Brainstorming Physics:

What to do about a patent policy that is hurting physics

Peter Horst Rehm, former patent attorney

Well-intentioned laws and policy that prohibit the patenting of laws of nature have had unintended consequences. Patenting, with its emphasis on novelty and an objective standard of non-obviousness, is a manner of scientific publication that excels at finding and filling literature gaps. The denial of these benefits to natural laws have slowed progress in the fundamentals of physics. Several specific literature gaps are identified herein, each suggesting material that would be both novel and non-obvious according to patent law standards. These examples include the need to reevaluate the role of the electrical force in the atomic nucleus, overlooked mechanics that may explain how mass bends spacetime, and why a very early theory of inertia deserves reconsideration.

1. Introduction

There is no argument herein that anyone should be able to patent a principle of nature to obtain an exclusive right to it. The very concept is meaningless in application. Patenting is more than just the granting of exclusive rights to inventors. The patent system also serves as a global brainstorming session that identifies promising new ideas, catalogs them, and recognizes their inventors.

Normally, an inventor "claims" an invention and obtains a time-limited head start for its commercialization. In the case of laws of nature, the discoverer should be able to "claim to contribute" the idea, to obtain the documentary benefits of patenting without claiming any exclusive right to nature.

This slight change, adding the concept and words "to contribute" to patent claim language, would enable science to advance by openly processing truly novel ideas through a vetting system that is not duplicative of peer review but complementary to it. The patent novelty requirement will prevent the issuance of multiple similar patents.

2. Patenting as a system of vetting ideas

Peer review has sometimes been regarded as hostile to certain types of ideas, particularly those that do not fit well into existing paradigms. Such ideas belong in the patent system, which evolved to give them a chance to thrive or not, without requiring them to prove acceptance or merit in advance. The more transformative, disruptive, or pioneering an idea is, the easier it is to patent.

The main criteria are an adequate description, novelty, and non-obviousness. The description is rarely a problem because patent attorneys and agents typically prepare descriptions and the requirements emphasize completeness. Novelty, loosely described, means that there is no older document that already teaches the patent applicant's idea. Non-obviousness is not a matter of opinion. Again, loosely described, it is an objective standard that means the patent examiner cannot show a reasonable number of older documents that, without applicant-guided hindsight in their selection or application, would likely have been considered together to teach the applicant's idea.

Patent systems excel at publishing new ideas because the issuance of a patent is not discretionary. New ideas often encounter opposition such as a failure of others to understand the invention, disbelief that the invention would work, opposing economic interests, or resistance to change. These forms of opposition hardly find traction in the patent system.

Patents claims are specific descriptions of the boundaries of the invention. Examiners make sure these claims are trimmed down to the truly new and non-obvious, and most patent applications then issue as a patent without a problem. In exceptional cases when there is a genuine difference of understanding about what something means or how to apply the law, the inventor has the right to appeal within the patent office and eventually in court. The inventor has a legally enforceable right to a patent so long as the well-defined criteria are met.

The patent examination process involves literature searches. The closest documents found help define the boundaries of the invention. The remaining material in the patent application that is new and non-obvious will be the first material to fill a newly documented literature gap.

Sometimes an older publication will make reference to an invention in a way that teaches against it, perhaps saying it does not work, is laden with serious disadvantages, or is impossible. If the invention is that the inventor has overcomes these issues then such a reference is useful to patent practitioners to show that their client's invention is non-obvious, since experts in the field teach against it. It actually helps the inventor to have critics.

3. How physics has been hurt

In normal patent practice, the inventor comes up with the idea first, and the evidence of its novelty and non-obviousness must be determined. However, herein the process will be worked backwards. A review of the history of physics, using mostly deductive reasoning, shows that it is littered with incidents where ideas could not be seen, were not explored, or they were discarded when not fully developed, or they were discarded for reasons that would no longer be sufficient in light of today's knowledge. In every case, this allowed the formation of the type of literature gap that would support the patenting of a "claim to contribute" if such claims were allowable in patents.

The following examples present ideas that would be both novel and non-obvious according to patent law standards. When presented casually outside the patent system, this level of non-obviousness often spurs informal disbelief and rejection. This is why patent examiners cannot rely on opinion but must find older documents that already teach the inventor's idea to objectively call into question the novelty or non-obviousness of the proposed invention.

4. Specific examples of literature gaps

The first example contains within it several examples that can be thought of as addressing the realities that cause the various phenomena of the bending of spacetime. In other words, they address the question of how a celestial body (any mass, really) reaches outside of its own boundaries to affect the properties of space elsewhere.

