

Verwechselt, vergessen, wieder gefunden

Referenzen – das fehlerhafte Gedächtnis der Wissenschaft(ler)

Manuel Cardona und Werner Marx

Schreibfehler in wissenschaftlichen Zitierungen (Referenzen) sind mehr als nur ärgerliche Zufälligkeiten, sondern werfen durchaus ernsthafte Fragen auf, etwa: Bei welchen Referenzen ist die Fehlerrate besonders groß? Wie hoch ist eigentlich der Anteil verschriebener Zitierungen und welche Folgen haben diese?

Jeder Wissenschaftler weiß aus eigener Erfahrung, dass Zitieren mit einem gewissen Maß an Willkür behaftet ist, etwa was Auswahl und Vollständigkeit betrifft. Das kann im Rahmen der Forschungsevaluierung, wo vielfach Zitierungen als zusätzlicher Indikator für die Wirkung (Impact) wissenschaftlicher Arbeiten herangezogen werden, zu Verzerrungen führen. Neben solchen (z. T. beabsichtigten) Verzerrungen gibt es eine ganze Reihe von Fehlern, die aus Nachlässigkeit oder Unkenntnis begangen werden und in Form von verschriebenen oder auch fehlenden Zitaten dokumentiert sind. Diese Fehler spielen weniger bei Impact-Abschätzungen eine Rolle, sondern erlauben vielmehr Einblicke in den Prozess des Zitierens und damit in die Rezeption wissenschaftlicher Arbeiten.

Stellt man die Varianten zusammen, mit denen einzelne hochzitierte

Arbeiten zitiert wurden¹⁾, so kann man leicht feststellen, dass rund fünf Prozent aller Zitierungen bezüglich der darin enthaltenen numerischen Angaben fehlerhaft sind. Die Bandbreite der Fehler ist dabei sehr groß: Von Zahlendrehern (vor allem bei mehrstelligen Seitenzahlen) über Vertauschungen von Bandnummern, Heftnummern und Seitenzahlen sowie falschen Jahreszahlen bis hin zu Zahlen-Verstümmelungen, die das Wiederauffinden unmöglich machen. Hinzu kommen Fehler bei den Namen der Erstautoren, vor allem bei komplexen oder aus anderen Sprachen transkribierten Namen.

Vielfalt der Varianten

Aufgrund der vielen Fehlermöglichkeiten kann es vorkommen, dass eine einzige Arbeit in hunderten von Varianten zitiert wurde. Das passiert vor allem bei ungewöhnlichen und/oder langen Zeitschriftennamen. Diese werden von ISI nur teilweise standardisiert und häufig so übernommen, wie sie in den Referenzen der zitierenden Arbeiten aufgeführt sind. Sie entsprechen deshalb nicht unbedingt den üblicherweise verwendeten Namen bzw. Kürzeln.

Die berühmte Arbeit des dänischen Physikers Jens Lindhard zur dielektrischen Funktion eines freien Elektronengases aus dem Jahr 1954

als zitierte Arbeit, wenn sie von mindestens einer der Arbeiten dieser Kernzeitschriften zitiert wurde. Für den Zeitraum ab 1974 stehen ca. 20 Millionen Source Items rund 200 Millionen Reference Items gegenüber, d. h. eine Arbeit wurde im Mittel 10-mal zitiert. In der Datenbank gelten für die zitierenden und die zitierten Arbeiten unterschiedliche Standards. Die zitierten Arbeiten werden nur in Kurzform abgespeichert: Das ISI-Referenzformat (eine Art Referenzcode) enthält lediglich den Erstautor, das Publikationsjahr, die Bandnummer und die Startseite. Im Rahmen von Zitierungsanalysen werden in der Regel nur diese Daten abgeglichen. Fehlerhafte Referenzen können nur im Falle der Untersuchung einzelner Arbeiten berücksichtigt werden, nicht jedoch bei der Bewertung größerer Ensembles (z. B. allen Arbeiten eines Wissenschaftlers oder eines Forschungsinstitutes).



