

Aharonov-Bohm-Effekt

Topologische Aspekte der Quantentheorie

Michael Köpf

5. Juli 2006

1 Einleitung

Im Jahre 1959 veröffentlichten die beiden Physiker Yakir Aharonov und David Bohm einen Artikel, in welchem sie verblüffende und ungeahnte Konsequenzen der Quantentheorie aufzeigten. Sie schlugen ein Experiment vor, bei dem ein Elektronenstrahl in zwei Teilstrahlen aufgespalten wird, die links und rechts an einer senkrechten Zylinderspule vorbei gelenkt und dahinter wieder superponiert werden (siehe Abb. 1). Dabei sei die Spule durch eine Metallfolie perfekt gegen die Elektronen abgeschirmt und zudem absolut "dicht", so dass nur im Inneren der Spule ein nichtverschwindendes Magnetfeld vorliegt. Obwohl die Elektronen sich deshalb ausschließlich im feldfreien Raum bewegen, entsteht bei der Überlagerung der Teilstrahlen ein Interferenzmuster, welches vom magnetischen Fluss durch die Spule abhängt. Diese verblüffende Tatsache bezeichnet man nach ihren Entdeckern als Aharonov-Bohm-Effekt.

2 Einfluss der Potentiale

Wie kann ein eingeschlossenes magnetisches Feld im feldfreien Außenraum wirken? Integriert man das \vec{B} -Feld über eine Teilfläche der Bildebene von Abb. 1, welche den Zylinderquerschnitt ganz einschließt, so erhält man über den Stokesschen Satz eine Beziehung zwischen dem magnetischen Fluss und einem Kurvenintegral über das Vektorpotential \vec{A} entlang des Flächenrandes. In Formeln:

$$\begin{aligned} \text{const.} = \Phi &= \int_F d\vec{f} \cdot \vec{B} = \int_F d\vec{f} \cdot (\nabla \times \vec{A}) \\ &= \oint_{\partial F} d\vec{r} \cdot \vec{A} \end{aligned}$$

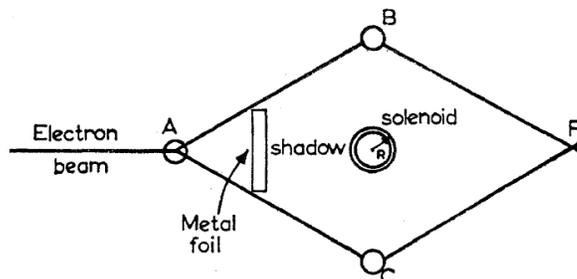


Abbildung 1: Der 1959 von Aharonov und Bohm vorgeschlagene Aufbau eines Experimentes zu dem von ihnen vorhergesagten Interferenzphänomen

Da dies für alle geschlossenen Wege gilt, die in irgendeiner Form um die Spule laufen, ist $\vec{A} \neq \vec{0}$ im Außenraum. Im Potential findet man also eine nichtverschwindende Größe im Außenraum, die in direktem Zusammenhang zum Magnetfeld innerhalb der Spule steht. In der klassischen Elektrodynamik wurden die Potentiale \vec{A} und ϕ als mathematische Hilfsgrößen eingeführt. Physikalische Realität wurde aber nur den Feldern \vec{B} und \vec{E} zugesprochen, die explizit im Lorentz-Kraftgesetz auftreten. Zwar gibt es auch von der klassischen Elektrodynamik eine Hamiltonsche Formulierung, in die die Potentiale explizit eingehen, doch stellt auch diese nur eine alternative mathematische Formulierung dar, welche keine neue Interpretation der Potentiale als reale Objekte erfordert. Die Situation in der Quantentheorie ist an dieser Stelle grundlegend anders. Egal, ob im Hamiltonian des kanonischen Formalismus oder im Wirkungsfunktional des Feynmanschen Pfadintegralformalismus: Die Potentiale treten an zentraler Stelle auf und spielen so eine viel direktere Rolle als in der klassischen Theorie.

