Conceptual Origins of Maxwell Equations and of Gauge Theory of Interactions

It is usually said that Coulomb, Gauss, Ampere and Faraday discovered 4 laws experimentally, and Maxwell wrote them into equations by adding the displacement current.

That is not entirely wrong, but obscures the subtle interplay between geometrical and physical intuitions that were essential in the creation of field theory.

19th Century

19.1

The first big step in the study of electricity was the invention in 1800 by Volta (1745-1827) of the Voltaic Pile, a simple device of zinc and copper plates dipped in seawater brine.

19.2

Then in 1820 Oersted (1777-1851) discovered that an electric current would always cause magnetic needles in its neighbor-hood to move.

Oersted’s experiment electrified the whole of Europe, leading to such devices as the solenoid, and to the exact mathematical laws of Ampere.

Ampere (1775-1836) was learned in mathematics. He worked out in 1827 the exact magnetic forces in the neighborhood of a current, as “action at a distance”.

Faraday (1791-1867) was also greatly excited by Oersted's discovery. But he lacked Ampère's mathematical training.

In a letter Faraday wrote to Ampère we read:

"I am unfortunate in a want to mathematical knowledge and the power of entering with facility any abstract reasoning. I am obliged to feel my way by facts placed closely together."

(Sept. 3, 1822)

Without mathematical training, and rejecting Ampere’s action at a distance, Faraday used his *geometric intuition* to “feel his way” in understanding his experiments.

- **In 1831 he began to compile his <*Experimental Researches*>, recording eventually 23 years of research (1831-1854). It is noteworthy that there was *not a single formula* in this whole monumental compilation.**

<u>19.3</u>

Then in 1831 Faraday discovered electric induction!

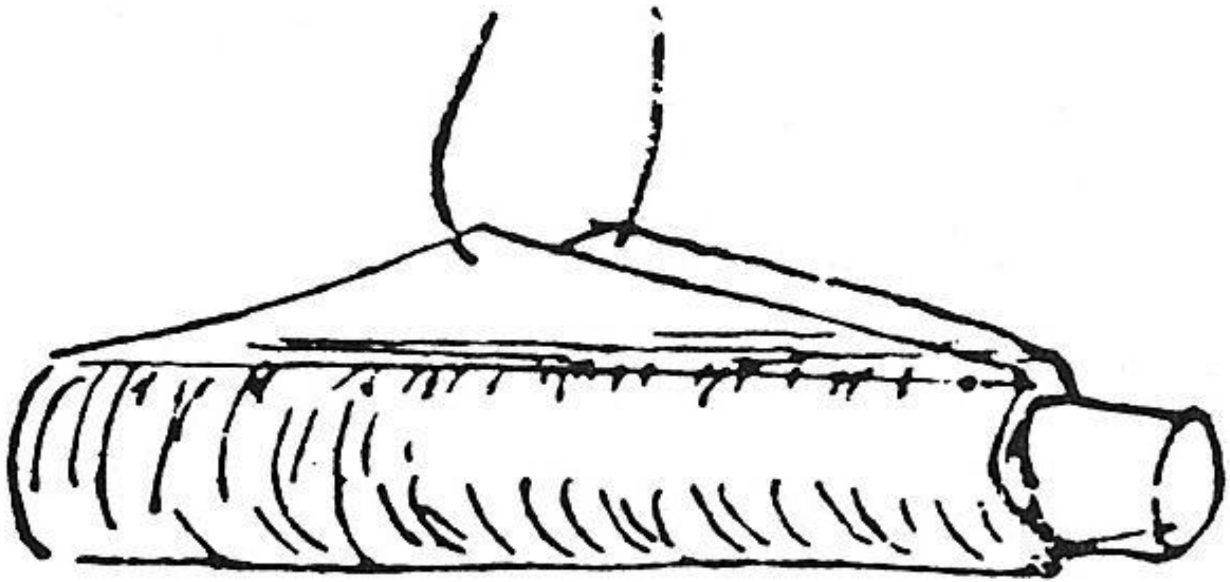


Fig. 2. A diagram from Faraday's Diary (October 17, 1831) (see Ref. 79). It shows a solenoid with coil attached to a galvanometer. Moving a bar magnet in and out of the solenoid generates electricity.

Faraday discovered how to convert kinetic energy to electric energy, thereby how to make electric generators.

- **This was of course very very important.**
- **But more important perhaps was his vague geometric conception of**
    - ***the electro-tonic state***

"a state of tension, or a state of vibration, or perhaps some other state analogous to the electric current, to which the magnetic forces are so intimately related."

