Physics – Uspekhi **52** (5) 495 – 498 (2009)

METHODOLOGICAL NOTES

PACS numbers: **01.70.+w**, **11.15.–q**, **12.60.–i**

Unsolved problems in fundamental physics

I F Ginzburg

DOI: 10.3367/UFNe.0179.200905d.0525

Contents

1. **Introduction** 495
2. **Do we understand correctly what we ostensibly think we know?** 495
3. **Why is the world the way it is? Weird numbers and the relations between them** 496
4. **Odd facts, manifestations of unknown interactions...?** 497
5. **What next?** 498
6. **Some concepts** 498
References 498

Abstract. Recent years have added very much to our knowledge of the structure of the Universe and elementary interactions which, combined with the critical rethinking of long-known results and ideas, gives considerable topical relevance to the questions listed in this paper. Compiling this list, the author had no notion of any targeting and the typical 'WHAT FOR' and 'WHY' questions were considered as mere abbreviations for "Does a (possibly yet unknown) law of Nature exist with which the property or phenomenon under study can be explained?" The list of problems reflects the scientific interests of the author and so does not claim to be comprehensive.

1. Introduction

The recent literature abounds in rather detailed reviews of our current knowledge of elementary particles and astrophysics, which describe many concepts and ask some questions that can be elucidated experimentally in the near or less near future (see, for example, Refs [1–8]).

Book [9] presents an extremely interesting list of problems whose solution requires either more than the current understanding of the properties of matter or creating substances with the desired properties.

The questions addressed in the present paper are of a different type in that, for them, the very existence of — let alone the time needed to find — the answers is in itself a question.

In this paper we utilize the standard notation e for the electron charge, c for the speed of light in vacuum, $\hbar$ for the Planck constant divided by 2π, and $G_{\rm N}$ for the Newtonian constant of gravitation in the law of gravity. Two combinations of these constants, with the respective dimensions of mass and length, are the Planck mass $M_{\rm Pl} = (\hbar c/G_{\rm N})^{1/2} = 1.2 \times 10^{19}$ GeV $= 2.2 \times 10^{-5}$ g and the Planck length $l_{\rm Pl} = \hbar/M_{\rm Pl}c = 1.6 \times 10^{-33}$ cm. For masses of particles the reader is referred to existing handbooks.

Some of the concepts used below are explained in Section 6, which is written at a level accessible to junior high school students and will be of interest to school teachers as well.

2. Do we understand correctly what we ostensibly think we know?

- Many theoretical constructions comprise asymptotic conditions when passing to the limit (say, for $x \to 0$ or $x \to \infty$). However, the world we live in is finite in space and time. According to the uncertainty principle, the minimum conceivable values for energy and momentum are, respectively, $\hbar/T_{\rm U}$ (where $T_{\rm U} \approx 14$ bln years is the age of the Universe) and $\hbar/cT_{\rm U}$. It goes without saying that at distances less than the Planck length $l_{\rm Pl} = 1.6 \times 10^{-33}$ cm (i.e., at energies larger than $M_{\rm Pl} = 1.2 \times 10^{19}$ GeV) — when the quantum gravitation effects become one hundred percent important — the modern picture of the world is no longer valid. For a material medium in which the divisibility limit exceeds the average intermolecular distance, such a lower bound is achieved even much earlier. Nevertheless, it is an often used argument that, say, "The quantity A cannot be a solution because it increases infinitely as $x \to \infty$ or $x \to 0$."

Given, then, the finiteness of the domain of variability of quantities, which of the results obtained by this kind of reasoning retain their truth?

The following 'arithmetic' example is meant to show that the above conclusion is not, at any rate, always substantial. Suppose it is argued that a certain property occurs at such large times t that $z = \ln(\ln t) \gg 1$. Now how much more is 'much more'? Suppose, for example, that 'much more' means that $z > 5$. Then it is easy to verify that $t > 10^{53}$. Whether we measure time in seconds or characteristic atomic times, even the age of the Universe is too short a time for this condition to be fulfilled.

- The fact that space and time cannot be divided infinitely into segments and intervals leads to quite unexpected

---

I F Ginzburg S L Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences,
prosp. akademika Koptyuga 4, 630090 Novosibirsk, Russian Federation
Fax (7-383) 333 25 98. E-mail: ginzburg@math.nsc.ru
Novosibirsk State University,
ul. Pirogova 2, 630093 Novosibirsk, Russian Federation

Received 24 December 2007, revised 12 March 2009
Uspekhi Fizicheskikh Nauk **179** (5) 525 – 529 (2009)
DOI: 10.3367/UFNr.0179.200905d.0525
Translated by E G Strel'chenko; edited by A Radzig

phenomena from a usual — 'continuous' — point of view. **In some cases even the very concept of continuity should be revisited and overhauled.** When we speak of causality, two different statements are usually implied.

