Helmut's Recipe Number 8:

How the electromagnetic field, the canonical momentum, and the quantum mechanical interaction between charged particles and the electromagnetic field emerges solely by demanding local gauge invariance of the Schrödinger equation

(Version 3, 25 Feb. 2020)

This "recipe" explains what happens when you demand that the Schrödinger equation should have a so-called "local gauge invariance". This means that the wave function's observables should not only remain invariant under a *global* phase transformation $\Psi(\vec{r},t) \to \Psi(\vec{r},t)\, e^{i\varphi}$, but also under a *local* phase transformation $\Psi(\vec{r},t) \to \Psi(\vec{r},t)\, e^{i\,\varphi(\vec{r},t)}$. It turns out that this demand forces the introduction of new fields, and that these new fields can be identified with the scalar potential ϕ of the electric field and the vector potential $\vec{A}$ of the magnetic field.

So we are demonstrating the following fascinating fact here: The existence of an electromagnetic field and the interaction of charged particles with this electromagnetic field emerge solely from the requirement of the invariance of the observables of the Schrödinger equation under a local phase transformation (a.k.a gauge transformation)!

The following knowledge is required: Basic knowledge of classical quantum mechanics and electrodynamics. In chapter 3 (which can also be omitted), additional knowledge on how to calculate with four-vectors in Minkowski space is required.

SI units are used in all calculations.

1 Gauge invariance and local phase transformation

So let's get started: As we know, the Schrödinger equation for a free (spin-free) particle is

$$i\hbar\frac{\partial}{\partial t}\Psi(\vec{r},t) = -\frac{\hbar^2}{2m}\vec{\nabla}^2\,\Psi(\vec{r},t) \tag{1}$$

Now we replace $\Psi(\vec{r},t)$ with a wave function with a global (i.e. constant) phase φ:

$$\Psi(\vec{r},t) \to \Psi'(\vec{r},t) = e^{i\varphi}\,\Psi(\vec{r},t) \tag{2}$$

In this case, the Schrödinger equation (1) gives the same physical result, because the constant phase factor $e^{i\varphi}$ is not affected by the differential operators, and therefore cancels from the equation:

$$\begin{gathered} i\hbar\frac{\partial}{\partial t}\Psi'(\vec{r},t) = -\frac{\hbar^2}{2m}\vec{\nabla}^2\,\Psi'(\vec{r},t) \overset{(2)}{\Rightarrow} \\ i\hbar\frac{\partial}{\partial t}\left(e^{i\varphi}\,\Psi(\vec{r},t)\right) = -\frac{\hbar^2}{2m}\vec{\nabla}^2\left(e^{i\varphi}\,\Psi(\vec{r},t)\right) \Rightarrow \\ i\hbar\cancel{e^{i\varphi}}\frac{\partial}{\partial t}\Psi(\vec{r},t) = -\frac{\hbar^2}{2m}\cancel{e^{i\varphi}}\vec{\nabla}^2\,\Psi(\vec{r},t) \end{gathered} \tag{3}$$

However, things look different when we assume this transformation to be carried out with a position- and time dependent (i.e. "local") phase factor $\varphi(\vec{r},t)$:

$$\Psi(\vec{r},t) \to \tilde{\Psi}(\vec{r},t) = e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t) \tag{4}$$

What happens if we insert this new $\widetilde{\Psi}(\vec{r},t)$ into the Schrödinger equation?

$$i\hbar\frac{\partial}{\partial t}\widetilde{\Psi}(\vec{r},t) = -\frac{\hbar^2}{2m}\vec{\nabla}^2\,\widetilde{\Psi}(\vec{r},t) \tag{5}$$

Will we get the original equation (as before) and thus unchanged physics? Let's try! If we insert (4) in (5), we get:

$$i\hbar\frac{\partial}{\partial t}\left(e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t)\right) = -\frac{\hbar^2}{2m}\vec{\nabla}^2\left(e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t)\right) \tag{6}$$