Where is the literature on the following issues:

4.1 Superposition of electric charges from "neutral" matter.

Coulomb's laws, Maxwell's equations, and the principles of electromagnetics were developed before it was fully appreciated that electrically neutral matter abounds with electric charges. Without anyone doing anything wrong, this has led to a subtle inconsistency in how the principle of superposition is applied. This in turn has led to questions about whether the vector fields **E** and **M** are real entities of nature or just mathematical constructs and whether it is meaningful to compute the value of the vector **E** at a point in space where there is no electric charge. Filling in the literature gaps around these issues may reveal answers to these questions and additionally offer explanations for how mass here bends spacetime elsewhere.

The electric field **E** and the principle of superposition that define it have been so successful it can be easily thought to be a well-settled matter. But they only claim to define the net field and the net forces on charges. Where is the discussion of the total forces on charges that add up to these net force? By this I mean following the first step of the principle of superposition literally, without inconsistency and without taking shortcuts. Taking shortcuts has practical utility when only the net effects are needed.

The first step of applying the principle of superposition to electric fields is to determine the Coulomb influence of every electric charge in the system on a test charge (or test point), each without regard to the existence, location, or influence of any other charges. For convenience, the name "fieldlet" will refer to the electric effect of a single fundamental electric charge on the universe. A fieldlet has the same sign and initial strength as its charge of origin from which it emanates and with distance wanes in intensity primarily according to Coulomb's law. According to the first step, the fieldlets of multiple charges overlap without mixing. Overlapping fieldlets do not influence one another directly. Every point on a fieldlet contains information that includes its electrical sign, scalar strength, and a vector pointing to the location of its charge of origin $\frac{x}{c}$ seconds ago. Where fieldlets overlap a fundamental electric charge, they each cast their influence upon it, pushing or pulling according to the information contained at that point and Coulomb's law. They do not sum up their influences in the absence of a charge.

Importantly, a literal interpretation of the principle of superposition requires the inclusion of all the fundamental electric charges in neutral matter, because, after all, nature does not have human *a priori* knowledge that only a mathematical net result is desired nor that the multiple electric charges in a neutron or an atom can be treated as a group and thought of as neutral.

A first consequence of this view is that every fundamental electric charge has an individual relationship with every other fundamental electric charge, without the need for any intermediate entity such as the classical vector fields **E** or **M**. Every such individual relationship is determined by two fieldlets: The fieldlet of the first charge is felt as a force by the second charge, and the fieldlet of the second charge is felt as a force by the first charge.

Compared to classical field theory, the calculation of the net effect on a test charge will be substantially unaffected by adopting this point of view. Any minor changes will only be getting closer to modeling nature. This suggests that the electric field (and magnetic field) are merely mathematical constructs. They may be, and will continue to be, very useful for engineering purposes without being real entities of nature.

A second consequence of this view is that it reveals a constant intense struggle imposed on every fundamental electric charge, wherever it is, that is always in almost perfect balance. According to the first step of the principle of superposition, the influence of every fundamental charge reaches substantially everywhere in a manner that overlaps the influences of all other charges.

For example, if Earth has N_{pos} fundamental positive charges (presumably from up quarks), then the N_{pos} overlapping positive fieldlets, from a point of view on Earth's surface, would present a cumulative total positive field strength of:

$$\hat{\mathbf{E}}_{\mathbf{pos}} = \sum_{i=1}^{N_{pos}} k\frac{q_{pos,i}}{{r_i}^2}\hat{\mathbf{r}}_i$$

Likewise, N_{neg} fundamental negative charges (presumably from electrons and down quarks), via their N_{neg} overlapping fieldlets, would present a cumulative total negative field strength of:

$$\hat{\mathbf{E}}_{\mathbf{neg}} = \sum_{i=1}^{N_{neg}} k\frac{q_{neg,i}}{{r_i}^2}\hat{\mathbf{r}}_i$$

Either of these cumulative fieldlet strengths, positive alone or negative alone, would apply incomprehensible force on a test charge. But as they also overlap each other, according to these formulas and assuming a neutral Earth, there will be no net force. Yet the presence of such intensity in overlapping fieldlets is an opportunity to look for emergent phenomena that are not cancelled by neutrality.

