020.11.11.20229872v2 https://doi.org/10.1101/2020.11.11.20229872

[9] https://ourworldindata.org/grapher/full-list-total-tests-for-covid-19

[10] https://covid19.who.int/WHO-COVID-19-global-data.csv

[11] Feng, S., *et al.* (2021) *Nature Medicine*, **27**, 2032.

[12] https://www.biospace.com/article/comparing-covid-19-vaccines-pfizer-biontech-moderna-astrazeneca-oxford-j-and-j-russia-s-sputnik-v/

[13] Ledberg, A. (2021) *Frontiers in Public Health*, **9**, 579948. https://doi.org/10.3389/fpubh.2021.579948

Appendix. Supplemental Material

In this supplement we discuss different countries, which adopted different strategies to contrast COVID-19. Sweden was one of the few countries letting the disease to spread freely and as a result had the largest number of fatalities as compared to nearby countries like Norway and Denmark, see **Figure S1**. When divided by the population, their death ratio is a factor 10 larger than Norway shedding some bad light on the choice of herd immunization. Probably that is also the reason why the Sweden data is incomplete, see the bottom part of **Figure S1**. This is the worst pandemic for Sweden since the S1918 [13]. S. Korea adopted the strategy of tracking, attention to hygiene and masks decreasing the fatality rate of a factor 2 respect to Norway. Thus there are better methods to contrast the virus even without vaccine and these are important lessons for future pandemics.

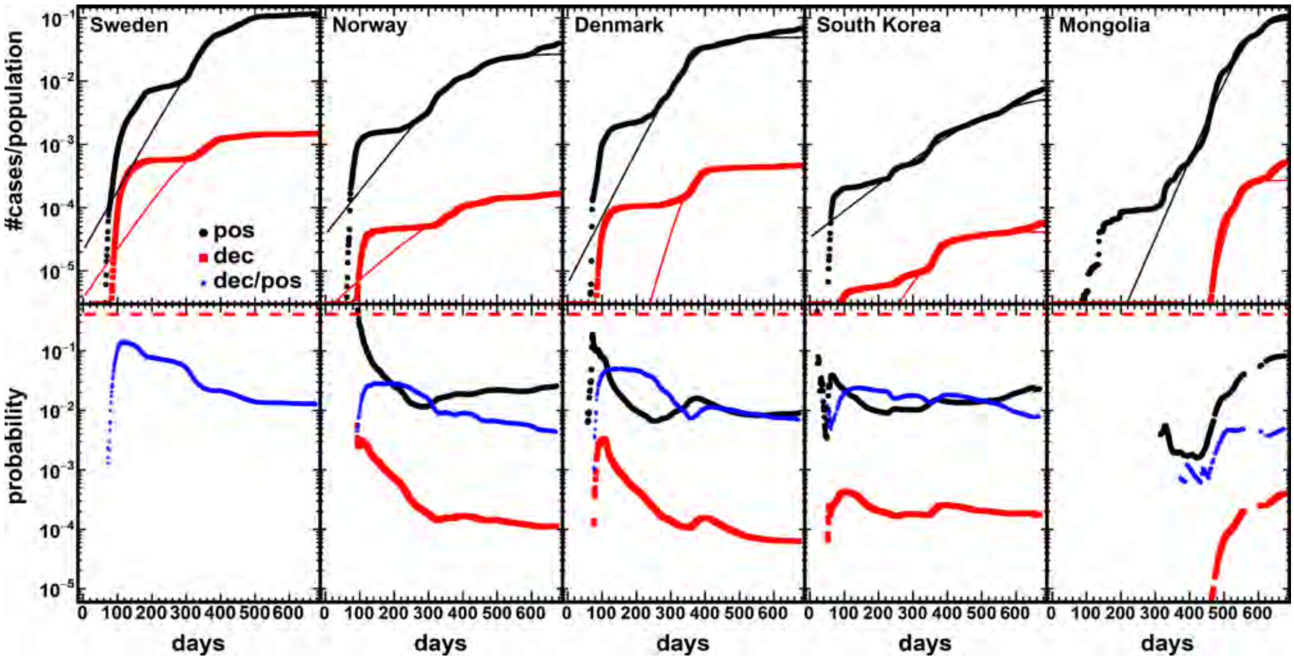


Figure S1. (Color online) Same as **Figure 1** for different countries. Updated to November, 2021.

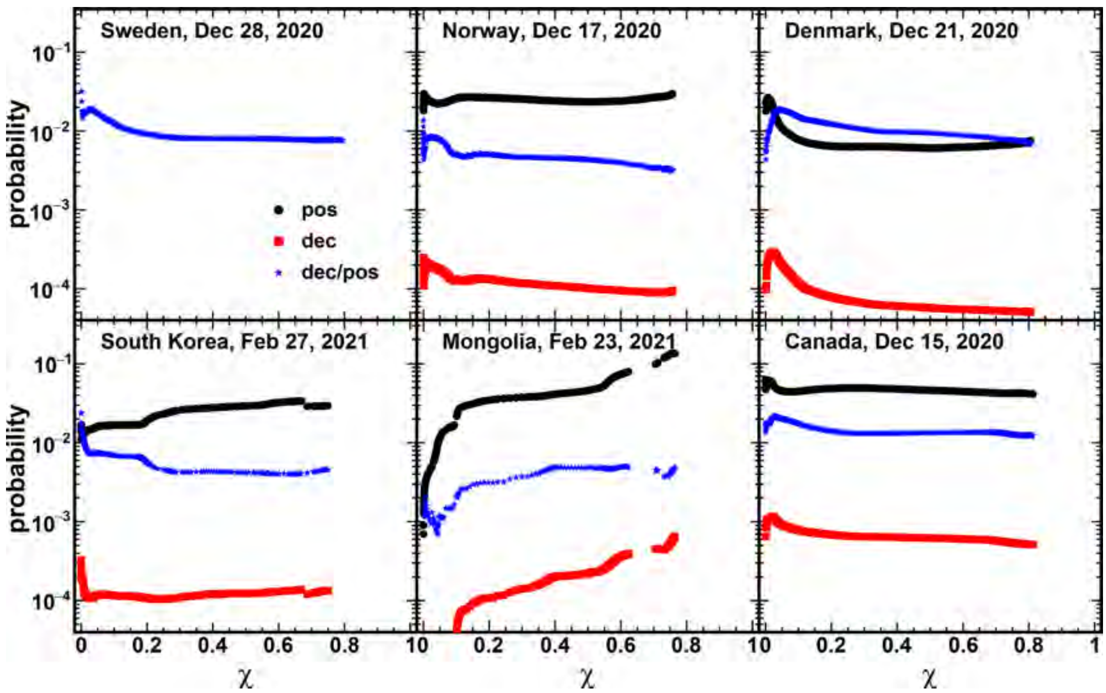


Figure S2. (Color online) The day 1 is the vaccination time which is shown in the legend. In **Figure S2** we plot the probabilities as function of vaccination rate for the same countries of **Figure S1**, compare to **Figure 3**.

Journal of Modern Physics, 2022, 13, 16-33
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Interaction between an Accelerated Mass in Straight Motion and a Hidden Energy Reservoir as a Strict Mathematical Consequence of Special Relativity

Andreas Trupp

Fachhochschule Muenster (University of Applied Science), Muenster, Germany
Email: andreas@andreastrupp.com

How to cite this paper: Trupp, A. (2022) Interaction between an Accelerated Mass in Straight Motion and a Hidden Energy Reservoir as a Strict Mathematical Consequence of Special Relativity. *Journal of Modern Physics*, **13**, 16-33.
https://doi.org/10.4236/jmp.2022.131002

Received: November 23, 2021
Accepted: January 17, 2022
Published: January 20, 2022

Abstract

A. Einstein and H.A. Lorentz had found that the mass of an accelerated body traveling at relativistic velocity appears to depend on whether the acceleration is performed in the direction of motion or in a transverse direction. E.P. Epstein rejected this result in the "Annalen der Physik"; he rather postulated an additional force that turns up when the body is accelerated in the longitudinal direction. It can be shown that the concept of an increased longitudinal mass is based on a simple mathematical error. When correcting this error, it turns out that Epstein's additional, hidden force is indispensable in order to avoid an inner inconsistency of Special Relativity. It does most of the total work absorbed by the moving object, and is thus responsible for most of the increase in its energy (=mass), given the speed attained is relativistic. In other words: While the total force on the body needed to maintain a constant acceleration a_0 is "$\left(1-v^2/c^2\right)^{-1} ma_0 = m_0\left(1-v^2/c^2\right)^{-3/2} a_0$", the technical force needed to maintain that acceleration amounts only to "$ma_0 = m_0\left(1-v^2/c^2\right)^{-1/2} a_0$". The total energy of two objects that undergo a symmetrical, elastic head-on collision is therefore not conserved during the collision, thus requiring the involvement of a hidden reservoir of energy. This result is confirmed by calculations that use the concept of momenergy. The phenomenon of an apparent disappearance of energy has been noticed in particle physics already (target-experiment), but its consequences have been ignored. Instead, an explanation has been given (reduced "energy of the center of mass") which is inconsistent and violates the relativity principle.

Keywords

Special Relativity, Longitudinal Mass, Transverse Mass, Momenergy, Target Experiment, Energy of the Center of Mass, Dark Energy

1. Introduction

1) In the past, Special Relativity has been subject to various attempts to modify it. Most attempts of a modification have also been an attempt of a refutation. This is because in those attempts the basic assumptions of Special Relativity, *i.e.*, the relativity principle and the invariance of the speed of light, have been cast into doubt as regards their postulated reach.

For instance, J. Wang [1] (pp. 1615-1644) recently revived the more-than-a-century-old idea according to which an ether could exist which "sticks" to heavy celestial bodies like a viscous and sticky fluid does to the surface of a spoon, so that no "ether wind" can be felt on their surfaces. As a consequence, the speed of light would not always be the same. At a location far away from earth where the "ether wind" could not be neglected, it would be different from the speed measured in experiments on earth. But since Special Relativity is based on the rigorous assumption of the speed of light being invariant for any observer without any exceptions, Special Relativity would not be a valid description of nature if the "viscous liquid" theory regarding an ether were true. However, experiments have been performed the outcome of which was unfavorable for the theory of a "viscous and sticky ether". The first of these experiments was the Michelson-Gale experiment in 1925.

Another recent attempt of modifying Special Relativity was undertaken by N.H. Abramson [2] (pp. 471-478). He modified the relativistic factor k which appears in the Lorentz-transformation by inserting the term $M/(M+m)$ in front of v^2/c^2 as a factor, and by defining v as the speed of the center of gravity of the system of the two observers involved (who are thought to have mass). But when working backward from the so modified equations of the Lorentz transformation, at least one of the two basic (empirical) equations from which the Lorentz transformation is derived is altered. However, no empirical evidence is presented which would compel us to prefer the new basic equation(s) over the old one(s).

My article which I am presenting here is of a different kind. It is by no means an attempt to modify and refute the basic assumptions of Special Relativity. What it does is the following: it reveals some strict, physically surprising mathematical consequences of these basic assumptions, which have so far been overlooked. The reason for this failure lies in the fact that a parameter which turns up in one of the equations of Special Relativity has commonly been interpreted in a way which contradicts a presupposition. This inconsistency leads to mathematically wrong consequences. They can be found in all textbooks which mention the subject.

2) According to the Theory of Special Relativity, the mass of a body has increased when it has gathered speed. Mass is resistance offered against acceleration. One should expect that the resistance a moving body offers against accele-

ration does not depend on the direction in which the acceleration is performed. But Einstein had to realize that the equations of Special Relativity DO make a difference regarding the direction of acceleration: the resistance is stronger when the acceleration is performed in the direction of motion, and is weaker in a transverse direction. Einstein therefore distinguished between a transverse and a longitudinal mass of a body in motion. He thus accepted (as did most physicists after him) the strange consequence that the directions of an applied force and that of the resulting acceleration deviate from each other. Whether or not this is justified shall be scrutinized in this article.

3) The nomenclature used is the following:

a = acceleration of the object in the lab's unprimed frame of reference

c = speed of light

E = total energy mc^2 of the moving object, which includes the energetic equivalent of its rest mass

E_{conv} = energy thought to be available for conversion when two particles collide with each other

F = force

h = parameter used to replace a complicated term

k = relativistic factor, equal to $\left(1-v^2/c^2\right)^{-1/2}$

m_0 = rest mass of an object

m = total mass of an object (rest mass plus relativistic mass)

p = linear momentum of an object in the unprimed frame of reference

s = distance (in the unprimed frame of reference) covered by a moving object

t = time in the unprimed reference frame in which an object is moving

tau = proper time of a moving object

u' = velocity of an object measured in the primed reference frame (rest frame) of an observer; the primed frame is moving in the unprimed reference frame at a velocity v

u = velocity of an object in the unprimed reference frame; the object is not necessarily at rest in the primed frame, but can be in motion in the primed frame, its velocity there being u'

v = velocity at which the primed frame itself or an object at rest in the primed frame is moving in the unprimed reference frame

W = work done (in the unprimed frame of reference) on an object when accelerating it

x = direction in which an object is moving in the unprimed frame of reference

y = direction perpendicular to the x- and the z-direction

z = direction perpendicular to the x- and the y-direction

2. The Distinction between Transverse and Longitudinal Mass

In case a body—moving in straight and steady motion with a velocity v_x along the positive x-axis of an unprimed coordinate system—is becoming subject to an accelerating force F_x in the positive x-direction, that force is (according to Special Relativity):

$$F_x=\frac{\mathrm{d}p_x}{\mathrm{d}t}=\frac{\mathrm{d}\left[m(t)v_x(t)\right]}{\mathrm{d}t}=\frac{\mathrm{d}\left[k(t)m_0v_x(t)\right]}{\mathrm{d}t}=\frac{\mathrm{d}\left[\dfrac{m_0v_x(t)}{\sqrt{1-v_x^2(t)/c^2}}\right]}{\mathrm{d}t}$$
$$=\left[\frac{m_0}{\left(1-v_x^2(t)/c^2\right)^{1/2}}+\frac{m_0v_x^2}{c^2\left(1-v_x^2(t)/c^2\right)^{3/2}}\right]a_x=\frac{m_0}{\left(1-v_x^2(t)/c^2\right)^{3/2}}a_x \quad (1)$$

The term p denotes the momentum of the body. The term m stands for the time-dependent relativistic (total) mass of that body, which is equal to the product of its rest mass m_0 and the relativistic factor $k=\left(1-v^2/c^2\right)^{-1/2}$. In purely mathematical terms, a presupposition made in Equation (1) is:

$$m=\frac{m_0}{\sqrt{1-v_x^2/c^2}} \quad (1a)$$

(In order to avoid confusion in connection with historical references, I will use the somewhat old-fashioned terms "rest mass" m_0 for objects at $v = 0$, and "mass" m for the total mass of an object, which, because of $m = E/c^2$, depends on rest mass and its kinetic energy.) The term $a = \mathrm{d}v/\mathrm{d}t$ denotes the acceleration of the body in the unprimed frame of reference. The sum that turns up in the second line is the result of an application of the quotient rule in differential calculus.

One realizes: when looking at the right-hand side of the second line of Equation (1), the total mass m of the body appears to be equal to $m_0\left(1-v^2/c^2\right)^{-3/2}$, and not $m_0\left(1-v^2/c^2\right)^{-1/2}$ as was presupposed. This appears to be an inner inconsistency, which, fortunately, will dissolve further below.

In case an acceleration is taking place only in a direction transverse to the original x-direction of motion, namely in the y-direction, we find for the force:

$$F_y=\frac{\mathrm{d}p_y}{\mathrm{d}t}=\frac{\mathrm{d}\left[mv_y(t)\right]}{\mathrm{d}t}=\frac{\mathrm{d}\left[km_0v_y(t)\right]}{\mathrm{d}t}=\frac{\mathrm{d}\left[\dfrac{m_0v_y(t)}{\sqrt{1-v_{total}^2/c^2}}\right]}{\mathrm{d}t}=\frac{m_0a_y}{\sqrt{1-v_{total}^2/c^2}} \quad (2)$$

The result falls short of what is gained for the x-component. The relativistic total mass m of the body and hence also the relativistic factor k do not depend on time here. This is because none of the terms that show up on the right-hand side of the equation depend on time. Although the total velocity v_{total} does increase a bit if the body is undergoing an accelation in the y-direction, this increase in total velocity vanishes because motion and acceleration are at right angle with respect to each other.

The same is true when it comes to an acceleration in the z-direction.

3. The Mass-Energy-Equivalence Principle

As regards the work W done when accelerating the body at a constant rate a_x

(assuming the validity of $v = at$ and of $s = 1/2at^2$), we get because of Equation (1):

$$\begin{aligned} \mathrm{d}W = F\mathrm{d}s &= \frac{m_0}{\left(1-v^2/c^2\right)^{3/2}} a_x \mathrm{d}\left(\frac{1}{2}a_x t^2\right) = \frac{m_0}{\left(1-v^2/c^2\right)^{3/2}} a_x \mathrm{d}\left(\frac{1}{2}a_x \frac{v^2}{a_x^2}\right) \\ &= \frac{m_0}{\left(1-v^2/c^2\right)^{3/2}} \frac{1}{2}\left(v\mathrm{d}v + v\mathrm{d}v\right) = \frac{m_0}{\left(1-v^2/c^2\right)^{3/2}} v\mathrm{d}v \end{aligned} \tag{3}$$

and therefore:

$$W = \int_0^{W(v)} \mathrm{d}W = \int_0^{s(v)} F\mathrm{d}s = \int_0^{v} \frac{m_0}{\left(1-v^2/c^2\right)^{3/2}} v\mathrm{d}v = \frac{m_0}{\left(1-v^2/c^2\right)^{1/2}} c^2 - m_0 c^2 \tag{4}$$

This equation—which constitutes the mass-energy-equivalence theorem—was set up by A. Einstein [3] (section 10) in 1905.

For a non-relativistic velocity of $v = 0.01$ (expressed in dimensionless units of a fraction of c), Equation (4) yields a work of 0.00005000375 per unit of rest mass, whereas Newton's law yields a work of exactly 0.00005.

It should be noted that the mass-energy-equivalence principle holds true on the basis of $\left(1-v^2/c^2\right)^{-3/2} m_0 a$ being the total force (exerted on the body in order to accelerate it), and would not hold true if the total force were just $\left(1-v^2/c^2\right)^{-1/2} m_0 a$.

The total energy of a moving body is obtained by adding the energy m_0c^2 of the rest mass to W. We then get:

$$E = W + m_0 c^2 = \frac{m_0}{\left(1-v^2/c^2\right)^{1/2}} c^2 \tag{4a}$$

4. R.C. Tolman's Contribution to the Discussion

1) The discrepancy between an acceleration in a longitudinal and in a transverse direction was recognized by A. Einstein [and, prior to that, by H.A. Lorentz [4] (Section 9)]. Einstein (see Ref. 3, Section 10) therefore distinguished between a "transverse" and a "longitudinal mass" (so did H.A. Lorentz). P.G. Bergmann [5] (Chapter VI, Eq. 6.70 and 6.71, p. 103/104) expressed this finding as follows:

> "The coefficients of the acceleration on the right-hand sides of eqs. (6.70) and (6.71) are occasionally referred to as 'longitudinal mass' and 'transversal mass', respectively."

N. Dragon [6] (Chapter 3.4, p. 53) recently wrote:

> "Inertia of fast particles is direction-dependent. In case the force acts at right angle with respect to v, the acceleration is $\mathrm{d}v/\mathrm{d}t = \left(1-v^2\right)^{1/2} F/m$; in the direction of the velocity the inertia of the particle is increased by the factor $1/\left(1-v^2\right)$."

The conflict between $m = m_0\left(1-v^2/c^2\right)^{-1/2}$ and $m = m_0\left(1-v^2/c^2\right)^{-3/2}$ was highligted by R.C. Tolman [7] (pp. 375-380) in 1912:

"On the basis of these accelerations, it has been usual to place the 'transverse mass' of a body moving with the velocity u as equal to $m_0\left(1-u^2/c^2\right)^{-1/2}$ and its 'longitudinal' mass as equal to $m_0\left(1-u^2/c^2\right)^{-3/2}$, where m_0 is the mass of the body at rest and c is the velocity of light. If, however, mass is a quantity to which a conservation law applies, the mass of a body cannot well be different in different directions; and it has been believed by Professor Lewis and the writer, that in general, without respect to direction, the expression $m_0\left(1-u^2/c^2\right)^{-1/2}$ is best suited for the mass of a moving body."

Tolman set out from the following equation:

$$-\left(m_1+m_2\right)v=m_1u_1+m_2u_2=m_1\frac{u'-v}{1-u'v/c^2}+m_2\frac{-u'-v}{1+u'v/c^2} \tag{5}$$

The terms m_1 and m_2 are the relativistic total masses of two bodies—that have equal mass when at rest—at a picked moment in time. At that very moment of interest, we imagine the two perfectly elastic masses to be in the process of head-on collision, and, since they are, at that moment, in contact with each other, they are decelerating. In the primed frame of reference, one of the two bodies has the momentary velocity $+u'$, and the other one has the momentary velocity $-u'$. In the unprimed frame of reference, the primed frame of reference moves to the left at constant speed v. If assuming that the sum of the two masses (=energies), viewed in the unprimed frame of reference, stays the same over time (kinetic energy is converted into what Tolman calls "elastic energy", and vice versa), the total momentum of the two bodies in the unprimed frame of reference must be $-(m_1 + m_2)v$ at the (somewhat later) moment when the two colliding objects will have come to rest in the primed frame of reference. Because of the assumed invariance of the sum of the two masses, this statement is valid despite the fact that m_1 and m_2 refer to the masses as they present themselves at a different moment in time (namely a moment *prior* to coming to relative rest). Because of an assumed conservation of momentum, $-(m_1 + m_2)v$ must equal $m_1u_1 + m_2u_2$, that is, the total momentum (sum of momenta) at the moment of deceleration. Finally, u_1 and u_2 are replaced in accordance with the relativistic theorem of addition of velocities.

By mere algebra, Tolman transforms his equation into:

$$\frac{m_1}{m_2}=\frac{\left(1-u_2^2/c^2\right)^{1/2}}{\left(1-u_1^2/c^2\right)^{1/2}}$$

When imagining that u_1 is zero (so that $m_1 = m_0$, with m_0 being the mass of one of the two objects when it is at rest in the unprimed frame, but is being equipped with some elastic energy), this equation turns into:

$$m_2=\frac{m_0}{\left(1-u_2^2/c^2\right)^{1/2}} \tag{6}$$

Tolman thus obtained the result that the mass of a body in motion is, at any moment, equal to $m_0\left(1-v^2/c^2\right)^{-1/2}$, regardless of whether or not it is, in the moment considered, being accelerated or in steady motion.

But this does not bring us any further (even if Tolman's assumptions of the conservation of total mass and of total momentum were correct). Instead, it brings us to where we started, that is, to the presupposition regarding the mass m that we made in the first line of Equation (1).

2) Some years later, Tolman [8] (Chapter 6, Section 69, Eq. 68, pp. 76/77) seems to have abandoned his opinion. He now states that F is equal to $m_0\left(1-u^2/c^2\right)^{-3/2}\mathrm{d}u/\mathrm{d}t$, and tells his readers that $m_0\left(1-u^2/c^2\right)^{-3/2}$ "*is sometimes spoken of as the expression for the longitudinal mass of a particle*."

5. The Need for an Additional, Hidden Force (First Postulated by P.S. Epstein)

1) But given that the relationship $m=m_0\left(1-v^2/c^2\right)^{-1/2}$ is a presupposition made in Equation (1) and hence in the mass-energy-equivalence principle (Equation 4), the only way to avoid the mentioned inner inconsistency in Equation (1) and Equation (4) is the following: when a body is accelerated by means of an external, technical force, an additional, hidden force turns up. Then Equation (1) can be re-formulated as follows:

$$F_x=\frac{m_0}{\left(1-v_x^2/c^2\right)^{3/2}}a_x=\frac{m_0}{\left(1-v_x^2/c^2\right)\left(1-v_x^2/c^2\right)^{1/2}}a_x=\frac{1}{1-v_x^2/c^2}ma_x \tag{7}$$

Now m is equal to $m_0\left(1-v^2/c^2\right)^{-1/2}$ even in the longitudinal direction, exactly as was presupposed in (1). The factor $\left(1-v^2/c^2\right)^{-1}$ which turns up on the right-hand side of (7) is an expression of the additional force, as it tells by which factor the net force is enlarged (in comparison with $F=ma$). One should therefore distinguish between a total force

$$F_{total}=\frac{m_0}{\left(1-v_x^2/c^2\right)^{3/2}}a_x=\frac{1}{1-v_x^2/c^2}ma_x \tag{8}$$

and a technical force F_{tech} which we exert on the object by technical means:

$$F_{tech}=\frac{m_0}{\left(1-v_x^2/c^2\right)^{1/2}}a_x=ma_x \tag{9}$$

The difference between these two forces gives the additional, hidden force

$$\begin{aligned}F_{hidden}=F_{total}-F_{tech}&=\frac{m_0}{\left(1-v_x^2/c^2\right)^{3/2}}a_x-\frac{m_0}{\left(1-v_x^2/c^2\right)^{1/2}}a_x\\&=\frac{m_0}{\left(1-v_x^2/c^2\right)^{1/2}}a_x\left(\frac{1}{1-v_x^2/c^2}-1\right)\end{aligned} \tag{10}$$

2) It was P.S. Epstein [9] (pp. 779, 783/784) who postulated such a force explicitly in the "Annalen der Physik" (1911):

"It is therefore of great advantage to introduce an additional force, ... This, is, for instance, accomplished by a compensation force $f_x^k = -\beta^2 f_x$ acting in the *x*-direction, by virtue of which the *x*-component is weakened by a factor of $1-\beta^2$, … Moreover, one is thereby getting rid of the tensor mass; given that force exists, … the mass is now $m_0\left(1-v^2/c^2\right)^{-1/2}$ in all directions."

Though Epstein conceived of a force whose direction was opposed to that of the technical force applied, he would surely not have raised concerns against a force that would act in the same direction as the technical force, since the "compensation force" he mentioned was only meant as an example. One has to confess that his example was a bad one. The force he took as an example would neutralize the accelerating force to some extent. As a net result, F/a would not equal $m_0\left(1-v^2/c^2\right)^{-3/2}$ though this is what Equation (1) is saying. This would bring Relativity into trouble. Therefore the additional force must be of a different kind, with its direction not being opposed to the direction of the technical force applied, but sharing it. This is to be expanded in greater detail below.

6. Determination of the Nature of Epstein's Additional, Hidden Force, and of the Source of the Energy It Conveys

1) But if the mass of a moving body is $\left(1-v^2/c^2\right)^{-1/2}$ in all directions (regardless of whether or not it is subject to acceleration), then the additional, work-performing force (whose existence must then be acknowledged) cannot feed energetically from the same source as the technical force does: given the technical force F_{tech} needed to maintain a fixed acceleration a_0 is

$ma_0 = m_0\left(1-v^2/c^2\right)^{-1/2} a_0$—and is not larger—then it falls short of

$\left(1-v^2/c^2\right)^{-1} ma_0 = m_0\left(1-v^2/c^2\right)^{-3/2} a_0$, which is the *total* force needed to maintain a fixed acceleration a_0. A work-performing hidden force—energetically fed from a hidden reservoir—is hence coming to the aid of the technical force.