To determine whether the Schrödinger equation remains unchanged, we first calculate how the differential operator $\frac{\partial}{\partial t}$ acts on $e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t)$:

$$\frac{\partial}{\partial t}\left(e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t)\right) = \frac{\partial e^{i\,\varphi(\vec{r},t)}}{\partial t}\Psi(\vec{r},t) + e^{i\,\varphi(\vec{r},t)}\frac{\partial\,\Psi(\vec{r},t)}{\partial t}$$

$$\frac{\partial}{\partial t}\left(e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t)\right) = i\frac{\partial\,\varphi(\vec{r},t)}{\partial t}e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t) + e^{i\,\varphi(\vec{r},t)}\frac{\partial\,\Psi(\vec{r},t)}{\partial t}$$

$$\underline{\frac{\partial}{\partial t}\left(e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t)\right) = e^{i\,\varphi(\vec{r},t)}\left(i\frac{\partial\,\varphi(\vec{r},t)}{\partial t} + \frac{\partial}{\partial t}\right)\Psi(\vec{r},t)} \tag{7}$$

Then we calculate in two steps how the Laplace operator $\vec{\nabla}^2$ acts on $e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t)$. First of all, we calculate the gradient …

$$\vec{\nabla}\left(e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t)\right) = i\left(\vec{\nabla}e^{i\,\varphi(\vec{r},t)}\right)\Psi(\vec{r},t) + e^{i\,\varphi(\vec{r},t)}\vec{\nabla}\,\Psi(\vec{r},t)$$

$$\vec{\nabla}\left(e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t)\right) = ie^{i\,\varphi(\vec{r},t)}\left(\vec{\nabla}\,\varphi(\vec{r},t)\right)\Psi(\vec{r},t) + e^{i\,\varphi(\vec{r},t)}\vec{\nabla}\,\Psi(\vec{r},t)$$

$$\underline{\vec{\nabla}\left(e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t)\right) = e^{i\,\varphi(\vec{r},t)}\left(i\vec{\nabla}\,\varphi(\vec{r},t) + \vec{\nabla}\right)\Psi(\vec{r},t)} \tag{8}$$

… and based on that, we calculate the effect of the Laplace operator:

$$\vec{\nabla}^2\left(e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t)\right) = \vec{\nabla}\cdot\vec{\nabla}\left(e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t)\right) \overset{(8)}{\Rightarrow}$$

$$\vec{\nabla}^2\left(e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t)\right) = \vec{\nabla}\cdot\left(e^{i\,\varphi(\vec{r},t)}\left(i\vec{\nabla}\,\varphi(\vec{r},t) + \vec{\nabla}\right)\Psi(\vec{r},t)\right)$$

$$\vec{\nabla}^2\left(e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t)\right) = \vec{\nabla}\cdot\left(ie^{i\,\varphi(\vec{r},t)}\left(\vec{\nabla}\,\varphi(\vec{r},t)\right)\Psi(\vec{r},t) + e^{i\,\varphi(\vec{r},t)}\vec{\nabla}\,\Psi(\vec{r},t)\right)$$

$$\begin{aligned}\vec{\nabla}^2\left(e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t)\right) = {} & i\left(\vec{\nabla}e^{i\,\varphi(\vec{r},t)}\right)\cdot\left(\vec{\nabla}\,\varphi(\vec{r},t)\right)\Psi(\vec{r},t) + ie^{i\,\varphi(\vec{r},t)}\left(\vec{\nabla}^2\,\varphi(\vec{r},t)\right)\Psi(\vec{r},t) + \\ & +ie^{i\,\varphi(\vec{r},t)}\left(\vec{\nabla}\,\varphi(\vec{r},t)\right)\cdot\left(\vec{\nabla}\,\Psi(\vec{r},t)\right) + \left(\vec{\nabla}e^{i\,\varphi(\vec{r},t)}\right)\cdot\left(\vec{\nabla}\,\Psi(\vec{r},t)\right) + e^{i\,\varphi(\vec{r},t)}\vec{\nabla}^2\,\Psi(\vec{r},t)\end{aligned}$$

$$\begin{aligned}\vec{\nabla}^2\left(e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t)\right) = {} & -e^{i\,\varphi(\vec{r},t)}\left(\vec{\nabla}\,\varphi(\vec{r},t)\right)^2\Psi(\vec{r},t) + ie^{i\,\varphi(\vec{r},t)}\left(\vec{\nabla}^2\,\varphi(\vec{r},t)\right)\Psi(\vec{r},t) + \\ & +ie^{i\,\varphi(\vec{r},t)}\left(\vec{\nabla}\,\varphi(\vec{r},t)\right)\cdot\left(\vec{\nabla}\,\Psi(\vec{r},t)\right) + ie^{i\,\varphi(\vec{r},t)}\left(\vec{\nabla}\,\varphi(\vec{r},t)\right)\cdot\left(\vec{\nabla}\,\Psi(\vec{r},t)\right) + e^{i\,\varphi(\vec{r},t)}\vec{\nabla}^2\,\Psi(\vec{r},t)\end{aligned}$$