The future cannot affect the past. In particular, it is this statement which is used in developing the Kramers–Kronig relation between the real and imaginary parts of the dielectric constant. There is no reason to question this statement unless there exists a certain supreme being who uses the results of today's game to change the rules of the game played yesterday.

The future is uniquely and continuously determined by the past (a possible example being the wave function with its probabilistic interpretation). This conception was used, for example, in Landau's course [10] in deriving the Schrödinger equation. In physical systems, which are inherently discrete, this is not always the case. **Sometimes even arbitrarily detailed knowledge about the state of a system is not a sufficient basis for uniquely predicting the future**, not even if quantum uncertainty is taken into account. In particular, going in discrete steps may result in jumping over the bifurcation point of the solution: small and usually negligible effects then become dominantly important, making it impossible to predict which solution branch will be followed in the future. It is these effects that produce well-known distortions in the results of large-scale numerical calculations in plasma physics and geophysics, prevent doing long term weather forecasts, etc. According to B V Chirikov, it is only along these lines that the origins of life, and indeed of the free will of humans, can be elucidated [11].

• Our understanding of what is going on in the microworld is to a large extent based on the perturbation theory approach. In this approach, which is in fact the method of consecutive approximations, physical quantities are expanded in powers of an (assumed-to-be-small) dimensionless coupling constant that measures the strength of interaction. In quantum electrodynamics this is the fine-structure constant $\alpha = e^2/\hbar c \approx 1/137$. Thus, it is in fact believed that physical quantities in the neighborhood of point $\alpha = 0$ are analytical functions of α amenable to a series expansions within a certain radius of convergence; a small variation in α within this radius should leave the world picture qualitatively unchanged. Such an assumption about analyticity is inherently and obviously flawed, though: with the replacement $e^2 \to -e^2$ $(\alpha \to -\alpha)$, like charges start to attract and unlike, to repel each other, atoms no longer exist, and all electrons swarm into a single huge cluster, thus leading to an entirely different world from ours (Dyson).

Nevertheless, quantum electrodynamics results obtained by perturbation theory in α turn out to be true to a fantastic accuracy (a record high of 12 digits after the decimal point was obtained for the anomalous magnetic moment of the muon).

Why does perturbation theory work so well? One fantastic hypothesis is that we simply are very lucky: *It is precisely and only for the observed values of the coupling constants that our series expansions miraculously happen to describe reality.* If so, these values of coupling constants should follow somehow from theory (as, for example, eigenvalues for some problems).

• One of the discoveries of recent decades is that electromagnetic and weak particle interactions, previously thought to be totally unrelated, are but different manifestations of a unified electroweak interaction. It is only at distances larger than 10^{-16} cm, where the weak interaction involved in nuclear decays is observed, that these interactions separate and behave totally differently. It is believed that, in principle at least, the theory of electroweak interactions enables the scattering amplitudes of leptons and gauge bosons to be calculated to any desired degree of accuracy.

According to the general principles of quantum field theory, account must necessarily be taken of all intermediate states of the system that occur in the limit of infinitely separated particles (*asymptotic states*). These are the states of stable particles that form a complete system of intermediate states, each of the physical states entering the system only once. Unstable particles do not have asymptotic states (they disappear as $t \to \infty$) and do not enter the complete system of the intermediate states of QFT.

In constructing an efficient perturbation theory for electroweak interactions it is necessary with current computational methods that the complete system of intermediate states also include the set of states of all gauge bosons and leptons. We note, however, that most of these particles are unstable, and including the states of unstable particles with no asymptotic states unjustifiably 'doubles' the complete set.

Why is it that the standard form of the theory of electroweak interactions, in which all gauge bosons and leptons are considered to be fundamental, works well even though most of these particles should be removed from the complete system of states as unstable?

May it be that the theory of electroweak interactions is in need of some modification?

3. Why is the world the way it is? Weird numbers and the relations between them

• The currently known masses of u and d quarks and coupling constants of nuclear and electromagnetic interactions suggest that the neutron is heavier than the proton by 1.3 MeV. If this excess were less than 0.5 MeV, the neutron would be stable and a considerable part of matter would exist in the form of neutron stars rather than ordinary, visible stars.