<ER> vol. III, p.443

This concept first appeared early, in Section 60, vol. I of <*ER*>, but without any precise definition.

(Sec. 66) All metals take on the peculiar state

(Sec. 68) The state appears to be instantly assumed

(Sec. 71) State of tension

Faraday seemed to be impressed and perplexed by 2 facts:

- **that the magnet must be *moved* to produce induction.**
- **that induction often produce effects *perpendicular* to the cause.**

- **Faraday was “feeling his way” in trying to penetrate electromagnetism.**
- **Today, reading his <Experimental Researches>, we have to “feel our way” in trying to penetrate his geometric intuition.**

Faraday seemed to have 2 basic geometric intuitions:

- **magnetic lines of force, and**
- **electrotonic state**

The first was easily experimentally seen through sprinkling iron filings in the field. It is now called *H*, the magnetic field.

The latter, the *electro-tonic state*, remained Faraday’s elusive geometrical intuition when he ceased his compilation of <*ER*> in 1854. He was 63 years old.

- **That same year, Maxwell graduated from Cambridge University. He was 23 years old.**
- **In his own words, he *"wish to attack Electricity"*.**

Amazingly 2 years later Maxwell published the first of his 3 great papers which founded

Electromagnetic Theory as a Field Theory.

<u>19.4</u>

- **Maxwell had learned from reading Thomson's mathematical papers the usefulness of**

$$\boxed{H = \nabla \times A}$$

- **Studying carefully Faraday's voluminous <*ER*> he finally realized that**

  Electrotonic Intensity = A

- **He realized that what Faraday had described in so many words was the equation:**

$$E = -\dot{A}$$

- **Taking the curl of both sides, we get**

$$\nabla \times E = -\dot{H}$$

This last equation is Faraday's law in differential form. Faraday himself had stated it in words, which tranlates into:

$$\int E \cdot dl = -\frac{d}{dt}\iint H \cdot d\sigma$$

Comment 1 Maxwell used Stokes' Theorem, which had not yet appeared in the literature. But in the 1854 Smith's Prize Exam, which Maxwell had taken as a student, to prove Stokes' theorem was question #8. **So Maxwell knew the theorem.**

<u>Comment 2</u> **Maxwell was well aware of the importance of his paper 1. To avoid possible controversy with Thomson about the origin of equation**

$$\boxed{H = \nabla \times A}$$

Maxwell carefully wrote:

With respect to the history of the present theory, I may state that the recognition of certain mathematical functions as expressing the "electrotonic state" of Faraday, and the use of them in determining electrodynamic potentials and electromotive forces is, as far as I am aware, original; but the distinct conception of the possibility of the mathematical expressions arose in my mind from the perusal of Prof. W. Thomson's papers…

5 years later,

1861	**paper 2, part I**
1861	**paper 2, part II**
1862	**paper 2, part III**
1862	**paper 2, part IV**

19.5

The *displacement current* first appeared in Part III:

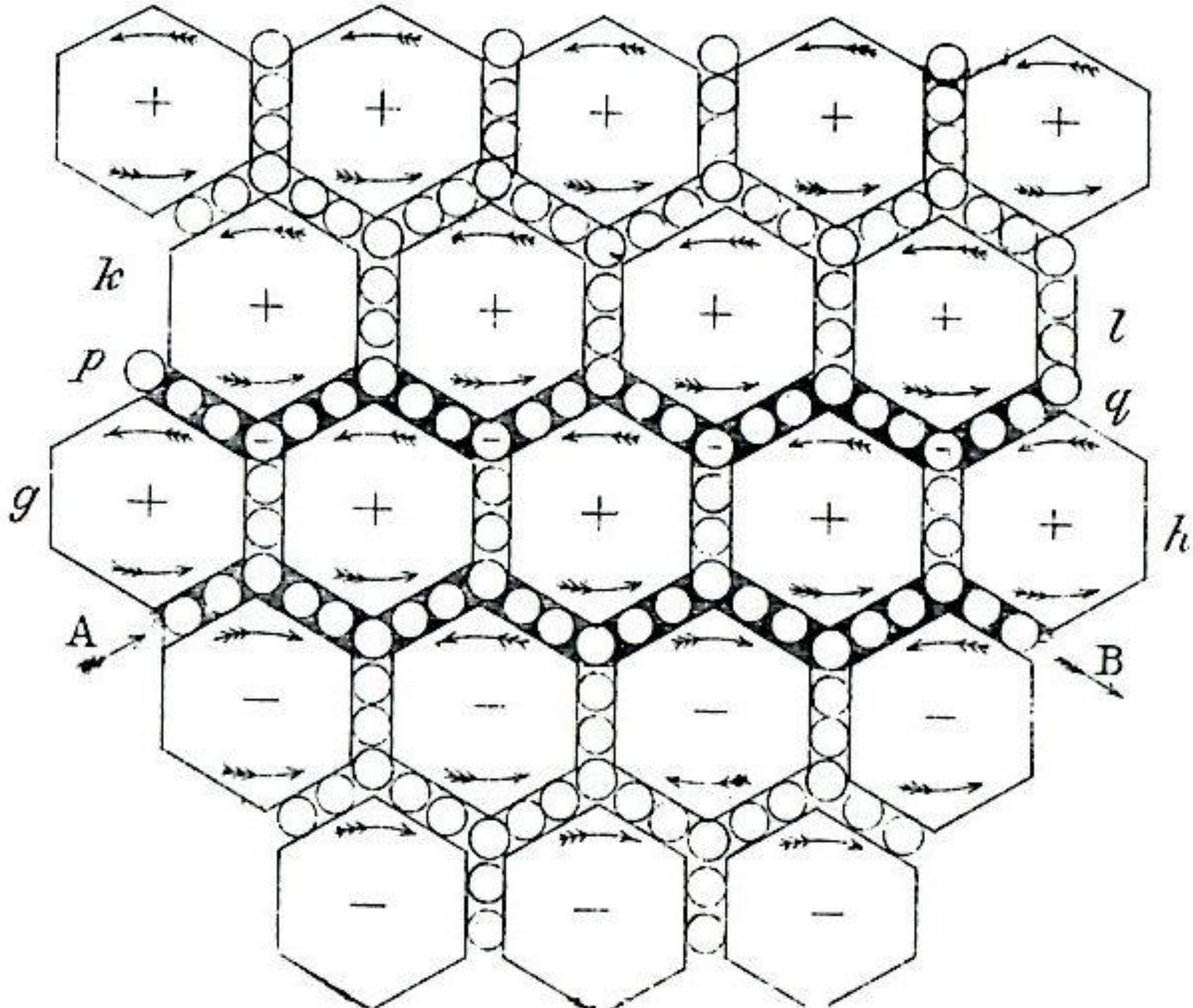
“Prop XIV – To correct Eq. (9) (of Part I) of electric currents for the effect due to the elasticity of the medium.”

I.e. He added the displacement current,

Maxwell arrived at this *correction*, according to his paper, through the study of a network of vortices.

I. PLATE VIII.

Fig: 2.



Maxwell *took this model seriously* and devoted 11 pages to arrive at the *correction*.

I made several attempts to understand these 11 pages. But failed.

With the correction, Maxwell happily arrived at the momentous conclusion:

"we can scarcely avoid the inference that *light consists in the transverse undulations of the same medium which is the cause of electric and magnetic phenomena*."

I.e. Light = EM waves.

<u>Comment</u> **Maxwell was a religious person. I wonder after this momentous discovery, did he in his prayers ask for God's forgiveness for revealing one of His Greatest Secrets.**

19.6

Paper 3 was published in 1865. It had the title: *A Dynamical Theory of the Electromagnetic Field*. In it we find the formula for energy density:

$$\frac{1}{8\pi}\left(E^2 + H^2\right).$$

Its Section (74) we read a very clear exposition of the basic philosophy of Field Theory:

"In speaking of the Energy of the field, however, I wish to be understood literally. All energy is the same as mechanical energy, whether it exists in the form of motion or in that of elasticity, or in any other form. The energy in electromagnetic phenomena is mechanical energy. The only question is, Where does it reside? On the old theories it resides in the electrified bodies, conducting circuits, and magnets, in the form of an unknown quality called potential energy, or the power of producing certain effects at a distance.

"On our theory it resides in the electromagnetic field, in the space surrounding the electrified and magnetic bodies, as well as in those bodies themselves, and is in two different forms, which may be described without hypothesis as magnetic polarization and electric polarization, or, according to a very probable hypothesis as the motion and the strain of one and the same medium."

That was historically

The first clear formulation of the fundamental principle of Field Theory

But Maxwell still believed there had to be an “aethereal medium”:

Comment Throughout his life time, M. ***always*** **wrote his equations with the vector potential *A* playing a key role. After his death, Heaviside and Hertz *gleefully* eliminated A.**

- **But with QM we know now that A has *physical meaning*. *It cannot be eliminated* (E.g. A-B effect).**

20th Century

Comment Thomson and Maxwell had both discussed what we now call the gauge freedom in

$$H = \nabla \times A$$

It was in the 20th century, with the development of QM, that this freedom acquired additional meaning in physics and mathematics, as we shall discuss below.

