

Halbleitertechnologie von A bis Z

Lithografie

Maskentechnik	3
<i>Maskentechnik</i>	3
<i>Maskenherstellung</i>	4
<i>Maskentypen</i>	4
<i>Belichtungsverfahren der nächsten Generation</i>	10

Maskentechnik

Maskentechnik

Die in der Fototechnik eingesetzten Masken enthalten ein Muster mit dem die jeweilige Schicht auf dem Wafer strukturiert wird. Ausgangsmaterial für die Masken sind Glasplatten, die ganzflächig mit Chrom und Lack beschichtet sind (Blanks). Über den für Elektronenstrahlen empfindlichen Lack wird die Chromschicht strukturiert, welche dann die lichtundurchlässigen Bereiche auf der Glasmaske darstellt.

Mit einem Elektronenstrahl werden die Masken direkt beschrieben. Die gesamte Apparatur – die Elektronenstrahlquelle, Fokussier- und Ablenkeinheit und der Blank – befindet sich im Hochvakuum (0,01–100 Pa; normaler Luftdruck ca. 100.000 Pa). Der Elektronenstrahl wird computergesteuert über die Maske gelenkt und belichtet den Lack. Bei dieser Methode lassen sich Strukturen bis weit unter 100 nm auflösen.

Ein einzelner Strahl schreibt eine hochaufgelöste Maske allerdings nicht in vertretbarer Zeit. Heutige Maskenschreiber arbeiten deshalb mit mehreren hunderttausend parallel geführten Einzelstrahlen (multi beam mask writer), die gemeinsam über den Blank geführt und einzeln ein- und ausgeschaltet werden. Erst damit sind Schreibzeiten von einigen Stunden pro Maske erreichbar – und erst damit lassen sich beliebig geformte, gekrümmte Strukturen genauso schnell schreiben wie rechtwinklige.

Korrektur der Abbildungsfehler

Auf Grund wellenoptischer Effekte (z.B. Beugung) kann es bei der Belichtung der Wafer zu Abbildungsfehlern kommen, welche durch die sogenannte *optical proximity correction* (OPC, etwa: optische Nahbereichskorrektur) korrigiert bzw. verringert werden.

Dazu können die eigentlichen Strukturen so abgeändert werden, dass die Abbildung auf dem Wafer dem gewünschten Muster entspricht. Außerdem können zusätzliche Hilfsstrukturen auf die Maske geschrieben werden welche nur dazu dienen, Abbildungsfehler zu minimieren, jedoch keine Funktion für die Schaltung haben.

Bei den kleinsten Strukturen genügt es nicht mehr, die Maske allein zu korrigieren. Optimiert werden Maske und Beleuchtung gemeinsam (source mask optimization, SMO): Auch die Winkelverteilung des einfallenden Lichts wird auf das jeweilige Muster zugeschnitten. Die weitestgehende Variante ist

die inverse Lithografie (inverse lithography technology, ILT). Dabei wird nicht mehr eine gezeichnete Maske korrigiert, sondern aus dem gewünschten Ergebnis auf dem Wafer rückwärts diejenige Maskenstruktur berechnet, die dieses Ergebnis erzeugt. Heraus kommen frei geformte, oft gekrümmte Konturen, die mit einer gezeichneten Vorlage keine Ähnlichkeit mehr haben. Die Rechenzeit dafür übersteigt inzwischen die Schreibzeit der Maske deutlich und wird auf großen Rechnerclustern erbracht.

Weil eine einzige Maske viele tausend Wafer belichtet, überträgt sich jeder Fehler auf ihr auf jeden Chip. Masken werden deshalb nach dem Schreiben mit hochauflösenden Verfahren geprüft und einzelne Defekte gezielt repariert – etwa durch Abtragen überschüssigen Absorbermaterials mit einem fokussierten Elektronen- oder Ionenstrahl.

Maskenherstellung

Grundsätzlich werden die Strukturen auf den Masken auf die gleiche Weise hergestellt, wie auf den Wafern. Im Gegensatz zum Belichtungsprozess in der Waferfertigung, bei dem die Strukturen der Maske durch Schattenwurf auf dem Lack abgebildet werden, werden die Masken aber mit einem Elektronenstrahl direkt beschrieben.