2.1 Berechnung im kanonischen Formalismus

Im Folgenden soll der Einfluss des Vektorpotentials auf das Interferenzmuster in einem Aufbau nach Abb. 1 berechnet werden. Der Hamiltonian in Anwesenheit von \vec{A} hat folgende Gestalt:

$$H = \frac{1}{2m} (\vec{p} + e\vec{A})^2$$

Die zugehörige Schrödingergleichung löst

$$\Psi = \Psi_0 e^{i\frac{S}{\hbar}} \quad \text{mit} \quad \begin{cases} \Psi_0 : \text{Lösung für } \vec{A} = 0 \\ \nabla \frac{S}{\hbar} = e\vec{A} \Rightarrow S = e \int_C d\vec{r} \cdot \vec{A} \end{cases}$$

Die Problemstellung legt nahe, die Wellenfunktion in zwei Teilwellen (Teilstrahlen) Ψ_1, Ψ_2 zu zerlegen. Jedes Ψ_i sei dabei eine Lösung der SGL mit den oben genannten Eigenschaften. Für die Interferenz ist der Phasenunterschied bei Superposition der Teilwellen $\Psi = \Psi_1 + \Psi_2$ entscheidend. Dieser berechnet sich zu

$$\begin{aligned} \frac{1}{\hbar}(S_1 - S_2) &= \frac{e}{\hbar} \oint_{C_1 - C_2 = \partial F} d\vec{r} \cdot \vec{A} = \frac{e}{\hbar} \int_F d\vec{f} \cdot \vec{B} \\ &= \frac{e}{\hbar} \Phi, \end{aligned}$$

und ist damit flussabhängig, obwohl der Strahl nie einem magnetischen Feld ausgesetzt ist.

2.2 Berechnung im Pfadintegralformalismus

Um den Einfluss von \vec{A} im Pfadintegralformalismus zu berechnen, ist es zweckmäßig, einen anderen Aufbau als den aus Abb. 1 zu betrachten. Denkt man sich stattdessen einen Doppelspalt, hinter dem sich - wieder perfekt abgeschirmt und absolut dicht - eine Zylinderspule befindet und einen Schirm, auf dem man das Interferenzmuster beobachten kann (siehe Abb. 2), dann ist die interessante Größe die Übergangsamplitude von der Elektronenquelle am Punkte \vec{x} zu einem Punkt \vec{y} auf dem Schirm. Diese berechnet man mit einem Pfadintegral

$$\langle \vec{y}, t_1 | \vec{x}, t_0 \rangle = K(\vec{y}, t_1; \vec{x}, t_0) = \int \mathcal{D}x e^{iS[\vec{r}]}$$

Das Wirkungsfunktional ist das Zeitintegral über die Lagrangefunktion $L = \frac{m}{2} \dot{\vec{r}}^2 + e\dot{\vec{r}} \cdot \vec{A}(\vec{r})$. Wir zerlegen es in einen Anteil S_0 , der dem Integral über den kinetischen Term von L entspricht und einen potentialabhängigen Teil:

$$S[\vec{r}(t)] = S_0[\vec{r}(t)] + e \int_{t_0}^{t_1} dt \vec{A}(\vec{r}) \cdot \frac{d\vec{r}}{dt} = \int_C d\vec{r} \vec{A}(\vec{r})$$

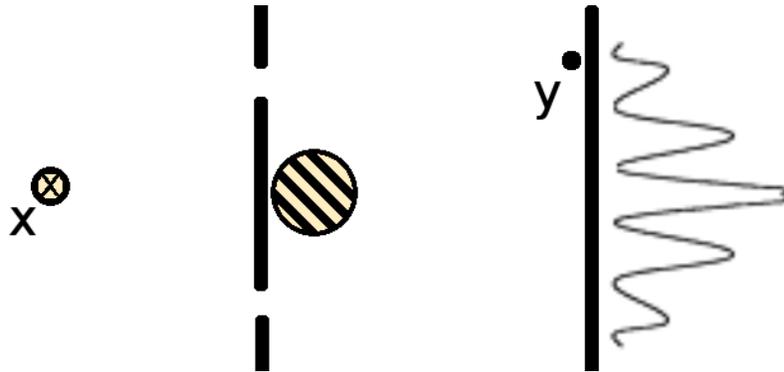


Abbildung 2: Ein modifizierter Yangscher Doppelspalt ist ein geeigneter Aufbau, um viele Eigenschaften des Aharonov-Bohm-Effektes- auch im Hinblick auf topologische Aspekte - zu demonstrieren. Im Punkt \vec{x} liegt die Elektronenquelle, der schraffierte Kreis hinter dem Doppelspalt ist der Querschnitt des Solenoiden

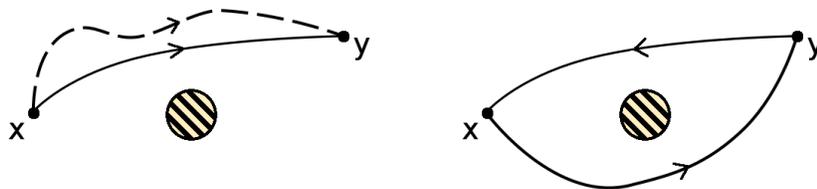


Abbildung 3: Links: Pfade vom Typ I, rechts: Die Kurve, entlang derer zur Berechnung des Phasenunterschieds integriert wird, entspricht dem Rand einer Fläche, durch den der magnetische Fluss Φ fließt