On the other hand, a small change in the proton–neutron interaction energy could result in there being no bound state in the proton–neutron system (which is the deuterium nucleus). In this case, the synthesis of heavier nuclei (proceeding through the deuterium stage) would be ruled out.

In either of these cases the world would be entirely different and life would hardly exist.

Are these energy values accidental?

In this connection, two possibilities are currently being discussed.

(1) There are multiple different universes with diverse sets of constants. We observe only that set at which there exists an observer (*anthropic principle*).

(2) The realized set of constants is for some reason the only one possible. *Why?*

• **The masses of truly elementary particles vary from 171 GeV (t-quark) to 0.5 MeV (electron) (a drop by a factor of 3.5×10^5) and to a few tenths or hundredths of an electron-volt (neutrino) — another six orders of magnitude down. What is the origin of these numbers?** Why are these values so small compared to the Planck mass?

• The Standard Model of electroweak interactions assumes a scalar particle known as the Higgs boson as dominantly instrumental in separating weak and electromagnetic interactions at distances larger than 10^{-16} cm and in the origin of particle masses. (Discovering the Higgs boson is an important objective of the Large Hadron Collider (LHC) that is to be launched soon.) Both the separation of the interactions and the origin of particle masses are attributed to the *nonzero vacuum expectation value of the Higgs field* v (similar to the average magnetization of a magnet). This expectation value is uniquely determined by the masses of the W and Z bosons and is equal to $v \approx 246$ GeV (the Higgs boson itself is similar to the field quantum of a spin wave in this magnet's analogy).

Why is the mass of the t quark so close to $v/\sqrt{2}$?

• All matter we know of is built up from electrons, protons, and neutrons. The last, in turn, consist of u and d quarks; neutrons generate ν_e neutrinos as they decay. However, the elementary set $(\mathrm{u}, \mathrm{d}, \mathrm{e}, \nu_e)$ is doubled twice. There exist other sets of 'building blocks', namely $(\mathrm{c}, \mathrm{s}, \mu, \nu_\mu)$ and $(\mathrm{t}, \mathrm{b}, \tau, \nu_\tau)$, which apart from being heavier (with the exception of neutrinos), are entirely similar to the electron–proton–neutron set. These sets are known as generations (or families).

Why is the first generation of quarks and leptons reproduced in more than one copy? It looks strange because one generation alone seems quite enough to build the visible Universe with.

Is it true that there are exactly three such copies?

If so, why precisely three?

• It is readily seen when treating electromagnetic phenomena quantum mechanically that the *electromagnetic interaction is a gauge one*, i.e., can be written in such a form that all observable results remain unchanged if the wave function of a charged particle is multiplied by $\exp\left[\mathrm{i}e\phi(\mathbf{x})\right]$ and the vector-potential of the electromagnetic field is simultaneously gauge-transformed: $A_i(\mathbf{x}) \to A_i(\mathbf{x}) - \nabla_i\phi(\mathbf{x})$.

The property that the observable results of theory are unaffected by similar (although somewhat more complex) transformations turns out to be shared by all fundamental interactions.

Why are all known interactions of a gaugelike nature?

• Electroweak interactions at distances less than 10^{-16} cm are highly symmetric, so that differently charged particles of the same kind interact in the same way. These interactions are said to have the symmetry $\mathrm{SU}(2) \times \mathrm{U}(1)$, which at a distance of about 10^{-16} cm breaks down, resulting in the well-known electromagnetic and weak interactions becoming separated.

It was found about forty years ago that at small distances strong (nuclear) interactions reduce to the quark–quark interaction — one that is mediated by gluons similar to the way the electromagnetic interaction is mediated by photons. One and the same type of quarks (for example, the u quarks) comes in three different forms known as *colors* — red, blue and green (the analogy with human color vision here goes quite far as seen, for example, from the fact that a state made up of quarks of three different colors is colorless). It is therefore natural that this theory — the fundamental theory of strong interactions — is called quantum chromodynamics. It is in accordance with there being three basic (color) states that quantum chromodynamics has the symmetry group SU(3).

In particle detectors, only colorless states of the quark–gluon system are registered directly. At the same time, processes occurring at distances less than 10^{-13} cm cannot be consistently described unless the color structure of matter is taken into account. The jargon for this is that color does not fly far. This phenomenon is known as color confinement.

Why are there precisely three colors — no more or no fewer — in quantum chromodynamics?

• **Why do the weak and strong interactions have the symmetry groups they have?**

Is it possible that the answers to the questions raised below are related to the fact that these two different *gauge* interactions have different symmetry groups?