Zitierungen sind der Leitfaden durch den Dschungel der wissenschaftlichen Publikationen. Allzu oft schleichen sich jedoch Fehler in die Referenzen ein, die nicht nur das Auffinden einer Arbeit erschweren, sondern oft genug darauf hindeuten, dass der Zitierende die Arbeit nie in den Händen gehabt hatte.

ist ein gutes Beispiel [1]. Die Arbeit wurde bisher in ca. 200 Varianten insgesamt rund 1660-mal zitiert, davon lediglich 219-mal (13 %) mit den korrekten numerischen Angaben. Abbildung 1 zeigt einen kleinen Auszug aus den zahlreichen Zitatvarianten.²⁾ Fehler im Namen und Publikationsjahr sind darin noch nicht enthalten. Viele der Varianten wurden durch den langen dänischen Titel der Quelle erzeugt (Matematisk-Fysiske Meddelelser Kongelige Danske Videnskabernes Selskab). Die meistzitierten Varianten unterscheiden sich jedoch vor allem durch die Startseite. Ganze 522 Zitierungen (31 %) entfallen auf die falsche Startseite 8 (statt 1). Offensichtlich wurde die Heftnummer 8 (Band 28, Heft 8) mit der Startseite verwechselt. Das geschah erstmals bereits 1959 in einer Arbeit von Ehrenreich und Cohen, die ihrerseits 738-mal zitiert wurde und damit als Ursprung des Fehlers angesehen werden muss [2]. 1962 folgten zwei weitere Arbeiten in Physical Review mit 505 bzw. 735 Zitierungen, die den Fehler dann weiter fortgepflanzt haben [3, 4]. Eine Arbeit, die 1968 in der renommierten Zeitschrift Reviews of Modern Physics erschienen ist und bisher 1213-mal zitiert wurde, hat sicherlich zur weiteren Verbreitung des Fehlers beigetragen [5].

1) Das erreicht man etwa durch das Expand-Kommando unter STN und Cited Reference Search unter WoS.

2) In der neuesten Version des Web of Science (WoS) sind die fehlerhaften Zitatvarianten weitgehend korrigiert und zusammengefasst worden. Damit stimmen jedoch die tabellierten Referenzen nicht mehr mit denjenigen in den Originalarbeiten überein und deren Zitatfehler sind nun verschleiert.

Prof. Dr. Manuel Cardona, Dr. Werner Marx, Max-Planck-Institut für Festkörperforschung, Heisenbergstraße 1, 70569 Stuttgart

Zitierindexe

Datengrundlage für die Ermittlung von Zitierungszahlen sind die Zitierindexe von Thomson ISI (Institute for Scientific Information, www.isinet.com/isi), hier insbesondere der Science Citation Index (SCI). Der SCI wird von ISI hergestellt und im Web of Science (WoS) oder über einen Datenbankanbieter wie STN International (www.stn-international.de) angeboten. Unter WoS reicht der SCI zeitlich bis 1945 zurück, unter STN nur bis 1974. Die Besonderheit des SCI besteht darin, dass nicht nur alle Arbeiten der erfassten Zeitschriften, sondern auch alle darin enthaltenen Literaturverweise (Referenzen, Zitate) abgespeichert werden. Ein und dieselbe Arbeit kann deshalb im SCI sowohl als zitierende wie auch als zitierte Arbeit erscheinen. Als zitierende Arbeit (Source Item), wenn sie in einer der vom SCI erfassten Kernzeitschriften (Source Journals) erschienen ist, und

3) Alte traditionsreiche Zeitschriften wie die Annalen der Physik haben mehrmals ihren Namen geändert, was zusätzliche Fehler provoziert und die Zuordnung von Referenzen beeinträchtigt.

Die Zitat-Varianten ermöglichen es, den Anteil der zitierten Arbeiten abzuschätzen, der von den zitierenden Autoren (vermutlich) nicht wirklich gelesen, sondern als Referenz von anderen Arbeiten abgeschrieben wurde. Im vorliegenden Falle haben mindestens 30 Prozent die Originalarbeit ganz offensichtlich nie in Händen gehabt, zumindest nicht zum Zeitpunkt des Zitierens. Die zeitliche Reihenfolge der Arbeiten mit einem bestimmten Zitatfehler dokumentiert, wer vermutlich die Referenz von wem abgeschrieben hat. Grundsätzlich gilt: Je älter eine Arbeit, desto größer die Wahrscheinlichkeit, dass eine „Mutation“ in der entsprechenden Referenz stattgefunden hat.³⁾