1. Belichten eines fotoempfindlichen Lacks zur Strukturierung einer Chromschicht auf dem Glassubstrat mittels Laser oder Elektronenstrahl



2. Entwickeln der Lackschicht



3. Ätzen der Chromschicht



4. Lackentfernen



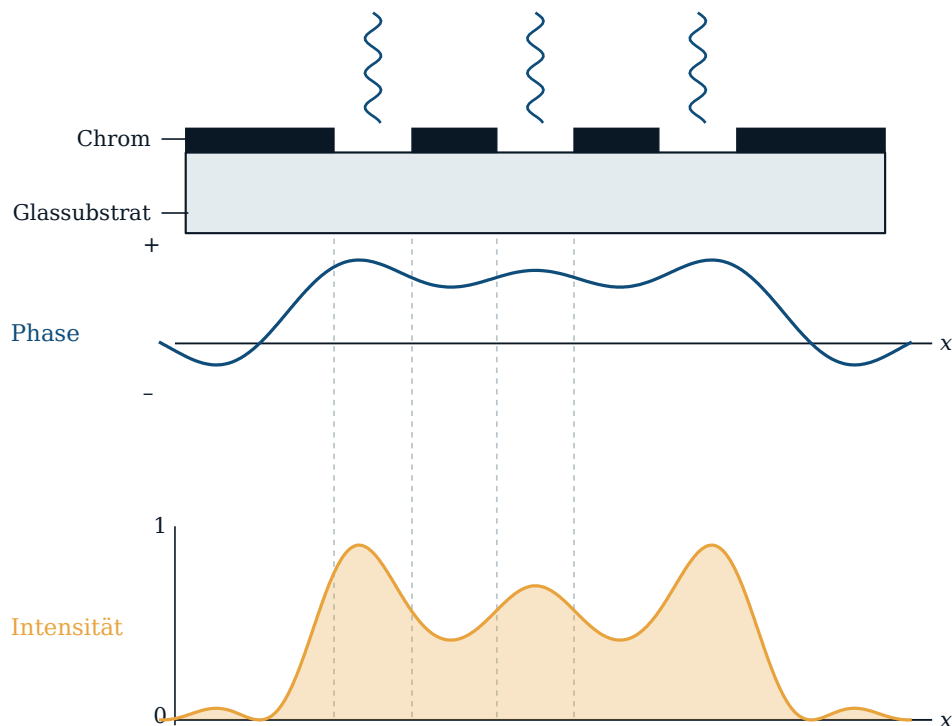
5. Anbringen des Pellicles



Maskentypen

Neben der klassischen Chrommaske (COG, Chrome On Glass) gibt es noch weitere Maskentypen, die eine verbesserte Strukturauflösung ermöglichen. Hauptproblem der COG-Maske ist die Beugung, die das Licht an den Strukturkanten erfährt. Dadurch fällt das Licht nicht nur senkrecht auf den Wafer, sondern wird auch in den Schattenbereich abgelenkt, wo der Fotolack nicht belichtet werden soll.