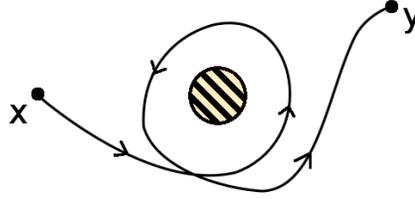


Abbildung 4: Auch um die Spule gewundene Pfade müssen berücksichtigt werden, hier: Ein Pfad vom Typ III (eine Windung)

Die Substitution zum Linienintegral macht auch den zweiten Term pfadabhängig. Das Pfadintegral ist die Summe über alle Pfade von \vec{x} nach \vec{y} . Fragt man sich, welche Arten von Pfaden in dieser Summation auftreten, so ist lediglich wichtig, wie sich die Pfade zur Spule verhalten. Bezeichnen wir alle Wege, die links am Solenoiden vorbei laufen als Wege vom Typ I, so finden wir für all diese Wege die Relation (siehe Abb. 3 links)

$$\begin{aligned} & \int_{C_a} d\vec{r} \cdot \vec{A} - \int_{C_b} d\vec{r} \cdot \vec{A} = \oint_{C_a - C_b} d\vec{r} \cdot \vec{A} \\ &= \int_F d\vec{f} \cdot (\nabla \times \vec{A}) = \int_F d\vec{f} \cdot \vec{B} = 0 \\ &\Rightarrow \int_{C \in \text{Typ 1}} d\vec{r} \cdot \vec{A} = \alpha_1 = \text{const.} \end{aligned}$$

Sowie eine analoge Beziehung für Wege, die rechts an der Spule vorbei gehen (Typ II):

$$\int_{C \in \text{Typ 2}} d\vec{r} \cdot \vec{A} = \alpha_2 = \text{const.}$$

Die Übergangsamplitude entspricht der Summe über alle Pfadklassen und ergibt sich somit (vorläufig, s.u.) zu

$$\begin{aligned} \langle \vec{y}, t_1 | \vec{x}, t_0 \rangle &= \int \mathcal{D}x e^{iS} = \int_{\text{Typ 1}} \mathcal{D}x e^{iS} + \int_{\text{Typ 2}} \mathcal{D}x e^{iS} \\ &= K_1 e^{ie\alpha_1} + K_2 e^{ie\alpha_2} \\ &= e^{ie\alpha_1} (K_1 + K_2 e^{ie(\alpha_2 - \alpha_1)}) \end{aligned}$$

Interessant für die Interferenz ist die Phasendifferenz $\alpha_2 - \alpha_1$. Diese berechnet sich zu

$$\begin{aligned} \alpha_2 - \alpha_1 &= \int_{C_2} d\vec{r} \cdot \vec{A} - \int_{C_1} d\vec{r} \cdot \vec{A} \\ &= \oint_{C_2 - C_1} d\vec{r} \cdot \vec{A} = \int_F d\vec{f} \cdot \vec{B} = \Phi \\ |\langle \vec{y}, t_1 | \vec{x}, t_0 \rangle| &= |K_1 + K_2 e^{ie\Phi}| \end{aligned}$$

Es gibt neben den Pfaden vom Typ I und Typ II noch weitere Typen, welche Wegen entsprechen, die sich einmal oder mehrfach um die Spule winden (siehe Abb. 4).

Unter Berücksichtigung aller Pfade mit höheren Windungszahlen erhält man schließlich das Ergebnis

$$\begin{aligned} \langle \vec{y}, t_1 | \vec{x}, t_0 \rangle &= K_1 e^{ie\alpha_1} + K_2 e^{ie\alpha_2} + K_3 e^{ie\alpha_3} + \dots \\ &= e^{ie\alpha_1} (K_1 + K_2 e^{ie(\alpha_2 - \alpha_1)} + K_3 e^{ie(\alpha_3 - \alpha_1)} + \dots) \end{aligned}$$

