

De finish van Moores marathon

Aan de schaling van transistoren komt eens een eind. Wat dan? Ramses Valvekens en Bart Dierickx beschouwen de toekomst van Moore. Dit artikel is het eerste deel van een tweeluik. In Bits&Chips 11 het vervolg, over arraygebaseerde chips en hun relatie met doorgedreven schaling.

Ramses Valvekens
Bart Dierickx

Consumenten zijn gewend geraakt aan steeds meer features voor dezelfde prijs, of dezelfde functionaliteit voor een lagere prijs. Daarbij verwachten ze een even groot of zelfs lager energieverbruik. Om hieraan te kunnen voldoen, is al decennialang een exponentiële schaling van chips in afmetingen en prestaties gaande. De wet van Moore beschrijft de wetmatigheid hoe steeds meer transistoren in eenzelfde chip kunnen worden geperst: vandaag tot meer dan een miljard per vierkante centimeter. Essentieel hiervoor tijdens de beginjaren waren de keuze voor silicium met zijn stabiele en makkelijk te groeien oxidelagen en de uitvinding van het planaire proces. De laatste jaren zijn de spectaculaire vooruitgang van de optische microlithografie en de introductie van nieuwe materialen de grootste drijfveren.

Aan de schaling volgens Moore zal eens een einde komen. Siliciumatomen zitten op een halve nanometer van elkaar en de kleinste lineaire afmetingen in de huidige chips zijn ongeveer twintig nanometer. Over enkele oppervlaktehalveringen zitten we op de absolute limiet van het atoom. Bij het huidige tempo van een halvering eens in de twee à drie jaar zijn we daar over twintig à dertig jaar. Lang daarvoor zal de schaling echter al stoppen door fysische en chemische grenzen, zoals de microscopische variabiliteit van de gedoteerde structuren, laagdiktes en de grootte van fotolakmoleculen. Hierdoor laat de *line-edge roughness* zich onderweg meer en meer voelen, tot de praktische schalingslimiet is bereikt.

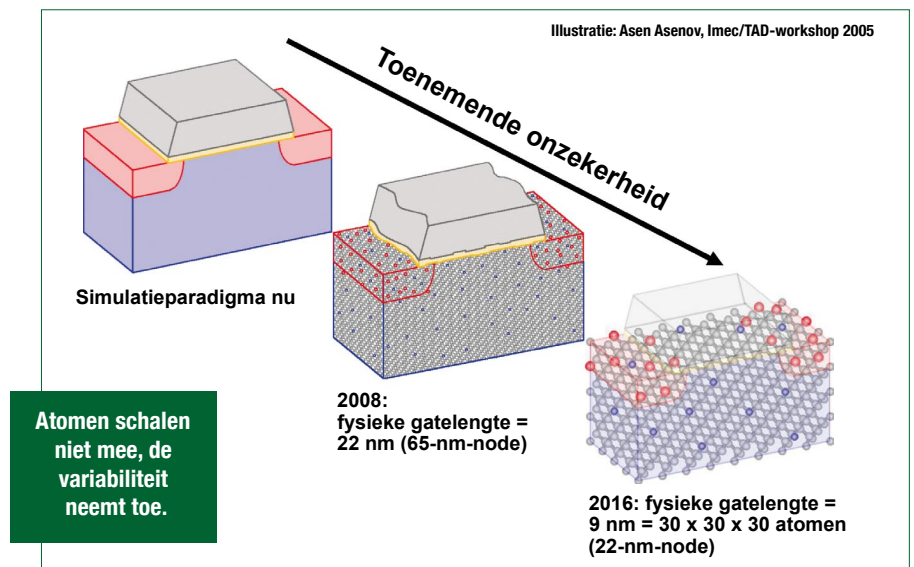
Nog veel eerder zullen de benodigde investeringen in de ontwikkeling van de 'ultieme' volgende generatie siliciumtechnologie en in de bijhorende productieapparatuur zodanig oplopen dat zelfs alle consumenten samen ze niet meer kunnen dragen met hun elektronica-aankopen. Volumes per chip kunnen dan slechts omhoog door minder verschillende chips uit te brengen: *one size fits all*.

Dit schreeuwt om nog meer programmeerbaarheid en betere softwareontwikkeltools. Deze trend is ingezet met systeemchips die volledig op software draaien. Denk aan de chips die we terugvinden in smartphones en tablets, aan GPU's in grafische kaarten en aan high-end FPGA's die onderdak vinden in allerhande embedded systemen.

Mogelijk bereikt de schalingstrein zijn eindstation zeer binnenkort. Kleine transistoren zijn gevoelige beestjes. Een groeiend aantal wordt doodgeboren (defect bij productie). En de overlevers worden onbetrouwbaar en lekken als een mandje. Ze vertonen meer en meer (ongewenste) onderlinge verschillen, hebben last van kwaaltjes

foutcorrigerende codering (*error correcting code*, ECC) in DRam, SRam en flash en aan *wear leveling* bij flash. Verder kreunen chips onder de massale populatie transistoren. Ze krijgen het warm met zovelen samen.