Why does the weak interaction (unlike the strong interaction) **prefer one of the two helicities? Why is it that only left-handed neutrinos are involved in this interaction?**

Why does color confinement occur in quantum chromodynamics and not in the electroweak interaction?

• **Why is it that only charges equal to the electron charge e or its multiples occur in Nature** or, in other words, why is electric charge quantized?[1]

Electric charge would necessarily be quantized if there existed a point Dirac monopole (a magnetic pole similar to an electric charge, two monopoles interacting by the Coulomb law g^2/r^2). In this case, an unambiguous description of our world would be possible only for $ge = 2\pi\hbar cn$ (where e is the electron charge, g is the magnetic charge of the monopole, and n is an integer). This means that the quantization of an electric charge can be explained if there is at least one point Dirac monopole somewhere in the Universe.

However, such a point monopole has not been discovered. Most likely, it does not exist in Nature, since its discovery would require a total revision of the current understanding of what our world looks like at very small distances. If it does not exist, it would be nice to know what exactly rules out its existence.

May there be other reasons why a charge is quantized?

• **Why is the total density of matter in the Universe so close to the critical density, so that the Universe as a whole turns out to be flat?**

4. Odd facts, manifestations of unknown interactions...?

• It has been long accepted by physicists that the description of elementary interactions remains unchanged (invariant) under each of the following transformations: the specular reflection of coordinates (P invariance), particle–antiparticle conjugation (C invariance), and time reversal (T invariance). W Pauli showed that under very general assumptions concerning the properties of the spacetime we live in, the observed physical picture of the world is invariant under the simultaneous application of these three transformations (CPT theorem), and individual interactions do not necessarily have to possess the above invariance properties separately.

In the mid-1950s, it was found that weak interactions do not possess mirror symmetry. This was soon followed by the discovery that, instead of this, CP invariance occurs, which is a specular reflection combined with particle–antiparticle transformation (combined parity conservation, *L D Landau*).

About ten years later it was found that the weak decays of neutral K and B mesons involve the violation not only of C and P invariances separately but also of CP invariance (CP symmetry violation). Other processes do not exhibit CP violation.

Why is CP invariance violated?

Why does it prefer to be conserved in most observed processes?

---

[1] Quarks and antiquarks that have fractal charges of $\pm e/3$, $\pm 2e/3$ cannot fly apart by a distance larger than the atomic nucleus size, 10^{-13} cm, and are not examined experimentally.

A parametrization of weak interactions has been proposed in which the properties listed above seem natural. But how is this parametrization related to the nature of elementary interactions? **Is the observed CP violation due to some new and unknown interaction or does the weak interaction have some properties that lead to this violation?**

- It has been discovered recently that only 5% of the energy of the Universe is in the form of ordinary matter: atomic nuclei (mostly protons), electrons, photons, and neutrinos. A further 20% is contained in dark matter, which is similar to ordinary matter but does not participate in interactions other than gravitational and so far has been detected only in astronomical observations. The remainder is dark energy, a mysterious substance for which the energy density versus pressure relation has the unusual form $\varepsilon = -p$.

What is dark matter? Existing models for the small-distance behavior of particles offer a range of promising dark matter candidates. Experiments are currently underway to search for signals from such candidates.

What is dark energy? Besides collecting and interpreting astrophysical data, is there any way to detect a signal from dark energy?

5. What next?

- As mentioned earlier, the electroweak interaction is decomposed into two interactions with lesser symmetry, the electromagnetic interaction and the weak interaction, at distances larger than 10^{-16} cm.

Do similar decompositions exist at smaller distances? What is the origin of the scales of such decompositions?

May it be that at small distances all interactions merge into a unified interaction with very high symmetry (Grand Unification), which with increasing distance (or with decreasing energy, which is the same thing) **is decomposed into separate interactions with lower symmetry** (ultimately, the strong, weak, and electromagnetic interactions)?

This is a very attractive scenario, and there is currently a wide discussion of the large list of its possible realizations: high symmetry groups, supersymmetries, string theory, etc. Possible experiments to confirm or refute some of these possibilities are being analyzed in detail. There is, however, no reliable way to specify the necessary energy scale for the relevant signals to be detected; all we do know is that such energies are beyond the reach of current experiment, and it is always safe to say *these signals cannot be observed today but will be observable when higher-energy accelerators come*. Current arguments for the existence of such unification are in fact just wishful thinking for giving beauty to short-distance physics, the meaning of beauty being, of course, author-specific.