1	LINDHARD J	KGL DANSKE VIDEN MFM	1954		
34	LINDHARD J	KGL DANSKE VIDEN MFM	28		1954
213	LINDHARD J	KGL DANSKE VIDEN MFM	28	1	1954
2	LINDHARD J	KGL DANSKE VIDEN MFM	8		1954
3	LINDHARD J	KGL DANSKE VIDEN SKA	28		1954
1	LINDHARD J	KGL DANSKE VIDENS FM	28		1954
3	LINDHARD J	KGL DANSKE VIDENSK 8	28		1954
1	LINDHARD J	KGL DANSKE VIDENSK M	28		1954
7	LINDHARD J	KGL DANSKE VIDENSK S	28		1954
1	LINDHARD J	KGL DANSKE VIDENSK S	23		1954
146	LINDHARD J	KGL DANSKE VIDENSKAB	28		1954
22	LINDHARD J	KGL DANSKE VIDENSKAB	28	3	1954
501	LINDHARD J	KGL DANSKE VIDENSKAB	28	8	1954
1	LINDHARD J	KGL DANSKE VIDENSKAB	28	30	1954
1	LINDHARD J	KGL DANSKE VIDENSKAB	28	104	1954
1	LINDHARD J	KGL DANSKE VIDENSKAB	28	113	1954
4	LINDHARD J	KGL DANSKE VIDENSKAB	8		1954
1	LINDHARD J	KGL DANSKE VIDENSKAD	28		1954
3	LINDHARD J	KGL DANSKE VIDENSKAP	28		1954
1	LINDHARD J	KGL DANSKE VIDENSKLA	28		1954
1	LINDHARD J	KGL DANSKE VIDENSKRA	28		1954
1	LINDHARD J	KGL DANSKE VIDED	28		1954
4	LINDHARD J	KGL DANSKE VIDENSKA	28		1954
1	LINDHARD J	KGL DANSKE VIDENSKA	28	29	1954

Abb. 1: Ein Auszug aus der langen Liste der Zitatvarianten der Lindhard-Arbeit [1]. Die orange hinterlegte Variante ist diejenige mit den korrekten numerischen Angaben. (Quelle: WoS, 24. Mai 2004)

Die falsche Arbeit zitiert
Schwerwiegender sind Fehler, die eine Verwechslung hinsichtlich der zitierten Arbeiten beinhalten. Vor allem im Falle alter Arbeiten, die nicht in Englisch abgefasst und auch nicht ohne weiteres verfügbar sind, ist diese Gefahr gegeben. Ein Beispiel dafür ist eine Arbeit von Rudolf Kohlrausch, die 1854 in Pogendorffs Annalen der Physik und Chemie erschienen ist [6]. Diese Arbeit beschreibt erstmals Relaxationsvorgänge in Form einer gestreckten Exponentialfunktion (stretched exponential function). Dieses Konzept war in Vergessenheit geraten und wurde mit Verweis auf Kohlrausch erst wieder 1974 [7, 8] und dann 1984 von Palmer et al. aufgenommen [9].

Unglücklicherweise verweist die Palmer-Arbeit jedoch auf die falsche Kohlrausch-Arbeit, eine frühere Arbeit, in der Relaxationsvorgän-

ge lediglich qualitativ beschrieben werden [10]. Die Kohlrausch-Arbeit von 1847 wurde erstmals seit 1945 im Jahr 1977 zitiert, hier jedoch zusammen mit der Arbeit von 1854 [11]. Frühere Verweise können nicht ermittelt werden, da kein Zitierindex für den Zeitraum vor 1945 existiert. In einer Arbeit von 1983 wird die frühere Kohlrausch-Arbeit erstmals alleine zitiert, doch die zitierende Arbeit wurde ihrerseits nur 13-mal zitiert und kommt deshalb als Ursache für die Verbreitung der unzutreffenden Zitierung nicht in Frage [12].

Mit dem Erscheinen der hochzitierten Palmer-Arbeit (917 Zitierungen) wurde jedoch die Kohlrausch-Arbeit von 1847 stark ansteigend zitiert (Abb. 2). Diese Arbeit erhielt bisher 630 Zitierungen, gegenüber 310 für die Arbeit von 1854. Im Zusammenhang mit dem Konzept der gestreckten Exponentialfunktion verweisen zumindest seit 1974 mehr Arbeiten auf die falsche Kohlrausch-Arbeit (117 Arbeiten) als auf die zutreffende Arbeit von 1854 (nur 59 Arbeiten). Die Vermutung, dass die Palmer-Arbeit die Ursache der Fehlerfortpflanzung ist, wird durch eine weitere Tatsache erhärtet: Diese Arbeit wurde alleine 123-mal mit der Kohlrausch-Arbeit von 1847 gemeinsam zitiert (cozitiert). Auch die Kohlrausch-Arbeit von 1854 wird in zahlreichen Varianten zitiert.⁴⁾

