

VERSLAG LEZING ONLINE CONFERENTIE 1986

DE VACATURE IN SILICIUM - EEN ONLINE ZOEKACTIE IN INSPEC

E.G. Sieverts,
Universiteitsbibliotheek,
Wittevrouwenstraat 7-11,
3512 CS Utrecht.

Samenvatting

In het bestand INSPEC is literatuur gezocht over een specialistisch onderwerp uit de halfgeleiderfysica: Theoretisch fysische berekeningen aan de vacature in silicium. Over dit onderwerp was met conventionele methoden al een tamelijk uitgebreide literatuurverzameling aangelegd. Hierdoor was het mogelijk de resultaten van online zoekacties met verschillende zoekstrategieën, bij de hosts ESA en DIALOG te evalueren. De problemen die algemene en specifieke facetten van de zoekvraag ieder voor zich opleveren zijn nader bekeken. Ook de diverse aspecten van de door INSPEC toegekende indexing konden beoordeeld worden. Tenslotte is onderzocht waarom bepaalde literatuur niet gevonden is.

1. Inleiding

Bij het evalueren van de resultaten van online zoekacties is het vaak moeilijk om een enigszins betrouwbare indruk te krijgen van grootheden als VANGST (in het Engels RECALL) en PRECISIE. Het is immers bijna altijd onbekend hoeveel relevante literatuur gemist is. Ook het beoordelen van de relevantie van de wel gevonden literatuur moet vaak aan een onderwerpsspecialist overgelaten worden. In dit geval lag dat wat anders omdat al over een reeks van jaren op conventionele wijze uit tijdschriften, uit de gedrukte Physics Abstracts, via conferenties en door literatuurverwijzingen een uitgebreide literatuurverzameling opgebouwd was.

Pas in een later stadium is die met online zoekacties nader aangevuld /1/.

De uiteindelijke verzameling bevatte 139 referenties. Verder was ik zelf niet alleen als intermediair werkzaam, maar ook direct inhoudelijk bij het betreffende onderzoeksproject betrokken, waardoor de beoordeling van de gevonden literatuur een stuk eenvoudiger werd.

Bij de beoordeling van de zoekresultaten kunnen verschillen in strategie nader bekeken worden. Ook problemen die elk van de facetten van de zoekvraag met zich meebrengt worden afzonderlijk besproken.

Omdat daarbij zeer algemene overwegingen aan de orde komen is deze behandeling van ruimere betekenis dan alleen voor de onderhavige specialistische zoekvraag. Zo komen aspecten van indexerings door de producent van het INSPEC-bestand aan de orde, zo goed als doorzoekbaarheid van bepaalde velden bij de hostcomputers ESA en DIALOG.

2. Het onderwerp

De zoekvraag betreft literatuur over theoretisch fysische berekeningen aan vacatures in silicium. Zeer specialistische fundamentele fysica wordt daarbij toegepast op een stofje waar bijna de gehele halfgeleidertechnologie om draait. Als gevolg van de uitgebreide toepassing kunnen silicium-kristallen met een zeer hoge graad van perfectie gemaakt worden. Toch kunnen in het regelmatige kristalrooster nog verstoringen optreden. De eenvoudigste verstoring van de regelmaat is het ontbreken van een enkel siliciumatoom. Men spreekt dan van de aanwezigheid van een vacature. Juist door zijn eenvoud is de vacature een geliefd object om er met theoretisch fysische methoden (en meestal met de computer) aan te rekenen. Wie experimenten aan zulke vacatures doet is vanzelfsprekend geïnteresseerd in de resultaten van die berekeningen /1,2/.

Voor het doen van een online zoekactie valt het onderwerp op nogal vanzelfsprekende wijze in drie facetten uiteen: (1) de vacatures waar het om gaat, (2) de stof silicium waar ze in moeten zitten, en tenslotte (3) hetgeen aan die vacatures gedaan is, de theoretische berekeningen. Ondanks de overzichtelijkheid van het onderwerp levert elk van die facetten zijn eigen problemen op bij het opzetten van een zoekstrategie.

3. De facetten

3.1 De vacature

INSPEC heeft voor vacatures een eigen thesaurusterm VACANCIES. Jammer genoeg wordt deze term frequent in een meer algemene zin gebruikt, ook voor allerlei complexen en structuren waarvan grotere aantallen vacatures deel uit maken. Voor een hoge vangst zal met graagte van deze term gebruik gemaakt worden, maar voor een heel preciese zoekstrategie is de thesaurusterm wellicht te ruim.