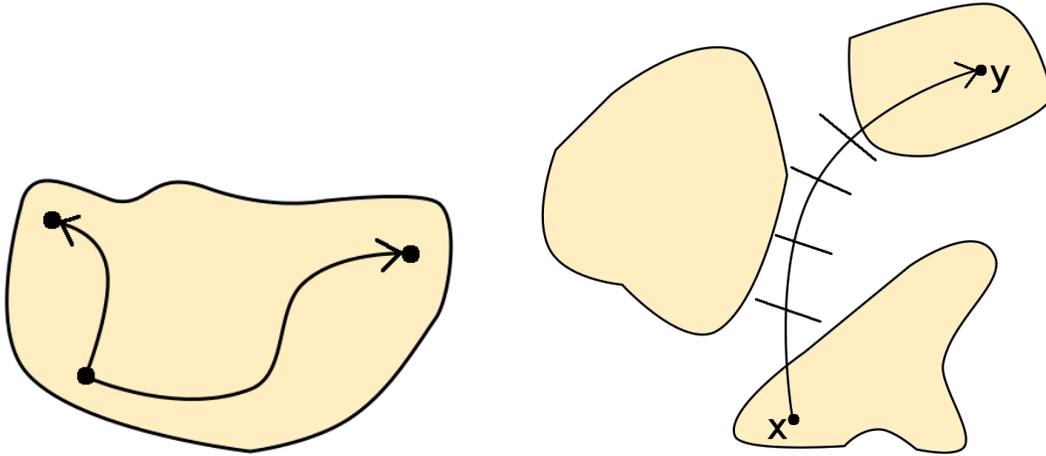


Abbildung 5: Links: In einem Gebiet lassen sich alle Punkte paarweise mit einer stetigen Kurve verbinden. Der schwarz eingezeichnete Rand gehört nicht mehr zu G , da dieses als Gebiet offen ist. Rechts: Ein Objekt aus disjunkten Teilmengen kann kein Gebiet sein, da sich zwei Punkte aus verschiedenen solcher Teilmengen nicht stetig innerhalb von G verbinden lassen

$$\Rightarrow |\langle \vec{y}, t_1 | \vec{x}, t_0 \rangle| = \left| \sum_{n \in \mathbb{Z}} K_n e^{ien\Phi} \right|.$$

In der Summation über \mathbb{Z} entsprechen negative Zahlen Pfaden, die sich rechts herum winden, während positive Zahlen Wege mit mathematisch positivem Drehsinn bezeichnen. Diese Klassifikation der Pfade nach Typ I, Typ II und ihrer Windungszahl stellt bereits, wie im folgenden Abschnitt deutlich werden wird, die Verbindung zur Topologie des Problems dar.

3 Topologische Betrachtungen

3.1 Mathematische Begriffe

Für eine topologische Beschreibung des Aharonov-Bohm-Effektes müssen einige mathematische Begriffe eingeführt werden. Wir definieren zuerst ein

Gebiet als einen *offenen* metrischen Raum G , der *zusammenhängend* und nicht die leere Menge ist.

Zusammenhängend bedeutet, dass sich je zwei Punkte aus G mit einer stetigen Kurve verbinden lassen, also $\forall x, y \in G \exists \varphi : [0, 1] \rightarrow G$ mit

- φ stetig
- $\varphi(0) = x, \varphi(1) = y$.

Der Begriff des Zusammenhangs lässt sich weiter verfeinern. Ein

einfach zusammenhängendes Gebiet G ist ein zusammenhängendes Gebiet, innerhalb dessen sich jede geschlossene Kurve *stetig* auf einen Punkt zusammenziehen lässt. Formal heißt das: Zu jeder stetigen Funktion $\varphi : [0, 1] \rightarrow G$ mit $\varphi(0) = \varphi(1) \exists$ ein fester Punkt $x_0 \in G$ und eine stetige Funktion $H = H(t, x) : [0, 1] \times G \rightarrow G$ mit $H(0, x) = \varphi(x), H(1, x) = x_0$. Man sagt dann: " φ ist homotop zu x_0 ".

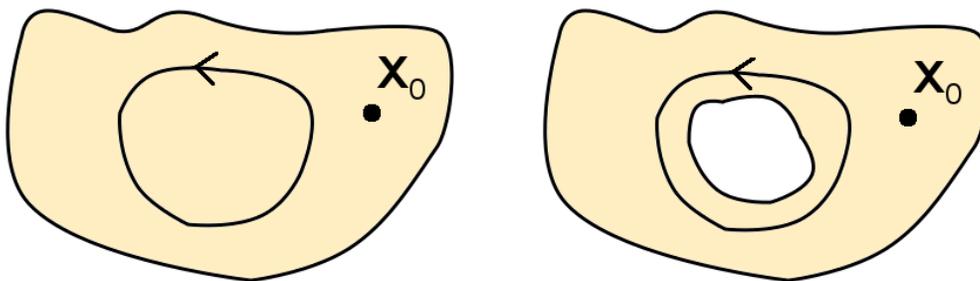


Abbildung 6: Links: Eine geschlossene Kurve in einem einfach-zusammenhängenden Gebiet, rechts: Dieselbe Kurve in einem mehrfach-zusammenhängenden Gebiet, welches durch "Herausknipsen" eines Bereiches aus einem einfach-zusammenhängenden Gebiet hervorgegangen ist

