

Zeitschrift: Mitteilungen der Naturforschenden Gesellschaften beider Basel
Herausgeber: Naturforschende Gesellschaft Basel ; Naturforschende Gesellschaft Baselland
Band: 20 (2022)

Artikel: Der Transistor : ein winziger Kristall verändert die Welt
Autor: Gassmann, Fritz
DOI: <https://doi.org/10.5169/seals-1008147>

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

Download PDF: 02.02.2026

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Der Transistor – Ein winziger Kristall verändert die Welt

FRITZ GASSMANN

Limmatstrasse 6, CH-5412 Vogelsang, gassmann@bluewin.ch

Zusammenfassung: Vor rund 160 Jahren gelang J. C. Maxwell eine präzise mathematische Beschreibung aller bekannten Eigenschaften elektrischer und magnetischer Felder. Trotzdem bedurfte es weiterer Jahrzehnte, um zu erkennen, dass negativ geladene Elektronen die Träger des elektrischen Stroms in Festkörpern sind. Ganz wesentlich dazu beigetragen hat der junge Doktorand E. Hall mit seinen Experimenten, bei denen er Strom durch hauchdünne Goldfolien leitete. Platzierte er die Goldfolien in ein starkes Magnetfeld, konnte er quer zur Stromrichtung eine kleine Spannung nachweisen, die heute «Hall-Spannung» genannt wird. Sie wurde später für viele weitere leitende chemische Elemente gemessen und dabei fiel eine Gruppe von Elementen auf, die besonders grosse Hall-Spannungen zeigten, obwohl sie den Strom meist nicht besonders gut leiteten. Zu dieser Gruppe, die wir heute als Halbleiter bezeichnen, gehören Silizium und Germanium. Damit können Hall-Sonden hergestellt werden, die heute in Alltagsgeräten nicht mehr wegzudenken sind. Sie übernehmen wichtige Funktionen im Auto, im Handy, in der Kaffeemaschine oder bei der Einbruchsicherung. Noch wichtiger sind jedoch aus Halbleitern zusammengesetzte elektronische Bausteine wie Dioden und Transistoren, die die Basis der heutigen Elektronik bilden und in Computern milliardenfach untergebracht sind. Die Zusammenschaltung von Milliarden von Computern im Internet hat unsere Arbeitswelt und unser soziales Verhalten in den vergangenen drei Jahrzehnten tiefgreifend verändert.

Schlüsselwörter: Hall-Effekt, Hall-Sonde, Halbleiter, Elektronen- und Löcherleitung, Dotierung, pn-Übergang, Photovoltaik, Diode, Transistor, Gatter, Computer

Abstract: The Transistor – A tiny Cristal changes the World. About 160 years ago, J. C. Maxwell found a precise mathematical description of all known properties of electric and magnetic fields. Nevertheless, it took several more decades to understand that negatively charged electrons were carriers of electric current in solids. An important contribution came from the young Ph.D. student E. Hall, whose experiments with electric currents through very thin gold foils showed a tiny voltage perpendicular to the currents after switching on a strong magnetic field crossing the foils. Later on, this so called Hall-voltage was measured for many other chemical elements and a group was discovered showing much larger voltages, though the materials were not always good conductors. Today, this group is called «semiconductors» including e.g. the elements silicon and germanium. They are used today for Hall-sensors controlling everyday systems as e.g. cars, smartphones, coffee machines or burglar alarms. But even more important are diodes and transistors as basic components of todays electronics: they can be found in numbers of billions in computers. Linking of billions of computers by the internet has profoundly changed our working environments and our social interactions during the last three decades.

Key Words: Hall-effect, Hall-sensor, semiconductor, electron- and hole-conduction, doping, pn-junction, photovoltaics, diode, transistor, gate, computer

Vorwort

Die Erfindung, Feuer zu entfachen, des Rades, der Schrift oder von Dampfmaschinen, Elektro- und Benzinmotoren haben das Leben aller Menschen tiefgreifend verändert. In dieselbe Liga gehört die Erfindung des Transistors Ende der 1940er-Jahre. Ohne diesen winzigen Kristall wären tragbare Computer für jedermann, Internet, Mobiltelefone, GPS, Industrieroboter und vieles andere mehr undenkbar. Der Transistor hat die Arbeitswelt, die zwischenmenschliche Kommunikation sowie den Handel grundlegend verändert und die Welt zum Dorf gemacht. Ich möchte im vorliegenden Artikel die unerwartete und spannende Entstehungsgeschichte des Transistors beleuchten, sein Funktionsprinzip erklären und seine Anwendung zum Aufbau von Computern zeigen.

Elektrische und magnetische Kräfte gehören zusammen

Der englische Wissenschaftler Michael Faraday (1791–1867) entdeckte 1831 das Induktionsgesetz und erkannte damit die innige Verbundenheit von Elektrizität und Magnetismus. Abb. 1 zeigt eines seiner Experimente, mit denen er erkannte, dass ein elektrischer Strom in einer Spule ein Magnetfeld erzeugt, wobei Änderungen dieses Magnetfeldes in einer zweiten Spule eine Spannung verursachen. Damit hat Faraday ein veränderliches Magnetfeld mit einem induzierten elektrischen Feld verknüpft und gezeigt, wie die Energie zur Bewegung von Magneten in elektrische Energie umgewandelt werden kann.

James Clerk Maxwell (1831–1879) gelang ein weiterer entscheidender Schritt, indem er 1862 alle bekannten Eigenschaften von elektrischen und magnetischen Feldern durch vier elegante Differenzialgleichungen mathematisch beschreiben konnte. Er löste damit eine Welle von Innovationen aus, die das neue Zeitalter der Elektrotechnik einläuteten. Elektromotoren, Dynamos, Glühlampen, Telegrafie, usw. führten zu tiefgreifenden Veränderungen im Leben der Menschen.

Experimente von Edwin Hall korrigieren Maxwells Ansicht

Trotz zahlreicher technischer Erfolge waren noch etliche grundlegende Fragen offen. So waren die Natur des elektrischen Stromes und der Kräfte auf stromdurchflossene Leiter in Magnetfeldern noch unklar. Obschon G. J. Stoney und H. Helmholtz 1874 ein «Atom der Elektrizität» postulierten und es «Elektron» nannten, war Maxwell der Ansicht, dass die Kraft auf einen stromdurchflossenen Draht in einem Magnetfeld nicht auf hindurchfliessende Elektronen wirke, sondern direkt auf die unbewegte Leiterstruktur. Der Physiker Edwin Herbert Hall (1855–1938), der 1879 mit seiner Doktorarbeit an der Johns Hopkins University in Baltimore (USA) beschäftigt war, sah dies genau umgekehrt. Er suchte deshalb nach einem Experiment, um zu beweisen, dass die Kraft auf die im Draht bewegten Elektronen wirkt und nicht auf die unbewegte Leiterstruktur.

Hall erwartete, dass die bewegten Elektronen im Magnetfeld abgelenkt würden und dadurch die Elektronendichte auf der einen Seite des Lei-

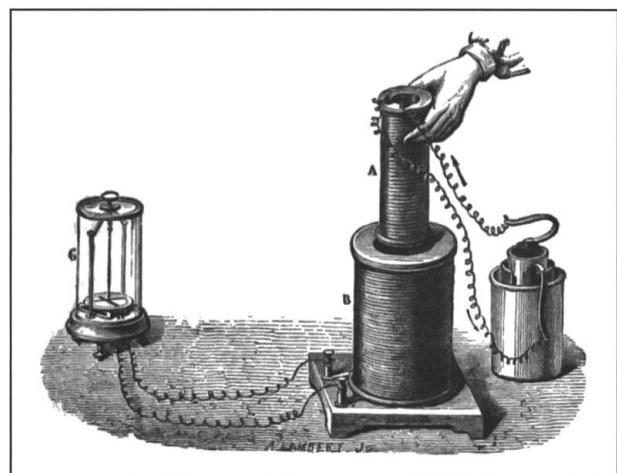


Abb. 1: Faradays Induktionsexperiment. Strom aus der Batterie rechts im Bild fließt durch die dünne Spule und erzeugt ein Magnetfeld. Wird diese Spule in eine zweite Spule geschoben, oder verändert sich der Strom durch die erste Spule, wird in der zweiten Spule eine Spannung induziert, die im Galvanometer links im Bild angezeigt wird. Diese Spannung entsteht jedoch nur, solange sich etwas verändert. Bei konstantem Strom und konstanter Lage der dünnen Spule wird keine Spannung induziert (Bildsignatur J. Lambert, Wikimedia Public Domain).

ters grösser würde als auf der gegenüberliegenden Seite. So entstünde ein elektrisches Feld quer zur Fliessrichtung der Elektronen, wodurch eine Spannung quer zum Leiter gemessen werden könnte. Diese Spannung müsste proportional zur Strom- und Magnetfeldstärke sein und das Vorzeichen bei Umkehrung des Stromes ebenfalls umkehren. Nach mehreren erfolglosen Versuchen verwendete er schliesslich als Leiter eine sehr dünne Goldfolie, die er auf eine Glasplatte klebte und mit vier Anschlüssen versah. Der Strom floss in Längsrichtung durch den Folienstreifen, und zwei Anschlüsse an gegenüberliegenden Rändern (in der Mitte des Streifens) verband Hall mit einem sehr empfindlichen Galvanometer (Gerät zur Messung kleinster Spannungen oder Ströme) und konnte so tatsächlich die erwartete Spannung nachweisen!

Heute ist die Hall-Spannung U_{Hall} mithilfe der von H. A. Lorentz 1895 abgeleiteten Beziehung sehr einfach bestimmbar (vgl. Herleitung in Abb. 2) und beträgt $U_{\text{Hall}} = R_{\text{Hall}} \times J \times B / d$, wobei J die Stromstärke in Ampère, B die Magnetfeldstärke in Tesla und d die Dicke der Goldfolie in Metern bedeuten. R_{Hall} ist der Hall-Koeffizient, der heute für viele Leitermaterialien bekannt ist (z.B. Koon 2000, D'Ans/Lax 1967). Für Gold beträgt er $0,71 \times 10^{-10} \text{ m}^3/\text{C}$ ($\text{C} = \text{Coulomb}$, die Masseinheit für die Ladung). Halls Goldfolie dürfte rund einen Mikrometer dick gewesen sein und seine Magnetfelder dürften um 0,01 Tesla betragen haben. Für eine Stromstärke von einem Ampère erhält man damit eine Hall-Spannung von 0,71 Mikrovolt. Erstaunlich, dass Hall diese Spannung genügend genau messen konnte! Die Polarität der Spannung war so, dass Hall auf sich bewegend negative Ladungsträger schliessen konnte, die er mit Elektronen gleichsetzte.

Der Hall-Effekt verrät Details zum Aufbau der Atome

Erstaunlicherweise kann der Hall-Effekt, ähnlich wie die Spektrallinien, Einblick in den Aufbau der Atome verschaffen. Das Geheimnis liegt im Hall-Koeffizienten, der, wie in Abb. 2 gezeigt, bei bekannter Ladung e der für die Stromleitung verantwortlichen Ladungsträger direkt mit ihrer Dichte N verknüpft ist: $N = 1/(e \times R_{\text{Hall}})$. Beruht

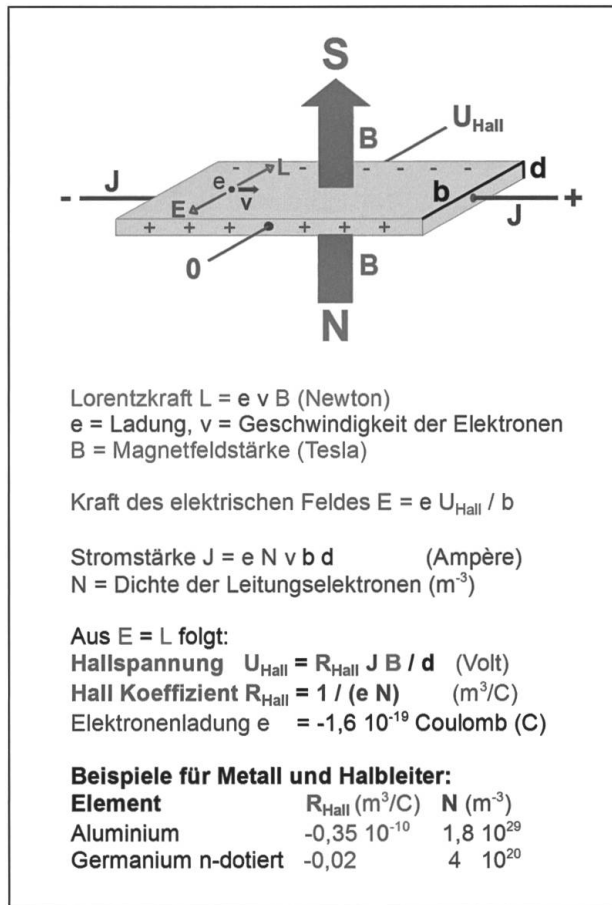


Abb. 2: Aufbau einer Hall-Sonde und physikalische Zusammenhänge. Erklärungen im Text (Bild: Fritz Gassmann).

die Stromleitung auf Elektronen, was anfänglich als selbstverständlich angesehen wurde, ist $e = -1,6 \times 10^{-19} \text{ C}$ (die Ladung der Elektronen) und der entsprechende Hall-Koeffizient ist negativ.