$$\begin{aligned}\vec{\nabla}^2\left(e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t)\right) = {} & -e^{i\,\varphi(\vec{r},t)}\left(\vec{\nabla}\,\varphi(\vec{r},t)\right)^2\Psi(\vec{r},t) + ie^{i\,\varphi(\vec{r},t)}\left(\vec{\nabla}^2\,\varphi(\vec{r},t)\right)\Psi(\vec{r},t) \\ & +2ie^{i\,\varphi(\vec{r},t)}\left(\vec{\nabla}\,\varphi(\vec{r},t)\right)\cdot\left(\vec{\nabla}\,\Psi(\vec{r},t)\right) + e^{i\,\varphi(\vec{r},t)}\vec{\nabla}^2\,\Psi(\vec{r},t)\end{aligned}$$

$$\vec{\nabla}^2\left(e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t)\right) = e^{i\,\varphi(\vec{r},t)}\left(\vec{\nabla}^2 + 2i\left(\vec{\nabla}\,\varphi(\vec{r},t)\right)\cdot\vec{\nabla} + i\left(\vec{\nabla}^2\,\varphi(\vec{r},t)\right) - \left(\vec{\nabla}\,\varphi(\vec{r},t)\right)^2\right)\Psi(\vec{r},t) \tag{9}$$

Claim: This is equivalent to

$$\underline{\vec{\nabla}^2\left(e^{i\,\varphi(\vec{r},t)}\,\Psi(\vec{r},t)\right) = e^{i\,\varphi(\vec{r},t)}\left(i\vec{\nabla}\,\varphi(\vec{r},t) + \vec{\nabla}\right)^2\Psi(\vec{r},t)} \tag{10}$$

If this claim is to be correct, the differential operator $\left(i\vec{\nabla}\,\varphi(\vec{r},t)+\vec{\nabla}\right)^2$ from equation (10) must have the same effect on $\Psi(\vec{r},t)$ as the operator $\left(\vec{\nabla}^2+2i\left(\vec{\nabla}\,\varphi(\vec{r},t)\right)\cdot\vec{\nabla}+i\left(\vec{\nabla}^2\,\varphi(\vec{r},t)\right)-\left(\vec{\nabla}\,\varphi(\vec{r},t)\right)^2\right)$ from equation (9).

Proof:

$$\left(i\vec{\nabla}\,\varphi(\vec{r},t)+\vec{\nabla}\right)^2\Psi(\vec{r},t)=\left(i\vec{\nabla}\,\varphi(\vec{r},t)+\vec{\nabla}\right)\cdot\left(i\vec{\nabla}\,\varphi(\vec{r},t)+\vec{\nabla}\right)\Psi(\vec{r},t)$$

$$\left(i\vec{\nabla}\,\varphi(\vec{r},t)+\vec{\nabla}\right)^2\Psi(\vec{r},t)=\left(i\vec{\nabla}\,\varphi(\vec{r},t)+\vec{\nabla}\right)\cdot\left(i\left(\vec{\nabla}\,\varphi(\vec{r},t)\right)\Psi(\vec{r},t)+\vec{\nabla}\,\Psi(\vec{r},t)\right)$$

$$\left(i\vec{\nabla}\,\varphi(\vec{r},t)+\vec{\nabla}\right)^2\Psi(\vec{r},t)=-\left(\vec{\nabla}\,\varphi(\vec{r},t)\right)^2\Psi(\vec{r},t)+i\left(\vec{\nabla}\,\varphi(\vec{r},t)\right)\cdot\left(\vec{\nabla}\,\Psi(\vec{r},t)\right)+i\left(\vec{\nabla}^2\,\varphi(\vec{r},t)\right)\Psi(\vec{r},t)+$$
$$+i\left(\vec{\nabla}\,\varphi(\vec{r},t)\right)\cdot\left(\vec{\nabla}\,\Psi(\vec{r},t)\right)+\vec{\nabla}^2\,\Psi(\vec{r},t)$$