Schüler mit Lehrer verwechselt

Statt Arbeiten werden manchmal auch Autoren verwechselt. Ein Beispiel: Der Physikochemiker und Nobelpreisträger Wilhelm Ostwald hat im Jahr 1900 in der von ihm gegründeten Zeitschrift für Physikalische Chemie eine Arbeit publiziert, die seit 1945 vermehrt zitiert wird (297 Zitierungen). Darin geht es um die nach ihm benannte Ostwald-Reifung [13]. Zusammen mit dieser Ostwald-Arbeit wird manchmal auch eine andere Arbeit von 1901 zitiert, die zwar Wilhelm Ostwald zugeschrieben wird, tatsächlich aber von dessen Schüler G. A. Hulett stammt [14]. Im SCI ist diese Arbeit mit 8 Zitierungen bezogen auf den korrekten Autor, jedoch mit 47 Zitierungen fälschlicherweise mit Ostwald als Autor nachgewiesen. Offenbar hat keiner der Autoren dieser 47 Arbeiten die vermeintliche Ostwald-Arbeit je gesehen, geschweige denn gelesen.

Wenn man wie oben die den falschen Autor zitierenden Arbeiten zeitlich ordnet, so findet man

als älteste Arbeit eine Publikation von H. E. Cline von 1971, welche damit als Ursprung in Frage kommt [15]. Diese Arbeit konnte mit 111 eigenen Zitierungen den Fehler fortpflanzen. Allerdings lässt sich nicht ausschließen, dass bereits vor

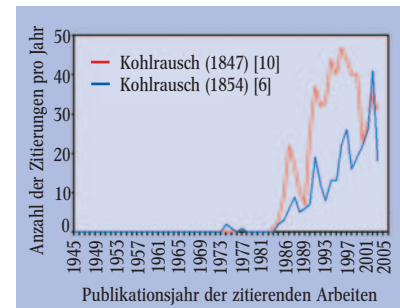


Abb. 2: Zeitkurven der Zitierungen der beiden Kohlrausch-Arbeiten [6] und [10]. Alle Zitatvarianten, die sich den Arbeiten zuordnen lassen wurden berücksichtigt. Statt einer Arbeit von 1854, in der Rudolf Kohlrausch eine gestreckte Exponentialfunktion zur Beschreibung von Relaxationsvorgängen eingeführt hat, wird eine andere Arbeit von 1847 fälschlicherweise häufiger zitiert.

1945 falsch zitiert wurde und Cline dort abgeschrieben hat. Unter den falsch zitierenden Arbeiten ist eine Arbeit von P. W. Voorhees von 1985 mit 312 eigenen Zitierungen diejenige mit dem größten Impact [16]. Diese Arbeit wurde innerhalb der 47 falsch zitierenden Arbeiten mindestens 18-mal mit der vermeintlichen Ostwald-Arbeit cozitiert, ein weiteres Indiz dafür, dass diese Arbeit zumindest Verstärker des Fehlers ist. Voorhees hat die Referenz möglicherweise bei Cline abgeschrieben.

Vergessen...

Manche frühen Arbeiten werden heute ausgiebig zitiert, obwohl deren Autoren nicht zu den eher prominenten Wissenschaftlern zählen. Doch einzelne frühe Arbeiten, die Fachkollegen als Pionierarbeiten ansehen, werden seit 1945 erstaunlich wenig zitiert. Ein Beispiel dafür ist eine Arbeit von Ewing und Humfrey aus dem Jahr 1903 (25 Zitierungen), die als erste die Ermüdungserscheinungen von Metallen auf mikroskopischer Ebene untersucht hat [17]. Als die moderne Materialwissenschaft in den fünfziger Jahren ihren Aufschwung nahm, war diese Arbeit nahezu vergessen. Nicht anders erging es einer Arbeit von Frenzel und Schultes von 1934, in der über den fotografischen Nachweis von durch Ultraschall erzeugter Lumineszenz in Wasser berichtet wird [18]. Auch diese Arbeit wurde bis heute wenig