Als Beispiel betrachten wir Aluminium mit $R_{\text{Hall}} = -0,35 \times 10^{-10} \text{ m}^3/\text{C}$. Mit der Ladung der Elektronen erhalten wir $N = 1,8 \times 10^{29}$ Leitungselektronen pro Kubikmeter. Mit der Dichte von Aluminium, seinem Molekulargewicht und der Avogadrozahl erhält man $0,60 \times 10^{29}$ Atome pro Kubikmeter: Jedes Aluminiumatom steuert also genau drei Elektronen zur Leitung von elektrischem Strom bei. Dies bedeutet, dass Aluminium drei nur leicht gebundene Elektronen in der äussersten Energieschale haben muss. Tatsächlich liegt Aluminium im Periodensystem drei Plätze nach dem Edelgas Neon, das zehn Elektronen in zwei sehr stabilen abgeschlossenen Energieschalen enthält. Bei den drei Lei-

tungselektronen von Aluminium muss es sich also um drei Elektronen in der äussersten dritten Energieschale handeln, die wesentlich schwächer an die Atomkerne gebunden sind.

Die Elektronen bewegen sich in Metallen sehr langsam

Berechnen wir nun die Geschwindigkeit der Leitungselektronen in einem Aluminiumdraht mit einer Querschnittsfläche von 1 mm^2 bei einer Stromstärke von $1\text{ A} = 1\text{ C/s}$. Es müssen also pro Sekunde $1/(1,6 \times 10^{-19}) = 0,63 \times 10^{19}$ Elektronen durch die Querschnittsfläche fliessen. Da sich in einem Drahtabschnitt von 1 mm Länge $1,8 \times 10^{20}$ Leitungselektronen befinden, beträgt deren Geschwindigkeit $0,63 \times 10^{19} / (1,8 \times 10^{20}) = 0,035\text{ mm/s}$; eine Schnecke wäre wesentlich schneller! Dies ist auch die Erklärung für die enorm kleinen Hall-Spannungen bei Metallen. Um grössere Hall-Spannungen zu erreichen, muss die Geschwindigkeit der Ladungsträger erhöht werden, was mit sehr dünnen Folien oder sehr kleinen Dichten der Ladungsträger erreicht werden kann. Hall hat mit sehr dünnen Goldfolien gearbeitet, doch heute weiss man, dass es wesentlich einfacher ist, Materialien mit sehr kleinen Ladungsträgerdichten zu verwenden.

Schnelle Elektronen in Halbleitern ergeben grosse Hall-Spannungen

Die technisch bedeutendsten Halbleiterelemente sind Silizium und Germanium, die im Periodensystem direkt unterhalb von Kohlenstoff liegen. Erst mit der in den 1920er Jahren entwickelten Quantentheorie konnte man verstehen, weshalb Elektronen in Metallen und Halbleitern sich fast wie im Vakuum durch die Gitterstrukturen sehr reiner Elemente bewegen können (für eine theoretische Analyse vgl. Feynman 1963). In einer ungestörten periodischen Struktur verhalten sich Elektronen im sogenannten Leitungsband (einem nach unten und oben begrenzten Energiebereich) wie Wasserwellen auf einer Seeoberfläche, die sich über grosse Distanzen fast reibungsfrei ausbreiten können. In Metallen liegt das Leitungsband im Bereich der Energien der äussersten Elektronen, weshalb diese leicht

in dieses übertreten können (bei Aluminium können alle drei Elektronen der dritten Schale ins Leitungsband übertreten). Bei Halbleitern liegt jedoch die untere Energiekante des Leitungsbandes etwas über der maximalen Energie der äussersten Elektronen. Diese Energielücke verhindert, dass Elektronen ins Leitungsband übertreten können, weshalb sich sehr reine Halbleiter bei Zimmertemperatur wie Isolatoren verhalten. Bei viel höheren Temperaturen wird die Bewegungsenergie der Elektronen so gross, dass sie die Energielücke überspringen und sich dann im Leitungsband frei bewegen können: Der Isolator wird zu einem Leiter!

Es geht aber auch anders: Bestrahlt man einen Halbleiter mit Licht von genügend kleiner Wellenlänge, können die Elektronen durch die Absorption von Photonen ausreichend Energie aufnehmen, um die Energielücke zu überspringen. Dieser Vorgang ist die Basis für die Umwandlung von Sonnenenergie in Elektrizität (z.B. Silizium-Photovoltaikzellen).

Es geht sogar noch anders: Durch gezielte Verunreinigungen (Dotierung) von Silizium oder Germanium durch benachbarte Elemente mit ähnlicher Gitterstruktur können Energiestufen in die Energielücke eingebaut werden. Diese Energiestufen wirken wie Treppenstufen, über die Elektronen, getrieben durch thermische Bewegungen, zum Leitungsband emporsteigen können. Zur Dotierung wählt man Atome, die leicht Elektronen abgeben können. Solche Donoren stehen im Periodensystem weiter rechts wie Phosphor, Arsen oder Antimon. Mit dem Dotierungsprozess kann man die Ladungsträgerdichte im Leitungsband auf sehr kleine Werte festlegen. Bei einer bestimmten Stromstärke bewegen sich dann die Elektronen viel schneller und es entsteht ein grosser Hall-Effekt. Übliche Dotierungen liegen in der Größenordnung von 0,01 Millionstel. Statt drei Elektronen pro Gitteratom wie bei Aluminium ist so pro 100 Millionen Gitteratomen nur etwa ein Elektron im Leitungsband, das sich jedoch 300 Millionen Mal schneller bewegt. Es erreicht bereits bei kleinen Stromstärken und in dicken Proben Geschwindigkeiten von 10 bis 1000 Metern pro Sekunde! Die Ablenkungskräfte sind dementsprechend auch in kleinen

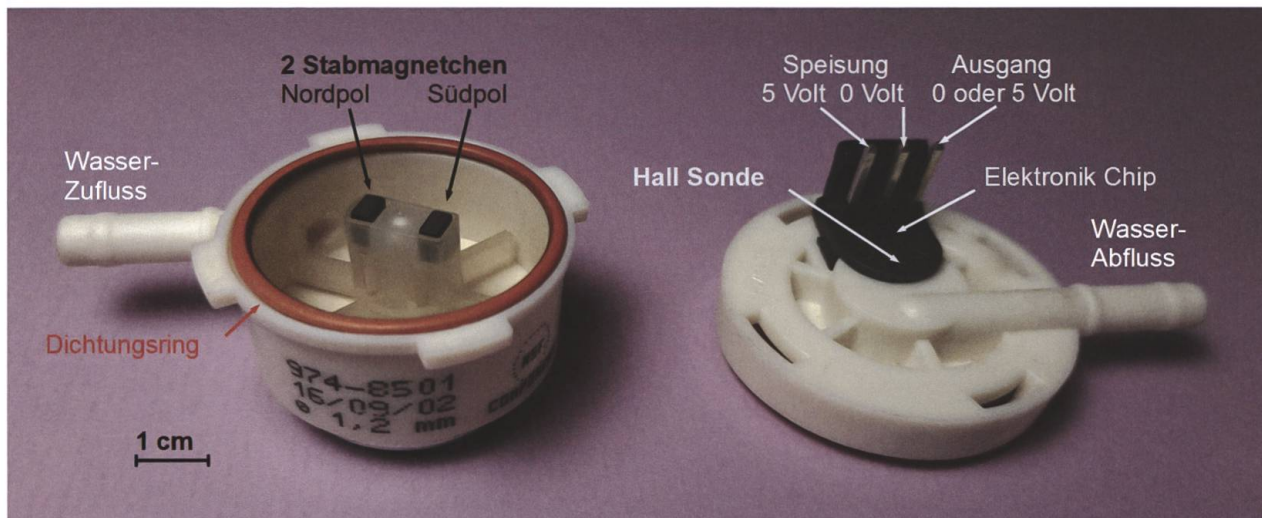


Abb. 3: Das Schaufelrad dreht sich proportional zur Strömungsgeschwindigkeit des Wassers. Zwei kleine, umgekehrt gepolte Stabmagneten verursachen bei jedem Durchgang unterhalb der Hall-Sonde eine abwechselnd positive und negative Hall-Spannung. Ein in die schwarze Komponente eingeschmolzener Elektronik-Chip verstärkt die Hall-Spannung und legt den Ausgang auf 0 oder 5 Volt, je nachdem, ob das zuletzt gemessene Magnetfeld von einem Nordpol oder Südpol ausging. Der Steuerungscomputer der Kaffeemaschine zählt die Spannungswechsel und stoppt die Pumpe bei einer zuvor bestimmten Zahl (Bild: Fritz Gassmann).

Magnetfeldern so gross, dass gut messbare Hall-Spannungen im Bereich von Millivolt entstehen.

Hall-Sensoren sind heute unverzichtbar

Für Hall-Sensoren eignen sich aus den oben erläuterten Gründen dotierte Halbleiter wesentlich besser als Metalle. In Alltagsgeräten werden die genial einfach aufgebauten und robusten Hall-Sensoren für drei Einsatzgebiete verwendet, die hier durch je ein Beispiel vorgestellt werden sollen.

1 Hall-Sensoren als Schalter: Mit Hall-Schaltern kann die Position von Hebeln und Türen (im Auto: Pedalpositionen, Türschliess-System, Gurtschloss, Bestimmung des Zündzeitpunktes), aber auch die Drehgeschwindigkeit von Rädern gemessen werden. Letzteres wird im ABS-System von Autos eingesetzt, um präzise Drehgeschwindigkeiten der Räder in Echtzeit erfassen zu können, und auch in der Getriebeschaltung. Ein einfaches Beispiel ist die Wasser-Durchflussmessung in einer Kaffeemaschine. Um für jede Kaffeesorte und unabhängig vom Verkalkungsgrad der Maschine immer dieselbe Wassermenge in eine Tasse zu geben, wird

der Wasserfluss mithilfe von Hall-Sensoren kontinuierlich gemessen (Abb. 3).

2 Hall-Sensoren zur Messung von Magnetfeldern: Schickt man einen bekannten konstanten Strom durch eine Hallplatte, ist die Hall-Spannung proportional zur Magnetfeldstärke senkrecht zur Hallplatte. Mit drei in einem Smartphone eingebauten Hallplatten lassen sich die drei Komponenten eines Magnetfeldes bestimmen. Die Software rechnet die Vektorkomponenten in der Geräteebene in eine entsprechende Richtungsanzeige einer Kompassnadel um (Abb. 4). Ebenfalls eingebaute Beschleunigungsmesser simulieren eine Wasserwaage, die dazu dient, das Gerät horizontal zu halten. Im Schweizer Erdmagnetfeld zeigt das Gerät eine Feldstärke von etwa $48 \mu\text{T}$ (Mikrotesla) als Summe der Vertikal- ($44 \mu\text{T}$) und der Horizontal- ($20 \mu\text{T}$) Komponente.