As a consequence, the technical *work* needed to bring the object to a velocity v only amounts to:

$$\begin{aligned} W_{tech} &= \int_0^{s(v)} F_{tech}\,\mathrm{d}s = \int_0^{s(v)} \frac{m_0}{\left(1-v^2/c^2\right)^{1/2}} a_0\,\mathrm{d}s = \int_0^{v} \frac{m_0}{\left(1-v^2/c^2\right)^{1/2}} v\,\mathrm{d}v \\ &= m_0c^2 - m_0c^2\left(1-v^2/c^2\right)^{1/2} \end{aligned} \tag{11}$$

For a non-relativistic velocity of $v = 0.01$ (expressed in dimensionless units of a fraction of c), Equation (11) yields an amount of work of 0.00005000126 per unit of rest mass. This falls short of what (4) is yielding, but exceeds the amount of work given by Newton's law by a small amount, quite as expected. The difference between the results of Equation (4) and Equation (11) (that is 0.00005000375 − 0.00005000126 = 0.00000000249, or 224.1 Newtonmeter per kg) is due to the work done by Epstein's hidden, additional force.

For a relativistic velocity that approaches c, Equation (4) yields an infinite amount of total work invested, whereas Equation (11) yields only m_0c^2 as the technical work spent on accelerating the object to nearly the speed of light. In that case, the hidden force has done much more work on the object than the technical force has, and is responsible for most of the increase in mass. Conversely, whenever a fast moving objects hits planet Earth, its kinetic energy cannot be larger than m_0c^2, and cannot possibly be above that (see below for further consequences in the field of Particle Physics).

From this follows: when a particle that has a rest mass m_0 and is traveling nearly at the speed of light is absorbed by a large, stationary system, the mass of that system thus increases by $2m_0$. When a photon (rather than an object or a particle that has a rest mass) is absorbed by that large, stationary system, the mass of that system increases by m. The difference between the two cases can be attributed to the fact that the photon's rest mass is missing and exists only virtually.

2) To summarize: we must recognize that relativity postulates the emergence of an additional, hidden force which comes to the aid of the technical force applied to a body in order to accelerate it. The energy of that hidden, work-performing force is fed from a hidden reservoir in space. The flow of that hidden energy has its sink in the positively accelerated body, as has the ordinary energy that is conveyed by the technical force (but whose source is in a battery or a similar device). The increase in strength of the gravitational field generated by the fast moving body is thus somewhat larger than it should be on the basis of the absorbed technical work alone. When the fast moving body is decelerated, the energetic process is reversed, and the relativistic mass of the object transforms into a flow of energy whose sink is partly in the hidden reservoir.

Epstein's hidden force thus doesn't do any technical work on material objects, but is, as has just been said, capable of strengthening the gravitational field generated by the moving object. It should also lead to an increase in the centrifugal force the moving mass is subject to if, after a linear acceleration, the electrically charged object is exposed to a Lorentz force which bends its trajectory. This is no indirect departure from the rule according to which the hidden force cannot do any technical work, since the centrifugal force is incapable of doing work (due to its direction being at right angle to the motion of the object).

Thereby both $F_{total}/a = m_0\left(1-v^2/c^2\right)^{-3/2}$ and $F_{tech}/a = m_0\left(1-v^2/c^2\right)^{-1/2}$ present themselves as valid descriptions of nature.

7. Neither Momenergy nor Energy Is Conserved during a Head-on Collision of Two Fast Moving Objects

One should note that the total relativistic mass (generated by the combined work-performing actions of the technical and the hidden force) is subject to change during an elastic head-on collision of two alike bodies: the relativistic mass or energy that was generated by Epstein's hidden force prior to the head-on collision is not converted into elastic energy (when the two colliding objects come to relative rest), but flows off into the unknown where it had come from. Otherwise F_{tech} would be identical with F_{total}.

In order to confirm this recognition, we turn our attention to momenergy (forgetting about collisions for a moment). Starting with Minkowski's line element of spacetime, we multiply both sides of Minkowski's equation by $m_0^2/\mathrm{d}\tau^2$ (with *tau* being proper time of the moving object). We then get:

$$m_0^2\frac{\mathrm{d}\tau^2}{\mathrm{d}\tau^2}=m_0^2\frac{\mathrm{d}t^2}{\mathrm{d}\tau^2}-m_0^2\frac{\mathrm{d}x^2}{c^2\mathrm{d}\tau^2}-m_0^2\frac{\mathrm{d}y^2}{c^2\mathrm{d}\tau^2}-m_0^2\frac{\mathrm{d}z^2}{c^2\mathrm{d}\tau^2} \tag{12}$$

As a next step, we replace $\mathrm{d}\tau^2$ by $\mathrm{d}t^2\left(1-v^2/c^2\right)$. This is permissible because the two point events at the ends of a temporal interval we are considering occur at the same spatial location in the reference frame of the moving object. We then get:

$$m_0^2=\frac{m_0^2}{1-v^2/c^2}-\frac{m_0^2v_x^2}{c^2\left(1-v^2/c^2\right)}-\frac{m_0^2v_y^2}{c^2\left(1-v^2/c^2\right)}-\frac{m_0^2v_z^2}{c^2\left(1-v^2/c^2\right)} \tag{13}$$

or (see Equation (4a)):

$$m_0^2=\frac{E^2}{c^4}-\left(\frac{p_x^2}{c^2}+\frac{p_y^2}{c^2}+\frac{p_z^2}{c^2}\right) \tag{14}$$

The term p denotes a momentum-component in the reference frame of the laboratory, E denotes the total energy of the moving body. The right-hand side of Equation (13) or Equation (14) is known as the "momenergy" of the moving body [see for instance: E.F. Taylor, J.A. Wheeler [10] (Chapter 7.7, Eq. 7.3, p. 211)].

Similar to Tolman's reflections, we now imagine two alike bodies of the same rest mass which travel on straight paths, strictly in the x-direction of a Cartesian system of coordinates, the first one in the positive x-direction at speed $+v_x$, and the other one in the negative x-direction at speed $-v_x$ (equal and opposite velocities). Near the origin of coordinates, the two objects undergo an elastic head-on collision with each other. As long as the two objects are in contact with each other, they first decelerate, come to a full stop ($v_x = 0$), and pick up speed again (strictly in the positive and negative x-directions). As Tolman remarked, the rest mass m_0 of each of the two bodies cannot be constant during the deceleration and acceleration, since both bodies are having what Tolman called "elastic energy" (see Reference 7, p. 377):

> "Since at this time [of rest with respect to each other] both bodies are moving with the velocity –v we might suppose that $m_1 + m_2$ equals $2m_0\left(1-v^2/c^2\right)^{-1/2}$. This is not the case, however, since the bodies now possess additional elastic energy beyond that which they possess when at rest and not in contact."

The amount of that "elastic energy" is a function of v in our case (with $0\le v<c$). This form of energy leads to an increase in the rest mass m_0 (of each of the two bodies), which, too, is therefore a function of v. Equation (13) thus

converts into:

$$m_0^2(v) = \frac{m_0^2(v)}{1 - v^2/c^2} - \frac{m_0^2(v)v_x^2}{c^2\left(1 - v^2/c^2\right)} - \frac{m_0^2(v)v_y^2}{c^2\left(1 - v^2/c^2\right)} - \frac{m_0^2(v)v_z^2}{c^2\left(1 - v^2/c^2\right)} \tag{15}$$

The equation applies to each of the two bodies separately, whose behaviour is symmetrical.

When assuming—for a short moment—that the energy of each of the two bodies (total mass m times c^2) stays constant during the collision even though its rest mass m_0 does not, we are setting up the following side-conditions:

$$\mathrm{d}v_y = 0, \quad \mathrm{d}v_z = 0, \quad \mathrm{d}\left[\frac{m_0^2(v_x)}{1 - v_x^2/c^2}\right] = \mathrm{d}E^2 = 0 \tag{16}$$

When forming the differential of Equation (15), we therefore get:

$$\mathrm{d}m_0^2 = 2m_0\mathrm{d}m_0 = -\mathrm{d}\left[\frac{m_0^2 v_x^2}{c^2\left(1 - v_x^2/c^2\right)}\right] \tag{17}$$

v_x shall now be considered as being a function of m_0. Re-arranging the last equation gives:

$$-2m_0 = \frac{\mathrm{d}\left[\frac{m_0^2 v_x^2}{c^2\left(1 - v_x^2/c^2\right)}\right]}{\mathrm{d}m_0} \quad \Rightarrow \quad \frac{v_x^2/c^2}{1 - v_x^2/c^2} = -1 \quad \Rightarrow \quad \frac{v_x^2}{c^2} \to \infty \tag{18}$$

To elucidate: Since $\mathrm{d}m_0^2/\mathrm{d}m_0 = 2m_0$, we know that $v^2/\left[c^2\left(1 - v^2/c^2\right)\right]$ must equal or approach -1. [When $v^2/\left(c^2\left(1 - v^2/c^2\right)\right)$ is substituted by the parameter $h[v(m_0)]$, Equation (18) converts into (using the product rule of differentiation): $-2 = 2h + h'm_0$. The variable m_0 doesn't have a fixed value during the collision, but can have any value within an interval. Moreover, the beginning (or end) of this interval can be as large or as small a number as desired, depending on how large or small a rest mass one chooses to start with. This makes sure that h can have no value other than −1.] But h can equal or approach −1 only when v is many, many times larger than c.

If we set up the additional side condition: $v < c$, we arrive at the conclusion that Equations (15), (16) and $v < c$ are incompatible with each other, and not all three of them can be valid in Special Relativity. Since we have no doubts concerning the physical validity of Equation (15) and of $v < c$, it is our assumption of $\mathrm{d}E^2 = 0$ (set up in Equation (16)) which we have to throw out. Explicitly, we arrive at:

$$\mathrm{d}E \neq 0 \tag{19}$$

We have thereby obtained a confirmation of our result according to which there is a flow of energy into the unknown when the two objects decelerate during their head-on collision, and a flow of energy from the unknown into the objects when they are gathering speed thereafter.

8. The Disappearance of Energy (Mass) during "Target Experiments" in Particle Physics

Particle physics has come across this phenomenon, too, but has so far not been able to explain it.

1) Let us consider particle collisions. The energy available for conversion in any collision (which includes the rest mass of the particle involved) shall be called E_{conv}. It is believed to be (c is set to unity and does not appear in the equation):

$$E_{conv} = \sqrt{\left(E_1 + E_2\right)^2 - \left(p_{1_x} + p_{2_x}\right)^2 - \left(p_{1_y} + p_{2_y}\right)^2 - \left(p_{1_z} + p_{2_z}\right)^2} \\ = const \tag{20}$$

Equation (20) can be found in standard textbooks on particle physics like that of B. Povh, K. Rith, Ch. Scholz, F. Zetsche, W. Rodejohann [11] (Chapter 9, p. 123, 124) or that of Ch. Berger [12] (Equations 1.39 and 1.40, p. 15). The term E_{conv} on the left-hand side is considered (by particle physicists) as being an invariant. The invariance (if it exists) cannot be rooted in Equation (14). It is not an expression of the sum of the two momenergies of the two particles. This is because $\left(E_1 + E_2\right)^2$ which appears in the equation is not equal to $\left(E_1^2 + E_2^2\right)$. Likewise, $\left(p_1 + p_2\right)^2$ is not equal to $\left(p_1^2 + p_2^2\right)$. The reason why E_{conv} is nevertheless thought to be invariant is the following: Equation (20) can be considered as being an expression of the supposed conservation both of energy and of momentum in the unprimed reference frame of the lab.

So far, so good. However, there is no reason why the alleged invariant should be identical with the energy available for conversion. Only ($E_1 + E_2$) alone can be an expression of that energy. It contains both the relativistic mass and the rest mass of the two particles involved in a collision. The total momentum of the two particles neither increases nor decreases that amount of energy, since the energy which could correspond to a momentum is already included in E_1 and/or E_2. Hence there is no justification for adding $-\left(p_{1_x} + p_{2_x}\right)^2 - \left(p_{1_y} + p_{2_y}\right)^2 - \left(p_{1_z} + p_{2_z}\right)^2$ under the root sign. We will come back to this point soon.

2) In what is called a "colliding-beam experiment", two particles of the same rest mass are moving at velocities $+u$ and $-u$ (equal and opposite), and collide with each (just as in Tolman's thought experiment). For such a collision, Equation (20) gives:

$$E_{conv} = \sqrt{\left(E_1 + E_2\right)^2} = 2E \tag{21}$$

The energy available in that collision is $2E$. The error contained in Equation (20), *i.e.*, the existence of the summands $-\left(p_{1_x} + p_{2_x}\right)^2 - \left(p_{1_y} + p_{2_y}\right)^2 - \left(p_{1_z} + p_{2_z}\right)^2$, is neutralized here, simply because these summands are all zero.

But when one of the two particles (of equal mass) is at rest (this is the case in a so called "target experiment"), the energy available in such a collision according to Equation (20) is not $1E + 1m_0$, but is much lower in amount. For Equation (20) then gives:

$$\begin{aligned} E_{conv} &= \sqrt{\left(E_1 + m_{0_2}\right)^2 - p_{1_x}^2 - p_{1_y}^2 - p_{1_z}^2} \\ &= \sqrt{E_1^2 + 2E_1 m_{0_2} + m_{0_2}^2 - p_{1_x}^2 - p_{1_y}^2 - p_{1_z}^2} \\ &= \sqrt{E_1^2 + 2E_1 m_0 + m_0^2 - E_1^2 + m_0^2} \\ &= \sqrt{2E_1 m_0 + 2m_0^2} < 1E_1 + 1m_0 \quad \left(\text{for } E_1 > m_0, \textit{i.e.}, \text{for } u_1^2 > 0\right) \end{aligned} \tag{22}$$

The replacement of $-p_x^2 - p_y^2 - p_z^2$ by $-E^2 + m_0^2$ in Equation (22) is justified by Equation (14). That is to say: our relativistic Equation (14) (which describes momenergy) now comes into play. Hence, the energy thought to be available for conversion is only $\left(2Em_0 + 2m_0^2\right)^{1/2}$.

As an example, we set m_0 = 1 and E = 10,000 for a single (moving) particle. The energy available in a "colliding-beam experiment" thus is 20,000. For comparison, the energy available in the "target experiment" according to Equation (22) is only 141, much less than the energy of the particle which had been in motion before collision—and had had an energy of 10,000 then.

We realize: Equation (20)—and hence also Equation (22)—is the result of a desperate search for a way to account for observations made when a fast moving particle was hitting a bulk of stationary matter, with the particles generated by the impact having altogether much less energy than the impacting particle had. The sole purpose of the insertion of the summands $-\left(p_{1_x} + p_{2_x}\right)^2 - \left(p_{1_y} + p_{2_y}\right)^2 - \left(p_{1_z} + p_{2_z}\right)^2$ into Equation (20) has been to model the equation after the observed facts, and to reduce the yielded sum in case of a "target experiment". But doing an adaption like this is pure nonsense, as the subtraction of p^2 from E^2 is no physically meaningful method of determining the maximum energy available for conversation.

3) A common illustration of what is thought to be the deeper reason for Equation (22) is the following: it is believed that only what is called the "energy of the center of mass" is available for a conversion when it comes to a collision. Different from the situation in a "colliding-beam experiment" in which almost all sub-particles can be affected by a collision, only some of the sub-particles a proton is made up are believed to collide with each other when the collision of protons is a "target-collision", *i.e.*, a collision with a target which is stationary. The rest of the sub-particles is thought of as not being affected by the "target collision".

There is, however, an obvious objection to this explanation: the distinction between a "colliding-beam experiment" and a "target experiment" is relative. In the reference of one of the two collision partners (in which that partner is at rest), all experiments are "target experiments". In another frame of reference which has just the right velocity v, both the particle at rest with respect to the lab and the particle in motion with respect to the lab have equal and opposite velocities $+u'$ and $-u'$. That is to say: in that frame of reference, all experiments are "colliding-beam experiments". But whether or not a fraction—which can exceed 90 percent or more—of the sub-particles penetrate each other (just as if the protons were made of intangible matter) is an absolute phenomenon which cannot depend on the frame of reference chosen for a description. Insofar as the existence of an absolute phenomenon is said to stand and fall with a certain reference frame (that of the lab), the relativity principle is violated.

Moreover, one should note that both Equation (21) and Equation (22) should be valid even when it comes to a small velocity of the moving particle. On top of this, Equation (22) has no upper limit for the rest mass of the particles. They can be cannon balls or bowling balls. But when a cannon ball moving at a speed of 10 m/sec hits another cannon ball which is at rest, all sub-particles of the moving bowling ball come to rest, and are therefore affected by the collision without any exceptions. Hence, the true amount of energy available for conversion is $1E + 1m_0$, and not less.

4) Qualitatively, there is indeed an energetic difference between a symmetric collision ("colliding-beam experiment") and a target experiment: in a "colliding-beam experiment", energy flows from the two particles into a hidden reservoir when both particles are decelerating. However, when the resulting particles re-accelerate thereafter, the flow of energy is reversed, and energy from the hidden reservoir returns to the particles. The energy available for conversion can thus be as high as $2E$. In a "target experiment", energy leaves the decelerating particle for the hidden reservoir. If the "target experiment" leads to a full stop of the moving particle without accelerating any particle other than in the form of thermal motion, the total energy available for a conversion into heat cannot be larger than $2m_0$ (see above).

9. The Directions of Applied Force and Resulting Acceleration

As a consequence of the inconsistent assumption according to which the inert mass of a fast moving body depends on direction, authors usually arrive at the (wrong) conclusion that an acceleration of the body does not point in the same direction as the (technical) force that caused that acceleration. If, for instance, the force vector forms an angle of 45 with the velocity vector, the acceleration vector forms an angle whose size is between 45 and 90. See for instance: N. Dragon (Reference 6, p. 53):

> "In general, acceleration does not point in the direction of the applied force, ..."

Since common opinion has so far refrained from the idea that a hidden energy can be in possession of an inert mass so big that it is capable of bringing about a correct counter-force, the assumed non-coincidence of the two vectors must present a serious challenge to the principle of force and counter-force (and also to momentum conservation).

This challenge has so far been ignored. It can be met in the following way: presupposing (for reason of simplicity) the longitudinal component $F_{tech\text{-}x} = m_0\left(1-v^2/c^2\right)^{-1/2} a_x$ of the applied technical force is equal in magnitude to its *transverse* component by arrangement, the longitudinal component of the acceleration a_x is as large as the transverse component of *a*. This follows from Equation (2), according to which the transverse component of the force is $F_{tech\text{-}y} = m_0 \left(1 - v^2/c^2\right)^{-1/2} a_y$. Hence the directions of the two vectors F_{tech} and *a* are one and the same, quite as they should.

10. Confirmation of Epstein's Hidden Force by Similar Cases in Electromagnetism, Revealed by the Recognition That the Principle of Force and Counter-Force Is a Local One

Surprisingly, the existence of a hidden force of that kind is confirmed by very simple reflections on some everyday phenomena:

Let us conceive of two equal spheres charged with electricity of the same sign. The two stationary spheres sit at a distance of 1 meter apart from each other. The electric field *E* of each of the two spheres then exerts a repulsive electrostatic force on the other. At first sight, this alone appears to meet the principle of force and counter-force: a force felt by the first sphere is accompanied by a force in the opposite direction felt by the second sphere.

But this is not enough for the principle of force and counter-force to be observed. This is because that principle (like the principle of energy conservation) is a local one: the counter-force must turn up at the same location at which the force itself is attacking. This requirement has become indispensable after Special Relativity showed that the simultaneity of two events which are spatially separated from each other is relative, that is, dependent on the frame of reference chosen. Given force and counter-force have to appear simultaneously, the principle of force and counter-force thus can be obeyed in all possible frames of reference only if the spatial distance between the two locations is zero.

In electromagnetism, the Poynting vector, that is, the cross product of the vectors *E* and *B*, plays an important role in connection with the principle of local momentum conservation and hence with that of force and counter-force. The Poynting-vector field makes flows of electromagnetic energy in space "visible". Moreover, when the Poynting vector undergoes changes with time, the momentum of the energy flow that it reveals does so, too, and thus a force (that may act as a counter-force with respect to another force) is revealed.

So far, so good. However, changes in the components of the Poynting vector cannot reveal a possible counter-force (exerted on the inert mass of the electromagnetic energy in the immediate vicinity of the sphere) if both the Poynting vector and the change in its components over time are zero. As regards our two charged spheres at rest, we simply have to imagine that their inert mass is very large. Despite the fact that a momentum transfer onto each of the two spheres is taking place, those spheres do not start to move visibly. Even after a considerably long interval of time, practically no motion of the spheres has been produced, and hence also no magnetic field *B* (which will turn up as soon as a sphere is moving). Consequently, the Poynting vector, that is the cross product of *E* times *B*, stays zero over time, and no change in momentum of the energy (=mass) of an electromagnetic field is brought about. That is to say: the counter-force must act on an object in the immediate vicinity of each sphere, but this object cannot be the mass of the electromagnetic field. The sphere that is receiving a momentum transfer must "push away" a hidden object (that exists in its immediate vicinity), instead.

The same is true if we replace the two charged spheres by two permanent magnets that repel or attract each other.

11. The Challenge of the Proximity Principle by Mach's Principle When It Comes to Rotations, and How This Challenge Can Now Be Met

Some important experiments in physics do not need structures like the LHC or LIGO, but require a cup of coffee only. If the coffee is brought into relative rotation with respect to the material of the cup, it is still undecided as to whether or not the liquid will rise along the walls of the cup. It will only rise in case the rotation of the liquid occurs relative to the distant stars as well. In that case, it does not even matter if there is a relative rotation between the liquid and the material of the cup.

We are thus confronted with "Mach's Principle", which reads: the inertial force on particles and bodies on earth and in the solar system is due to their acceleration relative to all matter residing outside the solar system. As regards the application of that principle to rotations, Mach's principle appears to be undoubtedly correct (as was pointed out by N. Graneau/P. Graneau [13], Chapter 7, p. 144). In E. Mach's [14] own words (Chapter II, sub-chapter VI 5, page 232):

> "Newton's experiment with the rotating vessel of water simply informs us, that the relative rotation of the water with respect to the sides of the vessel produces NO noticeable centrifugal forces, but that such forces ARE produced by its relative rotation with respect to the mass of the earth and the other celestial bodies."

But this is at odds with the proximity principle valid in physics as well. M. Planck [15] (Section 1, p. 2) formulated that principle as follows:

> "According to it, when predicting chains of events to happen at a given location one does not need to care about what is going on somewhere else at other, finitely distant places, but may restrict oneself to causes in the immediate neighborhood; whereas, in case of believing in actions at a distance, one is, strictly speaking, compelled to check the whole universe for causes that may have a noticeable effect on the events that are to be determined by calculations."

The conflict between these two principles has widely been ignored, obviously because there seems to be no way of saving the proximity principle. However, the proximity principle can be saved by the recognition that the coffee in the cup, when being stirred with a spoon, does not rotate with respect to distant stars and galaxies, but with respect to the local "hidden energy" described. Though this energy cannot be attributed a direction or a speed of translational motion, its state does provide a reference for *rotation*.

It should be mentioned that H. Reichenbach [16] (Chapter III, Section 38—The problem of rotation according to Einstein—p. 240) had tried to save the proximity principle in a similar, though very abstract way, by speaking of an "inertial field" that pervades the entire univserse:

> "In this conception the fixed stars are at rest and determine an inertial field which pervades the entire space and lends inertia to every moving mass point."

But he couldn't explain why this "inertial field" was more than just a name for what should hopefully be detected in the future.

12. Conclusions

The main results are the following: while the total force on a body needed to maintain a constant acceleration a_0 is

$$F = \frac{m}{1-v^2/c^2} a_0 = \frac{m_0}{\left(1-v^2/c^2\right)^{3/2}} a_0,$$

the technical force needed to maintain that acceleration amounts only to

$$F = \frac{m_0}{\left(1-v^2/c^2\right)^{1/2}} a_0$$

It is therefore no surprise to find that the total energy of two (alike) objects that undergo a head-on collision is not conserved during the collision, thus requiring the involvement of a hidden reservoir of energy. That is to say: when treating the summand E^2/c^4 which appears in the momenergy equation

$$m_0^2 = \frac{E^2}{c^4} - \left(\frac{p_x^2}{c^2} + \frac{p_y^2}{c^2} + \frac{p_z^2}{c^2}\right)$$

as a constant during the collision, we run into contradictions. The use of the equation

$$E_{conv} = \sqrt{\left(E_1 + E_2\right)^2 - \left(p_{1_x} + p_{2_x}\right)^2 - \left(p_{1_y} + p_{2_y}\right)^2 - \left(p_{1_z} + p_{2_z}\right)^2} = const$$

in particle physics, too, results in an apparent disappearance of energy, but is no valid way of describing this phenomenon.

The hidden energy which is indispensable in Special Relativity can be identified with the "dark energy" which turns up in General Relativity. Different from what has been believed since the existence of "dark energy" has been acknowledged in cosmology, "dark energy" thus does interact with other forms of energy at least indirectly (as it increases the strength of the gravitational field of fast moving objects), even though Special Relativity does not reveal any means how it could be harnessed as a source of technical work in cyclical processes. A.

Trupp [17] showed that "dark energy" even interacts with Newton's apple.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Wang, J. (2019) *Journal of Modern Physics*, **10**, 1615-1644. https://doi.org/10.4236/jmp.2019.1014107

[2] Abramson, N.H. (2018) *Journal of Modern Physics*, **9**, 471-478. https://doi.org/10.4236/jmp.2018.93033

[3] Einstein, A. (1905) *Annalen der Physik*, **322**, 891-921. https://doi.org/10.1002/andp.19053221004

[4] Lorentz, H.A. (1904) Electromagnetic Phenomena in System Moving with Any Velocity Less than That of Light. *Proceedings of the Academy of Sciences of Amsterdam*, Vol. 6, 809.

[5] Bergmann, P. (1976) Introduction to the Theory of Relativity. Revised Edition, Dover Publ., Mineola.

[6] Dragon, N. (2014) Geometrie der Relativitätstheorie. University of Hannover, Hannover.

[7] Tolman, R.C. (1912) *Philosophical Magazine*, **23**, 375-380. https://doi.org/10.1080/14786440308637231

[8] Tolman, R.C. (1918) The Theory of the Relativity of Motion. University of California Press, Berkeley.

[9] Epstein, P.S. (1911) *Annalen der Physik*, **341**, 779-795. https://doi.org/10.1002/andp.19113411404

[10] Taylor, E.F. and Wheeler, J.A. (1992) Spacetime Physics: Introduction to Special Relativity. 2nd Edition, W. H. Freeman, New York. https://doi.org/10.1119/1.17254

[11] Povh, B., Rith, K., Scholz, Ch., Zetsche, F. and Rodejohann, W. (2014) Teilchen und Kerne. 9th Edition, Springer, Berlin. https://doi.org/10.1007/978-3-642-37822-5

[12] Berger, Ch. (2014) Elementarteilchenphysik. 3rd Edition, Springer, Berlin. https://doi.org/10.1007/978-3-642-41753-5

[13] Graneau, P. and Graneau, N. (2006) In the Grip of the Distant Universe—The Science of Inertia. https://doi.org/10.1142/6087

[14] Mach, E. (1919) The Science of Mechanics. 4th Edition, The Open Court Publishing Company, Chicago and London.