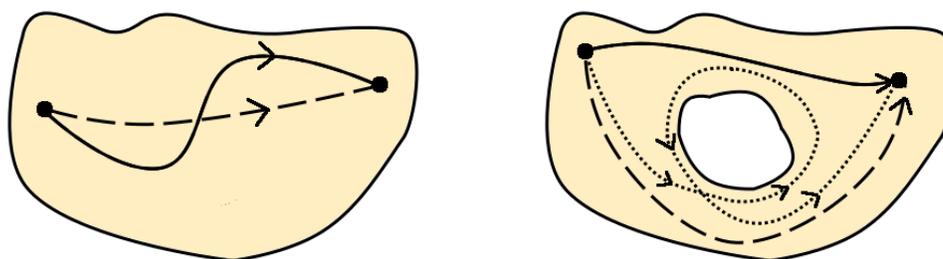


Abbildung 7: Funktionen mit identischem Anfangs- und Endpunkt in einem einfach-zusammenhängenden Gebiet und dem entsprechenden mehrfach-zusammenhängenden Gebiet mit einer Aussparung

Ein Gebiet, welches nicht einfach-zusammenhängend ist, bezeichnet man als **mehrfach-zusammenhängend**. Als einfaches Beispiel für ein solches Gebiet kann man sich ein einfach-zusammenhängendes Gebiet vorstellen, aus welchem man einen Bereich "herausknipst" (siehe dazu Abb. 6). Man kann sich einfach klar machen, dass man bei dem Versuch eine geschlossene Kurve, welche sich um die Aussparung windet, stetig in einen Punkt zu verformen, irgendwann am Rand der Aussparung hängen bleibt.

Der Begriff der Homotopie läßt sich auch auf nicht-geschlossenen Kurven verallgemeinern:

Zwei Kurven f und g sind zueinander homotop, falls sie sich stetig ineinander verformen lassen. Formal ausgedrückt: X, Y topologische Räume, $f, g : X \rightarrow Y$ stetig. f homotop zu g ($f \sim g$), falls

$$\forall x \in X \exists \text{ st. Funktion } h : [0, 1] \times X \rightarrow Y \quad \text{mit} \quad \begin{aligned} h(0, x) &= f(x) \\ h(1, x) &= g(x) \end{aligned}$$

Man kann sich anschaulich klar machen, dass in einem einfach-zusammenhängenden Gebiet alle Kurven zueinander homotop sind. Für unsere weiteren Überlegungen interessieren uns vor allem Kurven, die den selben Anfangs- und Endpunkt haben. Solche Kurven sind in einfach-zusammenhängenden Gebieten immer homotop zueinander. In mehrfach-zusammenhängenden Gebieten ist das grundsätzlich anders. Wir machen uns das wieder am Beispiel des Gebietes mit einer Aussparung deutlich. So sind dort zwei Kurven, von denen eine links und die andere rechts an

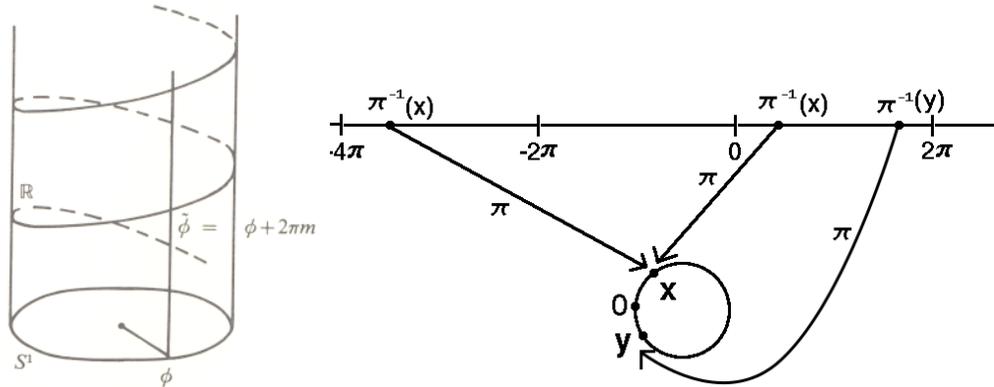


Abbildung 8: Der Einheitskreis und seine Universelle Überlagerung, der reelle Zahlenstrahl