3 Hall-Sensoren zur Messung von Strömen: Ein mit Gleichstrom durchflossener Draht erzeugt kreisförmige Magnetfeldlinien um den Draht herum. Ein Ampère ergibt im Abstand von einem Zentimeter vom Draht eine Feldstärke von $20 \mu\text{T}$. So können mit dem Smartphone auch Stromstärken gemessen werden. Praktischer sind jedoch Stromzangen, die durch einen

den Draht umschliessenden Ringkern das Erdmagnetfeld abschirmen und so etwa 0,1 bis 1000 A messen können. Gegenüber der Strommessung mit einem Ampèremeter hat eine solche Stromzange den grossen Vorteil, dass der Draht für die Messung nicht zerschnitten werden muss. Man braucht nicht einmal die Isolationsschicht zu entfernen oder den Draht zu berühren (Abb. 5).

Die grosse Überraschung: positive Hall-Koeffizienten

Bei vielen Metallen, aber auch bei anderen Elementen, rätselte man lange über gemessene «verkehrte Hall-Spannungen», also positive Hall-Koeffizienten. Beispiele sind bekannte Metalle wie Eisen, Chrom, Vanadium, Titan, Zink und Blei. Wie ist es möglich, dass in diesen Elementen freie positive Ladungsträger für die

Stromleitung aufkommen? Die einzig positiv geladenen Atomkerne kommen für die Stromleitung definitiv nicht infrage, da sie in der Gitterstruktur festsitzen.

Die rettende Idee war die folgende: «Normale» Leitermaterialien mit negativen Hall-Spannungen geben leicht Elektronen ab. So entstehen im Leitungsband negativ geladene Fehlstellen, die durch das elektrische Feld (d.h. durch die angelegte Spannung) zum positiven Anschluss geschoben werden. Für die in Abb. 2 gezeigte Anordnung bewegen sich diese negativen Fehlstellen nach rechts und werden durch das Magnetfeld nach hinten abgelenkt, sodass eine negative Hall-Spannung entsteht. «Nicht normale» Leitermaterialien mit «verkehrten» Hall-Spannungen nehmen leicht Elektronen auf. So entstehen im entsprechenden Leitungsband positiv geladene Fehlstellen mit fehlenden Elektronen, die «Löcher» genannt werden. Diese werden zum negativen Anschluss nach links geschoben und dabei durch das Magnetfeld ebenfalls nach hinten abgelenkt, weil sich gleichzeitig das Vorzeichen der Ladung und dasjenige der Geschwindigkeit geändert haben ($e \times v = -e \times -v$). Es entsteht so eine positive Hall-Spannung und folglich auch ein positiver Hall-Koeffizient.

n- und p-Halbleiter

Wir haben gesehen, dass Halbleiter wie Silizium oder Germanium mit Donoren wie Phosphor, Arsen oder Antimon gezielt so verunreinigt (dotiert) werden können, dass im Leitungsband eine gewollt niedrige Elektronendichte entsteht, wodurch sich ein grosser Hall-Effekt ergibt. Man nennt solchermassen dotierte Halbleiter mit negativen Hall-Koeffizienten n-Halbleiter. Man kann aber auch mit Atomen dotieren, die gerne Elektronen aufnehmen. Solche Akzeptoren stehen im Periodensystem weiter links, wie Aluminium, Gallium oder Indium. So dotierte Halbleiter mit positiven Hall-Koeffizienten werden p-Halbleiter genannt, weil in deren Leitungsband positiv geladene Löcher für die Stromleitung sorgen. Bei der Entwicklung der Halbleiter-Technologie in den 1940er-Jahren spielte der Hall-Effekt eine zentrale Rolle, weil mithilfe der Hall-Spannung die durch ver-

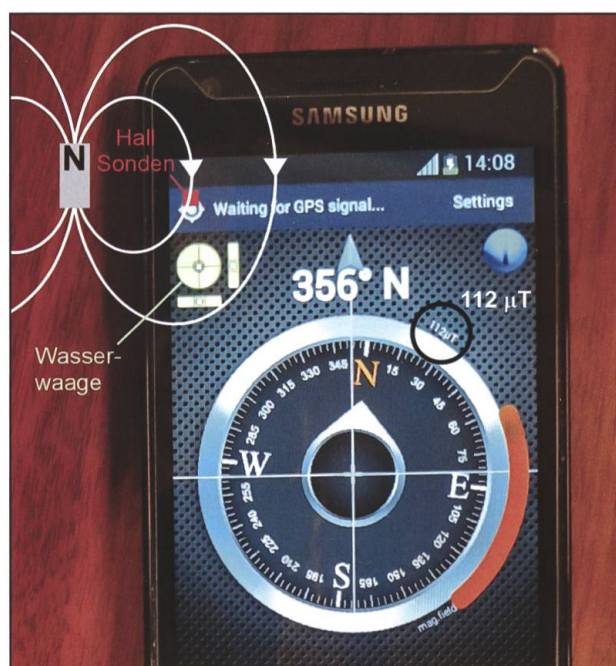


Abb. 4: Messung von Magnetfeldern mit dem Smartphone. Das rote Quadrat zeigt die Platzierung von 3 Hallplatten zur Messung der 3 Komponenten eines Magnetfeldes. Ein kleines Stabmagnetchen zeigt, wie sich der Kompass in seinem Magnetfeld ausrichtet. Die Magnetfeldstärke wird mit 112 μT angegeben (numerisch und roter Balken), was dem 2,3-Fachen des Erdmagnetfeldes entspricht. Um die genaue Feldstärke des Stabmagnetchens beim roten Quadrat zu erhalten, müsste der Erdfeldvektor subtrahiert werden (Bild: Fritz Gassmann).



Abb. 5: Stromzange zur berührungslosen Messung der Stromstärke in einem Draht (Bild: Harke public domain, Text Fritz Gassmann).

schiedene Verfahren erzeugten Dotierungen genau gemessen werden konnten.

pn-Übergänge

Interessant sind Kristalle aus Silizium oder Germanium, die man während des Wachstums oder nachträglich durch thermische Behandlung an gegenüberliegenden Seiten unterschiedlich dotiert, sodass in der Mitte ein Übergang von einem p-Halbleiter zu einem n-Halbleiter entsteht. In Abb. 6a ist ein solcher Übergang (engl. junction) schematisch dargestellt. Bei realen Halbleitern sind die Übergangszonen (in Abb. 6 durch überlappende Farben dargestellt) kleiner als Mikrometer oder sogar Nanometer.

Ohne Kontakt zwischen den beiden Hälften würden wir links positive Löcher vorfinden, deren Konzentration durch die Akzeptor-Dotierung bestimmt wird. Rechts wären negative Elektronen, deren Konzentration durch die Donor-Dotierung festgelegt wird. Sind die beiden Hälften in Kontakt miteinander, können Löcher nach rechts und Elektronen nach links diffundieren. Dieser Vorgang führt jedoch rasch zu einem dynamischen Gleichgewicht, weil die rechte Hälfte an Elektronen verarmt und Löcher gewinnt. Beides erhöht die Spannung auf der rechten Seite und der umgekehrte Prozess reduziert die Spannung auf der linken Seite. Der so entstehende Spannungsunterschied ist

gleichbedeutend mit einem elektrischen Feld, das dem Diffusionsvorgang entgegen wirkt und den Nettotransport von Ladungsträgern in diese sogenannte Raumladungszone stoppt. In dieser schmalen Zone bilden sich sehr starke elektrische Felder der Größenordnung von bis zu 1 V/nm, was einer Milliarde Volt pro Meter entspricht. Die Feldstärken in der Raumladungszone sind grösser als die bei Hochspannungsleitungen anzutreffenden!

Es ist jedoch nicht möglich, diese interne Spannungsdifferenz mit dem Anschluss eines Voltmeters an den Kristall zu messen, weil dadurch weitere Materialübergänge auftreten würden, die die interne Spannung exakt kompensieren. Wäre dies nicht der Fall, könnte man aus einem einzigen Wärmereservoir, das den Kristall auf konstanter Temperatur hält, elektrische Energie gewinnen. Dies ist nach dem zweiten Hauptsatz der Thermodynamik nicht möglich (wir hätten sonst ein Perpetuum mobile der zweiten Art).

Photovoltaik mit pn-Übergängen

Die interne Spannungsdifferenz eines pn-Übergangs macht sich jedoch bemerkbar, wenn man die Raumladungszone beleuchtet. Die Photonen erzeugen Elektronen-Loch-Paare, wenn ihre Energie grösser ist als die Energielücke (engl. energy gap) zwischen dem Elektronen- und dem

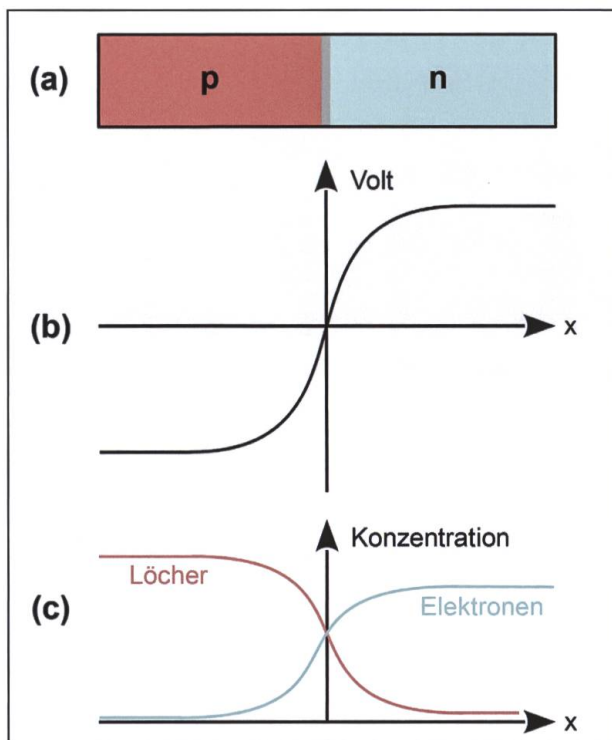


Abb. 6: (a) Übergang von einem p-dotierten Halbleiter (links) zu einem n-dotierten (rechts). (b) Durch Diffusion der Ladungsträger auf die jeweils andere Seite entsteht ein elektrisches Feld (eine Spannungsdifferenz in Volt), das die positiven Löcher nach links und die negativen Elektronen nach rechts drückt und so dem Diffusionsvorgang entgegenwirkt. Es entsteht ein dynamisches Gleichgewicht. (c) Konzentrationsverlauf von Löchern und Elektronen nach Erreichen des dynamischen Gleichgewichts. Es wurde angenommen, dass die Dotierung mit Akzeptoren auf der p-Seite etwas höher ist als diejenige mit Donatoren auf der n-Seite (Bild: Fritz Gassmann).

Löcher-Leitungsband. Dies ist der Fall, wenn die Frequenz ν der Photonen genügend hoch ist, sodass ihre Energie $h\nu$ grösser ist als die Energielücke ($h = 6,6 \times 10^{-34}$ Js ist das Planck'sche Wirkungsquantum). Mit geeigneten Dotierungen kann man erreichen, dass diese Bedingung mit Sonnenlicht erfüllt ist. Für rotes Licht ist $h\nu$ etwa 1,6 eV (Elektronvolt) und es ist grösser für grünes oder blaues Licht. Die Energielücke muss also kleiner sein als 1,6 eV, wenn man auch rotes Licht ausnützen will. Dies kann mit Silizium-Photovoltaikzellen erreicht werden, deren Energielücke typischerweise 1,1 eV beträgt. Eine Energie- oder Potenzialdifferenz von 1,1 eV entspricht einer Spannung von 1,1 V wenn es sich um Elektronen handelt. Die Ener-

gielücke wird deshalb vielfach physikalisch inkorrekt in Volt statt in Elektronvolt angegeben.

Sobald Elektronen-Loch-Paare durch die Absorption von Photonen erzeugt worden sind, werden Elektronen und Löcher durch das starke elektrische Feld in der Raumladungszone auseinandergerissen und können deshalb nicht rekombinieren. Die Elektronen gelangen auf die n-Seite und können über einen Lastwiderstand (z.B. Lampe, Motor) zur p-Seite geleitet werden, wo sie die dort ankommenden Löcher neutralisieren (Abb. 7). Die Leistung der Photozelle ist proportional zur Intensität des eingestrahnten Lichts.

Dioden ersetzen Gleichrichterröhren

In der Elektronik werden pn-Übergänge als Dioden verwendet, die den Strom nur in einer Richtung leiten und in der anderen sperren. Damit lässt sich Wechselstrom «gleichrichten», sodass der Strom wie aus einer Batterie immer in dieselbe Richtung fliesst. Gleichstrom ist zur Speisung aller elektronischen Geräte unumgänglich.