4) Die Autoren danken Herrn Prof. Ralph V. Chamberlin (Department of Physics and Astronomy, Arizona State University, Tempe, USA), der sie auf die Kohlrausch-Problematik aufmerksam gemacht hat.

zitiert (73 Zitierungen), obwohl die Forschungsdisziplin (Sonolumineszenz, Photoakustik) seit Anfang der achtziger Jahre einen enormen Aufschwung verzeichnet. Vermutlich hätte eine einzige Zitierung durch einen renommierten Wissenschaftler bzw. eine vielbeachtete Publikation Arbeiten wie diejenigen von Ewing oder Frenzel wieder in Umlauf gebracht. Offenbar verfolgen manche Wissenschaftler die Wurzeln der Ideen ihrer Forschungsdisziplinen nicht wirklich bis zu den Anfängen zurück. Zumindest Reviews sollten hier sorgfältig sein und auf die Ursprünge verweisen.

... oder wiederentdeckt

Wenn man die in den vergangenen Jahrzehnten meistzitierten Arbeiten eines bekannten Wissenschaftlers oder einer renommierten Zeitschrift aus dem Zeitraum vor 1945 untersucht, so findet man manche Arbeit, die erst nach vielen Jahrzehnten zitiert wurde. Einige Arbeiten aus der ersten Hälfte des zwanzigsten Jahrhunderts wurden in den vergangenen beiden Jahrzehnten stärker zitiert als viele Nobel-Arbeiten. Es handelt sich hierbei nicht um die Klassiker der Physik und Chemie wie z. B. die Arbeiten zu Einsteins spezieller und allgemeiner Relativitätstheorie. Einsteins heute meistzitierte Arbeit ist die EPR-Arbeit von 1935.

Solche klassischen Arbeiten werden eher nicht mehr explizit zitiert, sondern als bekannt vorausgesetzt und nur noch in Form der Entdeckernamen (Einsteinsche Gleichung) oder der entsprechenden Begriffe (Relativitätstheorie, relativistisch) im Text erwähnt. Bei den wiederentdeckten Arbeiten handelt es sich vielmehr um solche, mit denen die Fachkollegen zum Zeitpunkt der Publikation aus unterschiedlichen Gründen wenig anfangen konnten. Manchmal handelt es sich um Arbeiten, die ihrer Zeit voraus waren.

Ein Beispiel für solche Arbeiten ist die Publikation von Gustav Mie, welche im Jahre 1908 in den Annalen der Physik erschienen ist und die Lichtstreuung an Partikeln wie Aerosolen behandelt, welche größer als die Wellenlänge des Lichtes sind [19]. Im Gegensatz zur Rayleigh-Streuung ist die Mie-Streuung nahezu unabhängig von der Wellenlänge. Zumindest um 1945 wurde diese Arbeit kaum beachtet. Dann jedoch hat sie große Bedeutung im Zusammenhang mit der Untersuchung von

Kolloiden, Metallen in Suspensionen und interstellaren Molekülen erhalten. Die Arbeit wurde dementsprechend seit Ende der fünfziger Jahre verstärkt zitiert und gehört mit derzeit über 3000 Zitierungen seit 1945 zu den meistzitierten frühen Physik-Arbeiten (Abb. 3). Andere Arbeiten mit verspäteter Anerkennung wurden erst wieder stärker zitiert, nachdem in einer hochzitierten Arbeit darauf verwiesen wurde (siehe Kohlrausch). Bei den

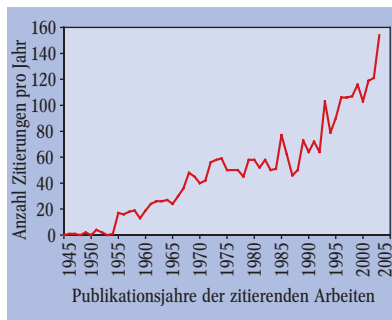


Abb. 3: Zeitkurve der Zitierungen der Mie-Arbeit von 1908. Diese Arbeit wurde seit 1945 rund 3000-mal zitiert und ist damit eine der heute meistzitierten frühen Physik-Arbeiten. Es wurden alle Zitatvarianten berücksichtigt, die sich der Arbeit zuordnen lassen.