der Aussparung vorbei gehen, nicht homotop. Allerdings: Alle Kurven, die rechts entlang laufen sind untereinander homotop und das selbe gilt für Kurven, die rechts vorbei laufen. Pfade, die sich mehrfach um das Loch winden, sind jeweils ausschließlich zu Pfaden mit derselben Windungszahl homotop (siehe dazu Abb. 7). Dies erlaubt eine Einteilung aller Kurven in sogenannte Homotopieklassen. Wir haben es hier mit einer abstrakten Version des zuvor betrachteten Doppelspalt-Experimentes zu tun. Um dies zu erkennen braucht man lediglich den Anfangspunkt der Kurven mit dem Ort der Elektronenquelle und den Endpunkt mit einem Punkt auf dem Beobachtungsschirm zu identifizieren. Die Aussparung entspricht dann genau der Spule, die für die Elektronen eine verbotene Zone ist. Offensichtlich entsprechen dann die Homotopieklassen genau den Klassen von Pfaden, die wir bereits bei der Berechnung im Pfadintegralformalismus eingeführt haben. So gibt es auch hier Pfade vom Typ I, die links am Loch vorbei laufen, Pfade vom Typ II, die rechts vorbei laufen sowie Pfade mit beliebigen Windungszahlen.

3.2 Quantentheorie auf mehrfach-zusammenhängenden Gebieten

Nachdem im letzten Abschnitt die Topologie eines Aharonov-Bohm-Experimentes untersucht wurde, soll jetzt die Physik ins Spiel gebracht werden. Ziel ist es, eine Quantentheorie auf mehrfach-zusammenhängenden Gebieten zu formulieren. Dazu führen wir einen weiteren wichtigen Begriff ein:

Die universelle Überlagerung (Universal Covering Space): Jeder topologische Raum Q besitzt einen UCS \tilde{Q} mit den Eigenschaften

- \tilde{Q} ist einfach-zusammenhängend
- \exists eine überdeckende Projektion $\pi : \tilde{Q} \rightarrow Q$
- \exists offene Überdeckung $\{U_i\}$ von Q , so dass $\forall U_i$ gilt:
 $\pi^{-1}(U_i)$ ist disjunkte Vereinigung offener Mengen in \tilde{Q} .
 Beschränkt auf eine dieser Mengen ist π Isomorphismus.

Tatsächlich besitzt ein UCS noch weitere Eigenschaften, die jedoch für die weiteren Überlegungen nicht wichtig sind und deshalb hier unerwähnt bleiben.

Zur Veranschaulichung betrachten wir das einfachste mehrfach-zusammenhängende Gebiet, den Einheitskreis, und seinen UCS den reellen Zahlenstrahl (Abb. 8). Man denke sich den Zahlenstrahl über dem Einheitskreis zu einem Helix aufgewickelt. Dann ist die Projektion π gerade die senkrechte Projektion auf den Kreis. Der Zahlenstrahl teilt sich in Intervalle der Breite 2π , die jeweils ein Urbild π^{-1} des Kreises darstellen. Jeder Punkt auf dem Kreis besitzt also unendlich viele

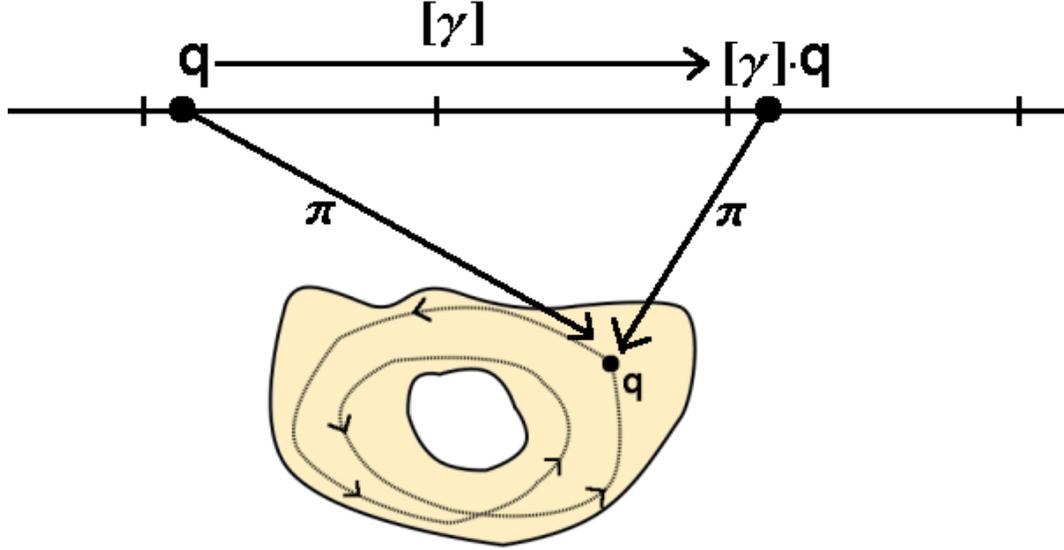


Abbildung 9: In einem mehrfach-zusammenhängenden Gebiet Q wird die Operation $[\gamma]$, die in \tilde{Q} wirkt, durch einen Pfad der Windungszahl γ repräsentiert. Die Darstellung von \tilde{Q} als Zahlenstrahl ist symbolisch zu verstehen. Natürlich wird der UCS eines Gebietes Q im Allgemeinen ein komplizierteres Konstrukt sein.