[15] Planck, M. (1922) Einführung in die Theorie der Elektrizität und des Magnetismus.

[16] Reichenbach, H. (1958) The Philosophy of Space and Time. Dover Publ., Mineola.

[17] Trupp, A. (2019) *Physics Essays*, **32**, 484-496. https://doi.org/10.4006/0836-1398-32.4.484

Journal of Modern Physics, 2022, 13, 34-45
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Mechanics of the Gyroscopic Precession and Calculation of the Galactic Mass

Jian Sun[1*], **Ting Xu**[1], **Zheng Jiao**[1], **Hui Wang**[2]

[1]Shanghai University, Shanghai, China
[2]Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai, China
Email: *sunjkl2@gmail.com

How to cite this paper: Sun, J., Xu, T., Jiao, Z. and Wang, H. (2022) Mechanics of the Gyroscopic Precession and Calculation of the Galactic Mass. *Journal of Modern Physics*, **13**, 34-45.
https://doi.org/10.4236/jmp.2022.131003

Received: December 19, 2021
Accepted: January 17, 2022
Published: January 20, 2022

Abstract

A spinning gyroscope precesses about the vertical due to a torque acting upon the wheel. The torque is generated by the shift of moment of force by gravity and it points to the vertical instead of the tangential direction of precession. This intuition offers an alternative and straightforward view of precession dynamics in comparison with the literature. It also presumes a dynamic balance of momentum between circular motions of the wheel spin and precession. Accordingly, the gyroscopic dynamics is then applied to the study of galactic motion of the solar system in space and the Galactic mass is calculated with the inclusion of gyroscopic effect of the solar planets. Results indicate that the gyroscopic effect of Mercury orbiting around the Sun can increase the calculated Galactic mass by 23% in comparison with the result obtained by the classic approach.

Keywords

Gyroscopic Precession, Solar System, Galactic Mass

1. Introduction

Newton's law of universal gravitation is arguably the first principle in astrophysics. It started with an apple falling down on his head and ended with the Moon falling sideways around Earth. Newton's comparison of acceleration of the apple to that of the Moon led to a rational understanding of the nature of gravity that came to be the fabric of our universe. All objects attract each other with a force that is proportional to the product of their masses and inversely proportional to their distance of separation. The laws of mechanics that govern the movement of objects on Earth also govern the motion of celestial bodies in space.

In the study of astrophysics, measuring the mass and motion of galaxies is one of our fundamental tasks. Luminous or baryonic matter in a galaxy such as stars and planets can be detected directly using optical and radio telescopes. And we can get an estimate of how much mass they contain. However, we cannot directly measure non-luminous matter, such as black holes that formed from stellar collapse, which we cannot see.

How can we measure the celestial mass that cannot be seen? According to the universal gravitational equation and Newton's second law of motion, if we know the mass of a body, we can work out how fast we need to go to orbit at distance. For instance, given the mass of the Sun and orbital radius of the solar planets, we can calculate the rotation curve of the solar system. Results show that the orbital or rotational speeds of the planets are inversely proportional to the square root of their orbital radii. The solar rotation curve falls. The calculation result is consistent with that of observation. Inversely, if we know the speed of a star orbiting on a circular orbit at distance from a galactic center, we can calculate the galactic mass interior to the orbit of the star. The gravitational forces from mass outside cancel out, assuming that the mass distribution is spherically symmetric. For instance, given the orbital speed and radius of the Sun orbiting about the Galactic center, the Galactic mass can be calculated as an equivalent of 95 billion solar masses. Consequently, suppose we measure the speeds of stars at various radii from a galactic center. In theory, each star allows us to calculate the mass interior to the orbit of that particular star. If we observe many stars, we can map out the rotation curve of the galaxy.

Accordingly, when we apply the rotation curve to our Galaxy, we expect results similar to that of the solar system. Specifically, within the inner galaxy, enclosed mass rises, so does the rotation speed. There are no stars and no more mass beyond the optical edge of the Galaxy, and rotation speed supposedly falls. Instead, we observe a flat rotation curve for the Galaxy in which speed virtually stays the same as we move further away from the center [1] [2].

There can be two ways of reasoning with the observation, as we go back to the equation derived from Newton's universal gravitation and second law of motion. If orbital speed is constant with increasing orbital radius, then mass is proportional to radius. Hence, there ought to be more and more mass as we go to larger and larger distances from the center, implying that there can be a substantial amount of mass like dark matter we do not see in space [3] [4].

Alternatively, we can compare the patterns of motion between the solar planets and galactic stars like the Sun. A solar planet moves along with other planets in groups around the Sun, while the Sun moves with a group of its planets orbiting around it. Orbital rotation of the solar planets produces a combined angular motion that would keep the solar plane moving along a straight line due to the conservation of angular momentum. In other words, the solar system acts like a giant gyroscope in space. Does the law of conservation of angular momentum affect orbital motion of the solar system in our Galaxy? If so, how? Perhaps we want to go back to the basic physics of gyroscopes.

2. The gyroscope

A gyroscope is made of a wheel or rotor on an axle that spins freely, usually mounted in a frame body. When spinning, the orientation of the spin axis is unaffected by tilting or rotating the body. Without its mounting body, a spinning gyroscope can stand by its axle like a top and move in counter-intuitive ways. It precesses about the vertical as if it defies gravity [5] [6] [7].

Newton's first law of motion tells us that a body in motion continues to move at a constant speed along a straight line unless acted upon by an unbalanced force. Since a spinning gyroscope precesses and changes its motion, there must be a net force acting upon it. And it has to be the force of gravity. **Figure 1** shows a precessing gyroscope with a description of precession dynamics and geometric illustration of angular motion [8]. When the wheel is not spinning at the shown position, it simply topples over by turning to the right around the x-axis, due to the moment of force created by its weight. When the wheel is spinning, it precesses by continuously falling sideways around the z-axis without actually falling over, tracing an invisible cone standing on its tip.

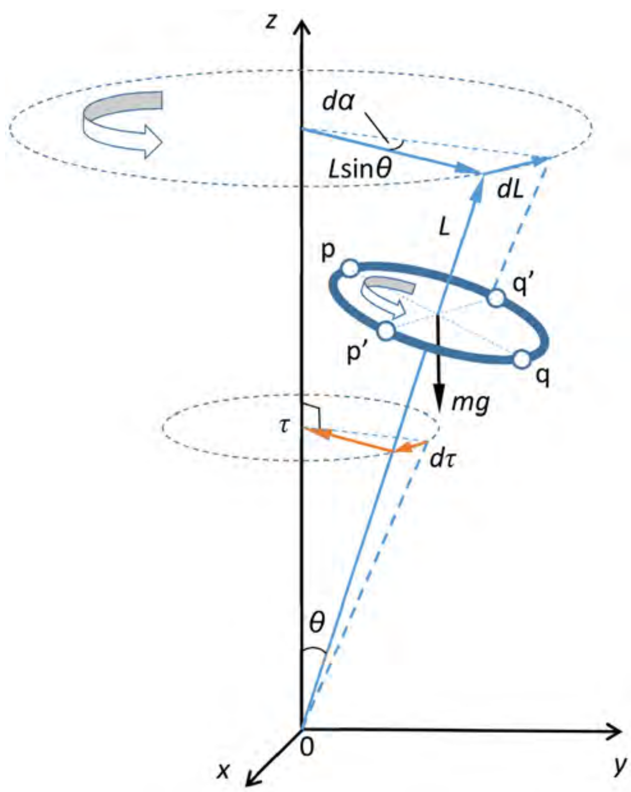


Figure 1. A spinning gyroscope falls sideways. The wheel consists of a solid ring connected to an axle with weightless spokes not shown. The gyroscope precesses counterclockwise as the wheel spins in the same direction. Positioned below the wheel for clarity, the torque of precession points to the z-axis due to a quarter-turn shift of the moment of force by gravity, according to the right-hand rule. The gyroscope will precess clockwise if the wheel changes its spin direction and the torque will be pointing away from the z-axis. Note that angular momentum and torque change in opposing directions. It points to a balance of momentum in horizontal between the wheel spin and precession.

How do we interpret precession? Assume our gyroscope is made of a solid ring for simplicity instead of a typical disk on an axle. Let's pick four points of mass p, q, p' and q' at top, bottom, left and right on the ring. Without spinning, all units of mass produce a total moment of force by gravity, with a moment arm being the distance from the wheel center of mass to the z-axis. This moment of force turns around the p'-q' horizontal median and the wheel falls. When spinning counterclockwise, all mass units shift in position continuously and the wheel precesses counterclockwise around the z-axis by turning around the p-q vertical median instead. It seems that the spinning of the wheel makes its turning shift by 90 degrees in the direction of precession. Since it is the force of gravity acting on the wheel that causes precession, the moment of force must also shift by 90 degrees due to spinning (see Appendix 1 for reasoning). In other words, the moment of force turns around the horizontal median without spinning. When spinning, it shifts by a quarter-turn counterclockwise and turns around the vertical median instead. The moment of force by gravity thus becomes the torque of precession, pointing to the z-axis instead of the tangential direction of precession [8]. This quarter-turn shift of moment/torque drives the wheel to make a quarter-turn twist in falling: it falls sideways around its vertical median instead of around the horizontal. As it continues to fall sideways, the gyroscope precesses. And the faster the wheel spins, the quicker the mass units shift in position, and inversely the easier and slower it precesses.

Following the geometric description in **Figure 1**, the precession of the gyroscope and spin of the wheel can be described as,

$$\begin{gathered} \mathrm{d}\alpha \cong \frac{\mathrm{d}L}{L\sin\theta} \\ \mathrm{d}L = L\sin\theta\,\mathrm{d}\alpha \end{gathered} \tag{1}$$

where $\mathrm{d}\alpha$ and $\mathrm{d}L$ are the changes of precession angle and angular momentum. θ is the tilt angle of the gyroscope from the vertical. The torque of precession that changes the angular momentum can be formulated as,

$$\tau = \frac{\mathrm{d}L}{\mathrm{d}t} = \frac{L\sin\theta\,\mathrm{d}\alpha}{\mathrm{d}t} = L\sin\theta\,\omega_p \tag{2}$$

where τ is the torque and ω_p is the rate of precession. Calculation of the precession rate is derived in Appendix 1.

In general, without a net force acting on a spinning gyroscope, the orientation of its spin axis remains unchanged. When standing by its axle, gravity force exerts on the wheel and turns into a shifting moment for precession. And the wheel changes its orientation by turning around its vertical median. When it completes a full circle of precession, it also finishes a complete cycle of orientation change at a constant speed. It is a result of the dynamic balance of momentum between the wheel spin and precession.

3. The Solar System

Suppose our solar system revolves around the Galactic center (GC) in a way similar to gyroscopic precession. Planets including Earth rotate around the Sun to form the spinning solar plane, which orbits about the GC or precesses like a gyroscope wheel around the vertical, as shown in **Figure 2** in three descriptive orbital positions. In fact, instead of a standing wheel, a spinning bicycle wheel in suspension is a better analogy for the solar scenario. The gyroscopic mechanism is the same regardless. Using Earth as an example, the rotation of Earth around the Sun generates an angular momentum pointing to the upper z-axis. The change of precession angle of the solar plane and angular momentum of the solar orbiting Earth follows the gyroscopic model,

$$\begin{gathered} \mathrm{d}\alpha \cong \frac{\mathrm{d}L}{L \sin \beta} \\ \mathrm{d}L = L \sin \beta \, \mathrm{d}\alpha \end{gathered} \tag{3}$$

where $\mathrm{d}\alpha$, L, and β are the change of precession angle, angular momentum of the solar orbiting Earth, and tilt angle of the solar plane from the Galactic plane. The torque that changes the angular momentum is formulated as,

$$\tau = \frac{\mathrm{d}L}{\mathrm{d}t} = \frac{L \sin \beta \, \mathrm{d}\alpha}{\mathrm{d}t} = L \sin \beta \, \omega_p \tag{4}$$

where ω_p is the precession rate of Earth or solar plane about the GC, and τ is the torque of precession. Earth is treated as a point of mass in our study so the angular momentum generated by its spin around its polar axis is ignored.

When the solar plane completes a Galactic circle, it also finishes a circle of precession around the z-axis and a complete horizontal turn by itself around a

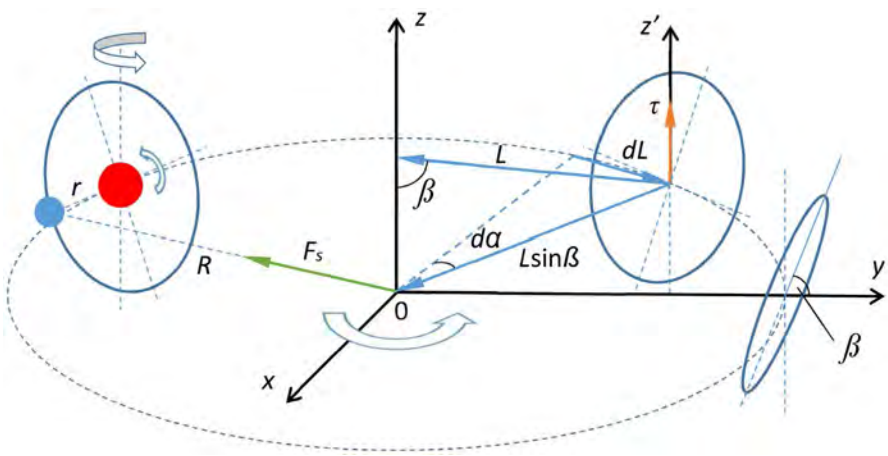


Figure 2. The spinning solar plane orbits or precesses about the GC in three select positions: Earth rotating around the Sun in the solar plane in the Galactic plane on the left; angular momentum of the solar orbiting Earth changing with its precession angle on the middle right; solar plane tilting from the Galactic plane on the right. The front side of the solar plane faces the z-axis at all times. Both the solar plane spin and Galactic rotation are shown counterclockwise for a clear geometric view of vectors. The torque of precession points upward, according to the right-hand rule. It will point downward if the solar system revolves clockwise around the GC. In space, the gravitational pull from Galaxy points to the GC, rather than downward, which is different from the case of the gyroscope on Earth in **Figure 1**.

moving z'-axis that goes through the solar center (**Figure 2**), just like a precessing gyroscope. This forced horizontal turn or inclination change of the solar plane is presumably due to an additional gravitational pull from the GC. Without this gravitational pull, the solar plane would turn its side to face the center in every half of its Galactic circle due to its conservation of angular momentum. This assumes that the solar plane keeps orbiting in the Galaxy as a point of mass, which is the implication in conventional calculations of the Galactic mass. It may not be a reasonable assumption as the solar planets would experience a substantial gravitational change when the solar plane moves into positions where it is aligned with the GC: the Galactic gravitational force becomes too strong when the planets position between the Sun and GC; the force is too weak when they reverse their positions with the Sun. Substantial gravitational changes as such can affect and even break the integrity of the solar system.

Likewise, assuming the solar orbit of Earth effectively forms a ring of uniform mass distribution in its rotation around the Sun, this Earth orbit-ring (EOR) completes a horizontal turn by itself around the moving z'-axis in one Galactic cycle. Therefore, a torque must be acting on the EOR during precession, which is generated by the (additional) Galactic gravitational pull. Let's call it the action torque. It exerts on the EOR so that it cannot turn its side while facing the GC in precession. In response, the EOR has to generate a reaction torque by its angular momentum to counterbalance the action torque. This results in a dynamic balance of torque on the EOR between its precession in circular motion and inertia of moving straight. Unlike gyroscopes on Earth where gravity is the lone force to drive precession, the inertia of the EOR drives precession along with the gravitational pull in space. And the precession direction of the EOR is untethered from its spin direction. When two torques are equally matched, so are their corresponding forces in opposing directions. The action force that produces the torque of precession must be balanced in horizontal with the reaction force generated by the angular momentum. And the latter can be formulated as a force pulling away from the GC, as shown in **Figure 2**,

$$F_s = \frac{\tau}{r} = \frac{L\omega_p \sin\beta}{r} \tag{5}$$

where F_s is the force of solar orbiting Earth and r is the solar orbital radius of Earth. This reaction force generated by the angular momentum of the EOR must be overcome in its Galactic motion and subtracted in calculating the Galactic mass using Newton's law of circular motion as follows,

$$F - F_s = ma = m\frac{v^2}{R} \tag{6}$$

where F is the gravitational force of the GC, m, a, v and R are the mass, Galactic centripetal acceleration, Galactic orbital speed and Galactic orbital radius (distance to the GC) of Earth. Combining with the equation of universal gravitation,

$$\frac{GMm}{R^2} - \frac{L\omega_p \sin\beta}{r} = \frac{mv^2}{R} \tag{7}$$

where G is the universal gravitational constant and M is the Galactic mass or mass of bodies interior to Galactic orbit of the solar system. Thus M can be calculated as,

$$M = \frac{Rv^2}{G} + \frac{R^2 L\omega_p \sin\beta}{rGm} \tag{8}$$

It reduces down to the conventional equation (the first term on the right) when the angular momentum of the EOR is treated as zero (in the second term on the right). The angular momentum can be estimated by assuming the solar orbiting Earth as a ring of uniform mass distribution.

$$L = I\omega = mr^2\omega \tag{9}$$

where I and ω are the moment of inertia and spin speed of the EOR. The mass of Earth is canceled out in calculation after combining Equations (8) and (9).

4. Discussion

In general, the Galactic mass can also be calculated with other solar planets in a similar fashion or with the Sun as a solid sphere, assuming that the Sun's spin axis points to the z-axis. It is rather ideal to treat an orbiting body as a ring of uniform mass in the calculation. In reality, orbital motion of the solar planets produces a non-uniform distribution of their angular momenta, presumably causing oscillations within the solar plane. The derivation above is to find the Galactic mass with the inclusion of gyroscopic effect. Inversely, given a galactic or solar mass, Equation (7) can also be used to calculate the potential orbital speed increase of stars or solar planets due to the gyroscopic effect of the planets or moons (Appendix 2).

Results of the Galactic mass increase calculated with gyroscopic motion of the solar planets and the Sun are presented in **Table 1**, shown as a list of mass calculated by the second term on the right in Equation (8). **Table 2** lists the parameters and constants used in the calculation. It appears that the gyroscopic effect falls with the solar orbital radius, presumably due to the decreasing solar gravitational pull that causes a decrease in the solar orbital speed. To calculate the total solar gyroscopic effect, the solar system can be approximated as a monolithic rigid plane of mass by adding up the angular momenta of all the solar bodies. However, estimation of the average radius and angular speed of the plane becomes a complicated issue. Hence, the gyroscopic effect is calculated and presented separately with each solar body in this study.

The Sun has a minimal gyroscopic effect on assessing the Galactic gravitational and mass increase as shown in **Table 1**, which is reasonable due to its small angular momentum. Among solar planets, Mercury has the highest precession impact on the Galactic gravitational increase that results in an additional 22 billion solar masses from the value of 95 billion solar masses obtained by the classic approach. It suggests that the effective force of gyroscopic precession of the solar

Table 1. Galactic mass increase calculated with the gyroscopic motion of each solar body in the solar system.

	Orbital radius m	Orbital period days	Solar mass equivalent, billion
Mercury	5.79×10^{10}	88	22.1
Venus	1.08×10^{11}	224.7	16.1
Earth	1.50×10^{11}	365.2	13.7
Mars	2.28×10^{11}	687	11.1
Jupiter	7.79×10^{11}	4331	6.03
Saturn	1.43×10^{12}	10,747	4.47
Uranus	2.87×10^{12}	30,589	3.15
Neptune	4.50×10^{12}	59,800	2.52
Pluto	5.91×10^{12}	90,560	2.19
Sun	6.96×10^{5}	27	0.00042

Table 2. Calculation parameters and constants.

R, m	2.60×10^{20}
v, m/s	2.20×10^{5}
ω_p, rad/s	9.06×10^{-16}
G, Nm^2/kg^2	6.67×10^{-11}
Solar mass, kg	1.99×10^{30}
β, rad	$\pi/2$

plane is comparable in magnitude to the gravitational force calculated by the classic approach. In other words, it is reasonable to assume that our solar system revolves around the GC like a precessing gyroscope as shown in **Figure 2**, instead of being a point of mass as implied in the classic approach.

5. Concluding Remarks

A new mechanical model of gyroscopic precession is derived and applied to the study of galactic motion of the solar system in space. The Galactic mass is calculated with the inclusion of gyroscopic effect of the solar planets. Results indicate that the gyroscopic effect of Mercury orbiting around the Sun can increase the calculated Galactic mass by 23% in comparison with the result obtained by the classic approach. In comparison with other theories, the maximal percentage increase of the Galactic mass calculated by the gyroscopic effect in this study is an order of magnitude smaller than those obtained by the dark matter models [4]. Nevertheless, the effect of gyroscopic precession introduces an empirical variable in the study of astrophysics. It can potentially lead to new understandings of our Galaxy.

Acknowledgments

Planetary data are obtained from NASA and Wikipedia websites. Financial support is graciously provided by Shanghai University. All calculations are performed with Microsoft Excel.

Conflicts of Interest

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

References

[1] Palma, C. (2020) ASTRO 801: Planets, Stars, Galaxies, and the Universe, Open Educational Resources Initiative of John A. Dutton e-Education Institute, The Pennsylvania State University.
https://www.e-education.psu.edu/astro801/content/l8_p8.html

[2] Kuhfittig, P. and Gladney, V. (2014) *Journal of Modern Physics*, **5**, 1931-1937.
https://doi.org/10.4236/jmp.2014.517187

[3] Mongan, T. (2017) *Journal of Modern Physics*, **8**, 919-922.
https://doi.org/10.4236/jmp.2014.517187

[4] Karukes, E.V., Benito, M., Iocco, F., Trotta, R. and Geringer-Sameth, A. (2020) *Journal of Cosmology and Astroparticle Physics*, **5**, 33.
https://doi.org/10.1088/1475-7516/2020/05/033
https://phys.org/news/2019-12-mass-milky-billion-sun.html

[5] Deimel, R.F. (1950) Mechanics of the Gyroscope. Dover Publications, New York.

[6] Scarborough, J.B. (2014) The Gyroscope Theory and Applications. Nabu Press, London.

[7] Usubamatov, R. (2021) *World Journal of Mechanics*, **11**, 177-185.
https://doi.org/10.4236/wjm.2021.119013

[8] Ling, S.J., Sanny, J. and Moebs, B. (2020) 11.5 Precession of a Gyroscope—University Physics Volume 1, Licensed by OpenStax University Physics.
https://openstax.org/books/university-physics-volume-1/pages/11-4-precession-of-a-gyroscope

Notations

α, dα—precession angle and its change
L, dL—spin angular momentum and its change
θ—tilt angle of gyroscope from the vertical
τ—torque of precession
ω_p—rate of precession
β—tilt angle of the solar plane from the Galactic plane
F_s—force by the angular momentum change of EOR
F—gravitational force of the GC
m—mass of the gyroscope wheel or Earth
a—Galactic centripetal acceleration of Earth
v—Galactic orbital speed of Earth
R—Galactic orbital radius of Earth
G—universal gravitational constant
M—Galactic mass
r—radius of the gyroscope wheel or EOR
I—moment of inertia
ω—rate of spin
mg—weight of the gyroscope wheel
h—length of the gyroscope axle

Appendix 1. Interpretation of the Shift of Moment of Force and Calculation of the Precession Rate

As the wheel tends to turn around the p'-q' horizontal median in precession due to the moment of force by gravity, the mass of the ring generates transversal acceleration. The upper half of the ring accelerates upward and the lower half does downward with the max acceleration at p and q. All mass units accelerate and travel along the ring counterclockwise. Unit p accelerates upward while traveling to p' and decelerates while continuing to q. Consequently, its upward transversal velocity reaches a max at p' and reduces back to zero at q. Similarly, unit q accelerates downward while traveling to q' and decelerates while continuing to p. Its downward transversal velocity reaches a max at q' and reduces back to zero at p. All mass units experience such a cycle of transversal velocity delay/shift for a quarter-turn from acceleration. As a result, the potential of the wheel turning around the p'-q' horizontal median becomes turning around the p-q vertical median instead. And the shift of transversal velocity practically twists the moment of force by a quarter-turn counterclockwise.

A precessing gyroscope does three circular moves at the same time: spinning the wheel, turning around its vertical median, and revolving around the vertical axis. The first movement makes the second that drives the third. Mechanically, the torque of precession comes from the force of gravity by twisting the moment/torque via shifting mass units on the spinning wheel. Following Equation (2),

$$L\omega_p \sin\theta = \tau = mgh\sin\theta \tag{10}$$

where mg and h are the weight of the wheel and length of the axle. Hence, the precession rate becomes,

$$\omega_p = \frac{mgh}{L} \tag{11}$$

where L is the angular momentum of the spin wheel, which can be calculated based on the wheel shape, a solid ring in this study,

$$L = I\omega = mr^2\omega \tag{12}$$

where I and ω are the moment of inertia and spin rate of the wheel.

Appendix 2. Calculation of Orbital Speed Due to Gyroscopic Effect

Equation (7) is reformulated by replacing the reaction force created by the angular momentum with the action force that creates the torque of precession,

$$\frac{GMm}{R^2} + \frac{L\omega_p \sin\beta}{r} = \frac{mv^2}{R} \tag{13}$$

Plug in the moment of inertia, replace the angular speed of precession with the linear speed of precession, and let $\beta = \pi/2$,

$$Rv^2 - r\omega Rv - GM = 0 \tag{14}$$

Solve for v. The potential orbital speed increase of Earth due to the gyroscopic effect of the Moon is calculated accordingly. Without changing its orbital radius and flying away from Earth, the Moon would need to increase its orbital speed by more than 130 thousand times in order to increase the orbital speed of Earth by 1%, provided that the Moon also changes the inclination of its orbit to squarely face the Sun. In other words, the gyroscopic effect of the Moon is negligible in comparison with those of the solar planets.

Journal of Modern Physics, 2022, 13, 46-65
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Many-Body Correlation Effects on the x_{Bjorken}-Dependence of Cross Section Ratios off Nuclei for $1 < x_{\text{Bjorken}} < 2$

Athanasios Petridis, Allen Barr, Drew Fustin

Department of Physics and Astronomy, Drake University, Des Moines, USA
Email: athan.petridis@drake.edu

How to cite this paper: Petridis, A., Barr, A. and Fustin, D. (2022) Many-Body Correlation Effects on the x_{Bjorken}-Dependence of Cross Section Ratios off Nuclei for $1 < x_{\text{Bjorken}} < 2$. *Journal of Modern Physics*, **13**, 46-65.
https://doi.org/10.4236/jmp.2022.131004

Received: July 17, 2020
Accepted: January 17, 2022
Published: January 20, 2022

Abstract

Many-body correlations in nuclei determine the behavior of Deep-Inelastic-Scattering (DIS) and Quasi-Elastic Scattering (QES) cross section ratios off heavy over light nuclei especially for $x_{\text{Bjorken}} > 1$, obtained at Jefferson Lab. They can be described in terms of quark-cluster formation in nuclei due to wave-function overlapping, manifesting itself when the momentum transfer is high so that the partonic degrees of freedom are resolved. In clusters (correlated nucleons) the quark and gluon momentum distributions are softer than in single nucleons and extend to $x_{\text{Bjorken}} > 1$. The cluster formation probabilities are computed using a network-defining algorithm in which the initial nucleon density is either standard Woods-Saxon or is input from lower energy data while the critical radius for nucleon merging is an adjustable parameter. The exact choice of critical radius depends on the specific nucleus and it is anti-correlated to the rescaling of the x_{Bjorken} needed for bound nucleons. The calculations show that there is a strong dependence of the cross section ratios on the x_{Bjorken} in agreement with the data and that four-body correlations are needed to explain the experimental results even in the range $1 < x_{\text{Bjorken}} < 2$. The dependence on the specific exponents of parton distributions in high-order clusters is weak.