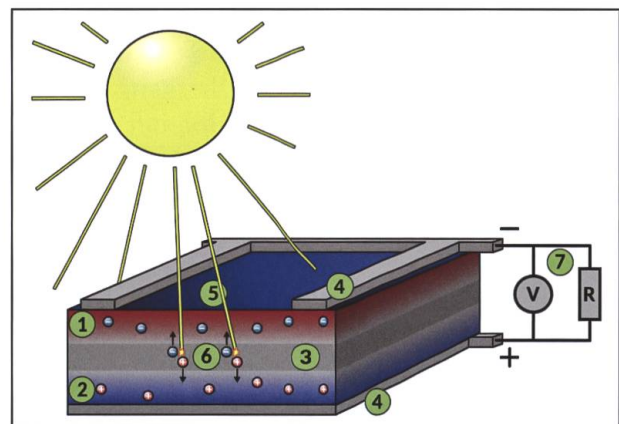


Abb. 7: Silizium-Photozellen zur Erzeugung von Solarstrom.

- (1) n-dotierte Schicht
- (2) p-dotierte Schicht
- (3) pn-Übergangsschicht
- (4) Stromsammelschienen
- (5) Photonen durchdringen n-Schicht
- (6) Photonen erzeugen Elektronen-Loch-Paare in der pn-Schicht
- (7) Die abgegebene Spannung beträgt im Leerlauf 0,64 V und bei optimaler Belastung etwa 0,5 V. Für Ausgangsspannungen von 15 V werden deshalb 30 Zellen in Serie geschaltet (Bild: Michael Paetzold, CC SA 3.0).

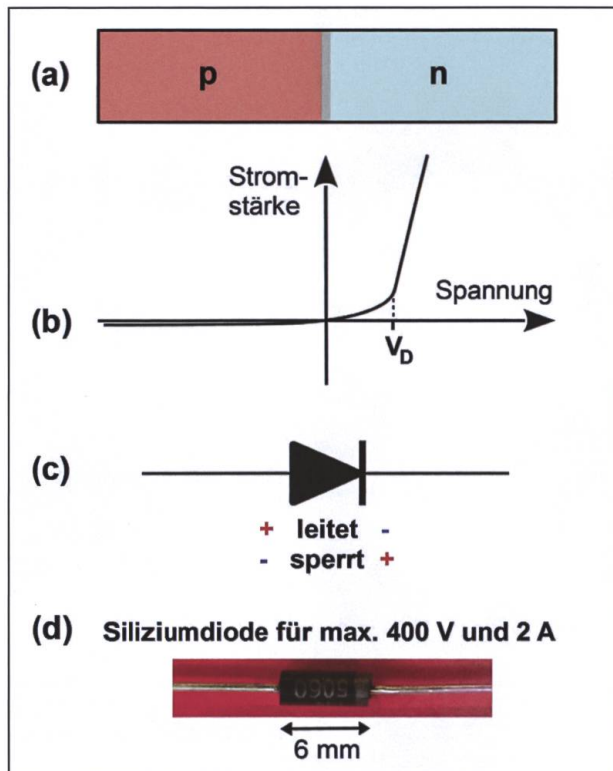


Abb. 8: pn-Übergang als Diode.

- (a) pn-Übergang
 (b) Kennlinie einer Siliziumdiode mit Durchlassspannung V_D
 (c) Diodensymbol *
 (d) Siliziumdiode 1N5060. Sie kann bis zu 400 V sperren und 2 A leiten. Der graue Streifen am rechten Ende bezeichnet den Sperrbalken (n-Seite).

* Der Pfeil zeigt die Durchlassrichtung für positive Ladungsträger, die vom Pluspol zum Minuspol fließen (technische Stromrichtung). Der Balken symbolisiert die Sperrrichtung (n-Seite) (Bild: Fritz Gassmann).

Solche Diodenkristalle haben gegenüber den früher verwendeten Gleichrichterröhren riesige Vorteile: Sie sind viel kleiner und leichter, sie haben keine geheizte Kathode und deshalb nur zwei Anschlüsse, sie sind robuster und lassen sich milliardenweise in integrierte Schaltungen einbauen.

Ohne eine zwischen der p- und n-Seite angelegte Spannung fließt kein Strom und die interne Spannungsverteilung ist wie in Abb. 6 dargestellt. Legen wir jedoch eine Spannung so an, dass die interne Spannungs Differenz verringert wird (Minuspole der Stromquelle an die n-Seite, Pluspol an die p-Seite), dann können Löcher immer besser «bergauf» auf die n-Seite gelangen,

je höher die externe Spannung ist. Wird die externe Spannung grösser als die in Abb. 6 dargestellte interne (bei Siliziumdioden ist dies um 0,6 V, bei Germaniumdioden um 0,35 V), wird der Spannungsberg für Löcher in ein Spannungstal verwandelt: Löcher fließen nun «bergab» zum Minuspol (n-Seite), wo sie durch zufließende Elektronen neutralisiert werden. Analog fließen Elektronen «bergab» zum Pluspol (p-Seite), wo sie die Diode verlassen (vgl. Abb. 8). Es fließt demzufolge ein starker Elektronenstrom vom Minuspol durch die Diode zum Pluspol: *Die Diode leitet.*

Polen wir die externe Spannungsquelle um (Minuspole an p-Seite), vergrößert sich der Spannungsberg für Löcher (und analog für Elektronen) und es fließt praktisch kein Strom: *Die Diode sperrt.*

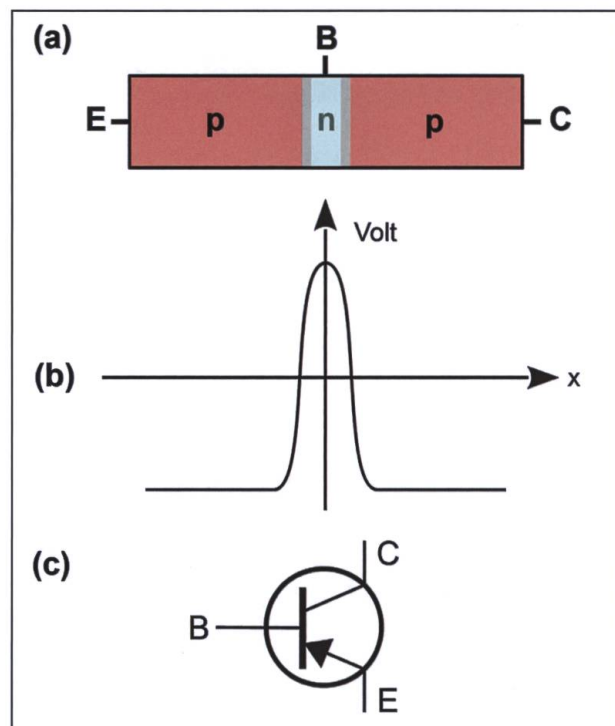


Abb. 9:

- (a) pnp-Übergang als Transistor.
 E = Emitter: emittiert positive Ladungsträger
 B = Basis: Grundplatte
 C = Collector: sammelt positive Ladungsträger
 (b) Interne Spannungsverteilung in einem Transistor ohne externe Spannungsquellen.
 (c) Das Symbol des Transistors zeigt die Emitter-Basis-Diode. Die umgekehrt gepolte Collector-Basis-Diode wird nicht dargestellt, um die asymmetrische Dotierung anzudeuten (Bild: Fritz Gassmann).



Abb. 10: Erfindung des Bipolartransistors 1948. Von links: John Bardeen, William Shockley und Walter Brattain in den Bell Telephone Laboratories in New York. Das Trio wurde 1956 mit dem Nobelpreis geehrt (Bild: AT&T, Jack St., public domain).

Transistoren ersetzen Radioröhren

Verbindet man einen pnp- mit einem np-Übergang auf einem einzigen Kristall, entsteht ein pnp- (oder npn-) Element, das man mit drei Anschlüssen versehen kann. Da sich die beiden Typen nur um das Vorzeichen der angelegten Spannungen unterscheiden, behandeln wir im Folgenden nur pnp-Transistoren. Abb. 9 zeigt die interne Spannungsverteilung sowie die Namen Emitter, Basis und Collector der drei Anschlüsse, ohne dass externe Spannungen angelegt sind. Es erstaunt vielleicht, dass Emitter und Collector verschiedene Bezeichnungen haben, obwohl der Transistor symmetrisch aussieht. Die Dotierungen von Emitter und Collector sind jedoch unterschiedlich, um die Eigenschaften des Transistors zu optimieren, weshalb die beiden Anschlüsse normalerweise nicht vertauscht werden dürfen. Für spezielle Hochfrequenz-Anwendungen ist es dennoch vorteilhaft, Transistoren «verkehrt» einzusetzen.

Es gibt heute verschiedene Transistortypen, hier soll jedoch nur der historisch erste und auch heute noch weit verbreitete Bipolartransistor behandelt werden, der 1948 in den Bell Telephone

Laboratories in New York erfunden wurde (Abb. 10–12). In integrierten Schaltungen werden heute fast ausschliesslich MOSFETs (Metal Oxide Semiconductor Field Effect Transistor) benutzt, deren Funktionsweise von derjenigen der hier vorgestellten Bipolartransistoren abweicht.

Die Emitter-Basisstrecke gleicht einer Diode, die mit einer negativen Basisspannung (in der n-dotierten Zone) auf Durchlassrichtung geschaltet wird (vgl. Abb. 13). Der Einfachheit halber betrachten wir nur die Löcher als (positive) Ladungsträger (die negativen Elektronen verhalten sich analog mit geändertem Vorzeichen). Die Collector-Basis-Diode wird durch die stark negative Collectorspannung in Sperrrichtung betrieben. Im Prinzip besteht also ein Transistor aus zwei entgegengesetzt gepolten Dioden und es ist vorerst unverständlich, dass vom Emitter zum Collector je ein Strom fließen könnte, da die Collectordiode sperrt. Das Geheimnis des Transistors liegt darin, dass die Basis so dünn ist, dass sie für die Ladungsträger quasi transparent wird.

Wir betrachten nun die Anordnung nicht nur in einer, sondern in zwei Dimensionen. Das in



Abb. 11: Germanium Bipolartransistor 2N1305 in Metallbecher-Gehäuse der 1960er-Jahre. Originalansicht und fünffache Vergrößerung nach Entfernung des Deckels. Der trapezförmige Ausschnitt der Bodenplatte wurde senkrecht nach oben gebogen und an die Transistoren-Basisplatte gelötet. So wurde die Verlustleistung des Transistors an sein Gehäuse abgeführt. Ins Zentrum der Basisplatte wurde ein Loch gebohrt und dessen Rand becherförmig ausgestülpt (vgl. Abb. 12). Die Anschlussdrähte an Emitter und Collector sind knapp erkennbar. Der Durchmesser des Gehäuses misst 8 mm (Bild: Fritz Gassmann).

Abb. 12 gezeigte graue Basisplättchen ist nur etwa $10\,\mu\text{m}$ dick, die Kontaktflächen der «Indiumpillen» haben jedoch einen Durchmesser von etwa $500\,\mu\text{m}$. Durch thermische Bewegungen gelangen umso mehr Löcher vom Emitter auf den Basis-Energiehügel, je negativer die an der Basis angelegte Spannung ist. Der Trick ist nun, dafür zu sorgen, dass die meisten Löcher nicht über den Basisanschluss abfließen, sondern zum Collector «hinunterfallen». Durch eine hohe n-Dotierung der Basis ist ihr Potenzial überall fast gleich und es entstehen fast keine radialen Felder, die die Löcher an den Rand der «Collectorpille» und zum Basisausgang leiten. Nur Löcher in einer etwa $1\,\mu\text{m}$ breiten Randzone fließen mit grosser Wahrscheinlichkeit über den Basisausgang ab. Alle sich weiter gegen die

Kreismitte befindenden Löcher geraten durch einen thermisch getriebenen Zickzackkurs in das sehr starke interne Collectorfeld und werden zum Collectorausgang «abgesogen». Der Collectorstrom wird deshalb mit den obigen Zahlenangaben rund 125-mal grösser als der Basisstrom. Transistoren sind deshalb ihrem Wesen nach Stromverstärker mit typischen Verstärkungsfaktoren von meist über 100 bis gegen 1000. Der Collectorstrom kann Milliampère (bei Hochfrequenztransistoren) bis viele Ampère (bei Leistungstransistoren) betragen und seine Steuerung geschieht vollumfänglich durch den viel kleineren Basisstrom.