De ruimtevaart is een voorbode van wat komen gaat wanneer de transistoren verder verkleinen: drievoudige redundantie met *majority voters* en stralingsharde flipflops die dubbel zo groot zijn als de gewone (door duplicatie en goed gekozen feedback). Transistoren in analoge circuits schalen al geruime tijd niet meer mee om de anders nefaste effecten van mismatch en ruis te vermijden. Ook hogespanningstoepassingen schalen niet mee. Bij geheugens



en verslijten. Steeds vaker laten ze steken vallen door omgevingsinvloeden (warmte of verdwaalde ioniserende deeltjes uit de atmosfeer) en vermoeidheid. Daarom zien we meer en meer redundantie verschijnen op chips om permanente productiedefecten en transiënte fouten op te vangen en te corrigeren of minstens in kaart te brengen. Denk aan redundante rijen en kolommen en

digitale schakelingen houdt het verhaal op wanneer de reserveonderdelen en andere overhead evenveel plaats innemen als de functionele stukken.

Hoge prijs

Intussen staat de wet van Moore onder hoogspanning omdat de kleinst af te beelden structuren vele malen kleiner zijn dan de

golfengte van het licht in de lithomachines. Die gebruiken al tien jaar hetzelfde UV-licht van 193 nanometer uit een ArF-eximeerlaser, maar combineren dat met steeds geavanceerdere optica, die zich allang niet meer louter gedraagt zoals in de basisnatuurkundehandboeken. UV-licht van kortere afmeting is mogelijk maar vraagt het onvoorstelbare, namelijk lithografie in hoog vacuüm.

Om de resolutie te kunnen verhogen bij gelijkblijvende golfengte en de optische diffractie en verstrooiing die daarbij optreden het hoofd te kunnen bieden, deden immersielithografie ('193I') en hogere-apertuur-

Maar de prijs is hoog. EUV wordt niet alleen geabsorbeerd door de atmosfeer, zodat de hele litho in vacuüm moet, maar ook door glas (en al zijn alternatieven), waardoor de ganse optiek reflectief moet. Dit betekent spiegels en reflectieve maskers (*reticles*) op basis van interferentiefilters, in plaats van de lenzen en doorschijnende maskers van de klassieke transmissieve optica.

E-beam is een alternatief dat geen maskers behoeft. Een elektronenbundel scant hierbij het ganse oppervlak af, wat erg tijdrovend is. Gebruik van meerdere parallelle bundels verhoogt de doorvoer, maar blijft

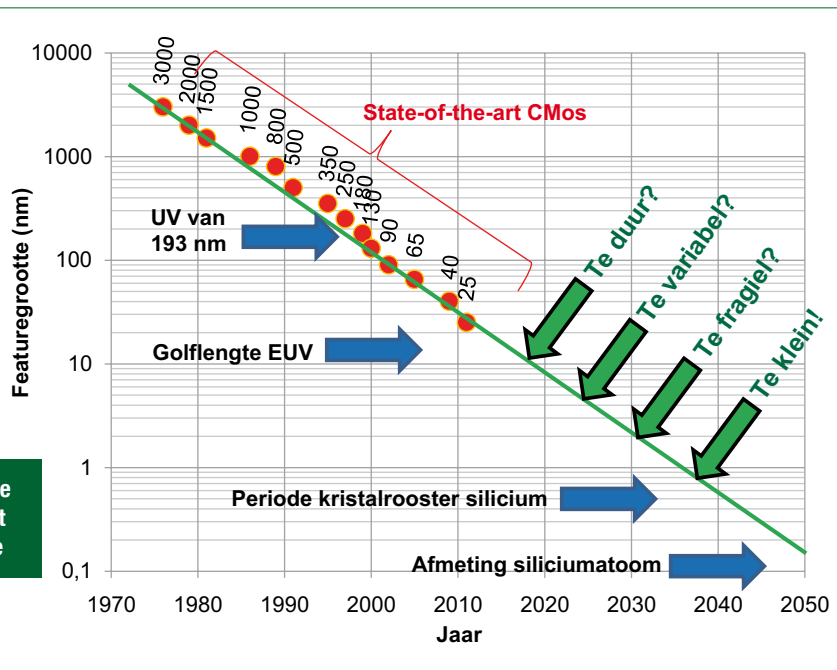
trap in GaAs of aan passieve componenten zoals spoelen, transmissielijnen en antennes. Zo ontstaat een *system-in-package*. Met *through-silicon vias* (TSVs) en *interposers* is driedimensionale stapeling van dies mogelijk. Dit vinden we nu al terug in NAND-flashgeheugenkaarten voor foto toestellen, in sommige DRAM-chips en in enkele grote FPGAs. Tweedimensionale die-*stitching* omzeilt de oppervlaktebeperking opgelegd door de maximale lithografische reticlegrootte door meerdere dies te combineren tot een grotere 'chip'. Meerwaardige spanningsniveaus vergroten bij flash de opslagcapaciteit per cel van één bit bij SLC (*single-level cell*) naar twee, drie en vier bits bij respectievelijk MLC (*multi-level cell*), 8LC en 16LC. MLC en 8LC zijn heel courant; 16LC blijkt te lastig.