Keywords

Deep Inelastic Scattering, Quasi-Elastic Scattering, Short-Range Correlations, Quark-Clusters

1. Introduction

The atomic nucleus is typically described as a system of protons and neutrons interacting via some potential that binds them together. Pions emerge as the quanta of these interactions. When the nucleus is probed by high-energy projectiles the parton-degrees of freedom of the nucleons and the pions are resolved and the picture of the quark model emerges. In this energy region the quarks appear to interact among themselves by means of gluons which in turn can give rise to quark-antiquark pairs, constituting the sea or ocean inside the nucleons and are distinguished from the valence quarks that make up the "core" of each nucleon. However, it is usually assumed that no matter how complicated the interactions among nucleons are or how high the energy scale of the probe is, the nucleons maintain their individuality and independence from one another, *i.e.*, they are uncorrelated. Data accumulated over the past two decades strongly suggest that this picture is overly simplistic.

The so called EMC-effect, named after the European Muon Collaboration [1] [2] [3] [4] [5], has shown that in the case of Deep Inelastic Scattering (DIS) of muons off nuclei the cross section does not scale with the number of nucleons and exhibits a strong dependence on the longitudinal momentum fraction carried by the struck quarks. This effect has been observed in many other processes involving the electromagnetic and the weak interactions of nuclei with probes at high energies, such as Drell-Yan pair production, neutrino scattering, and charmonium and bottomonium suppression [6] [7] [8] [9] [10] and predicted for direct photon production in hardon-nucleus and nucleus-nucleus collisions [11] [12] [13].

The many-body correlations can be attributed to a physical overlap of the bound nucleons leading to the formation of multi-quark clusters and the consequent exchange of quarks and gluons that results in softening of the parton distributions. The concept of multi-quark clusters was originally proposed by Pirner, Vary and Coon [14] [15] [16] and has had significant success in interpreting data on a variety of processes involving hardon-nucleus and nucleus-nucleus collisions such as those mentioned above [17]-[23]. This is the approach taken in this work. A model that describes the probability for such overlaps is developed along with reasonable assumptions about the parton distributions in clusters. The authors of Ref [24] have addressed the issue of correlations by estimating their probability to be formed and comparing the results with heavy-to-light nuclei cross section ratios for Quasi-Elastic Scattering. These results are in general agreement with the data but do not describe the explicit dependence of those ratios on the momentum fraction of the struck parton. This dependence contains a wealth of information on the nature of the correlations and their effect on the parton momentum distributions.

This article is organized as follows: in Section 2, the concept of many-body correlations is introduced and described in terms of multi-quark clusters in the nucleus. In Section 3, a network-based algorithm for calculating cluster probabilities given an initial spatial distribution of nucleons is presented. Those probabilities are expressed as functions of a critical distance (radius) between two nucleons. In Section 4, a method to calculate the numbers of various types of clusters (proton-proton, proton-neutron, etc.) in a given nucleus is introduced. In Section 5, the parton (quark and gluon) momentum distributions in clusters are postulated on the basis of existing data. They are expressed as functions of the longitudinal parton momentum fraction, x_{Bjorken}. In Section 6, it is argued that for bound nucleons, x_{Bjorken} may have to be rescaled by a multiplicative factor η. In Section 7, the structure functions of nuclei are calculated as sums of products of quark distributions weighted by the squares of the quark charges and summed over all types of clusters. In Section 8, the heavy-to-light nuclei cross section ratio is defined. In Section 9, comparison of the theoretical calculations with two sets of data is presented and the best-fit parameters (the critical radius and the rescaling factor) are extracted. The similarity of correlations between Deep Inelastic Scattering (DIS) and Quasi Elastic Scattering (QES) is discussed. Section 10 details the need for many-body correlations as indicated by data. Finally, in Section 11 conclusions are drawn and some directions of future continuation of this research are suggested.

2. Correlations as Quark-Clustering

Correlations among nucleons bound inside nuclei imply that the nucleons are not independent particles so that the nucleus is not a collection of mutually interacting, yet separate constituents. From a quantum-mechanical perspective this means that the wave-functions of nucleons overlap sufficiently to produce effective aggregates. These can be identified as clusters of various orders, from 1st-order, that is single nucleons, to 2nd-order, made of two nucleons and so on. When the protons and neutrons are so close together, they can easily exchange gluons and sea quarks. As the gluons propagate they can also produce quark-antiquark pairs which subsequently annihilate back into gluons. The quarks may emit and reabsorb gluons. The sharing of such particles leads to energy and momentum exchanges among the nucleons and modifications to the parton momentum distributions inside them.

In experiments using high energy electrons (or muons) as probes to peek into the structure of the nucleus the 4-momentum exchange between the projectile and the target determines the spatial and temporal resolution of the image of the nucleus and its constituents. When the square of the 4-momentum exchange, Q^2, exceeds 2 GeV^2 the sub-nucleonic (parton) degrees of freedom start being resolved, *i.e.*, the interaction is directly between the electron and the quarks since the exchange photon wavelength is $\lambda = h\big/\sqrt{Q^2}$. This corresponds to distances just under 1 fm. Given the average nuclear density it is at this distance that nucleon wave-functions begin to overlap. Consequently, at relatively low Q^2 the electron interacts with quarks in single nucleons and at high Q^2 with quarks that may be shared among nucleons in clusters. This is illustrated in **Figure 1**. Furthermore, if the (invariant) photon energy, ν, is also large (greater than 1 GeV) the quarks are probed in the asymptotically free region, that is the strong interaction among them is greatly weakened and the cross section becomes

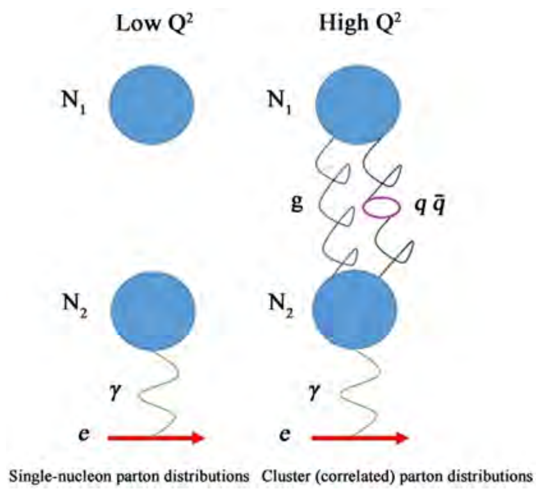


Figure 1. Electron scattering off nuclei. At low momentum transfer (low Q^2) the exchange photon resolves individual nuclei. As the momentum transfer increases the gluon (including quark-antiquark bubbles) exchange is probed. The nucleons are no longer independent and can be treated as clusters. The parton distributions in the clusters differ from those of single nucleons and become softer.

approximately independent of Q^2, depending only on the invariant quantity $x_{\text{Bjorken}} = Q^2/2M\nu$, where M is the nucleon mass (the Bjorken-x scaling variable, which, in a frame of reference where the nucleons move very fast and the transverse momentum components can be neglected, is the longitudinal momentum fraction of a nucleon carried by a struck quark). This kinematic region is referred to as Deep Inelastic Scattering (DIS), while the region with large Q^2 and lower ν in which the quarks are not completely resolved, is known as the Quasi-Elastic Scattering (QES) region. Completely elastic scattering corresponds to $x_{\text{Bjorken}} = 1$. Nucleon binding and the resulting Fermi motion lead to smearing of the cross section allowing x_{Bjorken} to deviate from unity. It has been reasonably argued [8] that Fermi motion and the formation of clusters, *i.e.*, the presence of correlations among bound nucleons are strongly overlapping concepts.

3. Calculation of Clustering Probabilities

The probabilities for quark-cluster formation can be calculated using a quasi-classical network-based algorithm. An initial distribution of particles, nucleons in this case, is produced. The coordinates of these particles are determined randomly from the given distribution in 3 dimensions. Then the array of distances between the particles is computed. The algorithm determines the particles whose distances are smaller than a critical radius R_c relative to one another and assigns these particles to a particular network (cluster). Special care is taken to ensure that each particle belongs to one and only one cluster. The particles are not physically displaced so that the center of mass remains fixed. The algorithm then counts the clusters of each order, the total number of clusters and the fraction that corresponds to each order. The process is repeated many times starting from different randomly-generated initial coordinates. The final averaged numbers of fractions are reported as the probabilities for each cluster-order for the chosen critical radius. The latter is then incremented and the algorithm is rerun to finally obtain the probabilities as functions of R_c. Specifically, the algorithm for N particles consists of the following steps:

(1) A spherically-symmetric distribution of N particles (nucleons) is produced for a range of different R_c (critical distance) values using either a Woods-Saxon nuclear density or a Fourier-Bessel nuclear density, fit to real experimental data. The distance array $d(i,j)$ for $i,j=1,\cdots,N$ is produced holding the values for the distances between each of the N particles in the distribution. Then, the initial network array $A_0(i,j)$ for $i,j=1,\cdots,N$ is produced so that $A_0(i,j)=1$ if $d(i,j)<R_c$ and $A_0(i,j)=0$ otherwise. The initial network array is reduced to the final network array $A(i,j)$ for $i,j=1,\cdots,N$ in the following manner.

(2) Each column $i=1,\cdots,N-1$ is compared with column $k=i+1,\cdots,N$.

(3) If two separate columns contain a 1 in the same position, the two columns are merged by replacing the column $A(i,j)$ with the result of ($A(i,j)$ OR $A(k,j)$) and the column $A(k,j)$ with the zero vector.

(4) The number of 1's in each column of the matrix $A(i,j)$ is then counted. This number corresponds to the order of the (quark) cluster that column is a part of. The total number of clusters of each order is then counted for the trial. These values are averaged over a large number of initial arrays for each R_c. The cluster probabilities f_m, where m is the order of the cluster, are the fractions of a cluster-type number over the total cluster number. The order m varies from 1 up to N.

The crucial steps in the above algorithm are (2) and (3). Step (2) ensures that all networks that are formed maintain their internal structure while step (3) guarantees that each particle is assigned to one and only one network, specifically the network from a member of which the said particle has the smallest distance. This step identifies all the networks upon cycling over the array of 1's and 0's that describe distances that pass the critical-radius test.

The initial particle distribution can be chosen to be a Woods-Saxon in the radial direction and uniform in the azimuthal and polar directions. Another option is to use a radial distribution derived from experimental data using a Fourier-Bessel method [25] [26]. In either case the results are consistent with larger cluster probabilities for larger and denser nuclei. In this approach the coalescence happens entirely in coordinate space. Even though the Fermi motion is implicit in the finite size of the wave-function, approximated by the critical radius, the relative momentum of the nucleons is not considered. Also differences between protons and neutrons are neglected at this stage. In **Figure 2** a schematic diagram of cluster formation is shown. In **Figure 3** the probabilities, f_m, for clusters of order $m=1,\cdots,N$ are plotted as functions of the critical radius R_c for various nuclei characterized by their mass number A which in this case is the total number of particles (N). The Woods-Saxon radial distribution is used, $P_r(r)\sim 1\big/\left[1+\exp\left((r-a)/b\right)\right]$, where r is the distance from the center of mass of

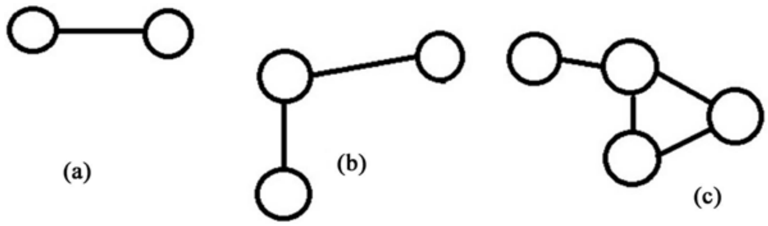


Figure 2. Three types of clusters produced by the network algorithm. (a) 2nd-order cluster (two nucleons or six valence quarks). (b) One type of 3rd-order cluster (three nucleons or nine valence quarks). (c) One type of 4th-order cluster (four nucleons or twelve valence quarks). The straight-line links indicate distances that are smaller than the critical radius.

the system, with an attenuation factor b = 0.55 fm and skin-thickness a = 0.33 fm. It is observed that as the critical radius increases larger clusters are favored, resulting in a non-monotonic behavior. In the case of A = 12 (carbon) results originating in a Fourier-Bessel initial radial distribution are shown for comparison (**Figure 4**). The difference in cluster probabilities between Woods-Saxon and Fourier-Bessel is small. For this work the Woods-Saxon distribution is used.

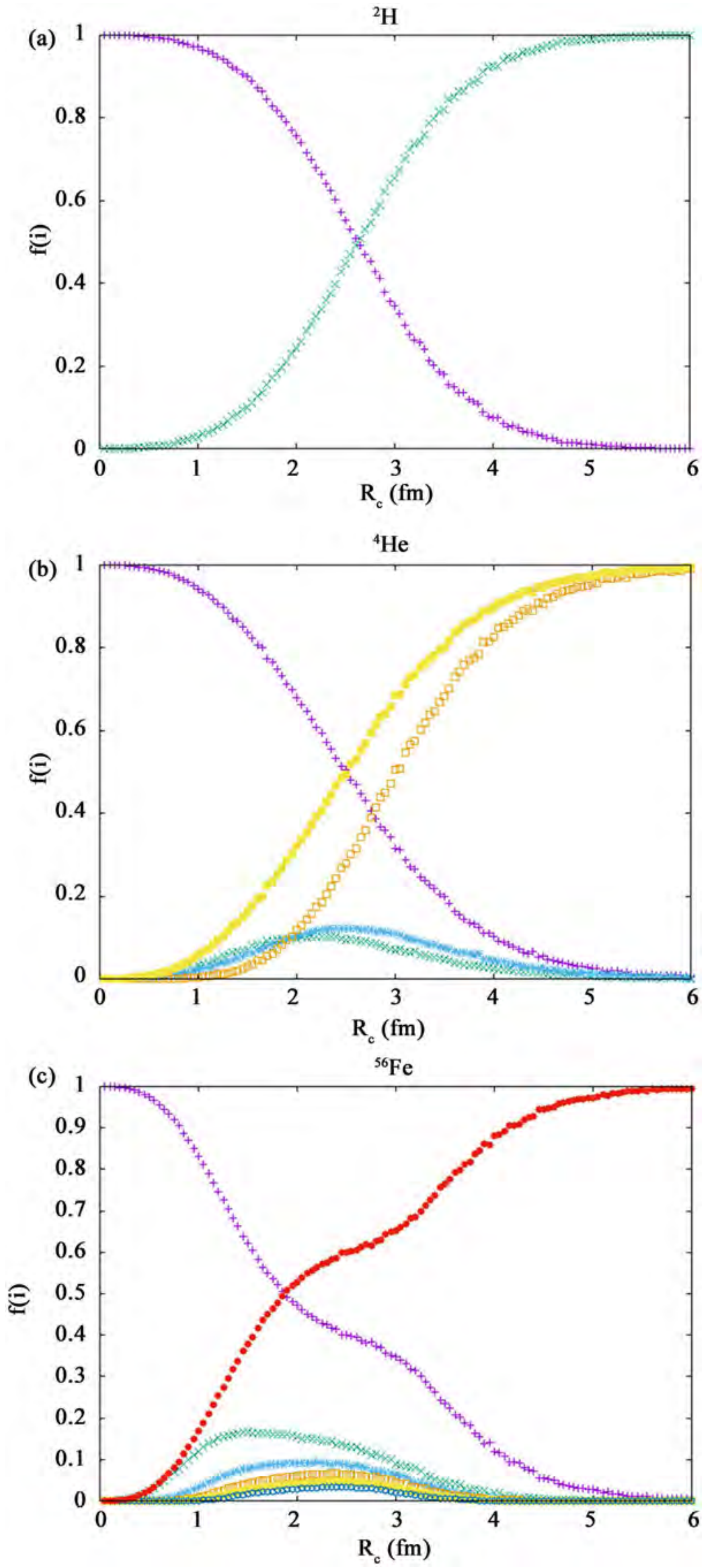
The free parameter in this calculation is the critical radius. This determines the relative values of cluster probabilities, hence the departure from single-nucleon physics. Earlier studies [7] [8] indicate that to account for diverse phenomena such as the EMC effect in DIS, nuclear, initial-state modifications of the Drell-Yan production, suppression of J/Ψ production in processes involving nuclei the cumulative cluster probability for deuterium should be about 4% and for heavy nuclei it could exceed 30% (**Table 1**). The approximate dependence of the cumulative probability on the mass number is logarithmic but for magic or doubly-magic nuclei this probability is higher. In practice, the critical radius will be treated as a parameter to be estimated using particle scattering data off nuclei.

It must be pointed out that the algorithm described in this section does not physically replace nucleons with a single object at the center of mass of the cluster while it is still running. It simply identifies networks of nucleons without changing the spatial distribution of matter (and charge) or moving the center of mass of the system.

4. Types of Clusters in Nuclei

The parton distributions in clusters depend on the type of valance quarks in the nucleons they are made of. The 2nd-order clusters may consist of two protons (pp-type), two neutrons (nn-type) or a proton and a neutron (np-type). The 3rd-order clusters can be of nnn, ppp, npp or nnp-type. A similar designation can be assigned to higher-order clusters. Therefore, a model is needed in order to calculate how many clusters of each type are present in a given nucleus at each value of the critical radius.

To derive the cluster numbers for a nucleus of mass number A and atomic number Z certain constraints must apply. 1) The total mass of the nucleus must be fixed. In other words, the total baryon number of all clusters must remain


(a)
²H
f(i)
R_c (fm)
(b)
⁴He
f(i)
R_c (fm)
(c)
⁵⁶Fe
f(i)
R_c (fm)

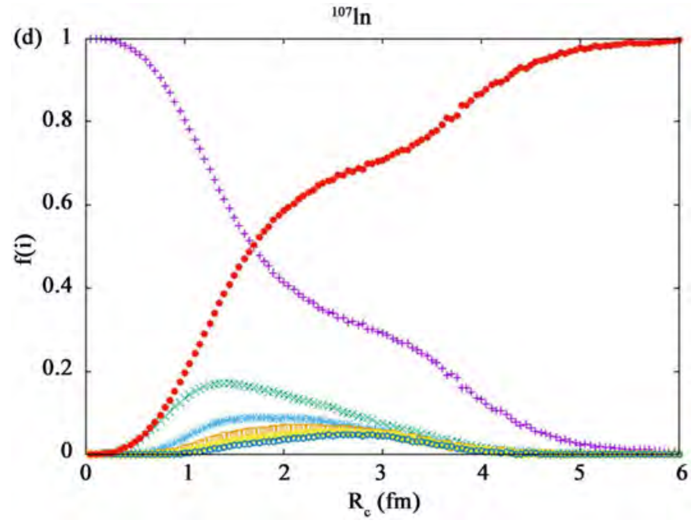


Figure 3. Cluster probabilities versus critical radius in fm. (a) A = 2. The brown points correspond to the two-nucleon cluster. (b) A = 4. The brown, green, black and purple points correspond to 1st, 2nd, 3rd and 4th-order clusters. The light-blue points indicate the cumulative cluster probability. (c) A = 56. The brow, green, dark-blue, purple and bright-green points correspond to 1st, 2nd, 3rd, 4th and 5th-order clusters. The black points indicate the cumulative cluster probability. (d) A = 107. The notation is the same as in the lower left panel. All these results are produced using Woods-Saxon radial probability density.

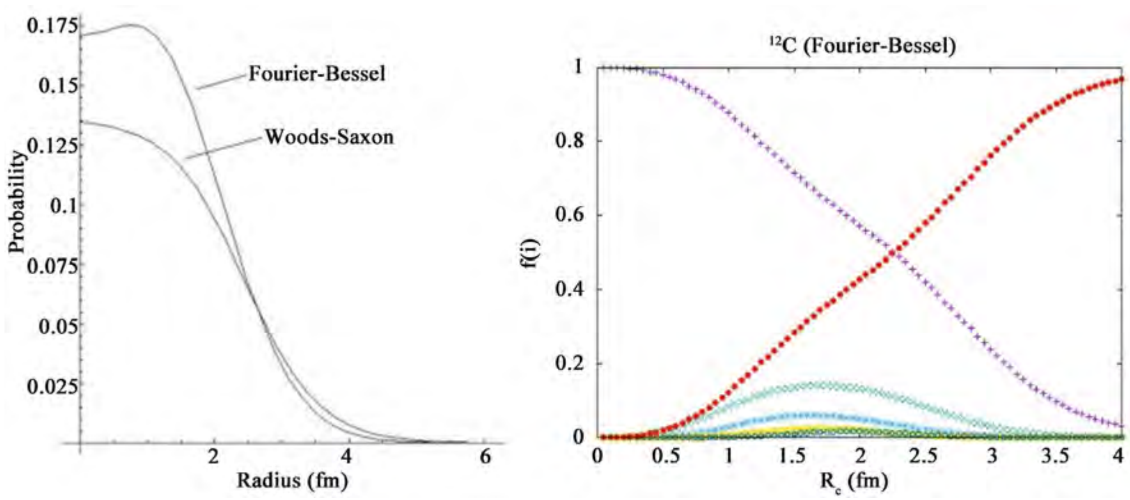


Figure 4. Left panel: the Woods-Saxon and the experimentally-derived Fourier-Bessel radial probability distributions versus the distance from the center of mass. For ^{12}C. Right panel: Cluster probabilities for ^{12}C versus the critical radius.

Table 1. A sample list of cumulative cluster probabilities for two values of the critical radius. For ^{12}C the results produced using Woods-Saxon (WS) and Fourier-Bessel radial distributions are shown for comparison. For ^{197}Au the probability can be more than 0.30.

Nucleus	Cluster Probability (f) for R_c = 0.85 fm	Cluster Probability (f) for R_c = 0.85 fm
^{2}H	0.04	0.09
^{12}C (WS)	0.06	0.13
^{12}C (FB)	0.08	0.17
^{197}Au	0.15	0.30

equal to A. 2) The total charge of the nucleus must be fixed. Thus, the sum of the charges (contributed only by proton components) is equal to Z. In addition, a further natural assumption is made. 3) The number of p (or n, pn, nn etc.)-type clusters must be proportional to the number of protons (or number of neutrons, number of protons times the number of neutrons, number of neutrons squared etc.) in the nucleus. In this work clusters only up to 4$^{\text{th}}$ order will be considered. The highest order clusters, containing 12 valence quarks, will be assigned an effective probability to occur equal to $f_4 = 1 - f_1 - f_2 - f_3$ and generic parton distributions (see next section). In order to accommodate for this truncation, the nucleus must be treated as having an effective mass number $A' = A(1 - f_4)$ and an effective atomic number $Z' = Z(1 - f_4)$. These equations guarantee mass and charge conservation. Then the numbers of 1$^{\text{st}}$, 2$^{\text{nd}}$ and 3$^{\text{rd}}$-order clusters of all the possible types can be calculated solving a linear system of 10 equations, resulting in

$$n_p = \frac{f_1 Z'}{1 + f_2 + 2f_3}, \quad n_n = \frac{f_1 N'}{1 + f_2 + 2f_3}, \tag{1.1}$$

$$\begin{aligned} n_{pp} &= \frac{f_2 Z'^2}{(1 + f_2 + 2f_3)(N' + Z')}, \quad n_{pn} = \frac{2 f_2 Z' N'}{(1 + f_2 + 2f_3)(N' + Z')}, \\ n_{nn} &= \frac{f_2 N'^2}{(1 + f_2 + 2f_3)(N' + Z')}, \end{aligned} \tag{1.2}$$

$$n_{ppp} = \frac{f_3 Z'^3}{(1 + f_2 + 2f_3)(Z' + N')^2}, \quad n_{ppn} = \frac{3 f_3 Z'^2 N'}{(1 + f_2 + 2f_3)(Z' + N')^2}, \tag{1.3a}$$

$$n_{pnn} = \frac{3 f_3 Z' N'^2}{(1 + f_2 + 2f_3)(Z' + N')^2}, \quad n_{nnn} = \frac{f_3 N'^3}{(1 + f_2 + 2f_3)(Z' + N')^2}, \tag{1.3b}$$

$$n_4 = \frac{A - n_1 - 2n_2 - 3n_3}{4}. \tag{1.4}$$

Here the notation is self-explanatory. For example, n_{pnn} indicates the number of 3$^{\text{rd}}$-order clusters of the type containing one proton and two neutrons. The effective number of neutrons is $N' = A' - Z'$. Also n_1, n_2, n_3 and n_4 are the total number of clusters for 1$^{\text{st}}$, 2$^{\text{nd}}$, 3$^{\text{rd}}$ and 4$^{\text{th}}$ order. The last equation involves the real mass number A. The 4$^{\text{th}}$-order clusters are not separated into various possible types.

5. Parton Distributions in Clusters

The parton distributions in nucleons or higher-order clusters describe the probability that a certain type of parton (valence or ocean quark or gluon) carries a certain fraction of the cluster momentum in a reference frame where the (longitudinal) momentum of the nucleus is very large. In this case motion is essentially one-dimensional and the transverse components of the parton momentum are usually and legitimately neglected. The momentum fraction for single nucleons (1$^{\text{st}}$-order clusters) is defined as the Bjorken-x (x_{Bjorken}) and ranges from 0 to 1.

For 2nd, 3rd or higher-order clusters the upper limit of the momentum fraction is 2, 3 or more, relative to single nucleons as a parton can carry up to 2, 3, or more times the nucleon momentum. The parton distributions in general also depend on the 4-momentum exchanged between the probe (lepton or quark) and the parton, denoted as Q^2. This is the invariant mass of the exchange quantum, be it a photon or gluon or an intermediate vector boson. However, when Q^2 is large, in excess of 2 GeV^2, this dependence becomes very weak, a phenomenon known as Q^2-scaling. The phenomena discussed in this article originate in the Q^2-scaling region, thus the distributions used will depend only on x_{Bjorken}. Scaling violations are relevant at very small x_{Bjorken} [22] [23] [27] [28] [29]. The calculations in this article intend to interpret data in the region $x_{\text{Bjorken}} > 1$ where the presence of higher-order clusters becomes prevalent. In the equations that follow x refers to the parton momentum fraction relative to the cluster in which the parton is found. Therefore, it is equal to $x_{\text{Bjorken}}/2$ for 2nd-order clusters, $x_{\text{Bjorken}}/3$ for 3rd-order and $x_{\text{Bjorken}}/4$ for 4th-order ones.