$$\left(i\vec{\nabla}\,\varphi(\vec{r},t)+\vec{\nabla}\right)^2\Psi(\vec{r},t)=\left(\vec{\nabla}^2+2i\left(\vec{\nabla}\,\varphi(\vec{r},t)\right)\cdot\vec{\nabla}+i\left(\vec{\nabla}^2\,\varphi(\vec{r},t)\right)-\left(\vec{\nabla}\,\varphi(\vec{r},t)\right)^2\right)\Psi(\vec{r},t)\quad q.e.d$$

Very good! If we now insert the intermediate results (8) and (10) into equation (6), we get:

$$i\hbar e^{\cancel{i\varphi(\vec{r},t)}}\left(i\frac{\partial\,\varphi(\vec{r},t)}{\partial t}+\frac{\partial}{\partial t}\right)\Psi(\vec{r},t)=-\frac{\hbar^2}{2m}e^{\cancel{i\varphi(\vec{r},t)}}\left(i\vec{\nabla}\,\varphi(\vec{r},t)+\vec{\nabla}\right)^2\Psi(\vec{r},t)\tag{11}$$

By eliminating the $e^{i\,\varphi(\vec{r},t)}$-terms and expanding the left side, we can write the equation like this:

$$\left(-\hbar\frac{\partial\,\varphi(\vec{r},t)}{\partial t}+i\hbar\frac{\partial}{\partial t}\right)\Psi(\vec{r},t)=-\frac{\hbar^2}{2m}\left(i\vec{\nabla}\,\varphi(\vec{r},t)+\vec{\nabla}\right)^2\Psi(\vec{r},t)\tag{12}$$

What we would have liked to see would have been a differential equation that leaves the observables of the wave function unchanged, no matter if we add a local phase to it (like above), or not (as in equation (1)). But this is obviously not the case, since $\varphi(\vec{r},t)$ remains in the operators both on the left and on the right side of the differential equation. Therefore, equations (1) and (12) are definitely not equivalent.

Our demand of "invariance" under a local phase transformation $e^{i\,\varphi(\vec{r},t)}$ is therefore not fulfilled. What to do now? Which modifications are required, and what do these modifications mean physically?

2 A gauge field saves the day and brings new physics

2.1 Left side of the Schrödinger equation

So we would like the observables of the wave function $\Psi(\vec{r},t)$ to remain unchanged when we introduce any position- and time-dependent phase factor $\varphi(\vec{r},t)$.

If we compare the left side of equations (1) and (12), we see the following difference:

$$i\hbar\frac{\partial}{\partial t}\;vs.\left(-\hbar\frac{\partial\,\varphi(\vec{r},t)}{\partial t}+i\hbar\frac{\partial}{\partial t}\right)\tag{13}$$

Obviously, we would like to “get rid” of the term $-\hbar\frac{\partial\,\varphi(\vec{r},t)}{\partial t}$.

Well, $\varphi(\vec{r},t)$ is a scalar, position- and time-dependent quantity. So we could try to invent another scalar, position- and time dependent quantity $\chi(\vec{r},t)$ and then add it to the operator in the following way to the left side of equation (12):

$$\left(-\hbar\frac{\partial\,\varphi(\vec{r},t)}{\partial t}+i\hbar\frac{\partial}{\partial t}+\hbar\frac{\partial\,\chi(\vec{r},t)}{\partial t}\right)\Psi(\vec{r},t)=\cdots\tag{14}$$

If we could now demand that $\chi(\vec{r},t)\overset{!}{=}\varphi(\vec{r},t)$, then we would get exactly what we need, right? The bothering term $-\hbar\frac{\partial\,\varphi(\vec{r},t)}{\partial t}$ in equation (14) would vanish.

But wait! We also need to add this correction to the original equation (1):

$$\left(i\hbar\frac{\partial}{\partial t}+\hbar\frac{\partial\,\chi(\vec{r},t)}{\partial t}\right)\Psi(\vec{r},t)=\cdots \tag{15}$$

At first, it looks like we've got a new problem. However, this is not the case, because equation (15) is simply a special case of equation (14) with $\varphi(\vec{r},t)=\chi(\vec{r},t)=0$.