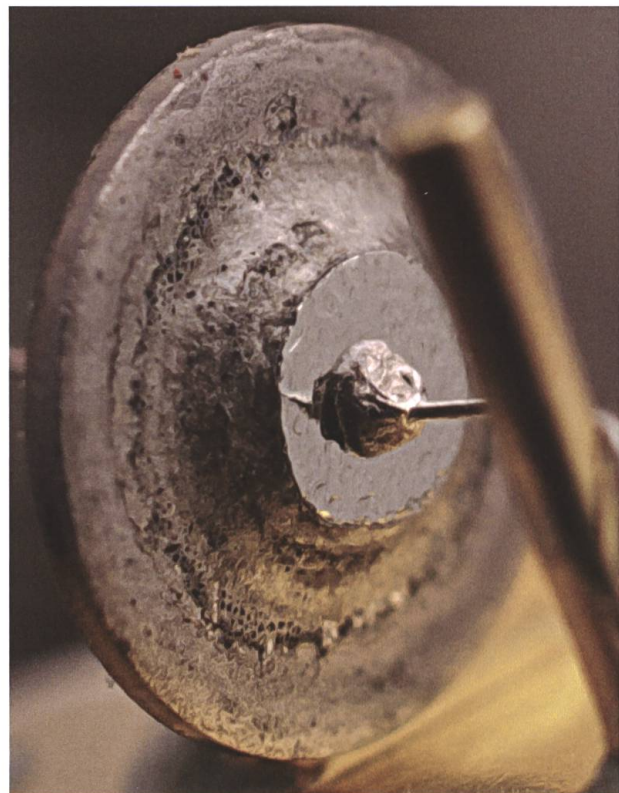


Abb. 12: Detailansicht der in Abb. 11 gezeigten Basisplatte von der Collectorseite her betrachtet. Ihr Durchmesser beträgt 3,5 mm, ihre Dicke 0,15 mm. Die flache, graue Kreisscheibe mit rund 1,2 mm Durchmesser ist das rund $10\,\mu\text{m}$ dicke n-dotierte Basisplättchen, das auf den ausgestülpten Rand des Loches im Zentrum der Basisplatte aufkontaktiert wurde. Die aufgesetzte p-dotierte «Indiumpille» ist durch einen dünnen Draht mit dem im Vordergrund sichtbaren Collector-Anschluss verbunden. Genau gegenüber sitzt eine analoge «Indiumpille», die mit dem Emitter-Anschluss verbunden ist. Die Herstellung war arbeitsintensiv, weshalb die Transistoren teuer waren (Bild: Rolf Süssbrich, Wikimedia CC BY-SA 3.0).

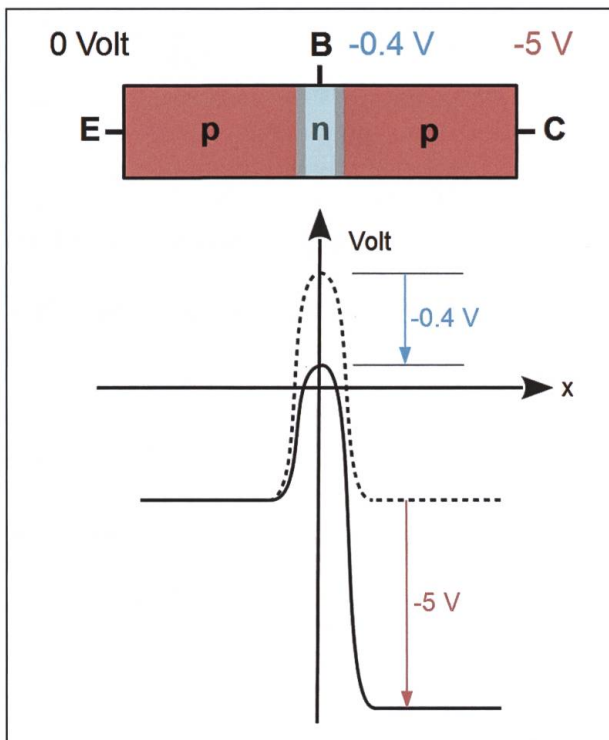


Abb. 13: Bipolar pnp-Transistor nach Anlegen einer kleinen negativen Steuerspannung an die Basis (blau) und einer grösseren negativen Spannung (rot) an den Collector. Je negativer die externe Basisspannung wird, desto kleiner wird der Hügel, über den die Löcher zum Collector gelangen können. Bei einer Schwellenspannung von ca. $-0,35\text{ V}$ für Germanium ($-0,6\text{ V}$ für Silizium) wird der Hügel so flach, dass ein Löcherstrom vom Emitter zum Collector einsetzt. Wird die Steuerspannung noch negativer, wächst der Collectorstrom proportional zum Basisstrom und es fliesst ein Löcherstrom «ähnlich einem Wasserfall den steilen Abhang» zum Collector hinab, der 100- bis 1000-mal grösser ist als der Basisstrom (Bild: Fritz Gassmann).

Die ursprüngliche Bedeutung des Kunstwortes **Transistor**, das aus dem Zusammenzug von **Transfer** (Übertragen) und **Resistor** (Widerstand) gebildet wurde, beschrieb das Element als «gesteuerten Widerstand» oder ausführlicher «durch die Basisspannung gesteuerter Emitter-Collector-Widerstand». Tatsächlich wird der Innenwiderstand der Emitter-Collectorstrecke R_{EC} durch die Spannung der Basis gegenüber dem Emitter gesteuert: Für Spannungsdifferenzen unterhalb der Dioden-Leitschwelle ($0,6\text{ V}$ für Silizium) ist R_{EC} sehr gross und die Spannung am Collectoranschluss ist gleich der über den Lastwiderstand angeschlossenen Betriebsspannung (für eine typische Verstärkerschaltung sei

auf Abb. 14b verwiesen). Sobald die Emitter-Basis-Diode zu leiten beginnt, fliesst ein Strom über den Basis-Vorwiderstand und ein viel grösserer Strom zwischen Emitter und Collector,

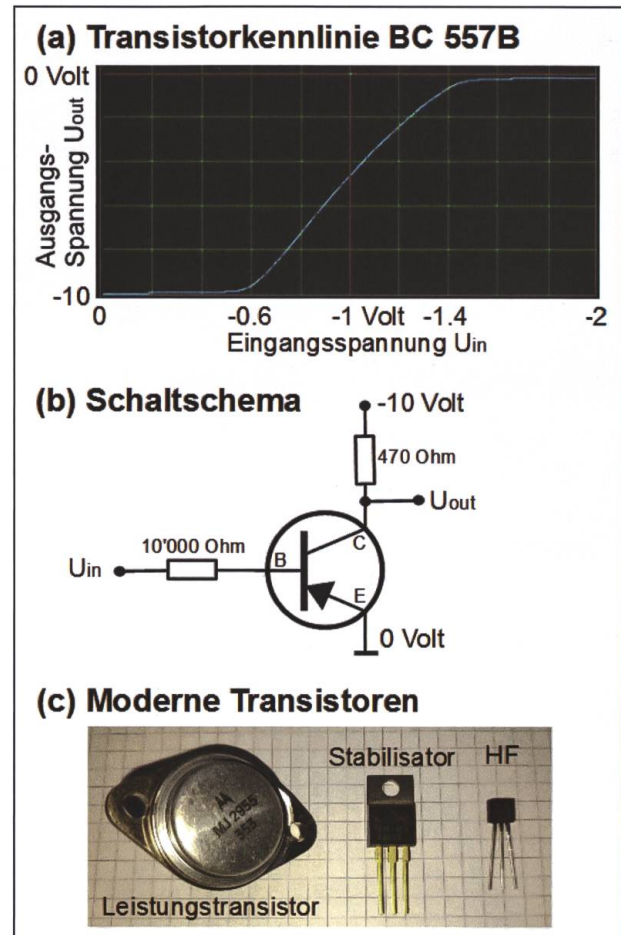


Abb. 14:

- (a) Gemessene Kennlinie des modernen pnp-Siliziumtransistors BC557B. Die Kurvensteigung im nahezu linearen Arbeitsbereich zwischen $0,6$ und $1,4\text{ Volt}$ ist fast konstant und beträgt -12 . Multipliziert mit dem Verhältnis von Basis- zu Collectorwiderstand ($10'000/470$) ergibt sich eine Stromverstärkung von rund 250 .
- (b) Schaltbild zur Aufnahme der Kennlinie.
- (c) Verschiedene moderne Transistoren auf 4-mm-Raster:
- MJ2955 für hohe Leistung: max. 15 A , max. 60 V , max. 115 W
Anwendung: Verstärker-Endstufe
 - MC78M15 Spannungsstabilisator 15 V : enthält 28 Transistoren
Anwendung: Stromversorgung für elektronische Geräte
 - BC557B für hohe Frequenzen (HF)
Kennlinie unter (a). Anwendung: Schalter in Steuerungen, Oszillatoren, Verstärker, usw. (Bild: Fritz Gassmann).

d.h. der Innenwiderstand der Emitter-Collectorstrecke ist kleiner geworden und die Spannung am Collectoranschluss wird ebenfalls kleiner (d.h. positiver). Der Transistor ist deshalb ein invertierender Spannungsverstärker: Wenn die Basisspannung zunimmt, nimmt die Collectorspannung ab. Ein am Basiswiderstand angelegtes Signal wird also gespiegelt (invertiert).

Für den Ersatz von Verstärkerröhren durch Transistoren war es wichtig, dass ein grosser Teil der Kennlinie möglichst linear war, d.h. eine fast konstante Steigung hatte. Dies ist bei der in Abb. 14a gezeigten Kennlinie weitgehend der Fall. Durch Gegenkopplungen kann die kleine Krümmung der Kennlinie kompensiert werden, was z.B. für hochwertige HiFi-Verstärkeranlagen für verzerrungsfreie Tonwiedergaben wichtig ist.

Transistoren als Schalter

Transistoren können auch als Schalter eingesetzt werden, indem man den nahezu linearen Arbeitsbereich überspringt und nur die beiden Sättigungsbereiche mit kleinem und sehr grossem Emitter-Collector-Widerstand benutzt. Abb. 15 zeigt, wie eine Schaltung aussieht, mit der man ein 1-Watt-Lämpchen mithilfe eines Stromes von nur 1,2 Milliampère bei 2 Volt einschalten kann, ähnlich wie dies mit einem Relais möglich wäre. Mit einer etwas aufwändigeren Thyristor-Schaltung habe ich erreicht, einen 3-kW-Töpferofen mit einem Computer zu steuern, ohne dass ein Relais zur Aufrechterhaltung einer konstanten Temperatur alle paar Sekunden mit lautem

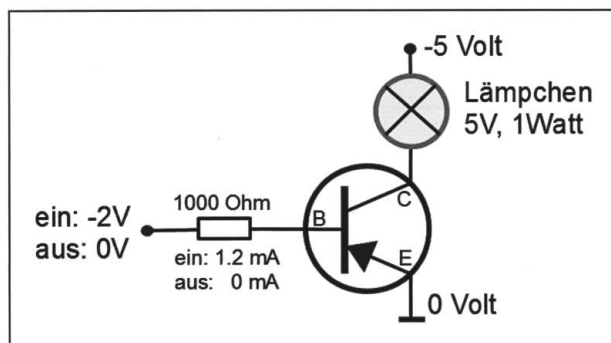
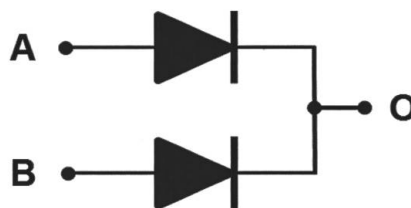


Abb. 15: Transistor als Schalter ersetzt Relais. Das 1-Watt-Lämpchen wird mit einem Strom von etwa 1,2 mA bei 2 Volt eingeschaltet. Die für das Einschalten aufzuwendende Leistung beträgt nur 2,4 mW, rund 400-mal weniger als das Lämpchen (Bild: Fritz Gassmann).

(a) OR-Gate



(b) Wahrheitstabelle

A \ B	0	1
0	0	1
1	1	1

Abb. 16:

(a) OR-Gate mit zwei Dioden.