Inside clusters there are valence quarks, ocean (or sea) quarks with equal numbers of ocean antiquarks and gluons. In order to determine the parameters of the parton distributions certain constraints must apply. 1) Since the number of valence quarks of each flavor is fixed, the valence distributions should be normalized to this number. 2) The total momentum fractions carried by all partons must add up to the momentum of the cluster. Additional model assumptions are also needed. 3) The gluon to total-ocean momentum ratio is fixed at 1/5. 4) The strange quark distribution in the ocean is chosen to carry one half of the up-quark (or antiquark) ocean distribution. Naturally the antiquark distributions in the ocean are the same as those of the ocean quarks of the same flavor. The parton momentum distributions, *i.e.*, the probability distributions multiplied by x are assumed to have the following forms:

$$V^{u,d}(x) = B^{u,d}\sqrt{x}(1-x)^{b^{u,d}} \quad \text{(valence up and down quarks)}, \tag{2.1}$$

$$S(x) = A(1-x)^{a} \quad \text{(sea quarks or antiquarks, one flavor)}, \tag{2.2}$$

$$G(x) = C(1-x)^{c} \quad \text{(total gluon)}, \tag{2.3}$$

where A, $B^{u,d}$ and C are constants. For the valence exponents, $b^d = b^u + 1$ as supported by scattering data [Ref]. If the total momentum fraction carried by one ocean quark flavor is x^s and that carried by the gluons is x^g then $A = x^s(1+a)$ and $C = x^g(1+c)$. This generic notation applies to clusters of all orders.

Isospin relations connect the distributions in various types of clusters of the same order.

$$V_p^u(x) = V_n^d(x), \tag{2.4a}$$

$$V_{pp}^u(x) = V_{nn}^d(x),\; V_{pn}^u(x) = V_{pn}^d(x) = \frac{1}{2}\left[V_{pp}^u(x) + V_{pp}^d(x)\right], \tag{2.4b}$$

$$V_{ppn}^u(x) = V_{pnn}^d(x) = \frac{5}{2}\left[\frac{1}{6}V_{ppp}^u(x) + \frac{1}{3}V_{ppp}^d(x)\right], \tag{2.4c}$$

$$V_{ppn}^{d}(x)=V_{pnn}^{u}(x)=2\left[\frac{1}{6}V_{ppp}^{u}(x)+\frac{1}{3}V_{ppp}^{d}(x)\right]. \tag{2.4d}$$

The ocean and gluon distributions are the same in all types of clusters of the same order. The exponents of (1 − x) in these distributions determine their overall behavior. Specifically, the higher the exponent, the "softer", *i.e.*, more concentrated towards lower x the distribution is. The choice of exponents is guided by the DIS data themselves. It turned out that the specific values of ocean and gluon exponents for 3rd and higher-order clusters are not as essential as is the actual presence of such clusters in cross-section calculations. Even though to a first approximation the gluon distributions do not enter the calculations of DIS of electrons off nuclei, the gluon distributions have to be considered because they affect the total momentum ratios carried by ocean quarks. The exponents for proton-like clusters that produce the best fits to the EMC and QES data (to be discussed) are shown in **Table 2**. The exponents in other types of clusters of the same order are determined by the isospin relations. For valence, the exponents related to up quarks are given. For valence down quarks the exponents are simply increased by one unit.

Application of the above mentioned constraints and isospin relations yields the multiplicative constants that appear in the distributions. The calculations involve extensive use of Beta-functions. The results are shown in **Table 3** for proton-like clusters. For the ocean and gluon distributions the total momentum fraction is presented as it is proportional to the respective coefficient. The isospin relations determine the coefficients in all other types of clusters of all orders. In the case of 4th-order clusters the up and down valence quark distributions are taken to be the same. It may be worth noting that the ocean and gluon total

Table 2. Exponents of parton distributions in proton-like clusters. The exponents in other types of clusters of the same order are determined by the isospin relations. For 4th-order clusters the valence distribution is generic.

	p	pp	ppp	pppp
Up-Valence (b)	3	9	15	21
Ocean (a)	9	11	18	23
Gluon (c)	6	10	16	20

Table 3. Coefficients in parton distributions for all orders of proton-like clusters used. The momentum fraction for the ocean pertains to one up quark or antiquark flavor. The 4th-order valence distributions are generic so that the up and down coefficients are the same.

	1st-order	2nd-order	3rd-order	4th-order
B^u (valence u-quark)	2.18750	7.04788	13.43519	31.57550
B^d (valence d-quark)	1.23046	3.70014	6.92752	31.57550
x^s (ocean, one flavor)	0.02289	0.02409	0.02442	0.02444
x^g (total gluon)	0.57239	0.60214	0.61039	0.61111

momentum fractions become saturated as the order of clusters increases.

6. Rescaled Bjorken-x

In a nucleus the physical size of bound nucleons (RMS radius of their wave-function) may be slightly altered compared to free nucleons [30] [31]. This is due to the attraction from surrounding nucleons and it increases their size. This phenomenon clearly affects correlations as it increases the probability for cluster formation amounting to an effective decrease in the critical radius. Furthermore, it results in a decrease of the parton momentum fraction in bound nucleons or higher-order clusters, relative to "free" ones. It can be accounted for by rescaling x_{Bjorken} by a factor η:

$$x_{bound} = \eta x_{free} \quad (\eta < 1), \tag{3}$$

where η is a constant parameter that does not appreciably depend on the energy scale.

7. Structure Functions of Nuclei

The structure function of a nucleus with mass number A and atomic number Z is defined as the sum of all quark momentum distributions over all nucleons and higher-order clusters multiplied by the square of the charge of the quark. It is denoted by $F_2^{(A,Z)}(x)$, $x = x_{\text{Bjorken}}$.

The presence of correlations, described here as clusters, modifies the structure function relative to the one calculated for uncorrelated nucleons. The forms of the distribution functions are given by Eqs. (2.1, 2.2 and 2.3). In the construction of the structure function the cluster numbers, strange-to-up quark ocean ratios and isospin relations are employed to express the result in terms of only proton-like distributions. A straightforward calculation, thus, yields:

$$F_2^{(A,Z)}(x) = \frac{4}{9}V_{A,Z}^u(x) + \frac{1}{9}V_{A,Z}^d(x) + \frac{11}{9}S_{A,Z}(x), \tag{4.1}$$

where the total up ($V_{A,Z}^u$) down ($V_{A,Z}^d$) valence quark and ocean ($S_{A,Z}$) contributions are

$$\begin{aligned} V_{A,Z}^u(x) &= n_p V_p^u(x) + n_n V_p^d(x) + n_{pp} V_{pp}^u(x) + n_{pn}\frac{1}{2}\left[V_{pp}^u(y) + V_{pp}^d(y)\right] \\ &\quad + n_{nn} V_{pp}^d(y) + n_{ppp} V_{ppp}^u(z) + n_{ppn}\frac{5}{2}\left[\frac{1}{6}V_{ppp}^u(z) + \frac{1}{3}V_{ppp}^d(z)\right] \\ &\quad + n_{pnn}\left[\frac{1}{3}V_{ppp}^u(z) + \frac{2}{3}V_{ppp}^d(z)\right] + n_{nnn} V_{ppp}^d(z) + n_4 V_{pppp}^u(w), \end{aligned} \tag{4.2}$$

$$\begin{aligned} V_{A,Z}^d(x) &= n_p V_p^d(x) + n_n V_p^u(x) + n_{pp} V_{pp}^d(x) + n_{pn}\frac{1}{2}\left[V_{pp}^u(y) + V_{pp}^d(y)\right] \\ &\quad + n_{nn} V_{pp}^u(y) + n_{ppp} V_{ppp}^d(z) + n_{ppn}\left[\frac{1}{3}V_{ppp}^u(z) + \frac{2}{3}V_{ppp}^d(z)\right] \\ &\quad + n_{pnn}\frac{5}{2}\left[\frac{1}{6}V_{ppp}^u(z) + \frac{1}{3}V_{ppp}^d(z)\right] + n_{nnn} V_{ppp}^u(z) + n_4 V_{pppp}^u(w), \end{aligned} \tag{4.3}$$

$$S_{A,Z}(x) = n_1 S_p(x) + n_2 S_{pp}(y) + n_3 S_{ppp}(z) + n_4 S_{pppp}(w), \quad (4.4)$$

and $y = x/2$, $z = x/3$ and $w = x/4$. In Equation (4) the variable x is the Bjorken-x rescaled by the parameter η.

8. Heavy-to-Light Nuclei Cross-Section Ratios

The interaction investigated in this work is Deep Inelastic Scattering (DIS) of electrons off nuclei in the Q^2-scaling region. The scattering cross section is proportional to the structure function. Therefore, the ratio of structure functions per nucleon for heavy over light nuclei studied as a function of x_{Bjorken} can reveal the presence and nature of correlations. Formally,

$$R(x) = \frac{F_2^{(A_1,Z_1)}(x)/A_1}{F_2^{(A_2,Z_2)}(x)/A_2}. \quad (5)$$

This is the ratio that revealed the EMC-effect. If no correlations or other modifications are present it should be equal to 1 for all x. However, data have shown that there is a strong, non-monotonic dependence of R on x even in the region $0 < x < 1$. Newer data show that R is measurable beyond $x = 1$. A step-wise behavior as x increases by a unit is a strong indication that the cluster model predicts correlations in the nucleus.

9. Comparison with Data

The calculations presented here are compared with data from JLab Experiment E02-019. The experiment used a 5.75 GeV electron beam and reported results for high Q^2. Final-state interactions are very small under this condition. To extract information on short-range correlations the data are focused on the QES region with cuts in the energy transfer imposed to remove DIS contributions. Therefore, the comparison may not expected to be complete. However, the x_{Bjorken}-dependence of cross section ratios must be similar for the following reasons: First, the quark-cluster model strongly overlaps with the Fermi-motion model [8]. Second, the nature of the correlations should be the same in both QES and DIS cases. There are strong indications for this. According to Ref. [32] [33] there is a remarkably linear relation between the EMC effect which manifests itself in the DIS region for $x_{\text{Bjorken}} < 1$ and the short-range correlations. In fact the A-dependence of the EMC effect closely matches that of the correlations effect according to Jefferson Lab data [34]. Because of these observations a comparison of the model derived using parton-distributions in the asymptotically-free region with the date in the QES region is possible. The comparison should be focused around $x_{\text{Bjorken}} = 1$ to evaluate the rise of the cross-section ratio and on the plateau that follows it.

First the 2007 data are addressed [35]. The free parameters in the model are R_c and η. The values of these parameters that produced the best fits to the 2007 data for various nuclei are shown in **Table 4**. Remarkably η turned out to have almost the same value for all nuclei studied. This parameter effectively controls the $x_{\text{Bjor-}}$

$_{ken}$ point at which the cross section ratio begins to rise.

The cluster formation probabilities for each order are shown in **Table 5** for the various nuclei studied. The values presented correspond to the critical radius shown in **Table 4**. The 4^{th}-order probability is the cumulative value for all orders higher than the 3^{rd}. In the model presented here even the dilute deuterium nucleus has a substantial probability for correlations.

In **Figure 5** the model calculations are compared with the experimental data from 2007. The curves represent the best obtained fits of the theoretical ratio, R, for four cases. It is immediately observed that the sharp rise of the data right after $x = 1$ is well described by the model. However, the dip around $x = 1$ is not. There is a pseudo-plateau in the data for $1 < x < 1.8$ that is almost followed by the curves. However, the shape of the data in the case of the ^{12}C over ^{3}He ratio is not reproduced well. It must be noted that the parameters used for each nucleus are maintained for all curves. A better fit to the ^{12}C over ^{3}He data would result in a worse fit to the ^{12}C over ^{2}H data. Overall the curves involving ^{2}H are closer to the data that those for ^{3}He. The experimental uncertainties in these plots are significant, though, and ^{3}He may be an unusual case.

The same experiment produced more detailed data published in 2012 [36]. Fitting the theoretical model to the new data with the two independent parameters, R_c and η, resulted in slightly different values, presented in **Table 6**. The resulting cluster formation probabilities are shown in **Table 7**. For ^{63}Cu, ^{12}C and ^{3}He a larger critical radius was needed with the same value of the rescaling parameter while for ^{2}H the rescaling parameter was smaller, equal to 0.76. Stronger correlations were, thus, found for these nuclei based on the 2012 data.

The fit to data in which ^{2}H is in the denominator of the ratio is better, as shown in **Figure 6**. It is clear, though, that the model overestimates the ratio at high x. This is to some extent the result of the truncation of the model to the 4^{th} –order clusters but mostly of the fact that the denominator, pertaining to ^{2}H,

Table 4. The critical radius R_c and the rescaling parameter η for the various nuclei studied. These values fit the data from E02-019 (2007) [35].

	^{63}Cu	^{12}C	^{3}He	^{2}H
R_c (fm)	0.75	1.20	1.50	1.40
η	0.80	0.80	0.80	0.80

Table 5. The cluster formation probabilities for the nuclei studied. They correspond to the critical radii shown in **Table 4**. These values fit the data from E02-019 (2007) [35].

	f_1	f_2	f_3	f_4
^{63}Cu	0.7549	0.1402	0.0512	0.0537
^{12}C	0.8148	0.1093	0.0438	0.0321
^{3}He	0.8564	0.1212	0.0224	---------
^{2}H	0.9176	0.0824	---------	---------

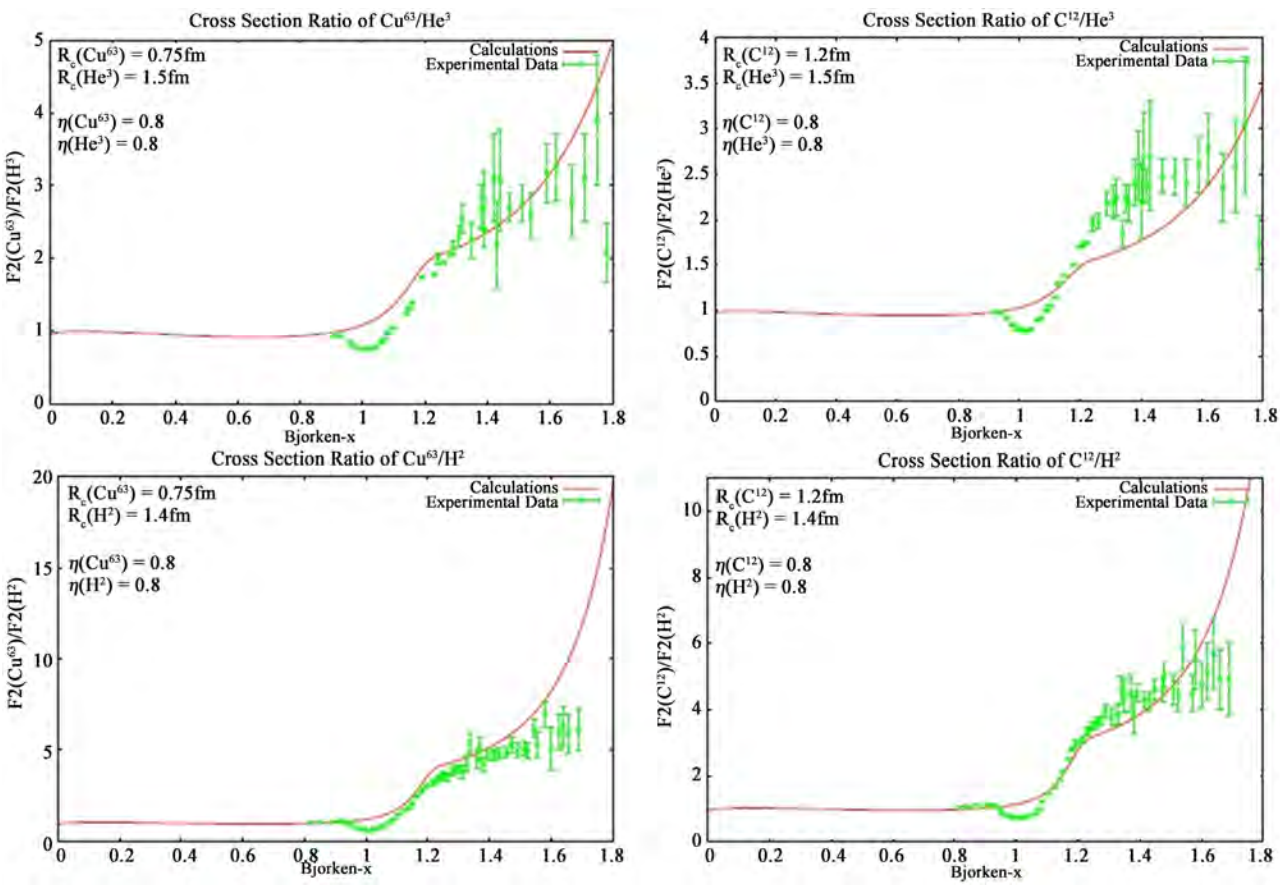


Figure 5. Comparison of the model calculations with data from E02-019 (2007) [35]. The nuclei involved are shown inside the plots. Upper left: ^{63}Cu/^{3}He. Upper right: ^{12}C/^{3}He. Lower left: ^{63}Cu/^{2}H. Lower right: ^{12}C/^{2}H. The model reproduces the sharp rise for $x > 1$ but fails to reproduce the dip at $x = 1$. The worst fit is for ^{12}C over ^{3}He data.

Table 6. The critical radius R_c and the rescaling parameter η for the various nuclei studied. These values fit the data from E02-019 (2012) [36].

	^{197}Au	^{63}Cu	^{12}C	^{9}Be	^{4}He	^{3}He	^{2}H
R_c (fm)	1.20	0.85	1.50	1.60	1.80	1.70	1.40
η	0.80	0.80	0.80	0.80	0.80	0.80	0.76

Table 7. The cluster formation probabilities for the nuclei studied. They correspond to the critical radii shown in **Table 4**. These values fit the data from E02-019 (2012) [36].

	f_1	f_2	f_3	f_4
^{197}Au	0.6772	0.1711	0.0704	0.0813
^{63}Cu	0.6960	0.1497	0.0629	0.0914
^{12}C	0.7151	0.1334	0.0597	0.0918
^{9}Be	0.7657	0.1179	0.0476	0.0688
^{4}He	0.7869	0.0741	0.0511	0.0879
^{3}He	0.7806	0.1692	0.0501	---------
^{2}H	0.9176	0.0824	---------	---------

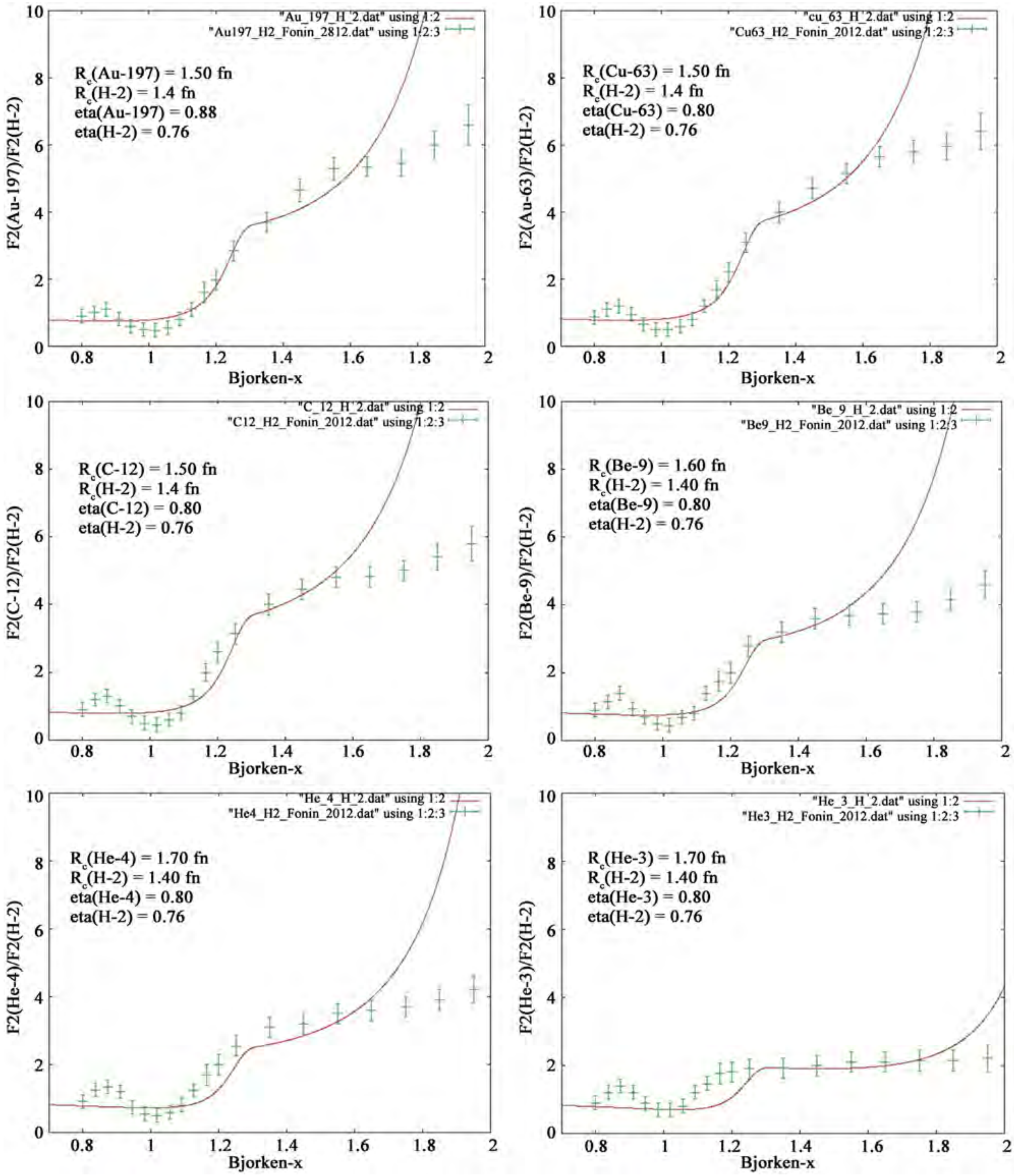


Figure 6. Comparison of the model calculations with data from E02-019 (2012) [36]. The nuclei involved are shown inside the plots. Upper left: $^{197}Au/^{2}H$. Upper right: $^{63}Cu/^{2}H$. Middle left: $^{12}C/^{2}H$. Middle right: $^{9}Be/^{2}H$. Lower left: $^{4}He/^{2}H$. Lower right: $^{3}He/^{2}H$. The model reproduces the sharp rise for $x > 1$.

approaches zero as x tends to 2 since there are no 3rd-order clusters in ^{2}H. The case of ^{3}He over ^{2}H ratio is most interesting in this regard. The level of the ^{3}He over ^{2}H plateau is described very well by the model as the curve remains flat up to $x = 1.8$. This indicates that the 2nd-order correlations are accounted for successfully. Combined with the other results the flat-behavior of the data may imply the presence of another mechanism that extends the 2nd-order correlations possibly beyond $x = 2$ or that the data are inaccurate near the kinematic limits of the experiment. Alternatively the theoretical ratio should be lower at high x by allowing the parton distributions in clusters to be softer. This, however, would make it difficult to account for the exact slope of the rise at $x = 1$.

The differences in the critical radii used to fit the two sets of data merits some discussion. It must be pointed out that the fits are done over the entire sets of data so that the parameters for each nucleus are kept the same for the various nuclei ratios.

The observation that the two sets of data required slightly different values of the free parameters of the model, most importantly the critical radius that controls the strength of correlations, should be discussed. Clearly the data presented in 2007 carry substantially more uncertainty. Furthermore, there are differences in the average scattering angle of the emerging electron, *i.e.*, in the average Q^2. The 2007 data used here correspond to an average $Q^2 = 4.1\ \text{GeV}^2$ while the data presented in 2012 are for $Q^2 = 1.5\ \text{GeV}^2$ and have the best kinematic range coverage and statistics. There may be a small scaling violation. The trends, however, are similar. As a result of these observations, the difference in the fitting parameters must result mostly from statistical differences in the data. The experiment has provided data for other scattering angles but with lower statistics. Members of the E02-019 Collaboration have elaborated on their measurements in Ref. [37].

An interesting extension and challenge is the comparison with data in the $2 < x < 3$ range. Such data were produced by the CLAS collaboration [38] and even though they are limited, they indicate the existence of a second plateau. The study of this range within the model is currently undertaken.

10. The Need for Many-Body Correlations

The existence of many-body correlations among bound nucleons is an irrefutable fact based on the data. To further elucidate this conclusion, the theoretical curves including 2nd, 3rd and 4th-order clusters are compared to 2007 data in **Figure 7**. The specific case of ^{63}Cu over ^{3}He is shown. It is clear that without up to 4th-order clusters the data cannot be reproduced. It is remarkable that even in the range $x < 2$ the higher-order correlations have a significant effect. The nucleon-nucleon clusters also affect the cross-section ratio for $x < 1$ but to a lesser extent.

11. Conclusion

In this work it was shown that many-body correlations in the nucleus are necessary in order to understand Quasi-Elastic Scattering (QES) data. This is in agreement with the need for such effects in Deep Inelastic Scattering (DIS) data.

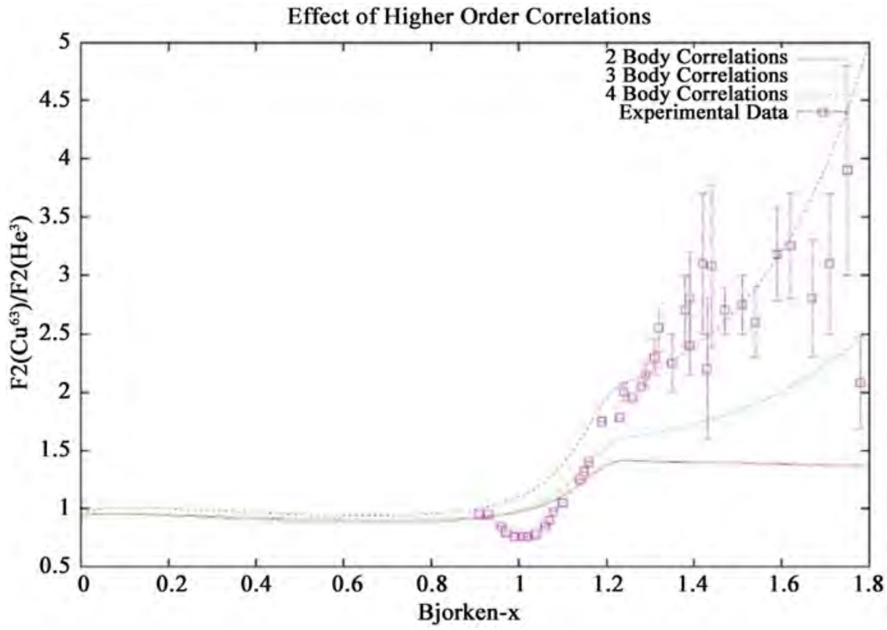


Figure 7. The ^{63}Cu to ^{3}He ratio. The curves correspond to theoretical calculations including 2nd (red), 3rd (green) and 4th (purple)-order clusters. The higher-order correlations are needed to reproduce the data for $1 < x < 2$. The data are from Ref. [35].