Een alternatief is niet schalen, enigszins eufemistisch 'More than Moore' gedoopt. Hieronder vallen alle creatieve devices die we kunnen maken in oude fabrieken, al dan niet na aanpassingen. Micro-elektromechanische systemen (Mems) zijn een mooi voorbeeld. Denk hierbij aan microspiegelarrays in beamers en gsm-microfoons. *Silicon photonics* is een andere opkomende meer-dan-Moore-technologie. De eerste toepassingen situeren zich in schakelapparatuur voor glasvezelnetwerken. Daarnaast is er organische en plastic elektronica, bijvoorbeeld voor RFID-tagging. Dergelijke schakelingen zijn in de toekomst wellicht goedkoop te drukken en hebben dus geen uitstaans met de klassieke lithogebaseerde productie. Wel zijn ze inherent minder performant dan silicium door de dramatisch lagere mobiliteit van de ladingstraggers.

Een mogelijk vervolg met (sub)nanometerstructuren is geen gewone verdere schaling. In hybride CMOS-/nanoarchitecturen worden grote aantallen kriskras uitgespreide koolstof nanobuisjes, nanoraden en consorten na een rudimentaire uitlijning aangesloten op een silicium basisarray die zelf niet verder schaal. Dit kan een eerste aanzet zijn tot zelfassemblage van circuits, à la DNA. Maar net als bij moleculaire schakelaars, kwantum- en spinelektronica is het pad naar echte applicaties en kostenefficiënte massaproductie en assemblage nog lang; een volwaardige opvolger voor lithografie en het planaire proces heeft zich nog niet aangediend.

Ramses Valvekens (ramses@easics.be) is CEO van Easics uit Leuven, dat hoogperformante en betrouwbare digitale systemen ontwerpt op basis van Asics en FPGA's. Bart Dierickx (bart.dierickx@caeleste.be) is CTO van Caeleste uit Antwerpen, dat geavanceerde Asic-beeldsensoren ontwerpt voor applicaties in de ruimtevaart, de medische wereld en nichemarkten.

Redactie Nieke Roos



lenzen hun intrede, samen met *resolution enhancement technologies* (RET) als *optical proximity correction*, *off-axis illumination*, *phase-shifting masks*, en recentelijk *double* of zelfs *multiple patterning*. Daarbij hebben chipontwerpers *restricted design rules* gekregen. Ook zijn er gesofistikeerde fotolakken gekomen, die door chemische *thresholding* het contrast verhogen van het beeld op de wafer.

Nu de 193I-plus-RET-hoed bijna is leeggetoverd, is er echter geen ontkomen meer aan: we verkorten de golfengte van het licht meer dan tien maal. Litho met extreem ultraviolet (EUV) kan de huidige resolutiepijnen spectaculair verlichten: 13,5-nanometer-fotonen zijn beter geschikt om de toekomstige kleine structuren te belichten. Hierdoor hoeft het RET-huzarenstuk niet op de spits gedreven, althans niet op korte termijn. Schalingskampioenen als ASML en hun klanten kijken reikhalzend uit naar een *hoogvermogen-EUV-bron*, die de fotolak op de wafers voldoende snel kan belichten. Zo kunnen ze het productieproces enigszins betaalbaar houden, tenminste voor massaproductie.

trager dan optische litho waarbij de hele chip ineens wordt belicht.

Geen volwaardige opvolger

Er zijn ook niet-lithogereleerde innovaties die essentieel zijn om de schaling op gang te houden. Zo zijn er enerzijds steeds betere lage-k-diëlektrica en koperen verbindingen ter vervanging van siliciumoxide diëlektrica en aluminium verbindingen in de metalen interconnectielagen, en anderzijds hoge-k-gatediëlektrica die toelaten om nog dunnere en toch efficiënte gateoxides te maken en metalen gates ter vervanging van siliciumoxide gatediëlektrica en polykristallijn silicium gates in de transistoren. Verder zijn er nieuwe technieken die de mobiliteit verhogen, zoals *strained silicon*, en nieuwe devices die de winst elders zoeken, zoals de niet-planaire Finfet-transistoren.

Een andere manier om Moore overeind te houden, naast het proppen van meer kleinere transistoren op één enkele *die*, is de functionaliteit per verpakte chip vergroten. Dat kan bijvoorbeeld door een digitale die in één verpakking te koppelen aan een RF-vermogen-