In eerste instantie is de thesaurusterm gebruikt en zelfs nog aangevuld door ook op de getrunceerde term VACANC? in de titels van publicaties te zoeken. In tweede en derde instantie is dit facet nader ingeperkt door in een AND-relatie te combineren met een reeks van termen die direct of indirect aangeven dat het om een enkele vacature gaat.

In het bijzonder voor het bedenken van de "indirecte" termen die bijna uitsluitend in samenhang met enkele vacatures gebruikt worden en die iets zeggen over verplaatsingen van siliciumatomen rondom de vacature, is de inbreng van de onderwerpsspecialist onontbeerlijk. De drie gebruikte omschrijvingen van dit facet van de zoekvraag zijn als A1, A2 en A3 gespecificeerd in tabel I.

3.2 Silicium

Ook de stof silicium wordt in INSPEC door een thesaurusterm aangegeven. Met SILICON/CT vindt men bij ESA evenwel ook alle literatuur waaraan de termen SILICON ALLOYS of SILICON COMPOUNDS toegekend zijn. Bij ESA zal die dus met een NOT combinatie uitgezonderd moeten worden. Bij DIALOG kan men in principe direct gebruik maken van de "full descriptor" aanduiding, SILICON/DF, om dit te bereiken. In de praktijk is het ook daar verstandig om de genoemde NOT combinatie te maken, om niet ook wat artikelen over eigenschappen van siliciumatomen in metaallegeringen of chemische verbindingen te krijgen.

Bij het gebruiken van deze omschrijving voor het stof-facet blijkt nog tamelijk veel literatuur niet gevonden te worden. Bij een nadere beschouwing blijkt bij veel berekeningen niet expliciet van silicium sprake te zijn, maar van een hele groep van halfgeleiders met soortgelijke kristalstructuur. Daartoe behoort in het bijzonder ook diamant, waardoor die kristalstructuur wel diamantstructuur genoemd wordt. Voor hogere vangst is dit facet daarom nog uitgebreid met de termen DIAMOND?, TETRAHEDRAL(1w)SEMICONDUCTOR? en COVALENT(w) SEMICONDUCTOR?.

Men kan hier van synoniemen of verwante termen van de term SILICON spreken. De beperkte en uitgebreide omschrijvingen van dit facet worden als B1 en B2 in Tabel I gegeven.

TABEL I

Gebruikte omschrijvingen voor de drie facetten van de zoekvraag. Zoektermen in de gegeven lijsten zijn met OR gecombineerd.

VACATURE -----	SILICIUM -----	THEORIE -----
A1: vacancies/ct OR vacanc?/ti	B1: silicon/ct NOT (silicon alloys OR silicon compounds)	C1: (theor? calculation? ? ab(w)initio self(w)consist? pseudopotent? molecular(w)orbital? green? ? wannier tight(w)binding xalpha mndo variational)
A2: vacanc?/ti OR (vacancies/ct AND D)	B2: B1 OR (diamond? tetrahedral(1w) semiconductor? covalent(w) semiconductor?)	C2: C1 OR (atomic(w)orbital? electronic(w)structure? defect(w)molecule electron(w)energy(w)state? ? electronic(w)state? ? 1cao wavefunction? function? ? many(w)electron jahn(w)teller)
A3: A1 AND D		C3: C2 OR (ideal undistort? unreconstruct? reconstruct?)
D: (relaxed unrelaxed reconstruct? unreconstruct? distort? undistort? ideal single isolated monovacanc?)		C4: TC=theoretical (not f) TC=experimental

3.3 Theoretische berekeningen

Het facet van de theoretische berekeningen blijkt verreweg het moeilijkst om kort te omschrijven. In het merendeel van de artikelen zullen de woorden theorie of berekening noch in titel of abstract voorkomen, noch als trefwoord toegekend worden. In plaats daarvan wordt een grote variëteit aan namen van theoretische methoden gebruikt. Een lijst van zoektermen die deze theoretische methoden aangeven, is in Tabel I onder C1 gegeven.