A, B = Inputs, O = Output

(b) Wahrheitstabelle des OR-Gate, die die Wirkung der Schaltung als Matrix-Operator darstellt (Bild: Fritz Gassmann).

Knall ein- und ausschaltet. Mit Thyristoren werden auch Lokomotiven gesteuert, weshalb die heutige S-Bahn ruckfrei beschleunigt, ohne dass knallende Hochspannungsschalter die Motorenleistung schrittweise hochschalten wie bei älteren Lokomotiven.

Halbleiter sind die Grundbausteine von Computern

Erstaunlicherweise genügen zwei Bausteine, um im Prinzip einen Computer bauen zu können, nämlich OR- und NOT-Gates (ODER- und NICHT-Gatter).

Das **OR-Gate** kann mit zwei Dioden realisiert werden wie in Abb. 16a gezeigt. Zur Vereinfachung stellen wir uns einen Strom positiver Ladungsträger (Löcher) vor, die vom Pluspol zum Minuspol fließen (sogenannte technische Stromrichtung). Das Pfeilsymbol für die Dioden gibt so die Durchlassrichtung an. Weiter betrachten wir nur zwei Spannungen und bezeichnen diese als 0 (0 V) und 1 (z.B. +5 V). Legt man

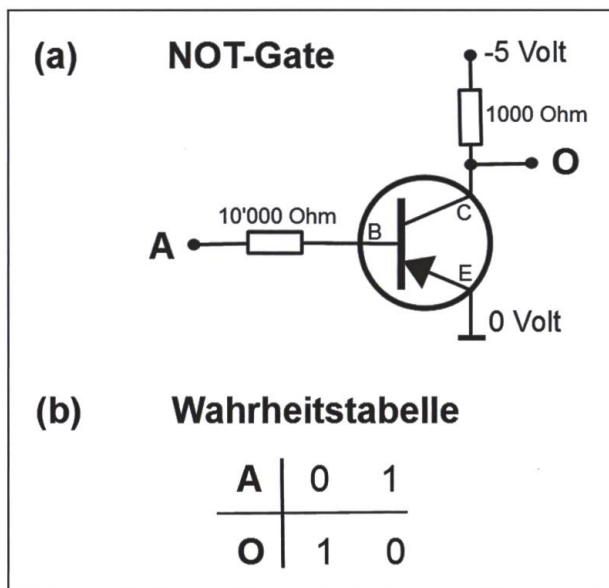


Abb. 17:

(a) NOT-Gate mit einem Transistor
(b) Wahrheitstabelle (Bild: Fritz Gassmann).

beide Eingänge A, B (inputs) des OR-Gate auf 0, ist die Ausgangsspannung O (output) ebenfalls 0. Liegt jedoch einer der beiden Eingänge oder beide auf 1, ist die Ausgangsspannung 1. O ist also 1, wenn **A ODER B** (oder beide) auf 1 gelegt werden (daher der Name ODER-Gatter). Man kann die Funktion eines OR-Gates auch mit der sogenannten Wahrheitstabelle (truth table) mathematisch beschreiben (vgl. Abb. 16b).

Das NOT-Gate (NICHT-Gatter oder Inverter) kann mit einem Transistor realisiert werden (Abb. 17a). Liegt 0 (0 V) am Eingang A, fliesst kein Strom in der Emitter-Collektorstrecke und der Ausgang O hat dieselbe Spannung wie die Betriebsspannung (z.B. -5 V), also den logischen Wert 1. Legen wir jedoch A auf 1 (-5 V), fliesst ein grosser Strom vom Emitter zum Collector und die Ausgangsspannung O fällt auf nahezu 0 V (etwa -0,2 V oder logisch 0; jede Spannung zwischen 0 V und etwa -2 V bedeutet logisch 0). Die Wahrheitstabelle ist sehr einfach (vgl. Abb. 17b). Es sei noch erwähnt, dass mit npn- statt mit pnp-Transistoren die Betriebsspannung +5 statt -5 Volt wäre und so das NOT-Gate mit dem in Abb. 16 gezeigten OR-Gate kompatibel würde (logisch 1 entspräche dann immer +5 V).

Mit NOT- und OR-Gates kann jeder Computer aufgebaut werden

Ich möchte mit drei Beispielen verständlich machen, dass mit den beiden Grundbausteinen NOT- und OR-Gate ein ganzer Computer aufgebaut werden kann (abgesehen von passiven Bausteinen wie Widerständen und Kondensatoren).

Speicherzelle mit zwei NOT-Gates: Der Arbeitsspeicher (Random Access Memory RAM) in modernen Computern ist 1 GB (Gigabyte) gross und braucht dafür mindestens 16 Milliarden Transistoren (1 Byte = 8 Bits und jedes Bit braucht 2 Transistoren). Die Grundschaltung einer Speicherzelle für 1 Bit (d. h. für eine 0 oder eine 1) ist in Abb. 18 wiedergegeben. Sie ist spiegelsymmetrisch aufgebaut und hat zwei stabile Zustände: Entweder ist der Ausgang des linken NOT auf 1 und folglich der Ausgang des rechten NOT auf 0, der den Eingang des linken NOT auf 0 festhält. Oder der zweite stabile Zustand ist

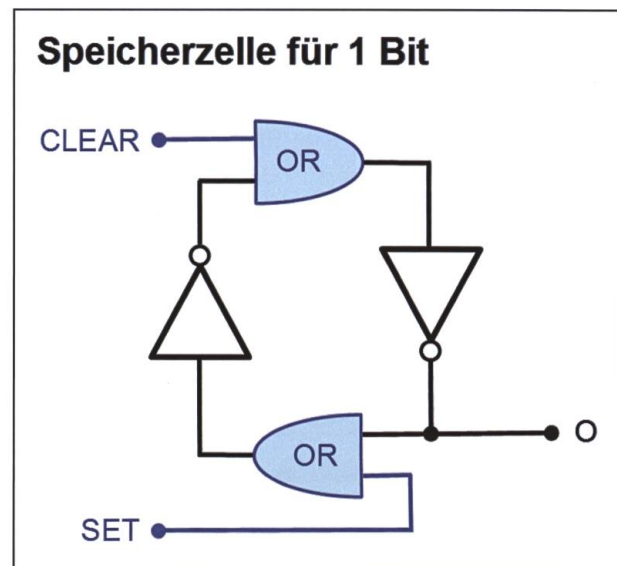
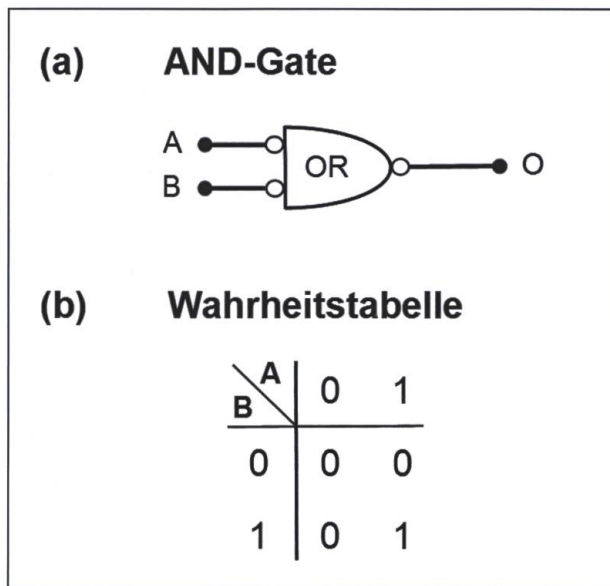


Abb. 18: Speicherzelle für 1 Bit. Damit die Schaltung übersichtlicher wird, wurden alle Widerstände und Speisespannungen weggelassen. Die Transistoren wurden durch das Symbol für Verstärker (Dreieck) dargestellt. Der kleine Kreis an der Dreiecksspitze bedeutet, dass der logische Wert invertiert wird (NOT). Die OR-Gates wurden durch ähnliche Symbole wie Verstärker, jedoch mit runder Ausgangsseite und blau dargestellt. Zur Verdeutlichung schrieb ich noch OR ins Symbol hinein. Die Logik der Schaltung wird so ohne weitere Kommentare klar ersichtlich (Bild: Fritz Gassmann).

**Abb. 19:**

(a) AND-Gate zusammengesetzt aus einem OR- und drei NOT-Gates.

(b) Wahrheitstabelle (Bild: Fritz Gassmann).

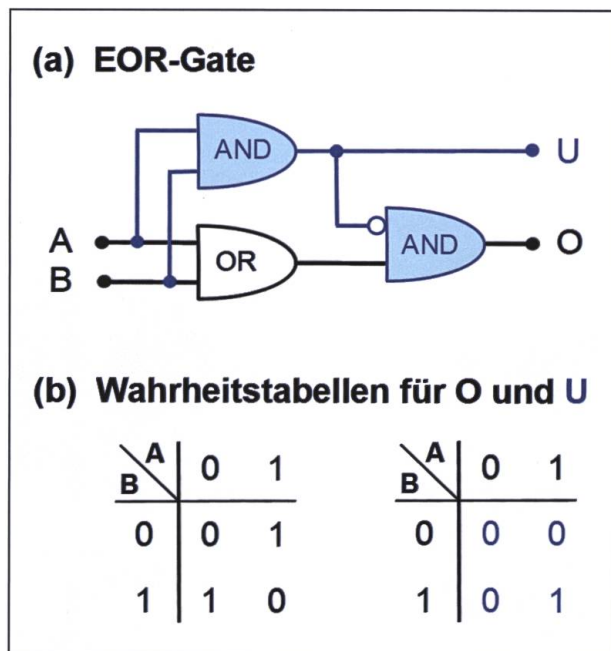
spiegelbildlich. Man kann den Ausgang O des rechten NOT als Inhalt der Speicherzelle definieren, der 0 oder 1 sein kann. Das Lesen der Speicherzelle ist kein Problem: Es wird einfach die Spannung bei O abgegriffen. Wie schreibt man aber eine 1 in die Speicherzelle, wenn anfänglich eine 0 drin war? Die in Abb. 18 gezeigte Schaltung hat zwei Eingänge, die normalerweise beide auf 0 liegen. Soll eine 1 in die Speicherzelle geschrieben werden, wird der SET-Eingang kurzzeitig auf 1 gelegt und für eine 0 wird der CLEAR-Eingang kurzzeitig auf 1 gelegt. Aus diesem Schema wird auch klar, weshalb ein Computer die Information im Arbeitsspeicher bei einem Spannungsunterbruch «vergisst».

Aus OR und dreimal NOT wird AND: Abb. 19 zeigt das AND-Gate mit seiner Wahrheitstabelle. Man erkennt, dass O genau dann 1 ist, wenn A UND B auf 1 gelegt werden (daher der Name UND-Gatter). Ich verwende fortan für AND-Gates das Symbol für OR-Gates, schreibe aber AND ins Symbol hinein.

Ein EOR-Gate kann Zahlen addieren: Ein OR-Gate addiert die beiden Bits an den Eingängen A und B fast korrekt: $0+0=0$, $0+1=1$, $1+0=1$, aber $1+1=1$. Korrekt wäre: $1+1=0$ mit einem Übertrag von 1 an das nächsthöhere Bit. Deshalb

werden für Additionen EOR-Gates (Exclusive OR) eingesetzt, die gemäss Abb. 20 aus den Grundbausteinen aufgebaut werden können. Der Übertrag kann bei U abgegriffen werden.

Dieser an drei Beispielen gezeigte Vorgang kann beliebig erweitert werden, um zusätzliche Operationen zu ermöglichen, die die Adressierung von Speicherzellen oder Multiplikationen erlauben. Mit einer kleinen Modifikation des 1-Bit-Speichers lässt sich zusammen mit dem Übertrags-Mechanismus ein Zähler realisieren, der es ermöglicht, dass ein Programm sequenziell abgearbeitet werden kann. Ein nichtflüchtiger Speicher (z.B. eine Hard Disk) dient der permanenten Sicherung von Programmen oder Texten, Bildern usw., die beim Abschalten des Computers nicht verloren gehen sollen. Weitere wichtige Komponenten eines Computers sind Eingabemöglichkeiten (vielfach ein Tastenfeld) und Mittel zur Ausgabe von Daten (vielfach ein LCD-Bildschirm und ein Druckerausgang), die hier nicht behandelt werden sollen.