The nature of the correlations appears to be the same. A quark-cluster model has been constructed to account for these correlations. This assumes overlapping of bound-nucleon wave-functions to the extent that they share partons whose momentum distributions are consequently softened compared to those in independent nucleons. A network-based algorithm was developed in order to calculate the formation probabilities for all orders of clusters. This was supplemented by a model that produces the actual number of clusters of each type in every nucleus as well as a set of reasonable parton distributions. Comparison with two sets of data from Jefferson Laboratory helped extract the model parameters but the values obtained were slightly different for the two sets. The model generally reproduces the quasi-plateau in cross section ratios for $x_{\text{Bjorken}} > 1.0$ and the slope of the rising curve in the range $1.0 < x_{\text{Bjorken}} < 1.2$. Even in the studied region up to 4-body correlations are needed. The obvious next step is to study the extension of the model for $x_{\text{Bjorken}} > 2$. There are also data by the CLAS Collaboration to compare the theoretical calculations to [38]. However, certain questions arise. Is the coordinate-space network algorithm adequate to compute the cluster probabilities? So far it has given results that agree with data but the effect of relative momentum among nucleons can be considered. The small differences between parameters fitting the two sets of data may be attributable to different average Q^2. Therefore, one can ask what is the exact Q^2-dependence of many-body correlations? The similarity of the correlations between DIS and QES is intriguing but what would their mathematical formulation be if the parton-degrees of freedom are not entirely resolved? These questions are the subject of current and future investigations.

Conflicts of Interest

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

References

[1] EMC, Aubert, J.J., *et al.* (1983) *Physics Letters*, **123B**, 275.

[2] Bodek, A., *et al.* (1983) *Physical Review Letters*, **50**, 1431. https://doi.org/10.1103/PhysRevLett.50.1431

[3] Bodek, A., *et al.* (1983) *Physical Review Letters*, **51**, 534. https://doi.org/10.1103/PhysRevLett.51.534

[4] EMC-NA38 Collaboration, Arneodo, M. *et al.* (1990) *Nuclear Physics*, **B333**, 1.

[5] NMC Collaboration, Amaudriuz, P., *et al.* (1991) *Zeitschrift für Physik*, **51**, 387.

[6] Kitagaki, T., *et al.* (1989) Results from FNAL E745 on Neutrino Nucleus Interactions (EMC Effect and Hadron Formation). *Proceedings of Electronuclear Physics with Internal Targets Conference*, Stanford, 95-115.

[7] Bordalo, P., *et al.* (1987) *Physics Letters B*, **193**, 368-372. https://doi.org/10.1016/0370-2693(87)91253-6

[8] NA3 Collaboration, Badier, J., *et al.* (1983) *Zeitschrift für Physik*, **20**, 101-116. https://doi.org/10.1007/BF01573213

[9] Alde, D.M., *et al.* (1991) *Physical Review Letters*, **66**, 133.

[10] Kowitt, M.S., *et al.* (1994) *Physical Review Letters*, **72**, 1318.

[11] Petridis, A., Lassila, K.E. and Vary, J.P. (1993) *Physical Review D*, **47**, 1906. https://doi.org/10.1103/PhysRevD.47.1906

[12] Jalilian-Marian, J. (2005) *Nuclear Physics A*, **753**, 307-315. https://doi.org/10.1016/j.nuclphysa.2005.02.156

[13] Betemps, M.A. and Gonçalves, V.P. (2008) *Journal of High Energy Physics*, **9**, 19. https://doi.org/10.1088/1126-6708/2008/09/019

[14] Pirner, H.J. and Vary, J.P. (1981) *Physical Review Letters*, **46**, 1376. https://doi.org/10.1103/PhysRevLett.46.1376

[15] Vary, J.P. (1984) *Nuclear Physics*, **418**, 195-214. https://doi.org/10.1016/0375-9474(84)90550-5

[16] Vary, J.P., Coon, S.A. and Pirner, H.J. (1985) Charge form Factors and Quark Clusters. *Proceedings of the International Conference on Hadronic Probes and Nuclear Interactions*, Tempe, 11-14 March 1985, 83-95.

[17] Lassila, K.E. and Sukhatme, U.P. (1988) *Physics Letters B*, **209**, 343-346. https://doi.org/10.1016/0370-2693(88)90959-8

[18] Harindranath, A. and Vary, J.P. (1986) *Physical Review D*, **34**, 3378. https://doi.org/10.1103/PhysRevD.34.3378

[19] Lassila, K.E., Petridis, A., Carlson, C. and Sukhatme, U.P. (1990) Multiquark Effects and Multiquark Fragmentation. *Proceedings of the Rice Meeting of the Division of Particles and Fields*, Houston, January 3-6 1990, 593.

[20] Lassila, K.E., Sukhatme, U.P., Harindranath, A. and Vary, J.P. (1991) *Physical Review C*, **44**, 1188. https://doi.org/10.1103/PhysRevC.44.1188

[21] Lassila, K.E., Petridis, A., Sukhatme, U.P. and Wilk, G. (1992) *Physics Letters B*, **297**, 191-196. https://doi.org/10.1016/0370-2693(92)91091-M

[22] Petridis, A. (1996) *Physical Review C*, **54**, 848.
https://doi.org/10.1103/PhysRevC.54.848

[23] Petridis, A. (2018) *International Journal of Applied and Natural Sciences*, **7**, 75.

[24] Pirner, H.J. and Vary, J.P. (2011) *Physical Review C*, **84**, 015201.
https://doi.org/10.1103/PhysRevC.84.015201

[25] Reuter, W., *et al.* (1982) *Physical Review C*, **26**, 806.
https://doi.org/10.1103/PhysRevC.26.806

[26] Anni, R., Co', G. and Pellegrino, P. (1995) *Nuclear Physics*, **584**, 35-59.
https://doi.org/10.1016/0375-9474(94)00508-K

[27] Gribov, L., Levin, E. and Ryskin, M. (1981) *Nuclear Physics*, **188**, 555-576.
https://doi.org/10.1016/0550-3213(81)90007-9

[28] Close, F.E., Qiu, J. and Roberts, R.G. (1989) *Physical Review D*, **40**, 2820.
https://doi.org/10.1103/PhysRevD.40.2820

[29] Eskola, K., Qiu, J. and Wang, X.-N. (1994) *Physical Review Letters*, **72**, 36.
https://doi.org/10.1103/PhysRevLett.72.36

[30] Li, G.L., Liu, K.F. and Brown, G.E. (1988) *Physics Letters B*, **213**, 531-536.
https://doi.org/10.1016/0370-2693(88)91305-6

[31] Close, F.E., Roberts, R.G. and Ross, G.G. (1984) *Physics Letters B*, **142**, 202-206.
https://doi.org/10.1016/0370-2693(84)91262-0

[32] Weinstein, L.B., Piasetzky, B.E., Higinbotham, D.W., Gomez, J., Hen, O. and Shneor, R. (2011) *Physical Review Letters*, **106**, 052301.
https://doi.org/10.1103/PhysRevLett.106.052301

[33] Hen, O, Piasetzky, E. and Weinstein, L.B. (2012) *Physical Review C*, **85**, 047301.
https://doi.org/10.1103/PhysRevC.85.047301

[34] Seely, J., *et al.* (2009) *Physical Review Letters*, **103**, 202301.

[35] Fomin, N. (2007) arXiv: 0808.2625.

[36] Fomin, N., *et al.* (2012) *Physical Review Letters*, **108**, 092502.

[37] Arlington, J., *et al.* (2012) arXiv: 1206:6343.

[38] CLAS Collaboration, Egiyan, K. S., *et al.* (2006) *Physical Review Letters*, **96**, 082501.

Journal of Modern Physics, 2022, 13, 66-80
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Multi-Energy Gamma-Ray Attenuations for Non-Destructive Detection of Hazardous Materials

Kaylyn Olshanoski, Chary Rangacharyulu

Department of Physics and Engineering Physics, University of Saskatchewan, Saskatoon, Canada
Email: chary.r@usask.ca

How to cite this paper: Olshanoski, K. and Rangacharyulu, C. (2022) Multi-Energy Gamma-Ray Attenuations for Non-Destructive Detection of Hazardous Materials. *Journal of Modern Physics*, **13**, 66-80.
https://doi.org/10.4236/jmp.2022.131005

Received: November 24, 2021
Accepted: January 23, 2022
Published: January 26, 2022

Abstract

We present a non-destructive method (NDM) to identify minute quantities of high atomic number (*Z*) elements in containers such as passenger baggage, goods carrying transport trucks, and environmental samples. This method relies on the fact that photon attenuation varies with its energy and properties of the absorbing medium. Low-energy gamma-ray intensity loss is sensitive to the atomic number of the absorbing medium, while that of higher-energies vary with the density of the medium. To verify the usefulness of this feature for NDM, we carried out simultaneous measurements of intensities of multiple gamma rays of energies 81 to 1408 keV emitted by sources ^{133}Ba (half-life = 10.55 y) and ^{152}Eu (half-life = 13.52 y). By this arrangement, we could detect minute quantities of lead and copper in a bulk medium from energy dependent gamma-ray attenuations. It seems that this method will offer a reliable, low-cost, low-maintenance alternative to X-ray or accelerator-based techniques for the NDM of high-Z materials such as mercury, lead, uranium, and transuranic elements etc.

Keywords

Non-Destructive Detection, Multi-Energy Photons, Radioactive Sources, Intensity Measurements, Safety and Security, XCOM Calculations

1. Introduction

Contaminations of high atomic number (*Z*) materials are of concern for both security and health reasons [1] [2]. Governments and industries are paying much attention to detecting these elements in goods during transportation or present in our surroundings, either water resources or agricultural environments. Currently, nuclear detection techniques mostly involve the use of X-ray beams of a few hundred keV, while high-energy gamma rays, or neutron beams from particle accelerators are considered as candidate tools [1] [2]. However, the latter technologies involve high-energy accelerators requiring high maintenance, large real estate, and the associated high costs of infrastructure and operations. They are also beset by the low yields of nuclear interactions. In this work, we suggest that multi-energy gamma-ray tomography, making use of photon emissions of radioactive isotopes of a few years of half-life, offers a promising tool. Below, we present the background physics, our choice of radioactive isotopes for the current work, and experimental data with a single detector to mimic a traveler's suitcase inspection and comparison of the experimental data with the standard theoretical XCOM database [3] results, which can be extended to a tomographic setting. We also present the results of our XCOM based calculation for plutonium contraband material smuggling and mercury contamination in water.

Background Physics

As is the case with X-rays, multi-energy gamma-ray tomography relies on the variation of photon intensity attenuation coefficients in material media. The intensity I of photons of energy E_γ as they pass through a thickness x in a material medium of density ρ and atomic number Z can be written as:

$$I\left(E_\gamma, Z, \rho, x\right) = I_0 \mathrm{e}^{-\mu_\rho\left(E_\gamma, Z\right)\cdot\rho x} \tag{1}$$

where I_0 is the intensity at the entrance of the medium, $\mu_\rho\left[\mathrm{cm}^2/\mathrm{g}\right]$ is the mass attenuation coefficient of the medium which varies with photon energy. It should be noted that the mass attenuation coefficient μ_ρ is specific to the atomic number of the medium and the photon energy. It does not depend on the density nor on the physical state (solid, liquid or gas) of the medium. The mass attenuation coefficients for pure elements can be found for energies in the range 1 keV - 100 GeV from the XCOM database [4], which can be easily extended to compounds and mixtures through Bragg's formula [5]. XCOM also provides the data for compounds and mixtures.

In **Figure 1**, we plot the mass attenuation coefficients of 1 keV - 10 MeV photons for six elements: carbon (Z = 6), oxygen (Z = 8), aluminum (Z = 13), copper (Z = 29), tin (Z = 50) and lead (Z = 82) covering nearly the entire periodic table of stable materials of interest. Two conspicuous features present themselves: 1) for low energies of $E_\gamma < 0.1\,\mathrm{MeV}$, the mass attenuation coefficients vary by a factor of about 10 - 100, and they exhibit discontinuities characteristic of absorption edges of the elements, 2) the coefficients are nearly constant for $1 < E_\gamma\left[\mathrm{MeV}\right] < 10$, with an overall change in the attenuation coefficients of less than three for the entire periodic table. These features indicate that the attenuation of gamma rays of $E_\gamma > 1\,\mathrm{MeV}$ in a medium of certain linear thickness is not sensitive to the elemental composition, but it is strongly affected by density

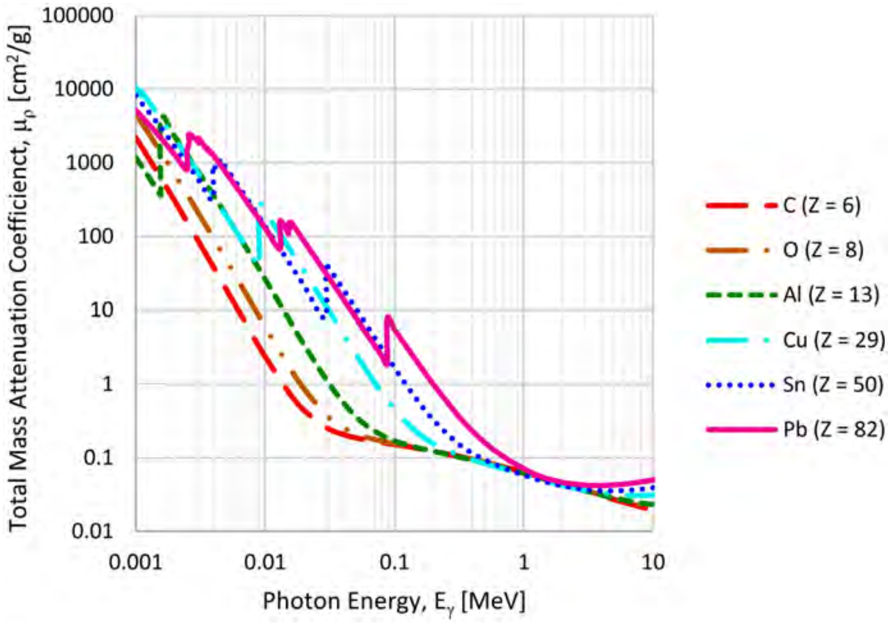


Figure 1. Photon mass attenuation coefficients for select elements of $Z = 6$ to 82 extracted from XCOM website [4].

of the medium. On the other hand, as is well known, low-energy photon attenuations are sensitive to the atomic number of the medium as the slope and discontinuities of attenuation coefficients are unique to each element. We, thus, conceive of a system where we may combine the density effect of high-energy photon intensity losses with the atomic number sensitivities of low-energy photons in a single measurement to detect small components of high-Z elements in a given medium. As we perform relative intensity measurements with respect to a reference medium devoid of the contaminants under investigation, our results are free from the variation of detection efficiencies, source strengths and/or source-absorber-detector geometries.

2. Experimental Method

2.1. Choice of Radio Isotopes

To be able to exploit the dual sensitivities, *i.e.* a region of high energies ($E_\gamma > 1\,\text{MeV}$) where the attenuations depend on the bulk of the medium and the low energies where the elemental constituents reveal themselves, an arrangement of radioactive source(s) emitting photons over a wide range of discrete energies is suggested. In this arrangement, a measurement of relative intensities of the full-energy peaks of the sample in question with a reference sample is made. A detector system of moderate energy resolution will suffice.

We used ^{152}Eu ($T_{1/2}$ = 13.5 years) and ^{133}Ba ($T_{1/2}$ = 10.55 years) as the pair of gamma-ray sources. Table 1 lists the energies and intensities (>2%) of the gamma rays emitted by these isotopes[1]. ^{152}Eu emits five gamma rays (779, 964, 1086,

[1]The nuclear data is available at the Brookhaven National Laboratory website. https://www.nndc.bnl.gov/. Accessed 29 July 2021.

Table 1. Energies and intensities ($I_\gamma > 2\%$) of gamma rays emitted by the ^{152}Eu and ^{133}Ba isotopes [6].

Isotope	Gamma Ray Emissions	
	Energy, E_γ [MeV]	**Intensity, I_γ [%]**
Ba-133	0.0810	35.6
Eu-152	0.1218	28.5
Eu-152	0.2447	7.5
Ba-133	0.2764	7.2
Ba-133	0.3029	18.3
Eu-152	0.3443	26.5
Ba-133	0.3567	62.0
Ba-133	0.3839	8.9
Eu-152	0.4114	2.2
Eu-152	0.4440	3.0
Eu-152	0.7789	12.9
Eu-152	0.8674	4.2
Eu-152	0.9641	14.5
Eu-152	1.086	10.1
Eu-152	1.112	13.7
Eu-152	1.408	20.9

1112, 1408 keV energies) of high intensities that can be used to judge the interaction thickness $\rho x\left[\text{g/cm}^2\right]$ of the material. ^{133}Ba and ^{152}Eu together emit 10 high-intensity gamma rays in the range of 81 to 444 keV which can be used to determine the Z of the material. The 81 keV gamma ray proved to be very useful as it, along with the 121 keV gamma ray of ^{152}Eu, overlap with the absorption edges of high-Z elements and their attenuations are sensitive to the presence of mercury, lead and transuranic elements, as we discuss below.

2.2. Experimental Setup

Figure 2 shows the top view of the experimental setup; consisting of a high-resolution HPGe (high-purity germanium) gamma-ray detector, a 50 cm long trough to hold absorber materials, and the ^{133}Ba, ^{152}Eu radioactive sources. Each source is point-like and of about 5 μCi (185 kBq) activity. The geometry is such that the detector sees only those gamma rays passing through the absorber medium. During data taking, the detector was shielded with 10 cm lead blocks to minimize the room background contributions.

To mimic a suitcase of interest for security checks at airports, etc., small bricks of organic material were made by compressing cotton cloths with a hydraulic press to approximately one third of their original size (blocks in **Figure 2**). Each

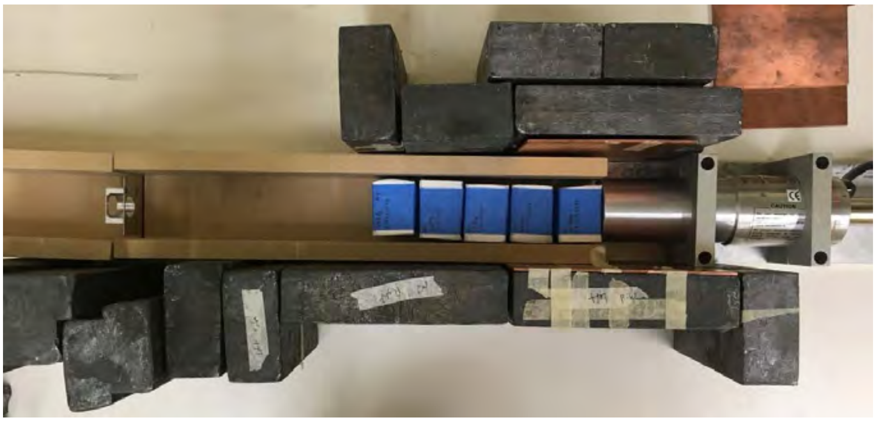


Figure 2. Top view of the Experimental Setup. Left to right: source holder, trough with organic blocks, HPGe detector.

block of organic material had an average density of 0.4 g/cm^3. The overall density thickness of blocks is chosen to correspond to an average suitcase of about 30 kg in weight and dimensions usually stipulated by the commercial airlines. Rough measurements indicate that moderately packed cotton material has a density of about 0.3 g/cm^3. A common dimension for a large passenger suitcase is 70 cm tall, 46 cm wide, and 30 cm in depth. A density of 0.3 g/cm^3 and dimensions of (70 × 46 × 30) cm corresponds to 29 kg. Most standard check-in baggage allowances are a maximum of 32 kg. Assuming moderate packing, the density thickness through the depth of the suitcase is 9 g/cm^2 and the density thickness through the width of the suitcase is 13.8 g/cm^2. The organic cloths used were made up of unknown proportions of cotton (cellulose $C_6H_{10}O_5$) and polypropylene (C_3H_6). During the experiment foreign elements of our interest were inserted in between the blocks. The data acquisition system comprised of conventional NIM electronics and an ORTEC—Maestro ADC based program[2].

3. Data Analyses

When calculating theoretical values of attenuation coefficients for the simulated suitcase materials from XCOM database, a 50% cotton (cellulose), 50% polypropylene mixture was assumed. To ensure that such assumption is not erroneous, we made XCOM [4] calculations for pure cellulose, pure polypropylene and 50% admixtures of each of them. As seen in Figure 3, for 81 to 1408 keV photon energies of our interest, the difference in mass attenuation coefficients between the three materials is monotonous and less than ~3%. Clearly, these small variations of attenuation coefficients for possible uncertainties in the relative compositions of cotton and polypropylene are of no consequence for the data analyses and interpretations.

The organic and absorber materials were measured to determine their linear $x[\text{cm}]$ and density-thicknesses $\rho x\left[\text{g/cm}^2\right]$. The reference spectrum was

[2]See ORTEC—Maestro User's Manual for information:
https://web.archive.org/web/20211002065539/https://www.ortec-online.com/-/media/ametekortec/manuals/a65-mnl.pdf

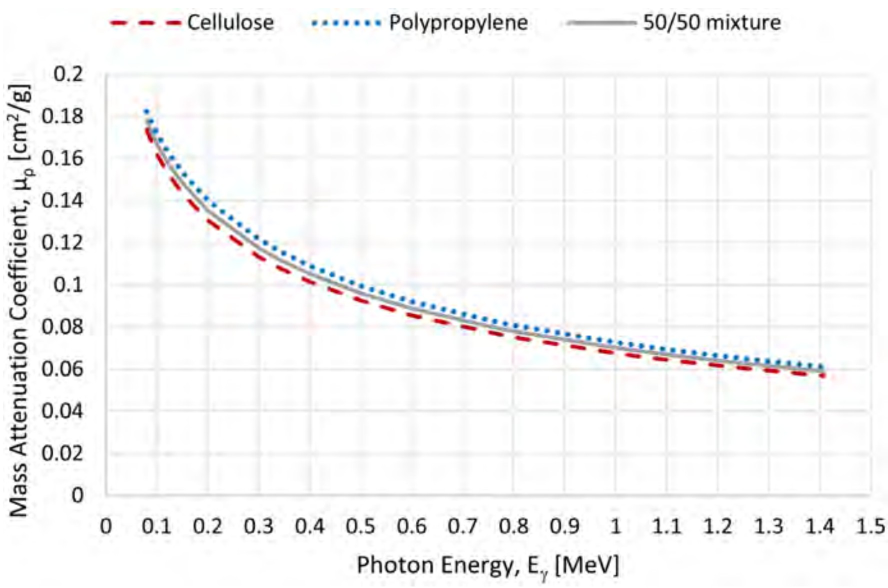


Figure 3. Mass Attenuation Coefficients for pure cellulose, pure polypropylene, and 50-50 admixtures in the organic material (calculated from the XCOM interactive database [4]).

obtained with only organic blocks between the radioactive sources and the detector. The absorber spectra were measured by placing various low-Z (graphite, aluminum), medium-Z (copper, tin), and high-Z (lead) materials in the trough along with the organic blocks. The spectra were collected such that the error in the counts of each peak measurement was less than 2%.

3.1. Determination of Absorber Role on Intensities

The number of gamma rays of a certain energy, E_γ, seen by the detector depends on three factors:

1) Yield, $Y(E_\gamma)$—characteristic of decay modes of the isotope, which determine the branching ratios for the intensities of photons of energy E_γ. It also depends on the activity of the source (expressed in Becquerels [Bq] or Curies [Ci]) and the source-medium-detector geometry. In our experiment, the 50 cm separation between the detector and source was maintained constant throughout.

2) Attenuation—characteristic of the absorbers placed between the detector and the source. This parameter depends on the material constituents of the medium, both for their elemental composition and their thickness x_i, density ρ_i, and attenuation coefficient $\mu_\rho(E_\gamma, Z_i)$, which as seen in **Figure 1** is specific to photon energy E_γ. The mass attenuation coefficient of any compound or mixture can be calculated from XCOM data for elements and by assuming Bragg additivity applies, written as [5]:

$$\mu_\rho(E_\gamma) = \sum_i w_i \cdot \mu_\rho(E_\gamma, Z_i) \tag{2}$$

where w_i is the fraction by weight of the i^{th} atomic constituent and $\mu_\rho(E_\gamma, Z_i)$ is the corresponding mass attenuation of the i^{th} element. The mass attenuation coefficient of a compound or mixture is the sum weighted by the composition. For a gamma ray of energy E_γ, passing through multiple absorbers, each of attenuation coefficient $\mu_\rho\left(E_\gamma, Z_i\right)$, the overall attenuation is given by:

$$\mathrm{e}^{-\sum_i \mu_\rho\left(E_\gamma, Z_i\right)\cdot x_i \rho_i} \tag{3}$$

where $x_i \rho_i$ is the density thickness of i^{th} absorber in the setup.

3) Full energy peak efficiency, $\varepsilon\left(E_\gamma\right)$—depends on the intrinsic detection efficiency of the detector, which varies with photon energy. The net area of an observed full-energy peak $Y\left(E_\gamma\right)$ is then:

$$Y\left(E_\gamma\right) = I\left(E_\gamma\right)\varepsilon\left(E_\gamma\right)\mathrm{e}^{-\sum_i \mu_\rho\left(E_\gamma, Z_i\right)\cdot x_i \rho_i} \tag{4}$$

The ratio of areas of two full-energy peaks at energy E_γ collected using absorbers 1 and 2 maybe written as:

$$\frac{\left[Y\left(E_\gamma\right)\right]_1}{\left[Y\left(E_\gamma\right)\right]_2} = \frac{I\left(E_\gamma\right)\varepsilon\left(E_\gamma\right)\left[\mathrm{e}^{-\sum_i \mu_\rho\left(E_\gamma, Z_i\right)\cdot x_i \rho_i}\right]_1}{I\left(E_\gamma\right)\varepsilon\left(E_\gamma\right)\left[\mathrm{e}^{-\sum_j \mu_\rho\left(E_\gamma, Z_j\right)\cdot x_j \rho_j}\right]_2} \tag{5}$$

which reduces to:

$$\frac{\left[Y\left(E_\gamma\right)\right]_1}{\left[Y\left(E_\gamma\right)\right]_2} = \frac{\left[\mathrm{e}^{-\sum_i \mu_\rho\left(E_\gamma, Z_i\right)\cdot x_i \rho_i}\right]_1}{\left[\mathrm{e}^{-\sum_j \mu_\rho\left(E_\gamma, Z_j\right)\cdot x_j \rho_j}\right]_2} \tag{6}$$

where the subscripts 1 and 2 refer to the two spectra with absorber 1 and 2, respectively. This ratio does not depend on the intensity $I\left(E_\gamma\right)$ of the gamma ray nor on the full-energy peak detection efficiency $\varepsilon\left(E_\gamma\right)$. Thus, it is simply the statistical errors of detected events which enter into the final result. The peak yield in the numerator of the ratio (spectrum 1) is referred to as the absorber spectrum: it is of the spectrum of gamma rays passing through the medium under investigation. The peak yield in the denominator (spectrum 2) is referred to as the reference spectrum, which is devoid of high-Z materials.

An experimental reference spectrum (spectrum 2 in Equations (5) and (6)) was measured by placing materials in the trough that are "expected" to be there. In the case of airport security, organic material (clothes, plastic, etc.) ought to be expected. The experimental ratio of the areas of peaks between the absorber and reference spectra were then calculated.