In een uitgebreidere omschrijving van dit facet kunnen ook nog grootheden of eigenschappen toegevoegd worden die specifiek met deze methoden worden bepaald. Hoewel je daarbij niet kunt uitsluiten dat ook experimentele studies gevonden worden waarin deze grootheden bepaald zijn, kan de extra vangst tegen dat bezwaar opwegen. Bij deze termen kan van "quasi-synoniemen" van het theorie-aspect gesproken worden; termen die in het verband van de zoekvraag indirect op het gebruik van een theoretische berekening (kunnen) wijzen. Onder C2 zijn deze extra termen in Tabel I opgenomen.

Voor een nog uitgebreider beschrijving kan met dezelfde redenering ook nog een aantal termen toegevoegd worden die eerder juist gebruikt zijn om het facet van de "enkele" vacature in te perken. Deze termen die iets over de omgeving van de vacature zeggen, geven bij de theoretische berekeningen een soort randvoorwaarden aan waarbinnen die berekeningen uitgevoerd zijn. Nu deze termen voor uitbreiding in plaats van inperking (zoals in 3.1) gebruikt worden, moeten woorden met een te algemene betekenis vermeden worden. In Tabel I is deze omschrijving van het theorie-facet als C3 vermeld.

Een geheel andere, veel globaler omschrijving van dit facet is bij DIALOG mogelijk. Daar is de door INSPEC toegekende TREATMENT CODE doorzoekbaar, zodat met TC=THEORETICAL volstaan kan worden.

Om alleen echte theorie te krijgen en geen experimentele verhandelingen, waarin ook nog met een theoretische formule vergeleken wordt, dient wel het tegelijkertijd voorkomen van de treatment code EXPERIMENTAL uitgesloten te worden. In Tabel I is deze omschrijving als C4 opgenomen.

4. Resultaten van de zoekacties

4.1 Algemene resultaten

Van verschillende mogelijke combinaties van de eerder genoemde omschrijvingen van de drie facetten is bekeken wat het resultaat is. Voor het bepalen van de vangst (recall) is het aantal gevonden relevante referenties gedeeld door het totale aantal relevante referenties dat op alle mogelijke manieren verzameld was. Opgemerkt moet worden dat ook daarbij nog uitsluitend van relatieve vangst gesproken kan worden, omdat niet gegarandeerd kan worden dat de gevormde verzameling werkelijk alle gepubliceerde literatuur bevat.

Voor het bepalen van de precisie is van elke combinatie het aantal gevonden relevante referenties gedeeld door het totaal aantal zo gevonden referenties.

Aantal gevonden referenties, vangst en precisie van elk van de gemaakte combinaties is opgenomen in Tabel II. Daarbij dient opgemerkt te worden dat bij ESA de periode 1971-heden en bij DIALOG 1977-heden doorzocht is. De totale aantallen relevante publicaties in de eigen verzameling voor die periode waren respectievelijk 139 en 102.

Voor een betere vergelijking zijn ook nog enkele ESA resultaten tot de kortere tijdsperiode 1977-heden ingeperkt en in de tabel opgenomen.

De ruimste combinatie A1*B1*C3 bevatte ook alle bij DIALOG met A1*B1*C4 gevonden relevante publicaties, plus nog 5 meer. Met de nauwere combinaties A2*B1*C3 en A2*B1*C4 werden zowel bij ESA als bij DIALOG drie relevante publicaties gevonden die er bij de ander niet bij zaten.

TABEL II

Resultaten van verschillende zoekstrategieën bij ESA en DIALOG. De omschrijving van de facetten is in Tabel I gegeven.

Combinatie	Host	Aantal	Relevant	Precisie	Vangst
A1 * B1 * C1	ESA	144	80	0,56	0,58
A1 * B1 * C2	ESA	211	93	0,44	0,67
A1 * B1 * C3	ESA	215	93	0,43	0,67
A1 * B2 * C3	ESA	259	120	0,46	0,86
A2 * B1 * C1	ESA	94	62	0,66	0,45
A2 * B1 * C2	ESA	125	75	0,60	0,54
A2 * B1 * C3	ESA	129	75	0,58	0,54
A3 * B1 * C1	ESA	58	41	0,71	0,30
A3 * B1 * C2	ESA	72	48	0,67	0,35
A3 * B1 * C3	ESA	76	48	0,63	0,35
A3 * B2 * C1	ESA	67	48	0,72	0,35
A1 * B1 * C4	DIALOG	183	71	0,39	0,70
A2 * B1 * C4	DIALOG	111	61	0,55	0,60
A1 * B1 * C1	ESA/77	124	66	0,53	0,65
A1 * B1 * C3	ESA/77	185	76	0,41	0,75
A2 * B1 * C1	ESA/77	79	51	0,65	0,50
A2 * B1 * C3	ESA/77	105	61	0,58	0,60