Intensitätsprofil einer Chrom On Glass-Mask



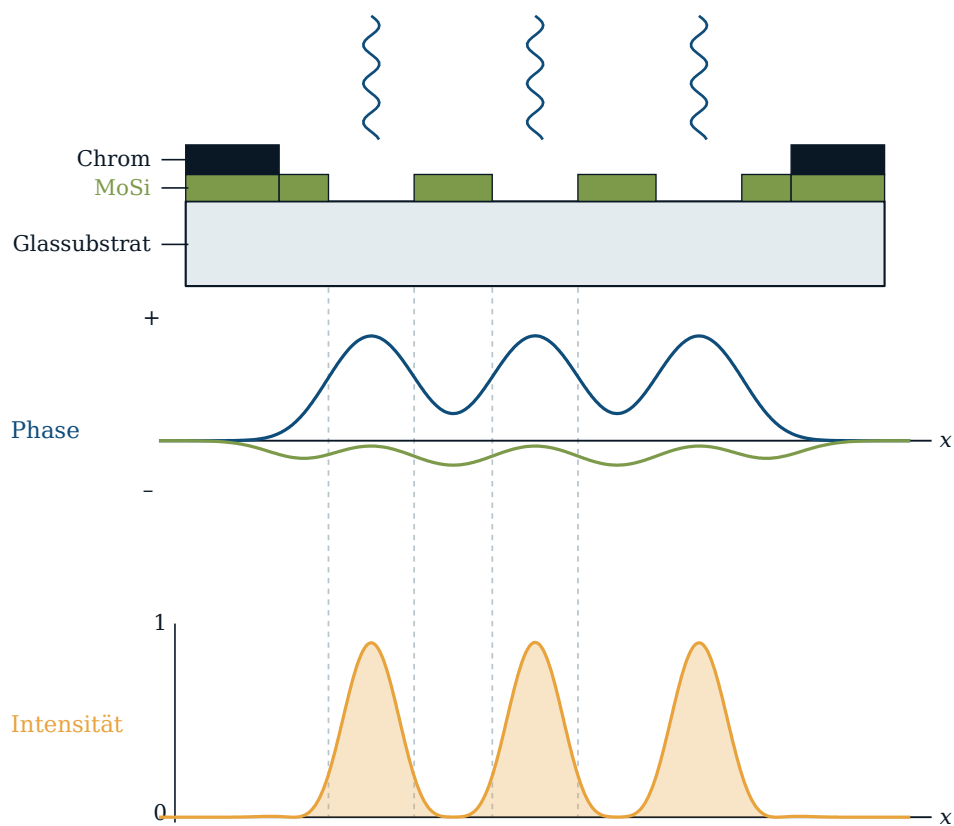
Mit verschiedenen Maßnahmen versucht man nun, die Intensität des gebeugten Lichts zu vermindern. Diese werden im Folgenden anhand der unterschiedlichen Maskentypen näher beschrieben.

Attenuated Phase Shift Mask (AttPSM)

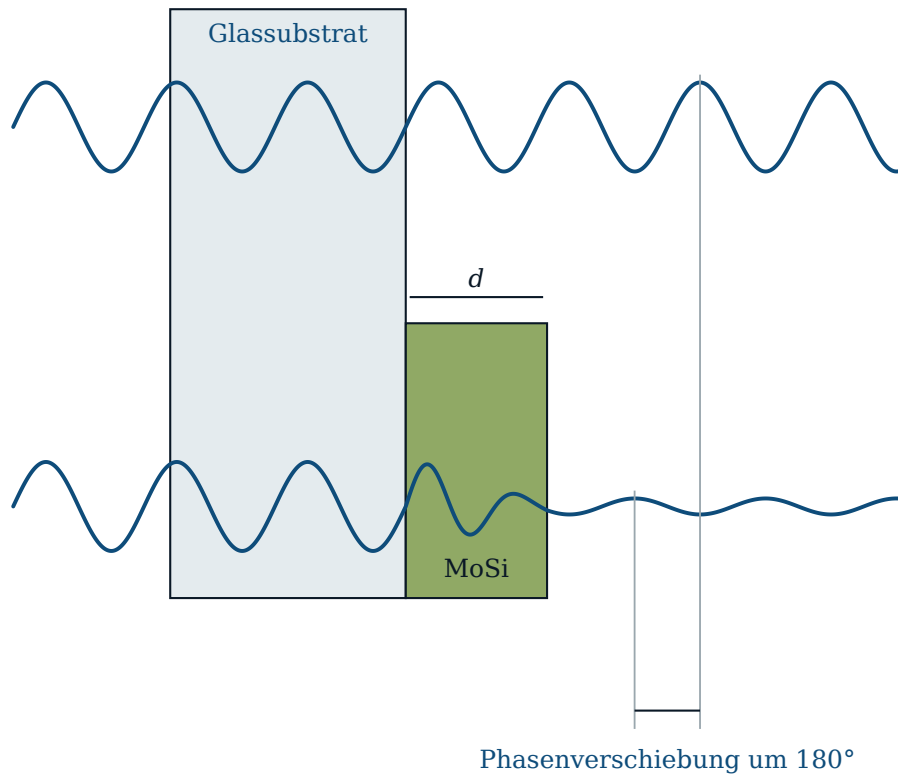
Bei der so genannten Halbtonmaske oder weichen Phasenmaske bildet eine Schicht aus Molybdänsilicid (MoSi) den strukturgebenden Teil, eine Chromschicht gibt es hier nicht. Die Dicke der MoSi-Schicht ist so gewählt, dass das Licht beim Durchgang eine Phasenverschiebung um 180° erfährt gegenüber dem Licht, das lediglich Glas durchläuft. Dies geschieht durch die