Figure 4 shows a spectrum with the reference material (upper) and an absorber spectrum which includes a 0.435 mm lead sheet (lower). A conspicuous feature is that the 121 keV gamma ray in the spectrum with lead absorber is more attenuated than the 81 keV gamma ray, as compared to their relative intensities in the reference spectrum. This is understood as due to smaller attenuation

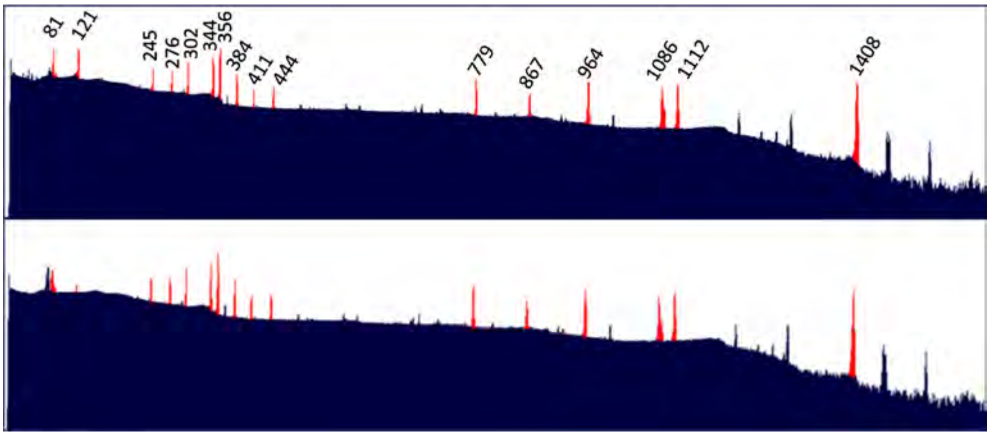


Figure 4. The ^{152}Eu + ^{133}Ba gamma-ray spectrum with reference organic material (upper) and a thin sheet of lead inserted (lower). The 81, 276, 302, 356, 384 keV full-energy peaks are of ^{133}Ba and the other energies are of ^{152}Eu decays. The unmarked peaks near 81 keV are due to Pb fluorescence, the rest of the unmarked peaks are due to background radiation.

coefficient for 81 keV gamma ray at the lead absorption edge (see Figure 1), while 121 keV gamma ray is outside the absorption edge. All other peaks at higher energies show a monotonous trend for the peak ratios.

The attenuation coefficients of the materials were obtained from XCOM and in some cases, calculated using the Bragg additivity formula (Equation 2). The theoretical ratio of the peaks between the absorber and the reference spectra were then calculated using Equation 6.

3.1.1. Z-Dependence of the Attenuation

To examine the Z-dependence of intensity variations, we measured two spectra one each with a copper sheet (0.508 mm = 0.455 g/cm^2) and one with a lead sheet (0.435 mm = 0.491 g/cm^2) inserted in the absorber sample. Note that the copper and lead sheets used are of nearly same density thickness. Figure 5 shows the ratio of intensities with respect to the reference absorber (Equation (6)). Also shown are the XCOM results as curves. Two things are worth noting. First, XCOM calculations are in excellent agreement with the data. Secondly, as expected, for energies above about 800 keV, there is no difference between the copper and lead absorbers. At lower energies, the intensities of lead absorber decrease much more steeply than with the copper absorber. Furthermore, the lead absorber shows a minimum ratio at 121 keV with an enhancement at 81 keV, not seen in the spectrum with copper sheet. This is understood as due to X-ray absorption edge of lead at 82 keV, which is not to be found in copper or other low Z-elements.

3.1.2. Density Thickness Dependence

It is of interest to examine the influence of density thickness of organic medium and the level of concentration of foreign materials. To this end, the ratios presented in Figure 6 were collected using absorber samples consisting of 22.1 cm (8.8 g/cm^2) and 31.0 cm (12.4 g/cm^2) of organic blocks each with a 0.435 mm thick lead sheet (0.493 g/cm^2) inserted, and a sample containing 44.9 cm (18.0

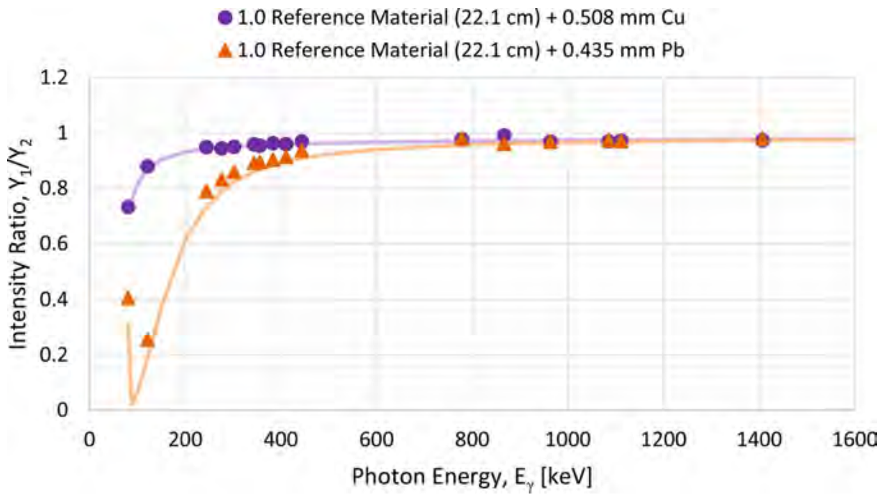


Figure 5. Measured intensity ratios of 22.1 cm reference material (8.8 g/cm^2) + 0.508 mm copper (0.455 g/cm^2), and reference + lead sheet (0.493 g/cm^2) with respect to the reference material. Also plotted are the XCOM calculation results. The sensitivity of the ratios to the Z-value of absorber at low energies is clear.

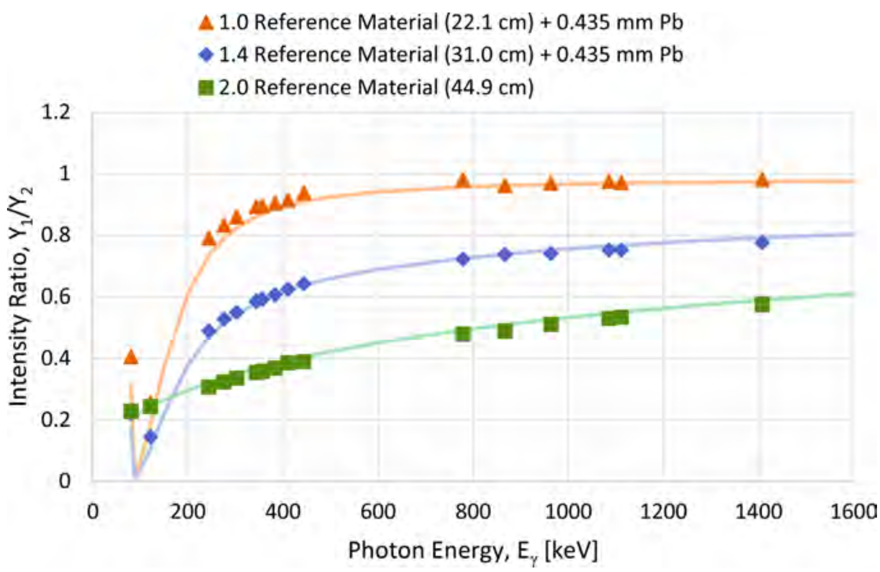


Figure 6. Experimental and theoretical (XCOM) intensity ratios of organic block medium (reference material, 22.1 cm, 8.8 g/cm^2) and the effect of a thin lead sheet (0.435 mm, 0.493 g/cm^2) inserted. At each energy, the data are normalized to yields with the reference material: triangles 22.1 cm organic medium (8.8 g/cm^2) + lead, diamonds 31.0 cm organic medium (12.4 g/cm^2) + lead, and squares 44.9 cm organic medium (18.0 g/cm^2). The solid points are experimental data. The curves are corresponding XCOM calculations.

g/cm^2) of organic blocks only. The spectrum of 22.1 cm organic blocks with no lead sheet is used as the reference spectrum. Superposed on the data in Figure 6 are the XCOM calculation results normalized to the 22.1 cm organic block reference spectrum at each gamma-ray energy.

A few observations of the results are in order:

1) The XCOM and experiment are in excellent agreement. The attenuations for all thicknesses with and without lead sheet are well accounted for.

2) For photon energies above about 800 keV, the ratio with respect to the reference spectrum decreases to 96% - 98% for 22.1 cm organic block + lead, and to 72% - 78% level for 31 cm organic block + lead and to 60% level with no lead.

3) At photon energies below 400 keV, the samples with lead show steeper decrease in the intensity with decreasing energy, while the 44 cm block without lead sheet has nearly continuous change with no steep decrease.

4) As indicated above, the organic blocks with lead sheet shows enhanced intensity of 81 keV peak relative to 121 keV, which is understood as due to an X-ray absorption edge.

It is thus clear that the presence of high-Z materials significantly attenuates the low-energy photons without much change at high energies, while increase of the mass density alone brings down the overall intensities. It is also seen that the high-energy gamma rays show the sensitivity to density thickness $x\rho$ of the absorber. XCOM database shows attenuation coefficients for atomic numbers Z = 1 - 100 behave very similarly around 1 MeV but with dramatic variations at low energies.

If the density thickness of two absorbers is same, then their ratio of peak intensities above about 1 MeV will be nearly the same. This is demonstrated in Figure 7, where we plot the XCOM results for 1 mm thick absorbers of graphite ($Z = 6$, $\rho = 2.27$ g/cm^3), aluminum ($Z = 13$, $\rho = 2.7$ g/cm^3), copper ($Z = 29$, $\rho = 8.96$ g/cm^3), tin ($Z = 50$, $\rho = 7.31$ g/cm^3), gadolinium ($Z = 64$, $\rho = 7.90$ g/cm^3) and lead ($Z = 82$, $\rho = 11.34$ g/cm^3). The data points in Figure 7 indicate the photon energies of the ^{133}Ba and ^{152}Eu sources. As seen in Figure 7, the intensities of spectra with copper ($Z = 29$, $\rho = 8.96$ g/cm^3), tin ($Z = 50$, $\rho = 7.31$ g/cm^3), and gadolinium ($Z = 64$, $\rho = 7.90$ g/cm^3) absorbers are within 1% at about 1 MeV, while the three spectra clearly separate the three distinct materials at low energies below about 200 keV. This confirms that the high-energy gamma rays

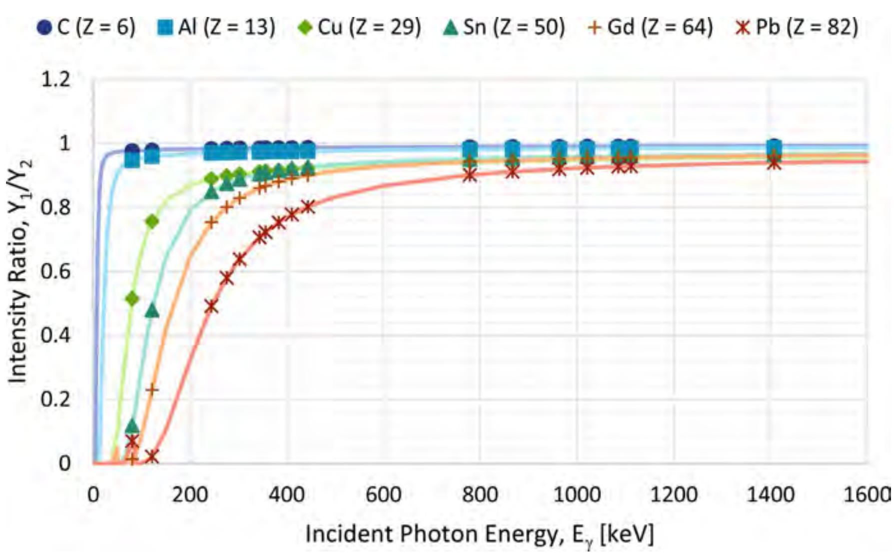


Figure 7. Theoretical XCOM calculations of ratios of peak intensities per 1 mm of absorber elements. Data points are ^{133}Ba and ^{152}Eu gamma-ray energies.

are insensitive to changes in Z. If the data is taken using solely high-energy or low-energy gamma rays, the effects due to different densities and different Z-numbers are indiscernible. The wide energy range of gamma rays emitted by the ^{133}Ba and ^{152}Eu sources make the evaluation of both the Z-number and density of the absorbing material possible.

4. Calculations for Transuranic Elements and Mercury

While we did not perform experiments with uranium, plutonium, or mercury for logistic reasons, we extend our calculations of XCOM for plutonium and mercury, representing a security and a health risk, respectively.

4.1. Presence of Plutonium as Trace Element

Plutonium ($Z = 94$) and uranium ($Z = 92$) are neighboring elements and solid materials. Our calculation for plutonium is readily applicable to uranium and nearby transuranic elements such as americium. We use the same reference material (organic block of 22.1 cm) to examine the response to plutonium as foreign element. In Figure 8 we see an indication of a high-Z material as an X-ray absorption edge kink, even with plutonium of 10-micron thickness. The 100-micron foil will show the characteristic X-ray absorption structure. For larger thicknesses, the steep drop of intensities below 500 keV is a clear indication of the presence of high Z-materials.

4.2. Presence of Mercury as a Contaminant

Globally, there is a serious concern of mercury contamination in water as a hazardous substance [6] likely to cause serious health effects. About 128 countries signed legally binding agreements to limit mercury concentrations. We explored

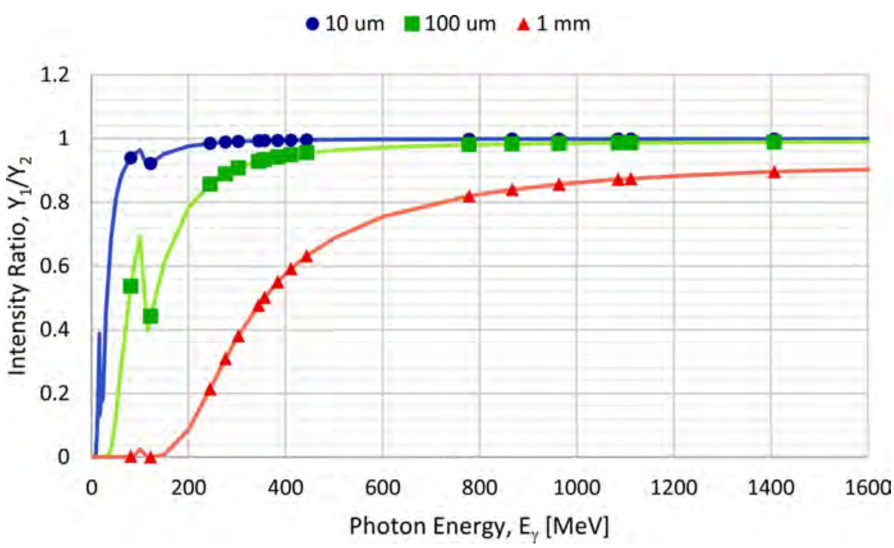


Figure 8. XCOM calculations for attenuation of ^{133}Ba and ^{152}Eu gamma rays for various plutonium thicknesses embedded in the reference material. The Pu content: circles (10 microns, 0.02 g/cm^2), squares (100 microns, 0.2 g/cm^2), and triangles (1 mm, 2 g/cm^2).

the possibility that multi-energy gamma-ray attenuation techniques could be a potential tool for detecting low concentrations of mercury. The occurrence of mercury in aqueous solutions, say water, is a volume effect and one must treat it as a mixture. One concerns here with concentrations less than one part per billion (ppb) in water. The X-ray absorption edge of mercury is close to ^{133}Ba's 81 keV gamma ray. Figure 9 shows the results of XCOM calculations for 1 cc water with admixtures of 0.001%-1% of mercury by volume. While the arrangement can detect concentrations of higher than 0.1% (one part in 1000), or better, it became obvious that this technique is not competitive against other technologies which can detect contamination levels of much less than a few parts in billion [6] [7].

5. Discussion

The experiment and XCOM calculations clearly indicate that a simultaneous measurement of gamma-ray intensity attenuations over a range of energies, about 80 to 1500 keV, can discern the presence of high-Z elements in otherwise dense medium. The test arrangement can be easily achieved with the simultaneous use of two radioactive isotopes of a few years lifetimes.

Below we compare the sensitivity of this method to that of more commonly used techniques such as the continuous spectrum of low-energy X-rays and nearly monoenergetic high-energy gamma rays. X-ray tubes provide a continuous energy spectrum of photons below about 200 keV and have long been used to identify materials. While the low-energy photons are very sensitive to the atomic number of a medium, they are easily attenuated. Table 2 compares the response of the low-energy photons to that of ^{152}Eu emissions for media of 20 cm thickness. The low transmission of X-ray energies is its limiting factor for use in

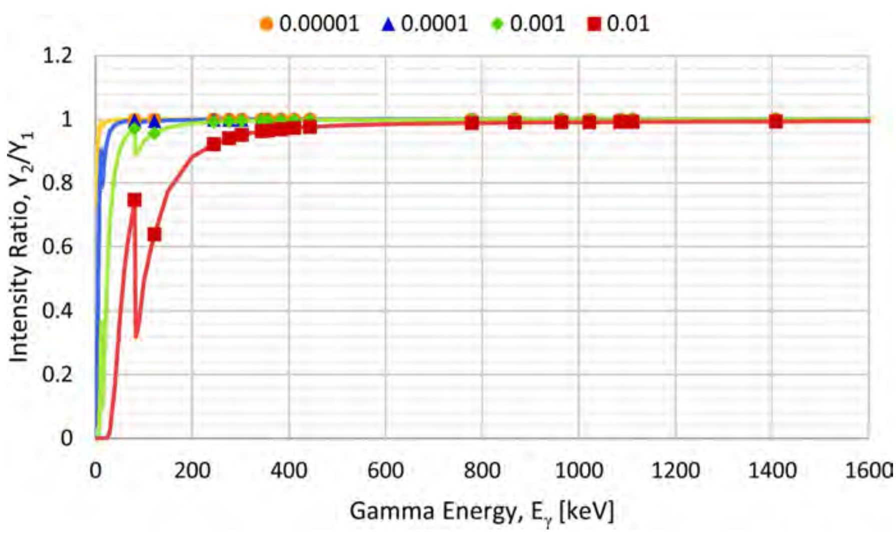


Figure 9. XCOM calculations for attenuation of ^{133}Ba and ^{152}Eu gamma rays through a 1 cc water column, mercury contaminants indicated by volume [mL/mL]. Hg concentrations: circles 10^{-5} [mL/mL], triangles 10^{-4} [mL/mL], diamonds 10^{-3} [mL/mL], square 10^{-2} [mL/mL].

bulky media.

In comparison to ^{152}Eu gamma rays, the X-ray tube would have to be run for almost a factor of three times longer for low-Z materials (C-6, Al-13) and many orders of magnitude longer for medium to high-Z materials in order to achieve the same sensitivity. For X-ray tubes producing photons at energies lower than 200 keV, the run times will be even longer. The US military uses a cobalt-60 (^{60}Co) source to probe vehicles[3]. This isotope emits two gamma rays of nearly the same energy (1173 and 1333 keV) which, as we have seen above, are insensitive to the atomic number Z of absorber elements. In Table 3, the transmission of gamma rays produced by ^{133}Ba, ^{152}Eu, and ^{60}Co sources through 1 mm of various materials is calculated. It is seen that the relative transmission for C:Al:Cu:Sn:Pb are 1.0:0.998:0.962:0.977:0.95 and 1.0:0.98:0.523:0.113:0.066 respectively for 1333 and 81 keV gamma rays. This observation clearly illustrates that the high-energy ($E_\gamma > 1\,\text{MeV}$) gamma rays are not sensitive to the medium's atomic number.

Table 2. Transmission of 200 keV X-rays and select ^{152}Eu gamma rays through a 20 cm thick sample of select materials.

Source	Transmission Through 20 cm of Material $e^{-(\mu(E,Z)\cdot\rho\cdot 20[\text{cm}])}$					
	Energy, E [keV]	C (Z = 6)	Al (Z = 13)	Cu (Z = 29)	Sn (Z = 50)	Pb (Z = 82)
X-ray Tube	200	3.80e−3	1.35e−3	7.36e−13	2.13e−21	4.36e−99
^{152}Eu	344	1.01e−2	4.84e−3	1.02e−8	2.06e−9	6.04e−31
^{152}Eu	779	1.48e−2	1.48e−2	1.16e−6	7.17e−6	5.10e−13
^{152}Eu	1086	6.25e−2	4.12e−2	3.96e−5	3.16e−4	3.16e−7
^{152}Eu	1408	6.47e−2	4.28e−2	4.50e−5	3.55e−4	4.35e−7

Table 3. Transmission of ^{60}Co and select ^{133}Ba, ^{152}Eu gamma rays through a 1 mm thick sample of select materials.

Source	Transmission Through 1 mm of Material $e^{-(\mu(E,Z)\cdot\rho\cdot 0.1[\text{cm}])}$					
	Energy, E [keV]	C (Z = 6)	Al (Z = 13)	Cu (Z = 29)	Sn (Z = 50)	Pb (Z = 82)
^{133}Ba	81	0.9642	0.9469	0.5048	0.1096	0.0644
^{152}Eu	122	0.9680	0.9596	0.7543	0.4725	0.0203
^{152}Eu	1086	0.9862	0.9842	0.9506	0.9605	0.9279
^{152}Eu	1112	0.9864	0.9844	0.9512	0.9611	0.9294
^{60}Co	1173	0.9868	0.9848	0.9525	0.9622	0.9324
^{60}Co	1333	0.9876	0.9857	0.9554	0.9647	0.9383
^{152}Eu	1408	0.9879	0.9861	0.9566	0.9657	0.9404

[3]See for example, Science Applications International Corporation (SAIC) website: https://web.archive.org/web/20071011181039/http://www.saic.com/products/security/pdf/mobile-VACIS.pdf

While the low-energy photons are sensitive to atomic number, the high attenuation renders them impractical for dense materials. The advantage of a multi-energy photon arrangement for energies of about 80 keV to 1.4 MeV to discriminate the contribution from high-Z materials versus high density materials is clear.

6. Conclusion

From experiment and XCOM calculations we demonstrated that a radioactive source arrangement of multi-energy photons can detect small quantities of high-Z contraband materials in otherwise low-Z media, such as suitcases. While the high-Z contaminants such as mercury in water or other environmental samples can also be detected, nano technology [7] and diverse analytical methods [6] may be more competitive than the gamma-ray attenuation methods. In comparison to X-ray tubes or particle accelerators, the radioactive source arrangement has practical advantages in that the sources are fail-proof, continuously operating systems, and require very little maintenance. Once procured, they can operate for years without additional investment due to their long half-lives. On the other hand, X-ray machines or accelerators require substantial maintenance, repairs, and elaborate power supply schemes; high-energy photon machines require large infrastructures. As an example, an X-ray scanner comprises of an X-ray tube and an X-ray detector, both of which consume power and require regular maintenance. An X-ray tube running at 100 kVp and 100 mA would consume at least 10 kW. It is safe to assume that the entire X-ray tube system would consume much more power. In our set-up of radioactive sources and a detector, only the detector and electronic assembly consume power, which can be drawn from a wall power outlet. The ^{133}Ba and ^{152}Eu sources only need to be replaced once or twice every half-life ($T_{1/2}$ = 10.55 y for ^{133}Ba, $T_{1/2}$ = 13.52 y for ^{152}Eu). Also, a combination of ^{154}Eu ($T_{1/2}$ = 8.59 y) emitting six gamma rays of 123 - 1205 keV energies along with ^{133}Ba can be a useful pair of isotopes. Of course, one can employ more than two sources without loss of information. Needless to say, fewer the number of sources, easier it is to handle. While we performed these tests with a single detector in a laboratory setting, the technique can be easily extended to a multi-detector ring system for tomographic reconstructions. In the field implementation, this system can be operated by a person or at most a few people.

Acknowledgements

This research was partially funded by a discovery research grant of Natural Sciences and Engineering Council (NSERC) of Canada to Chary Rangacharyulu.

Conflicts of Interest

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

References

[1] Rai, P.K., Lee, S.S., Zhang, M., Tsang, Y.F. and Kim, K. (2019) Heavy Metals in Food Crops: Health Risk, Fate, Mechanisms, and Management. *Environmental International*, **125**, 365-385. https://doi.org/10.1016/j.envint.2019.01.067

[2] Van Liew, S., Bertozzi, W., D'Olympia, N., Franklin, W.A., Korbly, S.E., Ledoux, R.J. and Wilson, C.M. (2017) Identification and Imaging of Special Nuclear Materials and Contraband Using Active X-ray Interrogation. *Physics Procedia*, **90**, 313-322. https://doi.org/10.1016/j.phpro.2017.09.023

[3] Ohgaki, H., Daito, I., Zen, H., Kii, T., Masuda, K., Misawa, T., Hajima, R., Hayakawa, T., Shizuma, T., Kando, M. and Fujimoto, S. (2017) Nondestructive Inspection System for Special Nuclear Material Using Inertial Electrostatic Confinement Fusion Neutrons and Laser Compton Scattering Gamma-Rays. *IEEE Transactions on Nuclear Science*, **64**, 1635-1640. https://doi.org/10.1109/TNS.2017.2652619

[4] Berger, M.J., *et al.* (2010) XCOM: Photon Cross Sections Database. NIST Standard Reference Database 8 (XGAM). NIST, PML, Radiation Physics Division. https://dx.doi.org/10.18434/T48G6X

[5] Rangacharyulu, C. (2014) Physics of Nuclear Radiations—Concepts, Techniques and Applications, Taylor & Francis. CRC Press, Boca Raton. https://doi.org/10.1201/b16330

[6] Bravo, A.A., Kothawala, D.N., Attermeyer, K., Tessier, E., Bodmer, P. and Amouroux, D. (2018) Clearing and Sampling Protocol for Analysis of Mercury and Dissolved Organic Matter in Freshwater Systems. *MethodX*, **5**, 1017-1026. https://doi.org/10.1016/j.mex.2018.08.002

[7] Wang, G., Lim, C., Chen, L., Chou, H., Choo, J., Hong, J. and de Mello, A.J. (2009) Surface Enhanced Raman Scattering in Nanoliter Droplets: Towards High Sensitivity Detection of Mercury (II). *Analytical and Bioanalytical Chemistry*, **394**, 1827-1832. https://doi.org/10.1007/s00216-009-2832-7

Journal of Modern Physics, 2022, 13, 81-88
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

The Preparation of Nano Silver by Chemical Reduction Method

Tevfik Raci Sertbakan[1], Emad K. Al-Shakarchi[2], Saif Sultan Mala[1]

[1]Kırşehir Ahi Evran University, Art and Science Faculty, Department of Physics, Kırşehir, Turkey
[2]Al-Nahrain University, College of Science, Physics Department, Baghdad, Iraq

Email: trsertbakan@ahievran.edu.tr, eks2000@hotmail.com

How to cite this paper: Sertbakan, T.R., Al-Shakarchi, E.K. and Mala, S.S. (2022) The Preparation of Nano Silver by Chemical Reduction Method. *Journal of Modern Physics*, **13**, 81-88.
https://doi.org/10.4236/jmp.2022.131006

Received: December 8, 2021
Accepted: January 25, 2022
Published: January 28, 2022

Abstract

A silver nanostructures prepared by using chemical reduction method. The silver nanoparticles were prepared with diameters of about (20 nm). Numerous techniques had been used to study the optical, structural like the UV-Vis absorption spectrometer, Ttransmission Electron Microscopy (TEM), Field-Emission Scanning Electron microscope (FESEM), and X-ray diffraction (XRD). The practical results exhibited the absorption spectrum of the prepared nanoparticles at (357 nm), it was found that there is a relationship between the positions of the optical absorption peak and the size of the silver nanoparticles. The analysis of TEM results showed the presence of nanoparticles in the range (20 nm). The analyzing of XRD results explained the crystal structure for silver nanoparticles. It is found a cubic unit cell have a lattice constants (a = 4.0855 Å), with the Miller indices were (111), (002), (002), and (113).