Well, yes. That's all well and good, but obviously, this approach is a little too trivial. Where should this magical $\chi(\vec{r},t)$ suddenly come from?

Unless…

Unless we invent yet another position- and time-dependent quantity $\Phi(\vec{r},t)$ that represents a scalar field. We now demand that it is also a gauge field. This means that the field $\Phi(\vec{r},t)$ also has a gauge freedom, similar to the wave function $\Psi(\vec{r},t)$:

$$\Phi(\vec{r},t)\rightarrow\widetilde{\Phi}(\vec{r},t)=\Phi(\vec{r},t)-\frac{\partial\,\chi(\vec{r},t)}{\partial t} \tag{16}$$

This means that whenever we follow this transformation (16), it should result in the same physics (the same observables). Assuming that the new gauge field $\Phi(\vec{r},t)$ is a potential of some sorts, we can easily integrate it into the Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}\Psi(\vec{r},t)=\left(-\frac{\hbar^2}{2m}\vec{\nabla}^2+\Phi(\vec{r},t)\right)\Psi(\vec{r},t) \tag{17}$$

This can be re-written into

$$\left(i\hbar\frac{\partial}{\partial t}-\Phi(\vec{r},t)\right)\Psi(\vec{r},t)=-\frac{\hbar^2}{2m}\vec{\nabla}^2\,\Psi(\vec{r},t) \tag{18}$$

It turns out that it makes sense to factor out a constant q (the coupling constant, which describes how strong the interaction between the particle described by Ψ and the field Φ is in the given unit system). We thus re-write the left side of equation (19) as:

$$\left(i\hbar\frac{\partial}{\partial t}-q\,\Phi(\vec{r},t)\right)\Psi(\vec{r},t)=\cdots \tag{19}$$

If we now carry out gauge transformations (4) and (16) simultaneously, we get

$$\left(-\hbar\frac{\partial\,\varphi(\vec{r},t)}{\partial t}+i\hbar\frac{\partial}{\partial t}-q\,\Phi(\vec{r},t)+q\frac{\partial\,\chi(\vec{r},t)}{\partial t}\right)\Psi(\vec{r},t)=\cdots \tag{20}$$

Now we can make the bothering term $-\hbar\frac{\partial\,\varphi(\vec{r},t)}{\partial t}$ disappear! We simply demand that it cancels out to zero with $q\frac{\partial\,\chi(\vec{r},t)}{\partial t}$. We can now really "demand" this, because $\varphi(\vec{r},t)$ and $\chi(\vec{r},t)$ are arbitrarily gauge parameters that leave physics unchanged!

$$-\hbar\frac{\partial\,\varphi(\vec{r},t)}{\partial t}+q\frac{\partial\,\chi(\vec{r},t)}{\partial t}=0 \tag{21}$$

We can also write this as follows:

$$\frac{\partial}{\partial t}(-\hbar\,\varphi(\vec{r},t))+\frac{\partial}{\partial t}(q\,\chi(\vec{r},t))=0 \tag{22}$$

This equation is easiest to fulfill if we equate the expressions to which the $\frac{\partial}{\partial t}$ operators act:

$$-\hbar\,\varphi(\vec{r},t)+q\,\chi(\vec{r},t)=0 \tag{23}$$

Hence:

$$\varphi(\vec{r},t) = \frac{q}{\hbar}\chi(\vec{r},t) \tag{24}$$

If we insert this into equation (20), we see that our strategy worked and that the term we wanted to make disappear actually cancels out:

$$\left(-q\cancel{\frac{\partial\,\chi(\vec{r},t)}{\partial t}} + i\hbar\frac{\partial}{\partial t} - q\,\Phi(\vec{r},t) + q\cancel{\frac{\partial\,\chi(\vec{r},t)}{\partial t}}\right)\Psi(\vec{r},t) = \cdots \tag{25}$$

Now that we have successfully eliminated the gauge-parameter dependent term, the left side of the equation looks like this:

$$\underline{\left(+i\hbar\frac{\partial}{\partial t} - q\,\Phi(\vec{r},t)\right)\Psi(\vec{r},t)} = \cdots \tag{26}$$

By introducing the gauge field $\Phi(\vec{r},t)$ and by cleverly choosing the associated gauge parameter $\varphi(\vec{r},t) = \frac{q}{\hbar}\chi(\vec{r},t)$, the left side of the Schrödinger equation after the gauge transformation, shown in Formula (26), looks exactly as before the gauge transformation in formula (19)!