A third consequence is that because of the abundance of neutral matter, a consistently applied imbalance, even if very slight, would have a noticeable emergent effect. For example, a little positive feedback would result in a slight bias for attraction among all matter. Such an imbalance could arise from the individual relationships formed at a charge. The charge would affect the many individual fieldlets that intersect it differently according to their relative signs, such that the scalar strength of individual attractive relationships would be slightly enhanced in intensity because of the attraction and the scalar strength of individual repulsive relationships would be slightly diminished in intensity because of the repulsion. To observers, this affect would appear to be charge-neutral because it would seem to be unrelated to net electric charge of macro-objects. According to the first step, it would penetrate all matter and be incapable of being shielded. In other words, it would behave just like gravity.

In the static case, the formula to describe the total electrostatic force between any two particles a and b, having fundamental charges q_a and q_b, respectively, and separated by distance r, would be Coulomb's law modified by an emergent gravity term. The second term implements the bias of gravity by slightly enhancing attraction and diminishing repulsion:

$$F_{ab} = k\frac{q_a q_b}{r^2} - k_g\frac{|q_a q_b|}{r^2}, \quad \text{where } 0 < k_g < 10^{-40}k$$

Over the years, the similarity between Coulomb's law and Newton's gravitational force equation have caused many people to wonder if gravity emerges from the electrical force. When only the first (Coulomb's law) term is considered, it makes no sense that gravity could be electrical. The second term, understood as occurring because of the penetration and range of all fieldlets of all matter, including neutral matter, and providing an opportunity for feedback, clarifies how gravity may emerge from the electrical force.

Applied to all electrostatic relationships between bodies A and B, having N_A and N_B fundamental electric charges, respectively, the total electrostatic force between them is:

$$\hat{\mathbf{F}}_{AB} = \sum_{a=1}^{N_A}\sum_{b=1}^{N_B} \left(k\frac{q_a q_b}{{r_{ab}}^2} - k_g\frac{|q_a q_b|}{{r_{ab}}^2}\right)\hat{\mathbf{r}}_{ab}$$

When dealing with large electric imbalances (a machine), the first term dominates. When dealing with large amounts of neutral matter (a heavy object and the earth), the first term adds up to zero and the second term dominates. The two terms result in very different phenomena and characteristics.

This is likely not the quantization of gravity that was sought after. One of the principles of non-obviousness is that difficult solutions are often not located where people are looking.

A fourth consequence is that whenever a charge accelerates, its fieldlet must adjust for the charge to stay concentric within it according to Coulomb's law. This adjustment would fly off from the charge in all directions like a wave. Such waves would be electromagnetic waves like light. This means that the overlapping fieldlets collective form an electromagnetic medium of light unlike any that has ever been considered.

According to Coulomb's law, this medium may be dominated by the fieldlets of the most proximate matter and thus would tend to appear stationary in the laboratory frame. The field-lets from the more distal parts of the earth would be weakened by the distance. The dominance of proximate fieldlets would be gradually lost as a test point gets farther from Earth's surface.

This medium of light is consistent with the null results of the famously influential Michelson-Morley type of experiments, all of which rely on Earth's motion to detect the wind of a supposed luminiferous aether. It should be noted that this medium of light is generated by Earth and is not the same as the concept of a luminiferous aether that is dragged by Earth.

At Earth's surface, a medium of light that is established by overlapping fieldlets would likely be most closely described as an Earth-centered, Earth-fixed non-inertial frame. It ought to be detectable by a mobile version of such tests in a vehicle that is moving relative to Earth's surface both rapidly enough and proximate enough to enable detection of anisotropy in the speed of light. Such tests have been proposed. Apparently, the reason they were not performed is the lack of a theory for the existence of such a medium of light. This now points out a literature gap surrounding such a theory and supporting such experiments.

A fifth consequence to be considered is whether the collective intensity of fieldlets, the absolute value sum of their charge's Coulomb influence, has other emergent effects such as the effects of slowing down light and processes happening to particles at relativistic velocities. This points out further literature gaps regarding the realities underlying gravitational lensing and time dilation.

4.2 Fractional charges and the justification of a separate nuclear force.

Prior to the discovery of the neutron in 1932, physicists were searching for how the intensively repulsive protons in an atomic nucleus could be held together. The excess mass of nuclei with more than one proton tempted the consideration that this excess mass was from proton-electron pairs in the nucleus, with the electrons somehow providing enough attraction to hold the entire nucleus together. It was thought that electrons lose their normal properties such as spin when inside the nucleus. This troubled line of thinking did not last and is hardly mentioned in texts today.

Experiments in the 1930s showed that protons shot at each other with low energies would scatter as predicted by Coulomb's law, but at higher energies that allowed them to approach within 5 femtometers, a new behavior arose that departed from the Coulomb predictions. This possible new force could neutralize Coulomb scattering before creating its own scattering, a fact that was treated as evidence that this force was strongly attractive even between protons.