Urbilder im UCS, und zwar in jedem Intervall genau eines. Wir definieren $[\gamma]$ als die Operation, die ein Urbild $\tilde{q} = \pi^{-1}(x)$ in das entsprechende $\pi^{-1}(x)$ eines $2\pi\gamma$ weiter rechts liegenden Intervalls überführt, z.B.:

$$[2] \cdot \tilde{q} = \tilde{q} + 4\pi$$

Die Menge der Verschiebungen $[\gamma]$ bilden im Übrigen eine Gruppe, die sog. 1. Fundamentalgruppe π_1 , auf die an dieser Stelle jedoch nicht weiter eingegangen werden soll. Betrachtet man $[\gamma]$ als eine stetige Verschiebung von \tilde{q} nach $\tilde{q} + 4\pi$ so entspricht die Projektion dieser stetigen Verschiebung in Q einem γ -fachen Umlauf des Kreises.

Die Grundidee ist nun, eine "projizierbare Quantenmechanik" auf UCS \tilde{Q} zu formulieren und diese auf Q zu projizieren. Was macht eine Quantentheorie projizierbar? Man bezeichnet eine Wellenfunktion $\tilde{\Psi}$ auf \tilde{Q} als projizierbar, falls $|\tilde{\Psi}(\tilde{q})|^2$, $\tilde{q} \in \tilde{Q}$, nur von $q = \pi(\tilde{q})$ abhängt, d.h.

$$\tilde{\Psi}([\gamma] \cdot \tilde{q}) = a([\gamma])\tilde{\Psi}(\tilde{q}) \quad \forall \tilde{q} \in \tilde{Q} \quad (1)$$

wobei: $|a([\gamma])| = 1 \quad \forall [\gamma] \in \pi_1(Q)$.

Betrachten wir anstatt des Einheitskreises ein beliebiges mehrfach-zusammenhängendes Gebiet Q mit einer Aussparung, so können wir auch auf dessen UCS \tilde{Q} eine Operation $[\gamma]$ definieren. Die Projektion einer solchen Verschiebung auf Q entspricht dabei einem Pfad, der die Aussparung γ -mal umläuft. Dabei kommt es nur auf die Windungszahl, nicht jedoch auf die genau Gestalt des Pfades an. Jedes γ entspricht also einer Homotopieklasse. Aus der Projizierbarkeitsbedingung (1) folgt, dass eine Wellenfunktion entlang eines Pfades, der sich einmal um die Aussparung windet, maximal einen Phasenfaktor $a([\gamma])$ aufnehmen kann. Wichtig für Interferenzphänomene ist, dass nur Wellenfunktionen superponieren können, deren Phasenfaktor bei jedem Umlauf übereinstimmt. Andernfalls würde mit

$$\Psi = \Psi_1 + \Psi_2 \quad \Psi \rightarrow a_1 \{ \Psi_1 + a_1^* a_2 \Psi_2 \}$$

ein beobachtbarer Einfluß auf das Interferenzmuster entstehen, was wiederum gegen die Projektionsbedingungen verstoßen würde. Das führt dazu, dass der Hilbertraum aller Zustandsfunktionen