2.2 Right side of the Schrödinger equation

Now let's have a look to the right side of equation (12):

$$\ldots = -\frac{\hbar^2}{2m}\left(i\vec{\nabla}\,\varphi(\vec{r},t) + \vec{\nabla}\right)^2\Psi(\vec{r},t) \tag{27}$$

If we compare this with the right side of equation (1), we see a difference here too:

$$\vec{\nabla}^2 \; vs. \; \left(i\vec{\nabla}\,\varphi(\vec{r},t) + \vec{\nabla}\right)^2 \tag{28}$$

Let us first move the $\hbar^2$ in formula (27) into the quadratic operator:

$$\ldots = -\frac{1}{2m}\left(i\hbar\vec{\nabla}\,\varphi(\vec{r},t) + \hbar\vec{\nabla}\right)^2\Psi(\vec{r},t) \tag{29}$$

Now we factor out an imaginary number i within the squared brackets:

$$\ldots = -\frac{1}{2m}\left(i\left(\hbar\vec{\nabla}\,\varphi(\vec{r},t) + \frac{1}{i}\hbar\vec{\nabla}\right)\right)^2\Psi(\vec{r},t) \tag{30}$$

Because of $\frac{1}{i} = -i$ we can write this as:

$$\ldots = -\frac{1}{2m}\left(i\left(\hbar\vec{\nabla}\,\varphi(\vec{r},t) - i\hbar\vec{\nabla}\right)\right)^2\Psi(\vec{r},t) \tag{31}$$

We can now pull i out of the outermost brackets, which reverses the sign because of $i^2 = -1$:

$$\ldots = +\frac{1}{2m}\left(\hbar\vec{\nabla}\,\varphi(\vec{r},t) - i\hbar\vec{\nabla}\right)^2\Psi(\vec{r},t) \tag{32}$$

In equation (24) we have determined how $\varphi(\vec{r},t)$ can be expressed by $\chi(\vec{r},t)$. We are now using this:

$$\ldots = \frac{1}{2m}\left(q\vec{\nabla}\,\chi(\vec{r},t) - i\hbar\vec{\nabla}\right)^2\Psi(\vec{r},t) \tag{33}$$

We now want to eliminate the vector expression $q\vec{\nabla}\chi(\vec{r},t)$. Because it worked so well before, we are introducing a gauge field again; but this time it is a vector field $\vec{\mathrm{A}}(\vec{r},t)$. Again we demand some gauge freedom shaped just after the term that we want to make disappear[1]:

$$\vec{\mathrm{A}}(\vec{r},t) \to \tilde{\vec{\mathrm{A}}}(\vec{r},t) = \vec{\mathrm{A}}(\vec{r},t) + \vec{\nabla}\chi(\vec{r},t) \tag{34}$$

We insert this gauge field into expression (33) as follows:

$$\ldots = \frac{1}{2m}\left(q\vec{\nabla}\chi(\vec{r},t) - i\hbar\vec{\nabla} - q\vec{A}\right)^2 \Psi(\vec{r},t) \tag{35}$$

If we now carry out the gauge transformation (34), we get, as required:

$$\ldots = \frac{1}{2m}\left(q\vec{\nabla}\cancel{\chi(\vec{r},t)} - i\hbar\vec{\nabla} - q\vec{A} - q\vec{\nabla}\cancel{\chi(\vec{r},t)}\right)^2 \Psi(\vec{r},t) \tag{36}$$

The expression $-i\hbar\vec{\nabla}$ is nothing else than the momentum operator $\hat{\vec{p}}$, and so we can ultimately write the right side of the equation as follows:

$$\ldots = \underline{\frac{1}{2m}\left(\hat{\vec{p}} - q\vec{A}\right)^2 \Psi(\vec{r},t)} \tag{37}$$

The new total equation, composed of the results (26) and (37), looks like this:

$$\boxed{\left(i\hbar\frac{\partial}{\partial t} - q\,\Phi(\vec{r},t)\right)\Psi(\vec{r},t) = \frac{1}{2m}\left(\hat{\vec{p}} - q\,\vec{\mathrm{A}}(\vec{r},t)\right)^2 \Psi(\vec{r},t)} \tag{38}$$