Keywords

Chemical Reduction Method, UV-Vis Absorption Spectrometer, Field-Emission Scanning Electron Microscope, Ttransmission Electron Microscopy

1. Introduction

The nano materials are rapidly developing recently through potential applications in a wide range of technological areas such as electronics, catalysis, ceramics, magnetic data storage, structural components [1]. Silver nanoparticles (AgNPs) are very important among different metallic nanoparticles, because it had a biomedical applications especially in cancer diagnosis and therapy. A nano spherical silver powder prepared using a chemical reduction method. The Turkevich procedure [2] used for synthesis AuNPs, it considered most probable many reasons. The procedure is very simple and reliably to produces AuNPs with diameters range (5 - 150 nm). The second thing, it is nontoxic and stabilized by citrate ions [3]. There is a fast nucleation in the synthesis leads to AuNPs with narrow size distribution [3] [4]. It was achieved by increasing the reactivity of the precursor. There are two procedure used to generate nanoparticles depend on a limited amounts of a material provided. Commonly, they applied in most researches and studies to produce nanostructure materials. These are classified as, the bottom-up and the top-down techniques, both mentioned in **Figure 1**. The bottom-up is assembled atoms together to produce nanoparticle or dis-assemble the bulk material producing nanostructure, the last one represent the up-down technique. The first procedure is more complicated because it needs high technology to assemble many atoms or molecule producing nanostructure. The second is simpler and it needs a simple equipment to perform laboratory because it depends on the decomposition of bulk material into nanostructure. There many chemical and physical process used to transfer the bulk material into nanostructure such as nanoparticles, but the reduction method is one of powerful technique used to perform nanoparticle in the procedure top-down [5] [6].

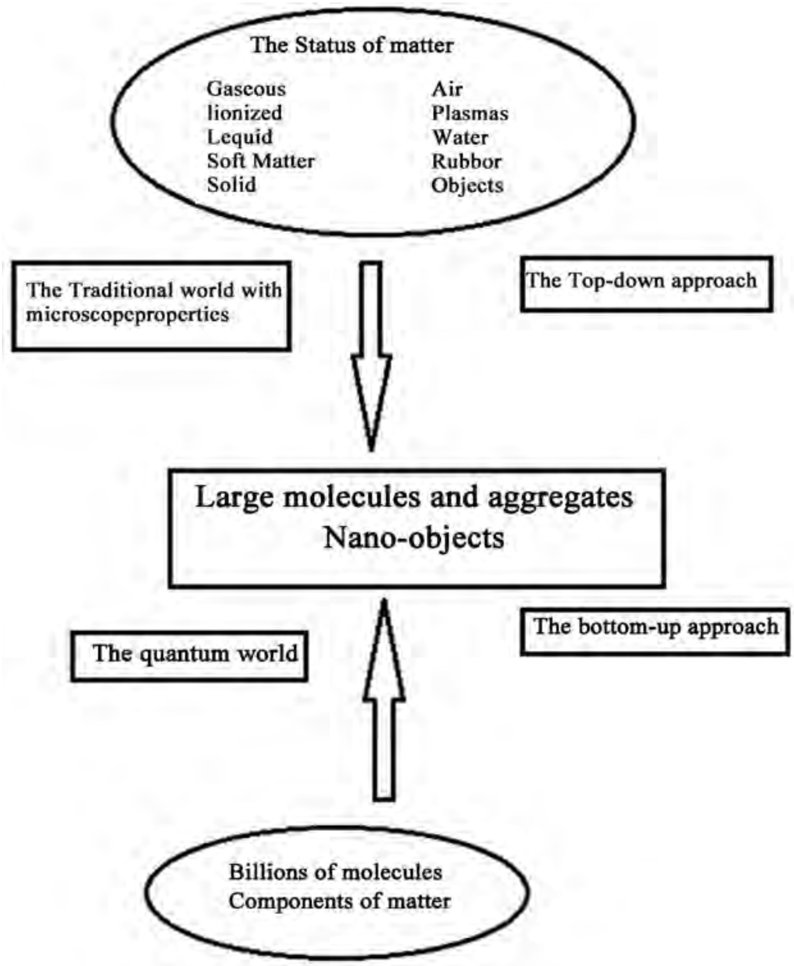


Figure 1. The technique of nanostructure [7].

In this research, it is related to prepare a silver nanoparticle experimentally by using the reduction method. It is necessary to apply in the laboratory and making a specification on the sliver nanoparticle represented by Transmission Electron Microscope (TEM), Field-Emission Scanning Electron microscope (FESEM), UV-VIS spectrophotometer, and X-ray diffraction. All these analysis give us a knowledge on the characterization Ag-nanoparticle prepared experimentally.

2. Experimental Procedure

The reduction method is one of the chemical methods those were used to prepare silver nanoparticles. It is considering a bottom-up procedure because the chemical reduction tends to produce Ag-atoms which are assembled to generate nanostructures. During the synthesis process, if the reaction rate is too fast, rapid formation of a large amount of metal nuclei will occur and result in particles that are too small. On the other hand, agglomeration of particles will take place if the reaction rate is too slow. The principle chemicals used in the experiment were the silver nitrate ($AgNO_3$), glucose and polyvinyl pyrrolidone (PVP), and sodium hydroxide. The Silver nanoparticles reduced from the silver nitrate in PVP aqueous solution. The glucose used as reducer and sodium hydroxide apply to accelerate the reaction. Firstly, the preparation of silver nitrate solution prepared by adding 3.4 g of $AgNO_3$ into 20 ml distilled water, the solution (a). The PVP solution prepared by dissolving PVP, the glucose and sodium hydroxide in 60 ml distilled water together, producing solution (b). Then the solution (b) heated to temperature (60˚C) by magnetic stirrer. After the temperature fixed for (15 min), the solution (a) added to (b) by drops. Then the mixed solution was stirred for (10 - 15 min). The particles were separated by centrifugation, and washing the products by distilled water many times to remove the possibility NO_3 that might be not reacted [8]. There are many technique used to examine the appearance of nanoparticles. A field emission scanning electron microscopy (FESEM) and transmission electron microscope were used to investigate the nanostructure production. Normally, FESEM-device is giving a high resolution rather than SEM. That's making a fine grain in the nano size. The main difference between FESEM and TEM is creating image by detecting reflected electrons, while TEM uses transmitted electrons to create an image. Then TEM-results offers a valuable information on the inner structure of the sample, while SEM provides information on the surface morphology and its composition. X-ray diffraction technique (XRD) is using to define the structural nature of the prepared sample. To investigate this purpose, the application of CrysDiff software is necessary to make a simulation on the XRD-pattern.

3. Results and Discussion

After the successive preparation had been done by the reduction method to produce a silver nanoparticles. These analysis and measurements are useful to prove the successful the preparation of silver nanoparticle experimentally. The absorption spectrum of silver nanoparticle by UV-visible spectrophotometer. It had been done for all prepared samples as a function of different concentration. The analysis of absorption spectrum state the first step of understanding the excited electronic states of molecules. It gives more information from state (S_o) to excited state (S_p) for the molecules under investigation, as shown in Figure 1 & Figure 2 for different concentration 1%, and 2% respectively. The spectrum is depending on the concept of surface Plasmon of silver nanoparticles which had been prepared by chemical reduction method. The surface Plasmon spectrum of large silver nanospheres those were mentioned by the sample with concentration (1%) as shown in Figure 2. It was clear that the maximum absorption peak occurred at (357 nm) with a peak intensity about (1.82 a.u). Whereas the surface Plasmon spectrum of silver nanostructure was denoted in Figure 3, for the concentration (2%). One can noted the maximum absorption peak occurred at (358 nm) with a peak intensity about (3.11 a.u).

It is clear that the maximum absorption peak of surface Plasmon absorbance of silver nanoparticles goes toward short wavelength. That is obvious the indication for the decreasing the size of nanostructure such as particles. These results

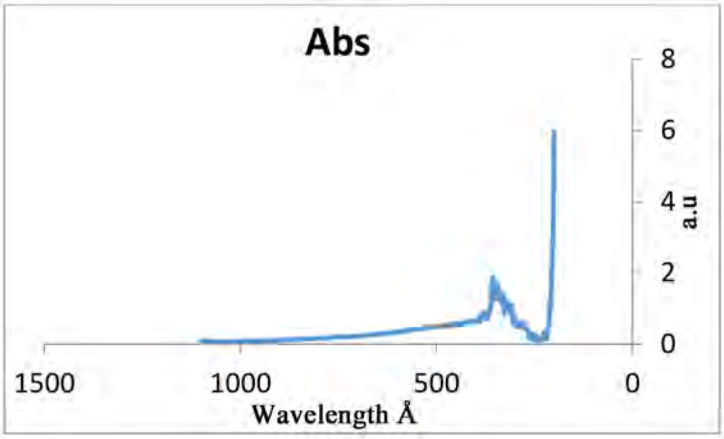


Figure 2. Indicate the UV-VIS spectrum for silver nanoparticle solution with 1% concentration.

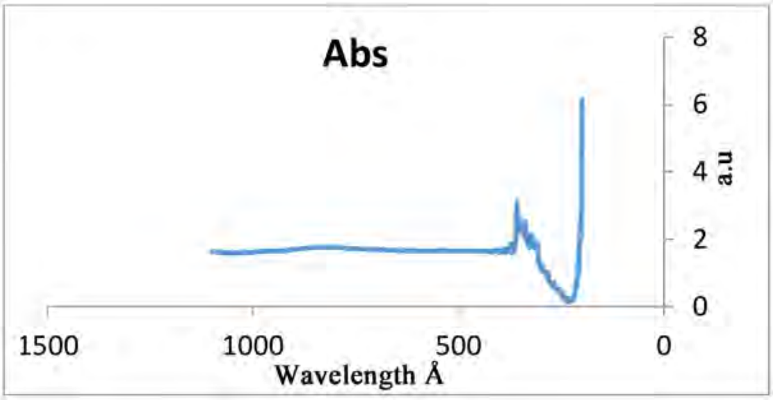


Figure 3. Indicate the UV-VIS spectrum for silver nanoparticle solution with 2% concentration.

is confirmed with the results obtained from transmission electron microscope previously [9].

The transmission electron microscopy (TEM) is a good technique to observe the size and the shape of the silver nanostructures, as shown in **Figure 4**. It was appeared as spherical shape with average particle size in the range (20 nm). The silver nanoparticles also showed high monodispersed. The results of FESEM is compatible with TEM-results. It was cleared the presence the nanoparticle, in addition it gives the morphology of surface and the homogeneity of grain size and grain boundaries, appeared in **Figure 5**. The results showed the presence of silver nanoparticle in different size. In addition, the agglomeration of silver nanoparticle is clearer through the increasing of silver nanoparticle rather than the

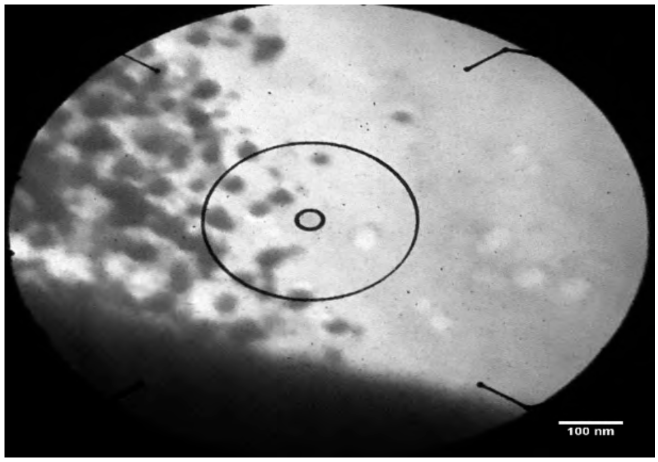


Figure 4. The TEM results for silver nanoparticle.

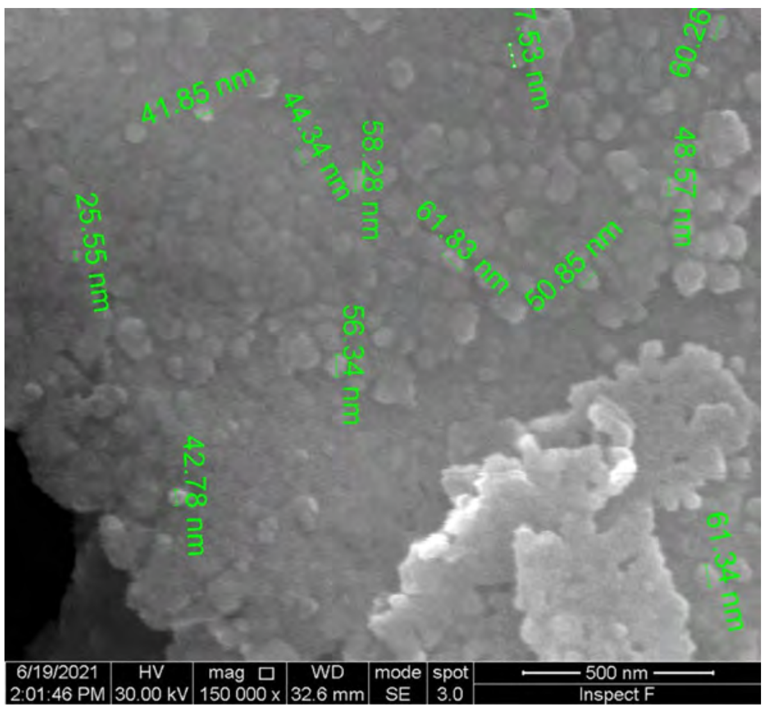


Figure 5. The FESEM results for silver nanoparticle with different position.

size of nanoparticle concluded in TEM results. The agglomeration is possible because of high conductivity of silver metal under the effect of the electron beam of FESEM.

The crystal structure during the XRD-pattern, as shown in **Figure 6**. The pattern analyzed by using specialized software to find the type of structure and the related lattice parameters, through the simulation by CrystalDiff. Software. The simulated results comparing with related database is the standard chart (1,100,136). The simulated data showed a cubic phase with lattice constants (a = b = c = 4.0855 Å), and the volume of unit cell is (68.1923 Å^3). The peaks are mentioned by the miller planes labeled by (111), (002), (022), and (113). It is agreement with the Lai *et al.* [1] and in contradiction with Mehta *et al.* [10].

The determination of lattice constants, and the crystal phase is related to position of the peaks at a certain (2θ), which are coincidence with the data base. The widen peaks in the X-ray pattern is considering the proof of nanoparticles appearance.

The appearance of many peaks is returning to polycrystalline structure, for this reason there are many orientation through the grains, and a predominant orientation is (111). Secondly, the peaks intensity is small enough comparable with X-ray diffraction of bulk sample. That is return the low density of Miller planes within the unit cell in comparable with bulk sample.

It is useful to determine the crystallite size inside the grains, and the elastic strain might be exhibited in the unit cell during the axis (a, b, c) by applying Williamson-Hall equation, as mentioned in **Figure 7**. The results showed the crystallite size is about (155.7808 Å) which is equal to (15.57 nm). That is exactly equal to the grain size appeared. It is smaller than the particle size that was discussed by TEM-results before. The important thing that is mentioned related to

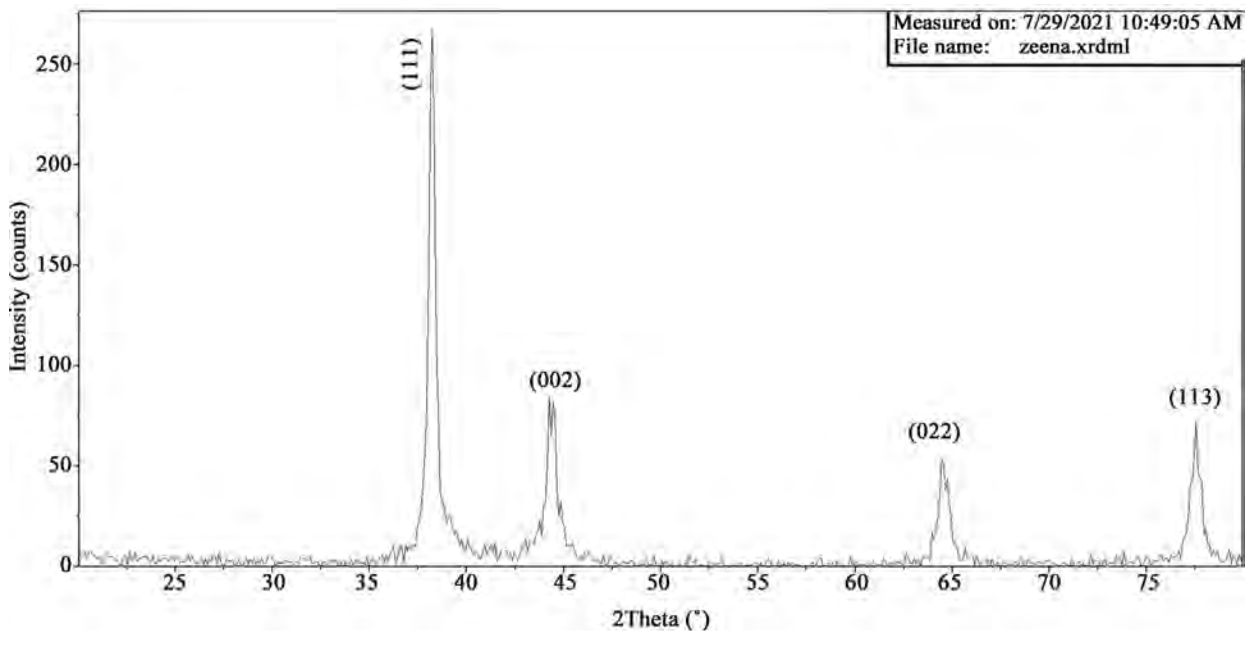


Figure 6. XRD pattern of silver nanoparticle.

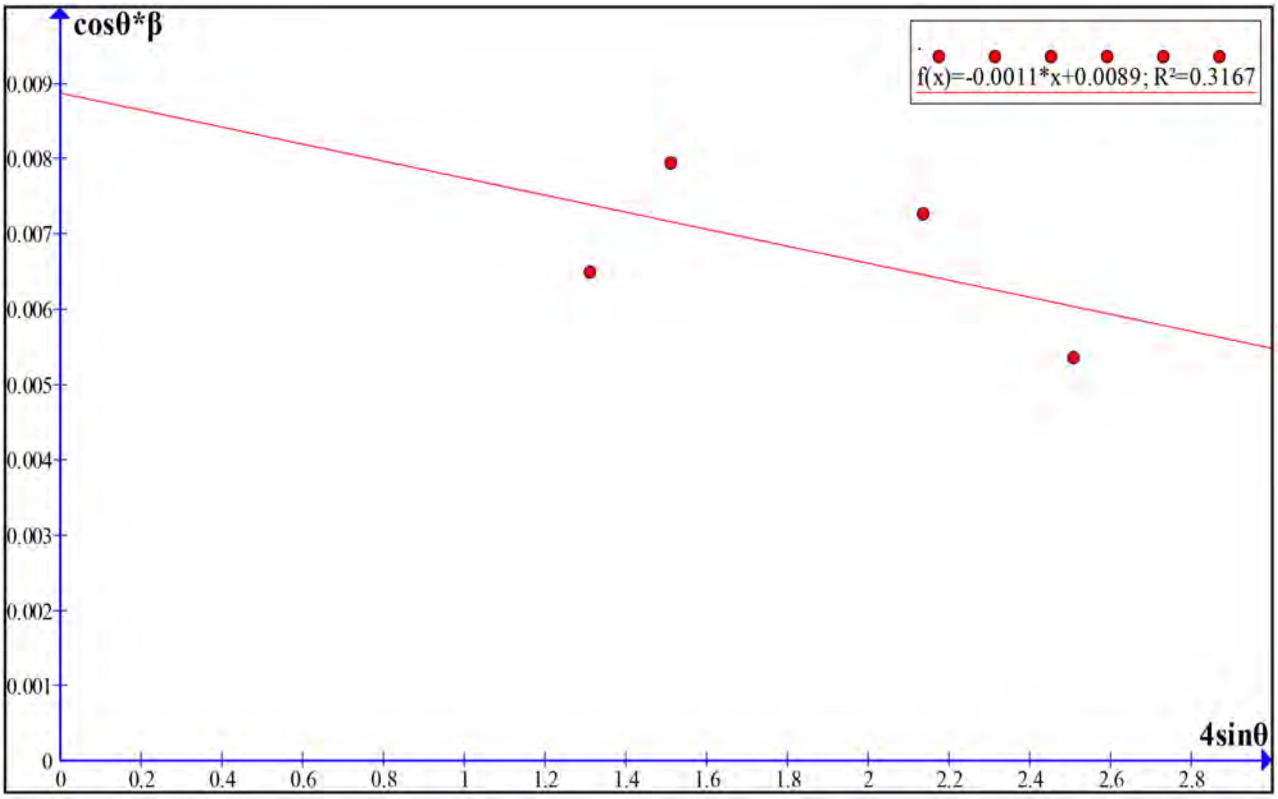


Figure 7. Explain the application of Williamson-Hall equation.

elastic strain, which is about (0.0011). It is a small enough, that is good because it is a function to the stability of the cubic phase produced.

4. Conclusion

The synthesis ability a metal nanoparticle is valid experimentally during the reduction method. The size of nanoparticle is about (20 nm), which is favor than that imported from the external company. The output nanoparticle powder is suitable to apply in the medical field through the obtainable results.

Acknowledgements

This research is done due to the cooperation of Al-Nahrain University/College of Science with Kırşehir Ahi Evran University/Art and Science Faculty/Turkey. It is done at Al-Nahrain University/College of Science/Physics Department during the supervision on M.Sc. student

Conflicts of Interest

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

References

[1] Lai, F., Lin, Z. and Zhang, P. (2018) Preparation and Characterization of Mi-

cro/Nano-Silver Powders. *AIP Conference Proceedings* 1995, pp. 0200101-0200105.

[2] Turkevich, J., Stevenson, P.C. and Hillier, J. (1951) A Study of the Nucleation and Growth Processes in the Synthesis of Colloidal Gold. *Discussions of the Faraday Society*, **11**, 55-75. https://doi.org/10.1039/df9511100055

[3] Zhao, P., Li, N. and Astruc, D. (2013) State of the Art in Gold Nanoparticle Synthesis. *Coordination Chemistry Reviews*, **257**, 638-665. https://doi.org/10.1016/j.ccr.2012.09.002

[4] Ji, X., Song, X., Li, J., Bai, Y., Yang, W. and Peng, X. (2007) Size Control of Gold Nanocrystals in Citrate Reduction: The Third Role of Citrate. *Journal of the American Chemical Society*, **129**, 13939-13948. https://doi.org/10.1021/ja074447k

[5] Zhang, H.X., Siegert, U., Liu, R. and Cai, W. (2009) Facile Fabrication of Ultrafine Copper Nanoparticles in Organic Solvent. *Nanoscale Research Letters*, **4**, Article No. 705. https://doi.org/10.1007/s11671-009-9301-2

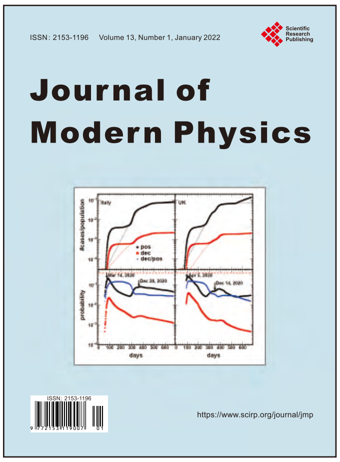
[6] Camacho-Flores, B.A., Martínez-Álvarez, O., Arenas-Arrocena, M.C., Garcia-Contreras, R., Argueta-Figueroa, L., de la Fuente-Hernández, J. and Acosta-Torres, L.S. (2015) Copper: Synthesis Techniques in Nanoscale and Powerful Application as an Antimicrobial Agent. *Journal of Nanomaterials*, **2015**, Article ID: 415238. https://doi.org/10.1155/2015/415238

[7] Nouailhat, A. (2006) An Introduction to Nanoscience and Nanotechnology. Wiley, New Jersey.

[8] Wang, H., Qiao, X., Chen, J. and Ding, S. (2005) Preparation of Silver Nanoparticles by Chemical Reduction Method. *Colloids and Surfaces A*: *Physicochemical and Engineering Aspects*, **256**, 111-115. https://doi.org/10.1016/j.colsurfa.2004.12.058

[9] Kanwal, Z., Raza, M.A., Riaz, S., Manzoor, S., Tayyeb, A., Sajid, I. and Naseem, S. (2019) Synthesis and Characterization of Silver Nanoparticle-Decorated Cobalt Nanocomposites (Co@AgNPs) and Their Density-Dependent Antibacterial Activity. *Royal Society Open Science*, **6**, 1821351. https://doi.org/10.1098/rsos.182135

[10] Mehta, B.K., Chhajlani, M. and Shrivastava, B.D. (2017) Green Synthesis of Silver Nanoparticles and Their Characterization by XRD. *Journal of Physics*: *Conference Series*, **836**, 012050. https://doi.org/10.1088/1742-6596/836/1/012050



Call for Papers

Journal of Modern Physics

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

Journal of Modern Physics **(JMP)** is an international journal dedicated to the latest advancement of modern physics. The goal of this journal is to provide a platform for scientists and academicians all over the world to promote, share, and discuss various new issues and developments in different areas of modern physics.

Editor-in-Chief

Prof. Yang-Hui He City University, UK

Subject Coverage

Journal of Modern Physics publishes original papers including but not limited to the following fields:

Biophysics and Medical Physics
Complex Systems Physics
Computational Physics
Condensed Matter Physics
Cosmology and Early Universe
Earth and Planetary Sciences
General Relativity
High Energy Astrophysics
High Energy/Accelerator Physics
Instrumentation and Measurement
Interdisciplinary Physics
Materials Sciences and Technology
Mathematical Physics
Mechanical Response of Solids and Structures
New Materials: Micro and Nano-Mechanics and Homogeneization
Non-Equilibrium Thermodynamics and Statistical Mechanics
Nuclear Science and Engineering
Optics
Physics of Nanostructures
Plasma Physics
Quantum Mechanical Developments
Quantum Theory
Relativistic Astrophysics
String Theory
Superconducting Physics
Theoretical High Energy Physics
Thermology

We are also interested in: 1) Short Reports—2-5 page papers where an author can either present an idea with theoretical background but has not yet completed the research needed for a complete paper or preliminary data; 2) Book Reviews—Comments and critiques.

Notes for Intending Authors

Submitted papers should not have been previously published nor be currently under consideration for publication elsewhere. Paper submission will be handled electronically through the website. All papers are refereed through a peer review process. For more details about the submissions, please access the website.

Website and E-Mail

https://www.scirp.org/journal/jmp   **E-mail: jmp@scirp.org**