unterschiedliche Lichtgeschwindigkeit in Luft und in MoSi. Gleichzeitig ist die Schicht je nach Molybdänanteil im Silicium zu 6 oder 18 % lichtdurchlässig (bei einer Belichtungswellenlänge von 193 nm), das Licht wird also abgeschwächt (attenuate, engl.: vermindern). Die gegenläufigen Lichtwellen löschen sich so unterhalb der MoSi-Strukturen nahezu aus, somit wird der Kontrast zwischen Hell und Dunkel erhöht. Zusätzlich kann in Bereichen, die nicht zur Belichtung benötigt werden, Chrom aufgebracht werden, um Licht vollständig auszublenden. Diese Masken werden als Tritonemaske bezeichnet.

Intensitätsprofil einer Attenuated Phase Shift Mask



Prinzip der Phasenschiebung mittels Molybdänsilicid



Chromfreie Phasenschiebermaske

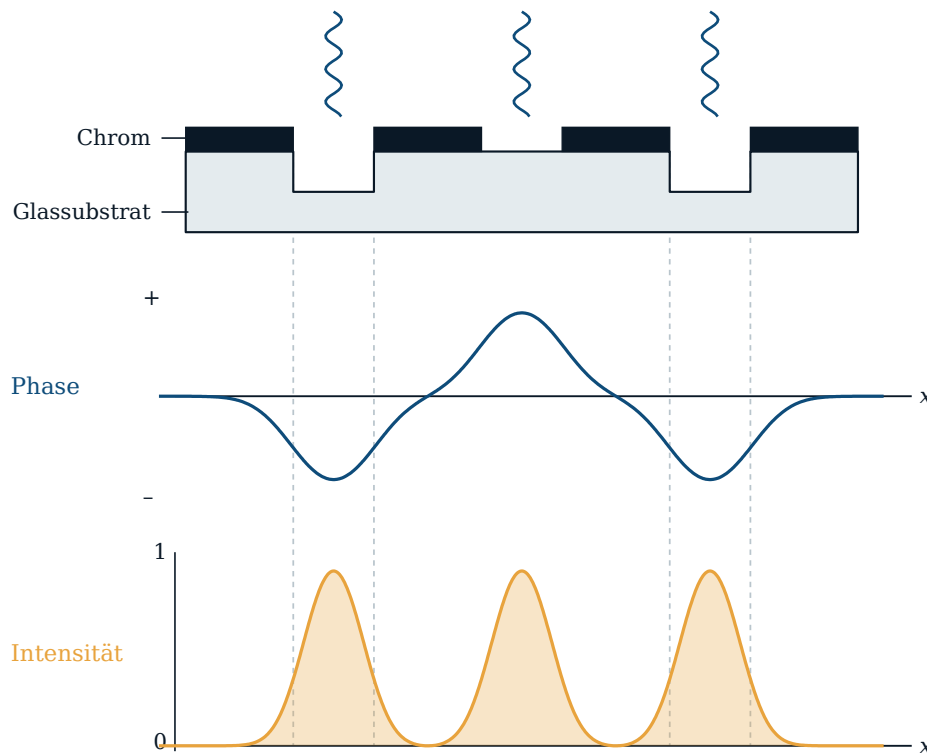
Die chromfreien Masken besitzen keine strukturgebende Beschichtung. Die Phasenverschiebung wird durch Gräben erzeugt, die direkt in die Glasplatte geätzt sind. Die Herstellung dieser Masken gestaltet sich schwierig, da der Ätzvorgang mitten im Glas gestoppt werden muss. Im Gegensatz zu Ätzprozessen, bei denen eine Schicht vollständig durchgeätzt wird und Änderungen im Plasma Auskunft darüber geben, wann die darunterliegende Schicht freigelegt ist, erhält man hier keine Information, wann die benötigte Tiefe erreicht ist.

Zusätzlich ergeben sich Probleme bei der Herstellung großer Strukturen. Dadurch, dass es keine abschattenden Bereiche gibt und das Licht überall mit derselben Intensität auf die Maske trifft, erfolgt nur an den Strukturkanten die gewünschte Interferenz und die daraus resultierende Abschwächung des Lichts. In der Mitte der Bereiche findet jedoch keine oder nur noch eine geringe Abschwächung statt. Um eine Belichtung dieser Bereiche zu verhindern, muss die Lichtintensität also von vorn herein reduziert werden, wodurch die Gefahr einer Unterbelichtung aller Strukturen entsteht.