Such experiments and the discovery of the neutron in 1932 painted a picture of the existence of a separate, non-electrical nuclear force. Three main reasons were given for why the cohesion of protons and neutrons in the atomic nucleus could not be explained with the electrical force. These three reasons are still given to this day, presented as providing absolute certainty that something other than electrical force is holding the atomic nucleus together.

The three reasons are:

(1) the overwhelming strength of nucleon bonding that overpowers the mutual electrical repulsion of the protons,

(2) the participation of electrically-neutral neutrons, and

(3) the short range or contact-like nature of the nuclear force that was unlike the infinite-range of the electrical force.

These three reasons marked the beginning of the nuclear force as a separate force of nature. They essentially ended the pursuit of theories of electrical bonding in the nucleus. Textbooks today tend to start nuclear theory with the discovery of the neutron in 1932.

However, these three reasons relied on limited views of these particles in the 1930s, such as the treatment of protons and neutrons as having single central-force charges of +1e and 0, respectively, with forces between protons measured between their respective centers. Also, the new nuclear force was given credit for the observed resilience (or hard core) of these particles, which would make a good argument in patent practice that such a belief may be conflating two different phenomena and would tend to inhibit proper consideration of the attractive phenomenon at distances closer than two radii of the particles.

What is pointed out herein is that since the 1970s, when both protons and neutrons were found to have fractional charges of both signs, these three reasons should have been reevaluated. Instead, we appear to have a literature gap.

If the three reasons had been properly reevaluated, it would have become apparent that they are no longer persuasive evidence of the existence of a non-electrical nuclear force.

If just one of the three reasons were to be compromised, it would be worth discussing and partially filling the literature gap. It turns out that all three reasons are compromised:

(1) As to the overwhelming strength of nucleon bonding, this can nevertheless be explained by the electrical force at closer distances. Plotting Coulomb's law all the way to zero distance shows that at some distance determined by the strength of the charges, the theoretically infinite-range characteristic of Coulomb's law gives way to a theoretically infinite-force characteristic. The curve transitions to a perpendicular asymptote. For the $+^2/_3$ e and $-^1/_3$ e fractional charges found in nucleons, nature conveniently places this distance at just under the radius of a nucleon. Thus, the overall electrical repulsion of protons at femtometer distances may be overcome by far more intense electrical attraction at sub-femtometer distances.

(2) As to the participation of electrically-neutral neutrons, it turns out that the neutrons are neutral only at distances large enough to support a net-central-force model. Up close, the neutrons are just as electrical as protons but have no net charge. Because the fractional charges are well-contained and also separate, the neutron may be capable of presenting a positive or negative fractional charge as the most proximate charge. When free to rotate, it would present an attractive fractional charge.

(3) As to the short range or contact-like nature of the nuclear force, these too can be explained electrically. In the contact case, any bonding strength observed between nucleons can be explained by Coulomb's law applied to sufficiently close proximity of oppositely-charged fractional charges in neighboring nucleons. For convenience, I will call this fractional charge bonding. The observed strength will suggest a bonding distance. In the non-contact but short-range case, free protons shot at each other should not be modeled as +1e central charges. Better models may include three separately-located fractional charges capable of rotation so that they present their $-^{1}/_{3}$ e fractional charges to each other, mitigating their overall mutual repulsion. Driven even closer, their negative fractional charges would threaten to become the dominant source of repulsion, so before that happens, they would likely rotate so that a pair of opposite fractional charges are the most proximate pair, with attraction starting to dominate over the overall Coulomb repulsion.

This shows that the above three reasons, as well as the conclusions drawn from high energy experiments, need to be reevaluated in light of the possibility that fractional charge attraction and bonding explains the observations. The evaluation should address whether the evidence continues to justify the existence of a separate, non-electrical nuclear force.

Exploring fractional charge bonding will lead to further literature gaps. Examples may include how closed rings and certain numbers of nucleons could form particularly stable structures (alpha particles, magic numbers), why only proton-neutron bonds seem to be found in nature (could it be spin?), how larger nuclei may be formed of complex structures of protons and neutrons, why smaller nuclei are most stable with similar numbers of protons and neutrons, why Z>20 nuclei start to have an excess of neutrons and whether the discarded idea of electrons in the nucleus could resurface here to make protons look like neutrons, the various types of radioactivity of isotopes, and more.