**Abb. 20:** EOR-Gate als Addierer.

(a) Aufbau mit OR-, AND- und NOT-Gates. O ist die Addition der Bits A und B, U ist der Übertrag ans nächsthöhere Bit. Die das OR korrigierenden Komponenten sind blau gezeichnet.

(b) Die Wahrheitstabellen für O und U zeigen, dass die Summe von $A=1$ und $B=1$ den Ausgang $O=0$ und den Übertrag $U=1$ ergibt. Im Binärsystem ist $1 + 1 = 10$ (Bild: Fritz Gassmann)

Schritte zum heutigen Computer

Der Relais-Computer der 1940er-Jahre

Alle oben erläuterten Bausteine können auch mit Relais realisiert werden. Der deutsche Bauingenieur Konrad Zuse baute im Jahr 1941 ohne Informatik- oder höhere Mathematikkenntnisse seinen dritten, verbesserten Relaiscomputer Z3, dessen CPU (Central Processor Unit = Zentrale Recheneinheit) mit rund 600 Telefonrelais Zahlen à 22 Bit im Gleitkomma-Format verarbeiten konnte. Die CPU konnte die vier Grundrechenoperationen ausführen sowie Quadratwurzeln ziehen. Der Arbeitsspeicher konnte mithilfe von 1408 weiteren Relais 64 Zahlen aufnehmen. Mit einer Taktfrequenz von bis zu zehn Schritten pro Sekunde (10 Hertz), einer elektrischen Leistung von 4'000 Watt und einem Gewicht von einer Tonne schaffte er eine Multiplikation in drei Sekunden. Das Steuerprogramm konnte mithilfe eines Lochstreifenlesers eingegeben werden. Der Z3 gilt als der erste programmgesteuerte Computer der Welt (sogenannte universelle Turing Maschine). Der in Abb. 21 gezeigte Nachbau steht im Deutschen Museum in München.

Der Röhren-Computer 1950–70

Um die Rechengeschwindigkeit zu erhöhen, wurden die Relais in den 1950er-Jahren durch Radioröhren (Abb. 22) und Magnetkernspeicher (Abb. 23) ersetzt. Dank eines Auslesezyklus dieser Speicher von rund 5 μ s konnte die Taktrate auf gegen 200 kHz erhöht werden (etwa 20'000-mal schneller als der Relaiscomputer Z3).

Die Firma IBM (International Business Machines, USA) stellte mit dieser Technologie die ersten «Grosscomputer» her, wobei sich aus heutiger Sicht «gross» vor allem auf die Dimensionen, das Gewicht, den Stromverbrauch und den Preis bezieht. Der MIT-Physiker Edward Lorenz (1917–2008) hat Anfang der 1960er-Jahre mit einem derartigen Gerät den Schmetterlingseffekt und damit die Chaostheorie entdeckt.

Neben den vielen Nachteilen der Röhrencomputer (Grösse, Gewicht, Leistung, Preis) kam noch deren Unzuverlässigkeit hinzu, weil bei der grossen Anzahl in wissenschaftlichen Computeranlagen verwendeter Radioröhren die Zeit zwischen Störungen durch den Ausfall einer einzelnen Röhre in die Grössenordnung von

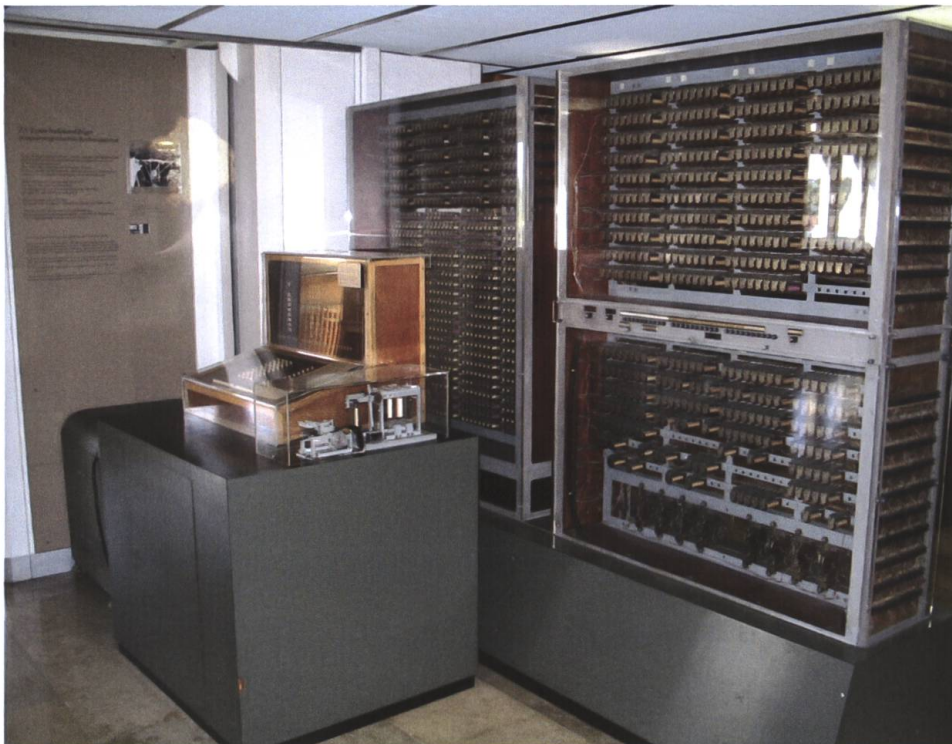


Abb. 21: Erster programmgesteuerter Computer der Welt. Ein Nachbau des Z3 von Konrad Zuse im Deutschen Museum in München. Die Maschine arbeitete mit Zahlen à 22 Bit. Mit einer Gleitkommaarithmetik schaffte sie eine Multiplikation in 3 Sekunden (Bild: Venusianer, CC BY-SA 3.0, Wikimedia Commons).

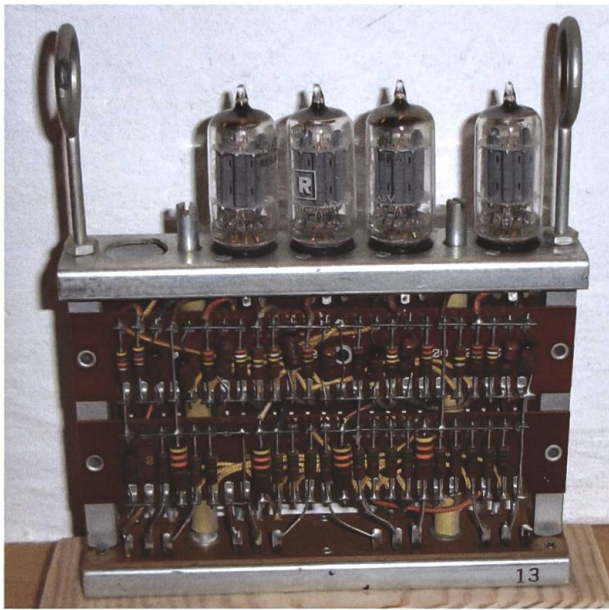


Abb. 22: Röhrenspeicher für 4 Bit mit vier Doppeltrioden NCR-5963 von 1956 (NCR = National Cash Register Company, Cincinnati, USA). Allein die Heizdrähte der vier Röhren haben eine Verlustleistung von 7,5 Watt. Dazu kommt noch etwa derselbe Betrag für die Anoden-Verlustleistung. Die Einheit war vermutlich Teil der Recheneinheit CPU (Central Processor Unit) einer elektronischen Registrierkasse (Bild: Fritz Gassmann).

Stunden kam und die Suche nach der defekten Röhre aufwändig war.

Transistor-Computer 1960–80

Der erste Transistor-Computer wurde 1953 an der Manchester Universität (England) gebaut. Die CPU hatte 92 Transistoren und 550 Dioden und war noch nicht sehr zuverlässig. Erst Anfang der 1960er-Jahre folgten erste kommerzielle Geräte von IBM und DEC aus den USA. Die PDP-1 von DEC enthielt 2'700 Transistoren und 3'000 Dioden und hatte einen Magnetkern-Arbeitspeicher für 4096 Zahlen zu je 18 Bit (9216 Byte), weil die dafür rund 150'000 notwendigen Transistoren zu teuer gewesen wären.

Studenten des Massachusetts Institute of Technology (MIT) verwendeten die PDP-1 auch für Computermusik und unter ihnen waren die ersten Hacker der Geschichte (telephone hackers), die mithilfe des Computers Gratistelefonverbindungen herstellen konnten. Als ich 1973 ins Eidg. Institut für Reaktorforschung EIR

(ab 1988 Paul Scherrer Institut, PSI) eintrat, waren Nachfolgemodelle der PDP-1 überall zur Steuerung von Anlagen anzutreffen, vor allem die legendäre PDP-8. Die ersten Einheiten dieses 12-Bit-Minicomputers wurden 1965 für 18'500 \$ (ca. 150'000 \$ auf heutigen Wert umgerechnet) verkauft. Die PDP-8-Serie war mit über 50'000 verkauften Einheiten das erfolgreichste kommerzielle Modell dieser Zeit.

Auf dem Pult meines Chefs von 1973 stand ein voll transistorisierter Tischrechner, auf den er echt stolz war (Abb. 24). Abb. 25 zeigt ein Detail einer Platine dieses Tischrechners und macht deutlich, dass die aktiven Elemente Transistoren und Dioden waren, die alle einzeln auf den Leiterplatten aufgelötet wurden.

Integrierte Schaltungen ab 1960

Der erste integrierte Schaltkreis war ein Speicher für ein Bit mit zwei Transistoren (vgl. Abb. 18), der 1958 von Jack Kilby (1923–2005), einem Ingenieur bei Texas Instruments (Dallas, USA) entwickelt wurde. Kilby gilt deshalb als Vater des Microchips, wofür er im Jahr 2000 mit dem Nobelpreis geehrt wurde.

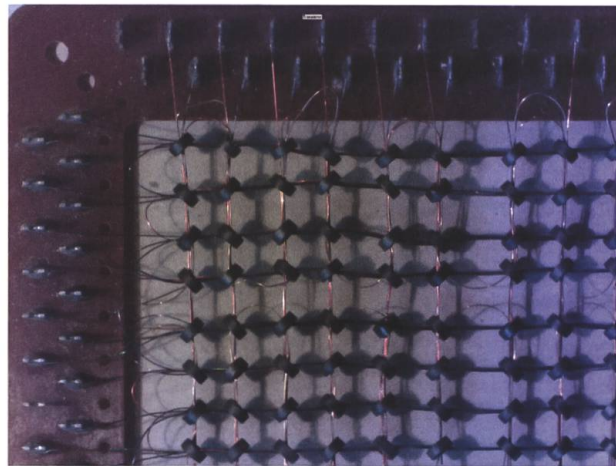


Abb. 23: Ringkern-Speichereinheit ca. 1961, vermutlich aus einem Rechner der Triumph-Adler-Büromaschinen-Vertriebs-GmbH in Nürnberg, die ab 1957 zur Max-Grundig-Gruppe gehörte. Erkennbar sind 64 Ringkerne aus einem 192-Bit-Speicher. Solche Speicher verlieren die Information beim Abschalten des Stroms nicht und entsprechen den heutigen Hard Disks oder Flash Memories (Memorysticks). Sie wurden anfänglich zusammen mit Röhren-Computern eingesetzt (Bild: Fritz Gassmann).



Abb. 24: Zweite Serie des weltweit ersten voll transistorisierten Tischrechners Friden Inc. San Leandro (California) Modell EC-132, das ab 1965 verkauft wurde. Mit einem Gewicht von 21 kg kostete er rund 20'000 \$ (auf heutigen \$-Wert umgerechnet). Er konnte die vier Grundoperationen mit 13-stelligen Festkomma-Zahlen durchführen und zusätzlich Quadratwurzeln ziehen und eine Zahl zur Wiederverwendung speichern. Er enthielt nur 324 Transistoren für die CPU und einige weitere für die Darstellung der 7-Segment-Zahlen, das Delay Memory und das Speisegerät. Um teure Transistoren zu sparen, wurde sein Arbeitsspeicher durch einen Metalldraht als Delay Line realisiert, in dem die gespeicherten Bits als mechanische Torsionswellen mit Schallgeschwindigkeit wanderten und für eine Speicherung durch die Elektronik laufend vom Drahtende gelesen und am Drahtanfang wieder eingegeben wurden (Bild: Children's Museum of Indianapolis, CC BY-SA 3.0).