Alternating Phase Shift Mask (AltPSM)

Bei der alternierenden Phasenmaske werden wie bei der chromfreien Maske Gräben direkt in das Glassubstrat geätzt, jedoch abwechselnd (alternierend) mit ungeätzten Bereichen. Zusätzlich werden Stellen mit Chrom beschichtet um die Lichtintensität an entsprechenden Stellen herabzusetzen.

Intensitätsprofil einer Alternating Phase Shift Mask



Dadurch ergeben sich jedoch Bereiche mit undefinierter Phasenverschiebung, so dass bei diesem Maskentyp, der eine sehr hohe Auflösung ermöglicht, grundsätzlich zweimal belichtet werden muss. Die erste Maske enthält dabei die Strukturen die in x-Richtung verlaufen, die zweite Maske die Struktur in y-Richtung.

Zielstruktur auf dem Wafer und entsprechende Maskenstruktur

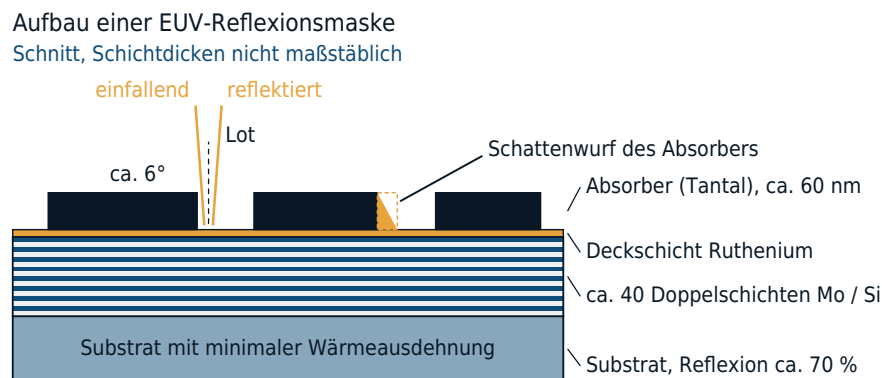


Reflexionsmasken für die EUV-Belichtung

Alle bisher beschriebenen Maskentypen sind Transparente: Licht tritt durch das Glassubstrat hindurch. Für EUV mit 13,5 nm funktioniert das nicht, weil jedes Material diese Strahlung absorbiert. Die EUV-Maske ist deshalb ein Spiegel.

Als Substrat dient ein Glas mit nahezu verschwindender Wärmeausdehnung, da sich die Maske unter der Bestrahlung sonst verzieht. Darauf liegen etwa vierzig Doppelschichten aus Molybdän und Silicium, deren Dicken so gewählt sind, dass sich die an den einzelnen Grenzflächen reflektierten Anteile konstruktiv überlagern. Selbst dieser Vielschichtspiegel reflektiert nur rund 70 % der auftreffenden Strahlung; da im Strahlengang mehrere solche Spiegel liegen, erreicht nur ein kleiner Teil der erzeugten Leistung den Wafer. Eine dünne Deckschicht aus Ruthenium schützt den Stapel, darüber bildet ein Absorber, meist eine Tantalverbindung, das eigentliche Muster.

Aufbau einer EUV-Reflexionsmaske



Aus dem Spiegelprinzip folgt eine Eigenheit, die es bei Transparentmasken nicht gibt: Der Strahl muss schräg einfallen, damit einfallendes und reflektiertes Licht getrennt werden können. Bei diesem Einfall von einigen Grad wirft der etwa 60 nm dicke Absorber einen Schatten, und zwar je nach Orientierung der Struktur einen unterschiedlichen. Diese Maskeneffekte müssen bei der Korrektur der Maske mitgerechnet werden.

Auch der Schutz vor Partikeln ist schwieriger. Ein **Pellicle** muss hier von der Strahlung zweimal durchlaufen werden und darf sie dabei kaum absorbieren; gleichzeitig muss die Membran der Wärmelast einer Hochleistungsquelle standhalten. Verwendet werden dafür extrem dünne Membranen, unter anderem auf Basis von Kohlenstoffnanoröhren.