4.3 Max Abraham's theory of inertia deserves a second look.

Another literature gap relates to the lack of processing of new evidence that may be favorable to physicist Max Abraham's 1903 theory of the inertia of the electron. He proposed that the electron was repelled by its own field, such that at any constant velocity it would be at rest in its own field. Any acceleration of the electron would be resisted by negative feedback from an induced temporary imbalance in the field.

When the proton was later discovered, it had the same amount of charge and 1870 times the inertia, which was difficult to explain. The inertia of the neutral neutron was even harder to explain.

It is not uncommon for an inventor to go through a time when things look bleak and nobody believes in the merits of the invention. Abraham never gave up on his inertia, but he passed away due to an aggressive tumor long before it was discovered that both protons and neutrons get their visible net charges from the sum of internal fractional charges of both signs. This fact broke the expectation that particle inertia must be proportional to the net charge that is visible.

Currently, it is becoming clearer that these heavier particles are quite complex. There may yet be additional electric charges within that contribute to inertia without affecting the net charge. These developments call into question the appropriateness of rejecting Abraham's concept of electromagnetic inertia.

Abraham tied his theory of inertia to the existence of the luminiferous aether even though it was already known that the Michelson-Morley experiment could not detect this aether. As more scientists of the day rejected the aether, it would be natural for them to reject Abraham's inertia with it. However, the model of overlapping fieldlets predicts a non-aether medium of light that is compatible with the null results of the Michelson-Morley experiment. In addition, the concept of fieldlets is highly compatible with concept of electrically-induced inertia.

Inertia that is electrically-induced by fieldlets may also offer an explanation of the mechanics behind mass defect. An electric charge's center is the location from which it sources its fieldlet. This center point defines the axis that Coulomb's law asymptotically approaches with its literature-gapped arm that is characterized by force approaching infinity as distance approaches zero. The distance from an electric charge's center, where it sources its fieldlet, to where the negative feedback occurs is not known, but it must be much shorter than the radius of a nucleon. Within this paradigm, it seems reasonable to suppose that if two opposing electric charges were separated by an ordinary bonding distance (say, just under 0.1 fm), they would be electrically bonded to each other such that in some ways they would start to act like one particle, being proximate enough to feel each other's negative feedback. Because they are oppositely-charged, feeling each other's negative feedback would have a neutralizing effect, interfering with each other's expression of inertia. In other words, the positive charge would be attracted to the negative fieldlet's distortions of acceleration, and vice versa, in a manner that just slightly reduces the negative feedback of acceleration for both of them. Thus, the presence of such bonds in atomic nuclei would present as inertial mass defect. This, of course, is another literature gap.

5. Summary and conclusion

These models of overlapping fieldlets, fractional charge bonding, and fieldlet-based electrostatic inertia are examples of novelty and non-obviousness as those terms are used daily in the practice of patent law. The well-intentioned policies of protecting natural laws from claims of ownership may have played a role in keeping them hidden for so long. To the extent that they did, it was certainly an unintended consequence.

Pointing out the above examples of non-obvious literature gaps is only one purpose of this letter. The examples will certainly lead to many more insights and literature gaps that I will not be pursuing.

The more important purpose of this letter is to point out that patent law standards of deciding whether a proposed invention merits patenting ought to be applied to natural laws as well. This is not to provide any exclusive rights regarding the natural laws to their discoverers, as if it were even possible to prevent a competitor from using a natural law, but rather to allow a patent office to vet the ideas, catalog them, and give appropriate credit to the discoverers. This is necessary because the conception of unknown laws of nature is an inventive process, and therefore is subject to all that has been learned about novelty and non-obviousness, the predictable hostility to unfamiliar ideas, and the need to nurture our ideas as we nurture our infants.

If the discoverers of natural laws are in a position to do further research and need external funding, having a patent on their ideas would naturally focus funding opportunities upon them, even if no exclusive right to such funding is granted.

If such natural law patents were allowable subject matter in the patent office, then all of the above ideas would qualify for a patent when first proposed. According to patent law practice, dislike is irrelevant, and incompatibility with accepted theory strengthens the case for patentability. The only way a patent examiner would be able to prevent the issuance of a patent, or to trim the scope of the “claim to contribute,” would be to show older documents in which others taught the same or significantly similar ideas.

A final reason patent practice should be restructured to support progress in the fundamentals of physics is that it is uniquely adapted to respect the value of novel and non-obvious ideas. This value spurs researchers to place fresh eyes where nobody is looking, which is often where inventions reside.

Peter Horst Rehm
Former patent attorney