Die Integration von Transistoren auf kleinen Germanium- und später ausschliesslich Siliziumplättchen führte zu einem exponentiellen Anstieg, bei dem sich die Anzahl Transistoren auf einem Chip etwa alle zwei Jahre verdoppelt hat (Moore's Law). Rechnen wir mit 60 Jahren ab 1958 bis 2018, ergeben sich mit diesem «Gesetz» $2 \times 2^{60/2} = 2^{31}$ Transistoren, also rund 2 Milliarden. Der Apple A12X Bionic Chip mit 10 Milliarden Transistoren zeigt, dass die Faustregel von Moore immer noch gilt, obschon bereits mehrere Male befürchtet wurde, dass die Entwicklung an ihre physikalischen Grenzen stösse. Mit dem Übergang von zwei zu drei Dimensionen (d. h. zwei oder mehr Lagen von Elementen übereinander), können die sicher existierenden physikalischen Grenzen weiter hinausgeschoben werden. Allerdings macht sich das Problem der Wärmeableitung immer stärker bemerkbar und treibt die Entwicklung zu immer verlustärmeren Transistoren.

Parallel dazu nahm auch die Rechengeschwindigkeit zu und erreicht heute TFLOPS (Tera Floatingpoint Operations Per Second = 1'000'000'000'000 Gleitkommamultiplikationen pro Sekunde). Umgekehrt verhielt sich der Preis, der von einigen Franken pro Transistor auf einige Nanofranken sank und den Computer für alle ermöglichte.

Der heute meistverkaufte Computer der Welt ist der ab 2012 in Wales (GB) produzierte Raspberry Pi, dessen Modell 3B+ um die Fr. 35.– kostet und rund 50 g wiegt (Abb. 26). Er sitzt auf einer einzigen 56 mm mal 86 mm grossen Platine und braucht bei einer Taktrate von 1400 MHz einen Strom von ca. 1 A bei 5 Volt. Mit Internet-Anschluss, WLAN, Bluetooth und 4 USB-Anschlüssen, Audio- und HDMI-Ausgang kann er mit einem Laptop mithalten. Für Tüftler, Elektronikfans und Physiker ist er jedoch dank seiner 26 digitalen Ein- und Ausgängen wesentlich interessanter als ein Laptop, weil er für beliebige Steuerungen und Erfassung von Daten verwendet werden kann. Er kann ohne Zusatzgeräte betrieben werden, braucht aber zur Programmierung den Anschluss eines Tastenfeldes und eines Monitors (z. B. Fernsehapparat) oder kann von einem Laptop aus via WLAN kontaktiert werden. Das Betriebssystem ist das Open-Source Linux, wofür alle wichtigen Programme wie z. B. Office existieren und gratis heruntergeladen werden können. Ein besonderes Highlight ist das weit verbreitete und leistungsfähige Mathematikpaket Wolfram-Mathematica, das dank der vier parallelen Prozessoren im Raspberry Pi auch anspruchsvolle Aufgaben bewältigen kann. Darunter fallen wissenschaftliche Berechnungen, Ingenieurproblemstellungen oder Projekte zur künstlichen Intelligenz.

Wie hat der Transistor die Welt verändert?

Ich denke, diese Frage wird weitgehend durch die Liste der heute alltäglichen Dinge beantwortet, die ohne den Transistor nicht möglich geworden wären:

- Computer, die sich alle leisten können
- Internet als billiges, weltweites Kommunikationssystem
- Mobiltelefon
- Geopositionssystem GPS
- Übertragung von Daten, Telefon, TV über dieselbe Leitung oder Glasfaser
- Drahtlosverbindung zu Computern, Druckern usw. (WLAN, Bluetooth)
- Entwicklungen, die viel Rechenleistung benötigen (Wettervorhersage, Grossraumflugzeuge, heutige astronomische Entdeckungen, usw.)
- selbstfahrende Autos
- künstliche Intelligenz

Diese Systeme haben alle Lebensbereiche grundlegend, aber leider nicht nur vorteilhaft verändert:

- Arbeitswelt der meisten Berufe
- einige Berufe sind obsolet geworden, andere wurden neu geschaffen
- Freizeitverhalten
- die Erde ist zum Dorf geworden
- Partnersuche im Internet

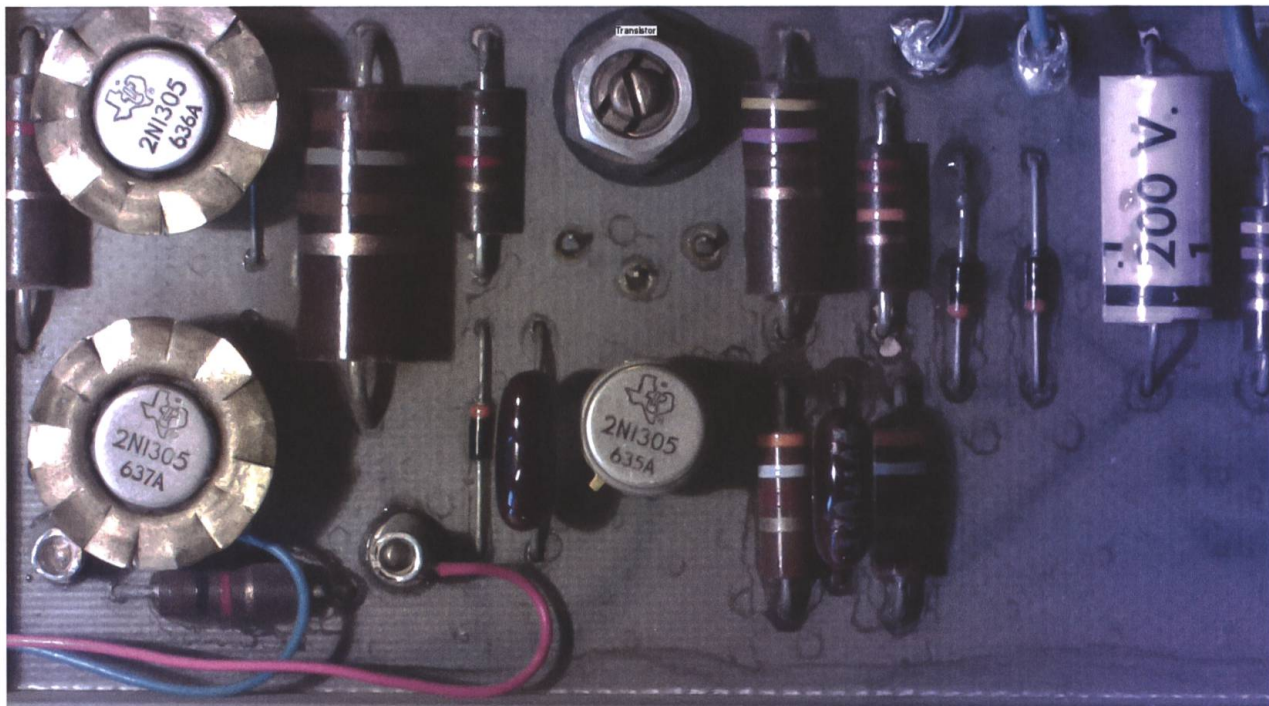


Abb. 25: Verstärker für das Delay Memory aus dem in Abb. 24 gezeigten Friden Tischrechner. Deutlich sind die drei Transistoren 2N1305 zu erkennen. Zwei davon tragen «Hüte», die die Abgabe der Verlustleistung erleichtern. Ebenfalls sichtbar sind drei Dioden mit der orangen Markierung des «Balkens» (vgl. Abb. 8). Alle Elemente mit mehreren Farbringen sind Widerstände. Beispielsweise bedeutet gelb-violett-braun (rechts der Bildmitte oben) 470 Ohm. Das links davon platzierte Element ist ein mit der Schraube justierbarer Widerstand (sogenannt Potentiometer). Das mit «200V» bezeichnete Element ist ein Kondensator (Ladungsspeicher) (Bild: Fritz Gassmann).

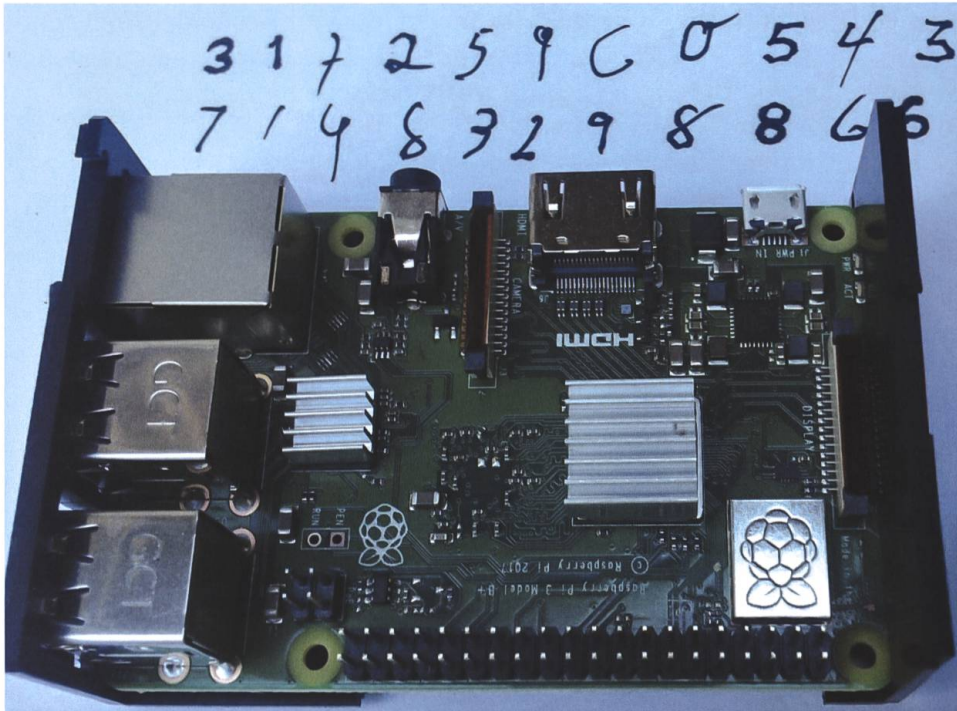


Abb. 26: Raspberry Pi Modell 3B+, der Billigcomputer für alle, die etwas mehr von Computern verstehen als Gaming und das Starten von fertigen Programmen. Das Wolfram - Mathematica-Software-Paket enthält Module zum Bau neuronaler Netzwerke. Diese können beispielsweise handgeschriebene Ziffern wesentlich schneller lesen als Menschen. Der CPU-Chip befindet sich unter den 7 Kühlrippen (Bild: Fritz Gassmann).

- unser Kommunikationsverhalten wurde konfrontiert mit:
 - jederzeitige Erreichbarkeit
 - Abhängigkeit vom Handy und Stress bei Jugendlichen
 - Parallelgesellschaften, die nur untereinander kommunizieren
 - Internet-Mobbing
 - Internet-Kriminalität (Verleumdung, Erpressung, usw.)
 - Aspekte von Kriegen werden ins Internet verlagert
 - Beschuss mit unerwünschten Angeboten, Spam-Mails, usw.
- Kleinkinder lernen die Welt durchs Tablet kennen
- Bücher und Zeitungen kommen aus der Mode, gelesen werden fast nur noch kurze, plakative Texte
- Jugendliche reagieren bevorzugt auf Bilder, leiden aber zunehmend an Reizüberflutung

Ist die Welt durch all dies einfacher oder komplizierter geworden? Ich denke, es schlummert ein gewaltiges positives Potenzial in den neuen Möglichkeiten. Um dies auszuschöpfen, muss jedoch die Psyche des Menschen den Entwicklungen folgen können – und hier scheint ein hoffentlich nur transientes Defizit vorhanden zu sein.

Literatur

- D'Ans/Lax (1963): Taschenbuch für Chemiker und Physiker. Springer, Berlin, Heidelberg, New York: 113-165.
- Feynman RP et al. (1963): The Feynman Lectures on Physics, Vol. III, Kap. 14-4. Addison-Wesley, Reading (Mass.)
- Hall EH (1879): On a New Action of the Magnet on Electric Currents. Am. J. Mathem. 2 (3): 287-292.