Phasenschiebende EUV-Masken, die den Kontrast wie bei 193 nm zusätzlich erhöhen würden, sind Gegenstand der Entwicklung, in der Fertigung aber noch

nicht etabliert. Die Auflösung wird bei EUV daher vor allem über die numerische Apertur, die Beleuchtung und den Lack gewonnen, nicht über die Maskentechnik.

Belichtungsverfahren der nächsten Generation

Der Wechsel, den dieses Kapitel jahrelang als bevorstehend beschrieben hat, ist vollzogen: Die EUV-Lithografie mit 13,5 nm ist seit 2019 in der Serienfertigung und belichtet die kritischen Ebenen aller führenden Logik- und Speicherprozesse. Sie arbeitet, wie vorhergesagt, im Vakuum, bildet über Spiegel ab und nutzt Masken mit spiegelnder statt transparenter Oberfläche.

Gleichzeitig ist die 193-nm-Immersionsbelichtung nicht verschwunden. Sie belichtet weiterhin die Mehrzahl der Ebenen eines modernen Bauelements, weil sie schneller und je Belichtung deutlich günstiger ist. EUV wird nur dort eingesetzt, wo es sich rechnet: Wo eine EUV-Belichtung zwei oder mehr Immersionsschritte samt Zwischenprozessen ersetzt, ist sie im Vorteil.

High-NA-EUV

Die aktuelle Ausbaustufe erhöht die numerische Apertur von 0,33 auf 0,55 und erreicht damit rund 8 nm Halbperiode in einer Belichtung. Erkauft wird das mit einer anamorphen Optik, die in den beiden Achsen unterschiedlich stark verkleinert und damit nur noch das halbe Feld belichtet, sowie mit erheblichem Aufwand: Eine Anlage kostet etwa das Doppelte einer herkömmlichen EUV-Anlage und braucht am Standort Monate bis zur Qualifikation. Erste Ebenen eines Serienprodukts wurden im Juli 2026 freigegeben, der breite Einsatz ist für 2027 und 2028 zu erwarten. Nicht alle Hersteller gehen diesen Weg – teils wird die 0,33er-Generation mit Mehrfachbelichtung als wirtschaftlicher eingeschätzt.

Wo die Grenze inzwischen liegt

Die optische Auflösung ist nicht mehr der eigentliche Engpass. Begrenzend wirken:

- Statistik: die stochastischen Defekte durch die geringe Photonenzahl pro Struktur (siehe [Fotolack](#))
- Überlagerung: der zulässige Fehler bei der Justierung mehrerer Ebenen zueinander liegt im Bereich weniger Nanometer und wird bei geteilten Feldern und Mehrfachbelichtung weiter aufgezehrt
- Lichtleistung: der Durchsatz einer EUV-Anlage hängt direkt an der Leistung der Zinn-Plasmaquelle und an der nötigen Belichtungs-dosis
- Kosten: Anlagen und Masken sind so teuer, dass sich feinste Strukturen nur bei sehr hohen Stückzahlen rechnen

Alternative Ansätze

Neben der weiteren Steigerung der Apertur werden Verfahren verfolgt, die ohne abbildende Optik arbeiten:

Nanoimprint

Eine Vorlage mit dem Negativ der Struktur wird in eine flüssige Lackschicht gepresst, die anschließend aushärtet. Ohne Beugungsgrenze und ohne teure Optik, dafür mit den Problemen jedes Kontaktverfahrens: Defekte und Verschleiß der Vorlage. Im Einsatz für Speicherbauelemente und optische Komponenten.

Direktschreiben mit Elektronen

Vielstrahlssysteme schreiben die Struktur maskenlos direkt auf den Wafer. Für Prototypen, Kleinserien und kundenspezifische Bausteine attraktiv, für die Massenfertigung nach wie vor zu langsam.

Gerichtete Selbstorganisation

Blockcopolymere ordnen sich innerhalb einer lithografisch grob vorgegebenen Führungsstruktur selbst zu regelmäßigen feinen Mustern. Geeignet für streng periodische Strukturen wie Kontaktlochraster, nicht für beliebige Layouts.

Keiner dieser Ansätze ersetzt die optische Projektion in der Logikfertigung. Die Entwicklung der vergangenen zwanzig Jahre hat vielmehr gezeigt, dass ein etabliertes Belichtungsverfahren durch Maskentechnik, Beleuchtung, Rechenkorrektur und Mehrfachbelichtung sehr viel weiter getragen werden kann, als es die Wellenlänge zunächst erwarten lässt.