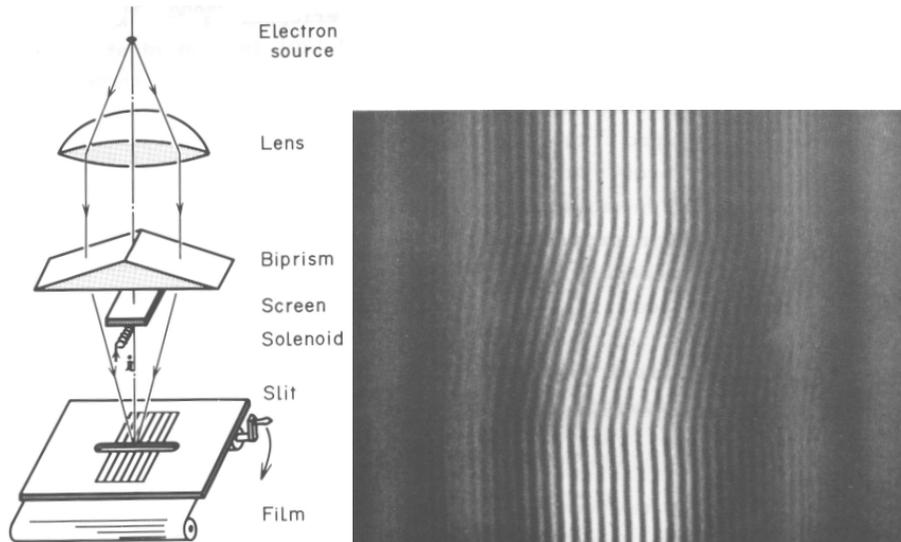


Abbildung 10: Links ist der schematische Aufbau des Experimentes von Möllenstedt und Bayh dargestellt. Mit diesem wird die zeitliche Entwicklung des Interferenzmusters auf einem Film aufgenommen. Rechts ist ein solcher Film zu sehen: Am unteren Bildrand fließt kein Strom durch die Spule, dann wird dieser hochgefahren (schräge Linien) und schließlich konstant gehalten.

in die direkte Summe von Unterräumen zerfällt, die jeweils zu einer festen Phase gehören. Superposition ist dann nur innerhalb eines solchen Unterraumes erlaubt.

4 Experimente zum Aharonov-Bohm-Effekt

Zum Aharonov-Bohm-Effekt sind zahlreiche Experimente angestellt worden. An dieser Stelle sollen zwei davon vorgestellt werden. Das Experiment von Möllenstedt und Bayh aus dem Jahre 1962 war der erste experimentelle Nachweis des AB-Effektes. Der Aufbau des Versuches und ein mit der Apparatur aufgenommenes Interferenzmuster sind in Abb. 10 dargestellt. Als Zweites sei hier das Experiment von Tonomura genannt, welches im Jahre 1986 durchgeführt wurde. Der technische Aufwand war dabei ungleich höher als bei Möllenstedt und Bayh. So wurde anstelle einer Zylinderspule ein Toroid verwendet, dessen Außenhülle aus einem Supraleiter bestand. Dieser fungierte einerseits als Abschirmung gegen die Elektronen und schloß andererseits - aufgrund des Meißner Effektes - das Magnetfeld dicht in seinem Inneren ein. Das Interferenzmuster wurde per Elektronenholographie aufgenommen. Der Aufbau und zwei Aufnahmen des Interferenzmusters sind in Abb. 11 dargestellt. Dieses Experiment, das den AB-Effekt mit sehr hoher Genauigkeit zeigt, zog einen Schlussstrich unter eine jahrzehntelange Diskussion über die Existenz des Effektes.

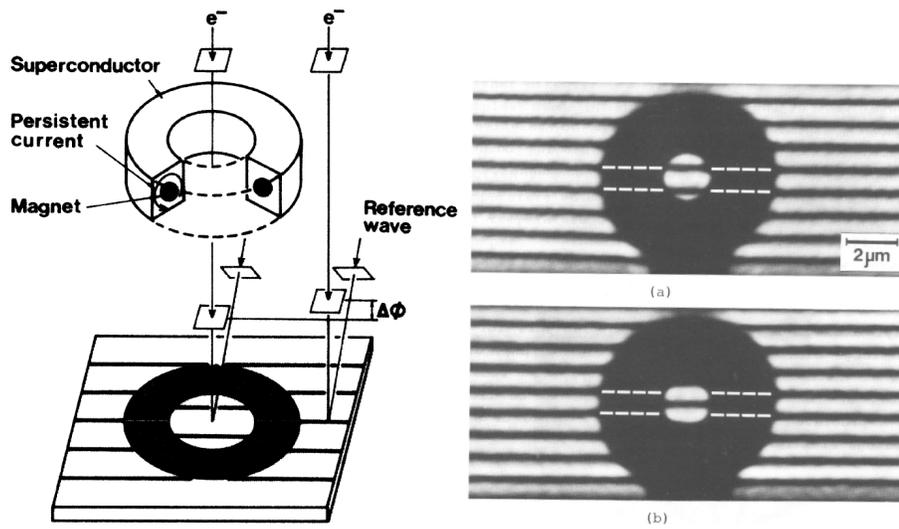


Abbildung 11: Links ist der Aufbau des Experimentes von Tonomura zu sehen. Rechts sind zwei Aufnahmen des Interferenzmusters gezeigt, wobei das obere Bild bei $T > T_C$ und das untere bei $T = T_C$ aufgenommen wurde. Der Maßstab macht deutlich, welche hohen Anforderungen der Aufbau an die Fertigungskünste der Experimentatoren seiner Zeit gestellt haben muss.