Arbeiten, die frühe Arbeiten wieder in Umlauf bringen, handelt es sich oft um Review-Arbeiten oder Arbeiten in renommierten Zeitschriften wie Nature oder Science. Verzögerte Arbeiten sind gemessen an der Menge aller publizierter Arbeiten zwar selten, doch durchaus keine Einzelfälle. Die Regel ist, dass Arbeiten in den ersten Jahren nach ihrer Publikation zitiert werden, oder nie. Doch die Wissenschaft wird allem Anschein nach von relativ wenigen Pionierarbeiten angetrieben und man kann nie wissen, wie eine bestimmte Arbeit später bewertet wird.

Schlussfolgerungen

Die hier beispielhaft aufgeführten Fehler sind durchaus keine seltenen Einzelfälle. In Anbetracht dessen sollten Wissenschaftler sich generell und insbesondere beim Zitieren älterer Arbeiten nicht dazu verleiten lassen, Referenzen bei anderen Arbeiten abzuschreiben. Andernfalls tragen sie mit dazu bei, solche Fehler endlos fortzupflanzen. Sie sollten sich darüber im Klaren sein, dass es bei Referenzen nicht nur um die Wiederauffindbarkeit der Arbeiten geht. Referenzen werden in maschinenlesbarer Form abgespeichert (inzwischen nicht nur im SCI, sondern auch in weiteren Datenbanken) und

sind Grundlage für die Ermittlung des in Form von Zitierungen dokumentierten Impact wissenschaftlicher Arbeiten. Außerdem werden Zitierungen dazu genutzt, Sachverhalte zu recherchieren, das ursprüngliche Ziel bei der Erfindung des SCI. Fehlerhafte Referenzen bedeuten in diesem Zusammenhang immer auch Informationsverluste. In einigen Fällen kann der Informationsverlust massiv sein (siehe Lindhard). Jeder Wissenschaftler sollte auch beim Zitieren die in der Wissenschaft allgemein geforderte Sorgfalt walten lassen und dazu beitragen, die Archive der Wissenschaft nicht unnötig mit Fehlern zu belasten.

Literatur

- [1] J. Lindhard, Matematisk-Fysiske Meddelelser Kongelige Danske Videnskabernes Selskab **28** (8), 1-57 (1954)
- [2] H. Ehrenreich und M. H. Cohen, Physical Review **115** (4), 786 (1959)
- [3] S. L. Adler, Physical Review **126** (2), 413 (1962)
- [4] D. R. Penn, Physical Review **128** (5), 2093 (1962)
- [5] U. Fano und J. W. Cooper, Reviews of Modern Physics **40** (3), 441 (1968)
- [6] R. Kohlrausch, Poggendorffs Annalen der Physik und Chemie **91** (1), 56/179 (1854) [Startseite ist 56, S. 179 verweist auf den Teil, der die gestreckte Exponentialfunktion behandelt und wird meistens zitiert.]
- [7] M. U. Farooq und J. Hirsch, Solid State Communications **15** (8), 1417 (1974)
- [8] J. Thorson und M. Biederman-Thorson, Science **183** (4121), 161 (1974)
- [9] R. G. Palmer, D. L. Stein, E. Abrahams und P. W. Anderson, Physical Review Letters **53** (10), 958 (1984)
- [10] R. Kohlrausch, Poggendorffs Annalen der Physik und Chemie **72** (11), 353/393 (1847) [Startseite ist 353, S. 393 verweist auf den Teil, der die Relaxation behandelt und wird manchmal zitiert.]
- [11] H. Markovitz, Transactions of the Society of Rheology **21** (3), 381 (1977)
- [12] K. L. Ngai, A. K. Rajagopal, R. W. Rendell und S. Teitler, Physical Review B **28** (10), 6073 (1983)
- [13] W. Ostwald, Zeitschrift für Physikalische Chemie **34**, 495 (1900)
- [14] G. A. Hulett, Zeitschrift für Physikalische Chemie **37**, 385 (1901)
- [15] P. W. Voorhees, Journal of Statistical Physics **38** (1-2), 231 (1985)
- [16] H. E. Cline, Acta Metallurgica **19** (6), 481 (1971)
- [17] J. A. Ewing und J. C. W. Humfrey, Philosophical Transactions of the Royal Society A **200**, 241 (1903)
- [18] H. Frenzel und H. Schultes, Zeitschrift für Physikalische Chemie, **B27**, 421 (1934)
- [19] G. Mie, Annalen der Physik **25**, 377 (1908)