With some basic knowledge in electrodynamics it is easy to recognize the following facts:

- The expression $\Phi(\vec{r},t)$ with gauge freedom $\Phi(\vec{r},t) \to \tilde{\Phi}(\vec{r},t) = \Phi(\vec{r},t) - \frac{\partial\,\chi(\vec{r},t)}{\partial t}$ is obviously the electric potential;
- The expression $\vec{\mathrm{A}}(\vec{r},t)$ with gauge freedom $\vec{\mathrm{A}}(\vec{r},t) \to \tilde{\vec{\mathrm{A}}}(\vec{r},t) = \vec{\mathrm{A}}(\vec{r},t) + \vec{\nabla}\chi(\vec{r},t)$ is obviously the magnetic potential.
- The expression $q\,\Phi(\vec{r},t)$ describes the particle's potential energy, and
- substituting $\hat{\vec{p}}$ with $\hat{\vec{p}} - q\vec{A}$ corresponds to the substitution of the kinetic momentum $\hat{\vec{p}}$ with the canonical momentum $\hat{\vec{\pi}}$.

It turns out:

- Equation (38) describes the interaction of a charged particle in an electromagnetic field in (non-relativistic) quantum mechanics! This is extremely remarkable because we only started with the Schrödinger equation and had nothing to do with Maxwell or electromagnetism! Only the requirement for gauge invariance led us to this result!

The mechanism just described is the special case of a general principle:

Whenever a physical law satisfies some global symmetry (here: invariance of the observables of the Schrödinger equation against global phase transformations), the stronger demand for invariance against local transformations (here: invariance against local phase transformations of the symmetry group $U(1)$ can only be achieved by introducing new fields, ie. a new interaction mediated by these gauge fields emerges.

---

[1] We will explain in chapter 3 why we – seemingly arbitrarily – have chosen a positive sign in that gauge transformation!

From what we know today, all known fundamental interactions can be formulated as gauge theories! Not only can the electromagnetic, strong and weak interaction be described with gauge theories based on the symmetry groups $U(1)$, $SU(2)$ und $SU(3)$, but you can even get to the gravity theory of general relativity by extending the (global) coordinate transformations of special relativity to local transformations[2]!

3 We close a small gap with the help of four-vectors

In the previous chapter we completely abstain from the usual four-vector representation in gauge theory. The idea behind this is to make the otherwise so abstract gauge theory easier to understand by example of the Schrödinger equation in conventional notation.

In addition, the use of four vectors is only really appropriate in relativistic covariant theories. Unlike the Dirac equation, the Schrödinger equation, on basis of which we have shown the principle of local gauge invariance, is not relativistic covariant.

A quick look at four vectors is still useful. After all, it should be possible to write down above results in four-vector notation even in the limit of non-relativistic velocities. Also, four-vector representation will help us to close a small gap in the above explanation.

3.1 General definitions

So far we have written the wave function as a function of position $\vec{r}$ and time t:

$$\Psi = \Psi(\vec{r}, t) \tag{39}$$

We are now switching to four-vectors, so we are writing from now on

$$\Psi = \Psi(x^\alpha) \tag{40}$$

where we define the four-coordinate vector x^α as

$$x^\alpha \mathrel{\widehat{=}} \begin{pmatrix} ct \\ \vec{r} \end{pmatrix} \tag{41}$$

with $\alpha = 0,1,2,3$. We assume the following Minkowski metric:

$$\eta_{\mu\nu} = \mathrm{diag}(1, -1, -1, -1) \tag{42}$$

Thus:

$$\partial_\mu \mathrel{\widehat{=}} \begin{pmatrix} \frac{1}{c}\partial_t \\ \vec{\nabla} \end{pmatrix};\ \partial^\mu \mathrel{\widehat{=}} \begin{pmatrix} \frac{1}{c}\partial_t \\ -\vec{\nabla} \end{pmatrix} \tag{43}$$

The electrical potential $\Phi(\vec{r}, t)$ and the magnetic vector potential $\vec{\mathrm{A}}(\vec{r}, t)$ can be merged into the four-potential A^μ:

$$A^\mu \mathrel{\widehat{=}} \begin{pmatrix} \frac{1}{\mathrm{c}}\Phi(x^\alpha) \\ \vec{\mathrm{A}}(x^\alpha) \end{pmatrix} \tag{44}$$

---

[2] UTYIAMA, R., *Phys. Rev.* 101 (1956) 1597.