4.2 Niet relevante publicaties

Van de gevonden niet-relevante literatuur is ook bepaald in welk opzicht zij niet aan de vraag voldeed. Daarbij valt een aantal categorieën te onderscheiden:

1. Het gaat niet over het inwendige van kristallijn silicium.
2. Het gaat over vacature clusters.
3. Het gaat in feite over experimenteel werk.
4. Het gaat om eigenschappen of verschijnselen waarin we niet geïnteresseerd zijn, hoewel dat niet expliciet in de zoekstrategie tot uitdrukking komt.

Van enkele van de eerdere combinaties is gekeken welke categorieën van toepassing zijn. Die gegevens zijn in Tabel III vermeld, respectievelijk onder de hoofdes "stof", "cluster", "exper" en "eigenschap". Daarbij kunnen soms meer redenen voor niet-relevantie tegelijk van toepassing zijn, zodat optellingen meer dan 100% kunnen opleveren.

TABEL III

Verdeling van gevonden niet-relevante publicaties over vier globale redenen voor niet-relevantie. (Zie ook tekst).

Combinatie	Host	Niet-relev.	Stof	Cluster	Exper	Eigen- schap
A1*B1*C4	DIALOG	112	12%	36%	11%	67%
A2*B1*C4	DIALOG	50	12%	38%	6%	60%
A1*B1*C1	ESA/77	58	26%	50%	19%	33%
A1*B1*C3	ESA/77	109	16%	53%	47%	27%
A2*B1*C1	ESA/77	28	28%	50%	11%	32%
A2*B1*C3	ESA/77	44	20%	39%	30%	34%
A1*B2*C3	ESA	139	15%	52%	45%	21%
A3*B1*C1	ESA	17	29%	76%	6%	18%
A3*B1*C3	ESA	28	25%	68%	32%	25%
A3*B2*C1	ESA	19	32%	68%	5%	21%

4.3 Niet gevonden publicaties

Zelfs in de meest uitgebreide zoekactie werden 19 documenten uit de eigen verzameling niet gevonden. Hiervan bleken er 14 wel in INSPEC aanwezig te zijn, de andere 5 niet. Van de vijf niet aanwezige bleken er vier uit eenzelfde conferentieverlag afkomstig te zijn; verslagen van andere jaren van die zelfde conferentie bleken wel opgenomen. De vijfde was een corrigendum op een eerder wel opgenomen artikel, op zich dus ook voldoende belangrijke informatie.

Redenen voor het niet vinden van de 14 wel aanwezige waren globaal:

- (1) het artikel was algemener onder "defecten" geïndexeerd, omdat het daar in het algemeen over ging, zij het dat ook vacatures behandeld werden (7x);
- (2) vacatures kwamen alleen als vrij trefwoord voor (4x);

- (3) Het artikel was algemener onder "semiconductors" geïndexeerd, hoewel het ook over silicium ging (7x);
- (4) er was helemaal geen stofje geïndexeerd, alleen het chemisch symbool Si als vrije term of een vergelijkbaar ander stofje, germanium (7x);

Daarnaast was eenmaal (in een review) geen zinnige term voor het theorie aspect gegeven, was eenmaal in de indexering en de titel helemaal geen sprake van defecten en was eenmaal van een frans artikel de titel onvolledig vertaald en ook onvolledige indexering toegekend.

Een speciale opmerking verdienen de acht review artikelen in de basis-collectie, waarvan er drie niet gevonden werden als gevolg van indexering met ruimere termen.