Or is it possible that, on the contrary, at small distances symmetry is low, and the higher *approximate* symmetry observed at distances larger than the nuclear scale arises for reasons similar to those discussed in Refs [12, 13]?

- **Why is the dimension of spacetime precisely four?**

Two groups of models are underway in which the dimensionality of the full spacetime is higher, $d > 4$, and our four-dimensional spacetime is a subspace of this full space.

A nice analogy for phenomena occurring in the first group of models is quasi-one-dimensional crystals. In these, the shear elasticity modulus along one axis is much less than its perpendicular counterparts, resulting in motion, in fact, occuring along the low-rigidity axis. In this group of models, our spacetime appears as a branch of the full multidimensional space, akin to the direction of possible motion in a quasi-one-dimensional crystal. Motion in the direction of 'extra' variables is essentially impossible because it takes too much energy to escape our world.

In the models of the second group, all the 'extra' variables are assumed to be compacted to a very small spatial volume. A world within a long thin cylinder nicely illustrates what this means. If we observe motions on a scale larger than the radius of the cylinder, we see a one-dimensional world; transverse motion occurs only at distances smaller than the radius of the cylinder, meaning that the transverse coordinate is compacted.

Some models go so far as to calculate the total dimensionality d of space (superstring theory, in particular, yields $d = 11$) — but the arguments the models rely on are, in my view, less than convincing.

Whatever the model, though, the question remains — whence the four-dimensionality of the *observed* world?

Activities along these lines of research show that the calculation of the dimensionality of space is a worthy subject of scientific study.

6. Some concepts

By the mass of a particle is usually meant its rest energy mc^2, which is measured in electron-volts ($1\ \text{eV} = 1.6 \times 10^{-12}$ erg) or multiples thereof, such as megaelectron-volt ($1\ \text{MeV} = 10^6$ eV), gigaelectron-volt ($1\ \text{GeV} = 10^9$ eV), and teraelectron-volt ($1\ \text{TeV} = 10^{12}$ eV).

Leptons are spin-$1/2$ particles not participating in strong (nuclear) interactions. They include electrons, μ mesons (muons), τ leptons, and various types of neutrinos.

Gauge mesons (of electroweak interactions) are carriers of elementary interactions in electroweak theory. They comprise photons, W bosons, and Z bosons, the last two being, respectively, about 85 and about 95 times the weight of the proton (the nucleus of the hydrogen atom).

This work was carried out under the support of RFBR (grant No. 08-02-00334-a) and NSh (grant No. 1027.2008.2).

References

1. Chernin A D *Usp. Fiz. Nauk* **178** 267 (2008) [*Phys. Usp.* **51** 253 (2008)]
2. Lukash V N, Rubakov V A *Usp. Fiz. Nauk* **178** 301 (2008) [*Phys. Usp.* **51** 283 (2008)]
3. Ryabov V A, Tsarev V A, Tskhovrebov A M *Usp. Fiz. Nauk* **178** 1129 (2008) [*Phys. Usp.* **51** 1091 (2008)]
4. Okun L B *Usp. Fiz. Nauk* **177** 397 (2007) [*Phys. Usp.* **50** 380 (2007)]
5. Rubakov V A *Usp. Fiz. Nauk* **177** 407 (2007) [*Phys. Usp.* **50** 390 (2007)]
6. Ioffe B L *Usp. Fiz. Nauk* **176** 1103 (2006) [*Phys. Usp.* **49** 1077 (2006)]
7. Tsukerman I S *Usp. Fiz. Nauk* **175** 863 (2005) [*Phys. Usp.* **48** 825 (2005)]
8. Krasnikov N V, Matveev V A *Usp. Fiz. Nauk* **174** 697 (2004) [*Phys. Usp.* **47** 643 (2004)]
9. Ginzburg V L *O Fizike i Astrofizike* (The Physics of a Lifetime) 3rd ed. (Moscow: Byuro Kvantum, 1995) [Translated into English (Berlin: Springer, 2001)]
10. Landau L D, Lifshitz E M *Kvantovaya Mekhanika*: *Nerelyativistskaya Teoriya* (Quantum Mechanics: Non-Relativistic Theory) (Moscow: Nauka, 1984) [Translated into English (Oxford: Pergamon Press, 1977)]
11. Chirikov B, physics/0503072
12. Ginzburg I F *Yad. Fiz.* **25** 421 (1977) [*Sov. J. Nucl. Phys.* **25** 227 (1977)]
13. Forster D, Nielsen H B, Ninomiya M *Phys. Lett. B* **94** 135 (1980)