3.2 Four-vectors explain a sign

Let's look again at the Schrödinger equation for a free particle that we started from:

$$i\hbar \frac{\partial}{\partial t} \Psi(\vec{r},t) = -\frac{\hbar^2}{2m} \vec{\nabla}^2 \Psi(\vec{r},t) \tag{45}$$

We can rewrite this as follows:

$$i\hbar \frac{\partial}{\partial t} \Psi(\vec{r},t) = \frac{1}{2m} \left(-i\hbar\vec{\nabla}\right)^2 \Psi(\vec{r},t) \tag{46}$$

Below, in comparison, is a rewritten version of equation (38), which is invariant under local gauge transformation (whereby we replace $\hat{\vec{p}}$ with $-i\hbar\vec{\nabla}$):

$$\left(i\hbar \frac{\partial}{\partial t} - q\,\Phi(\vec{r},t)\right) \Psi(\vec{r},t) = \frac{1}{2m} \left(-i\hbar\vec{\nabla} - q\,\vec{\mathrm{A}}(\vec{r},t)\right)^2 \Psi(\vec{r},t) \tag{47}$$

If we look at the differential operators highlighted in red and blue, the following substitutions apply:

$$\begin{gathered} i\hbar \frac{\partial}{\partial t} \rightarrow i\hbar \frac{\partial}{\partial t} - q\,\Phi(\vec{r},t) \\ -i\hbar\vec{\nabla} \rightarrow -i\hbar\vec{\nabla} - q\,\vec{\mathrm{A}}(\vec{r},t) \end{gathered} \tag{48}$$

This we can write as

$$i\hbar \begin{pmatrix} \frac{1}{c}\partial_t \\ -\vec{\nabla} \end{pmatrix} \rightarrow i\hbar \begin{pmatrix} \frac{1}{c}\partial_t \\ -\vec{\nabla} \end{pmatrix} - q \begin{pmatrix} \frac{1}{c}\Phi(x^\alpha) \\ \vec{\mathrm{A}}(x^\alpha) \end{pmatrix} \tag{49}$$

Or, taking into account (43) and (44), we can write in four-vector notation:

$$\boxed{i\hbar\partial^\mu \rightarrow i\hbar\partial^\mu - qA^\mu} \tag{50}$$

This is called <u>minimal substitution</u>. We can now close a small gap in the explanation from Chapter 2:

In (16), we introduced the following gauge freedom:

$$\Phi(\vec{r},t) \rightarrow \tilde{\Phi}(\vec{r},t) = \Phi(\vec{r},t) - \frac{\partial\,\chi(\vec{r},t)}{\partial t} \tag{51}$$

The negative sign highlighted in red was chosen arbitrarily. However, this choice means that we must (as shown in (48)) <u>substract</u> $q\,\Phi(\vec{r},t)$ from $i\hbar\frac{\partial}{\partial t}$.

In (34) we then decided (apparently arbitrarily and without explanation) to define the gauge freedom with a positive sign for the vector potential (highlighted in blue below), so that we ultimately must also <u>substract</u> $q\,\vec{\mathrm{A}}(\vec{r},t)$ from $-i\hbar\vec{\nabla}$.

$$\vec{\mathrm{A}}(\vec{r},t) \rightarrow \tilde{\vec{\mathrm{A}}}(\vec{r},t) = \vec{\mathrm{A}}(\vec{r},t) + \vec{\nabla}\,\chi(\vec{r},t) \tag{52}$$

We could have justified this at this point by already guessing that apparently the electromagnetic field is "arising" here, and it follows from the Maxwell equations that the gauge transformations of $\Phi(\vec{r},t)$ and $\vec{\mathrm{A}}(\vec{r},t)$ require a different sign.

We don't have to bother Maxwell, though! A look at (49) and (50) reveals that this choice was necessary so that the minimal substitution $i\hbar\partial^\mu \rightarrow i\hbar\partial^\mu - qA^\mu$ can be formulated covariantly in four-vector notation!