20.1

The first important development in the 20th century in physicists' understanding of interactions was Einstein's 1905 special relativity, according to which:

There is no aethereal medium. The EM field *is* the medium.

20.2

The next important development was the 1930-1932 discovery of the positron, which led to

Dirac's sea of negative energy particles, to

QED

QED was very successful in the 1930s in low order calculations, but was reset with infinities in higher order calculations.

20.3

1947-1950 Renormalization

$a=(g\text{-}2)/2$

Accuracy one pair in 10^9!

1950-1970

- Efforts to *extend* filed theory.
- Efforts to find *alternatives* of field theory.
- Return to field theory, to non-Abelian gauge theory.

20.4

1919 H Weyl:

"…the fundamental conception on which the development of Riemann's geometry must be based if it is to be in agreement with nature, is that of the infinitesimal parallel displacement of a vector. ..."

If in infinitesimal displacement of a vector, its direction keep changing then:

"Warum nicht auch seine Länge?" (Why not also its length?)

Based on this idea Weyl introduced a Streckenfacktor or Proportionalitätsfacktor,

$$\exp\left[-\int eA_\mu dx^\mu / \gamma\right]$$

where *γ* is real

Then in 1925-1926 Fock and London independently pointed out that in QM

$$\left(p - e\,\mathrm{A}\right) \quad \text{becomes} \quad -i\hbar\left(\partial_\mu - \frac{ie}{\hbar}\mathrm{A}_\mu\right),$$

I.e. Weyl's *γ* should be imaginary in QM

20.5

In 1929 Weyl published an important paper, accepting that *γ* should be imaginary, arriving at:

(a) A precise definition in QM of gauge transformation *both* for EM field, *and* for wave function of charged particles.

(b) Maxwell equations are invariant consider this combined gauge transformation.

Weyl’s gauge invariance produced no new experimental results. So for more than 20 years, it was regarded as an elegant formalism but not essential.

After WWII many new strange particles were found.

How do they interact with each other?

<u>20.6</u>

This question led to a generalization of Weyl's gauge invariance, to a possible new theory of interactions beyond EM. Thus was born non-Abelian gauge theory.

Motivation for this generalization was concisely stated in a 1954 abstract:

…the electric charge serves as a source of electromagnetic field; an important concept in this case is gauge invariance which is closely connected with (1) the equation of motion of the electromagnetic field, (2) the existence of a current density, and (3) the possible interactions between a charged field and the electromagnetic filed. We have tried to generalize this concept of gauge invariance to apply to isotopic spin conservation. ...

Non-Abelian gauge theory was very beautiful, but was not embraced by the physics community for many years because it seemed to require the existence of massless charged particles.

20.7

Starting in the 1960s the concept of spontaneous symmetry breaking was introduced which led to a series of major advances, finally to a

U(1) x SU(2) x SU(3) gauge theory of electroweak interactions and strong interactions called the *Standard Model*.

In the forty some years since 1970 the international theoretical and experimental physics community working in “particles and fields” combined their efforts in the development and verification of this model, with spectacular success, climaxing in the discovery of the “Higgs Boson” in 2012 by two large experimental groups at CERN, each consisting of several thousand physicists.

Comment Despite its spectacular success, most physicists believe the standard model is not the final story. One of its chief ingredients, the symmetry breaking mechanism, is a phenomenological construct which in many respects is similar to the four ψ interaction in Fermi's beta decay theory. That theory was also very successful for almost 40 years after 1933. But it was finally replaced by the deeper U(1) x SU(2) electroweak theory.

Entirely independent of developments in physics there emerged, during the first half of the 20th century, a mathematical theory called *fiber bundle theory*, which had diverse conceptual origins: differential forms (Cartan), statistics (Hotelling), topology (Whitney), global differential geometry (Chern), connection theory (Ehresmann), etc.. The great diversity of its conceptual origin indicates that fiber bundle is a *central* mathematical construct.

<u>20.8</u>

It came as a great shock to both physicists and mathematicians when it became clear in the 1970s that the mathematics of gauge theory, both Abelian and non-Abelian, is *exactly* the same as that of fiber bundle theory. But it was a *welcome shock* as it served to *bring back* the close relationship between the two disciplines which had been interrupted through the increasingly abstract nature of mathematics since the middle of the 20th century.

<u>Comment:</u> **In 1975 after learning the rudiments of fiber bundle theory from my mathematician colleague Simons, I showed him the 1931 paper by Dirac on the magnetic monopole. He exclaimed** ***"Dirac had discovered trivial and nontrivial bundles before mathematicians."***