5. Conclusies

Al met de eenvoudigste bij ESA gemaakte combinatie A1*B1*C1 werden heel bevredigende resultaten voor precisie en vangst bereikt (respectievelijk 56% en 58%). Toch moest voor het meest generieke facet "theorie" al een hele lijst van 12 alternatieve of synonieme termen bedacht worden. Het is niet verwonderlijk dat het gebruik maken van een simpele "treatment code" voor dit facet, zoals bij DIALOG mogelijk is, niet alleen de zoekactie aanzienlijk vereenvoudigt, maar bovendien een veel grotere vangst van 70 % oplevert. Dat daar ook een lagere precisie tegenover staat is begrijpelijk, aangezien allerlei theoretische methoden die in dit verband niet interessant zijn, in de gespecificeerde ESA-lijst eenvoudig weggelaten zijn. Pas met het toevoegen van nog uitgebreider lijsten van 10 of 14 quasi-synoniemen in C2 en C3 kan bij ESA een vergelijkbaar hoge vangst (en lage precisie) bereikt worden. Daarbij leverde C3 trouwens geen extra relevante publicaties meer op. Opmerkelijk resultaat van een nauwkeuriger vergelijking met de DIALOG resultaten over dezelfde tijdsperiode is dat met de "treatment code" bij DIALOG toch nog vijf relevante publicaties gemist worden, die er bij ESA wel uitkomen. Bij ESA was de vangst dus zelfs 75%.

Kijken we naar de "ten onrechte" gevonden publicaties (Tabel III), dan blijkt dat met de treatment code (C4) vooral veel niet interessante eigenschappen gevonden worden, eigenschappen die bij een expliciete omschrijving als in C1 t/m C3 gemakkelijk vermeden konden worden. Doordat in omschrijving C3 ook te berekenen eigenschappen en grootheden waren inbegrepen, leverde dat, zoals te verwachten was, relatief veel experimentele studies op. Waarom met C4 relatief zo weinig over vacature-clusters gevonden werd, is niet zonder meer duidelijk.

We mogen uit het voorgaande concluderen dat het gebruik van een "treatment code" voor een generiek begrip een on-line zoekactie zeer vereenvoudigt. Hoewel het gebruik ervan nog geen maximale vangst garandeert, zal deze beperking in de praktijk meestal nauwelijks van belang zijn. Het is dus een duidelijk gemis dat bij ESA niet in dit veld gezocht kan worden.

Ondanks de te ruime betekenis van de thesaurusterm VACANCIES leverde het gebruik ervan een acceptabele precisie op. Alleen in de combinatie bij DIALOG kwam de precisie onder de 40%. Pogingen om de zoekactie expliciet tot enkele vacatures in te perken met omschrijvingen A2 en A3, leverden wat de precisie betreft het gewenste resultaat op. In de meeste gevallen daalde de vangst daardoor echter beneden een aanvaardbaar niveau. Bij de gevonden niet-relevante publicaties (Tabel III) is het opmerkelijk dat met A2 en A3 de relatieve bijdrage van vacature-clusters toeneemt, terwijl die daarin juist uitgesloten werden. In absolute zin neemt hun aantal echter af.

Ook in dit geval mogen wij concluderen dat de bij indexering gegeven term VACANCIES, hoewel in feite te ruim, op eenvoudige wijze een uitputtende zoekactie mogelijk maakt. In dit geval werd zelfs de precisie niet zeer nadelig beïnvloed.

Uitbreiding van het stof-facet silicium tot de hele groep halfgeleiders met vergelijkbare kristalstructuur (omschrijving B2) verhoogde de vangst nog aanmerkelijk. Opmerkelijk is dat zelfs de precisie er iets beter van werd; kennelijk werden bijna uitsluitend relevante publicaties toegevoegd. Bij de niet gevonden (maar wel aanwezige) literatuur bleek in bijna alle gevallen dit facet problemen op te leveren. Een uitbreiding tot de hele groep van alle halfgeleiders zou de precisie te sterk omlaag brengen.

Een laatste opmerking betreft de indexering van overzichtsartikelen. In de praktijk blijkt dat de verschillende aspecten die daarin aan de orde komen bij INSPEC niet tot op het meest specifieke niveau geïndexeerd worden, maar alleen met ruimere termen worden aangegeven. Omdat gebruik van die ruimere zoekterm meestal te veel ruis zou opleveren, vormt dit een onnodige beperking op het terugvinden van dergelijke artikelen voor specifieke vragen.

- /1/ M. Sprenger, Magnetic resonance studies on defects in silicon,
Proefschrift, Universiteit van Amsterdam, 1986.
- /2/ M. Sprenger, S.H. Muller, E.G. Sieverts and C.A.J. Ammerlaan,
Hyperfine interactions of the negatively charged vacancy in silicon,
Physical Review B (1986, in druk).