Epitaxial growth of highly mismatched III-V materials on (001) silicon for electronics and optoelectronics

Qiang Li\textsuperscript{a,b,*}, Kei May Lau\textsuperscript{a,b}

\textsuperscript{a} Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
\textsuperscript{b} Institute for Advanced Study, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

\textbf{A B S T R A C T}

Monolithic integration of III-V on silicon has been a scientifically appealing concept for decades. Notable progress has recently been made in this research area, fueled by significant interests of the electronics industry in high-mobility channel transistors and the booming development of silicon photonics technology. In this review article, we outline the fundamental roadblocks for the epitaxial growth of highly mismatched III-V materials, including arsenides, phosphides, and antimonides, on (001) oriented silicon substrates. Advances in hetero-epitaxy and selective-area hetero-epitaxy from micro to nano length scales are discussed. Opportunities in emerging electronics and integrated photonics are also presented.

1. Introduction

Mismatched hetero-epitaxy of semiconducting materials allows for tailoring of heterojunctions, strain manipulation, and device integration for a wide range of applications. One exemplary topic of study is the growth of cubic III-V semiconductors on lattice-mismatched (001) Si substrates. Si is the backbone material in digital integrated circuits for logics and signal processing. It possesses some attractive properties, like larger wafer size, better thermal conductivity, low-cost and mature manufacturing, and natural-abundance. III-V semiconductors, by contrast, have their niches in applications from optoelectronics to high-frequency and high-speed devices, for their superior electron mobilities and direct bandgap properties. The concept of monolithic III-V/Si integration began as a simple notion, that the best features and utility of III-V materials and devices could be combined with the benefits of Si manufacturing \cite{1}. Despite the tremendous effort put into this field, progress had been slow until this topic regained momentum a few years ago, driven by the development of emerging transistors \cite{2} and Si photonics technology \cite{3}.

The purpose of this review article is to update recent advances in the epitaxial growth of large lattice-mismatched III-V materials on (001) Si substrates. Nearly lattice-matched hetero-epitaxy (e.g., GaP on Si \cite{4–6}) or pseudomorphic growth \cite{7,8} is important but will not be discussed here. We focus on the growth performed on (001) oriented Si, rather than those on (111) Si (e.g., the vapor-liquid-solid growth method), for compatibility with the mainstream complementary-metal-oxide-semiconductor (CMOS) technology. Both wafer-level hetero-epitaxial thin film growth and selective-area hetero-epitaxy will be presented. Their respective device applications are discussed in detail.

2. Fundamental challenges in III-V hetero-epitaxy on (001) silicon

Growing fully relaxed, lattice-mismatched materials on a foreign substrate is often referred to as “metamorphic growth” \cite{9}. There are several fundamental challenges that limit the quality of metamorphic III-V layers on (001) Si, including a large mismatch in the lattice constants and thermal expansion coefficients, as well as the growth of polar materials on non-polar substrates. In III-V/Si hetero-epitaxy, the introduction of dislocations is necessary to accommodate the structural mismatch. For arsenides and phosphides material systems, films grown past the pseudomorphic critical thickness generally relax through the nucleation of dislocation half loops at the surface, which glide down on the (111) plane towards the hetero-interface, creating 60° misfits and threading dislocations in the process. These threading dislocations, with a density up to $10^{8}–10^{10}$ / cm$^2$, act as non-radiative recombination and carrier scattering centers, change local strain, and lead to device early failures. Achieving a low threading dislocation density (TDD) is therefore one of the most important metrics for metamorphic epitaxial growth. The thermal expansion mismatch limits the maximum buffer thickness that can be used. Macroscopic cracks emerge if a thick buffer accumulates too much strain energy upon temperature changes. With linear thermal expansion coefficients of Si, GaAs, and InP being $2.59 \times 10^{-6}$ /K,

\[ \text{E-mail address: eeqli@ust.hk (Q. Li).} \]

https://doi.org/10.1016/j.pcrysgrow.2017.10.001

Available online 14 November 2017
0960-8974/ © 2017 Elsevier Ltd. All rights reserved.
5.73 × 10⁻⁶ /K, and 4.6 × 10⁻⁶ /K respectively, the thickness of hetero-epitaxial III-V thin films on Si is typically limited to about 10 µm and below [11,12]. Experimentally observed critical thickness for the onset of crack formation in GaAs epilayers on Si is approximately 7 µm for a ΔT of 575 °C, 5.1 µm for a ΔT of 675 °C, and 4.9 µm for a ΔT of 725 °C [13], where ΔT represents the change in temperature in the cooling-down stage after growth. It is interesting to note exceptions in III-Sb on GaAs and III-Sb on Si hetero-epitaxy. With the extremely large lattice mismatch and the softness nature of antimonides, strain relaxation can occur by the formation of 90° pure edge-type misfit dislocations (MDs) [14–17]. These sessile MDs, arranged in a two-dimensional (2D) network and confined at the hetero-interfaces, provide the most efficient strain relaxation and do not easily thread upward into the layers grown atop. This fundamentally different growth mode is referred to as the interfacial misfit (IMF) growth mode. High quality GaSb thin films have been reported without resorting to several-µm-thick metamorphic buffer layers [18].

The prevailing approach to perform IMF growth of III-Sb on Si is to begin with an ultra-thin AlSb nucleation layer [19–21], such that a periodic array of 90° misfit dislocations, akin to those in the IMF growth of GaSb on GaAs, can be created. Apart from the advantages of the IMF growth mode, the thermal expansion coefficient of AlSb (2.55 × 10⁻⁶ /K at 300 K) is very close to that of Si (2.59 × 10⁻⁶ /K at 300 K). With a relatively small thermal expansion coefficient mismatch, absence of microcracks or wafer bending even in very thick (10 µm) AlSb on a Si wafer has been reported [22]. In the past few years, III-Sb-based quantum-well edge-emitting laser diodes monolithically grown on Si substrate have been demonstrated [23–25], covering wavelengths from near 2 µm to 1.5 µm telecom wavelength range.

A unique type of defect associated with III-V/Si hetero-epitaxy is antiphase-domains (APDs), which arise from the lack of inversion symmetry of III-V materials. As a group IV semiconductor, Si crystalizes in a diamond structure, with two face-centered-cubic (FCC) sublattices occupied by the same type of atom species. Each Si atom is bonded to four neighboring Si atoms in a tetrahedral arrangement, and the bonding electrons are shared equally between the nuclei. As a result, Si is a non-polar material. In contrast, for cubic III-V semiconductors with a zinc-blende crystal structure, the two FCC sublattices are occupied by different atom species. The bonds are polar due to the difference in the ionicity of the constituent atoms. The cubic zinc-blende structure’s symmetry forbids spontaneous polarization but allows piezoelectric polarization [26].

APDs are inherent to polar-on-non-polar growth. In practice, any real (001) Si surface always exhibits steps. Single-layer steps (or odd
layer height steps) produce two domains in the III-V overlayer with opposite sublattice allocation (see Fig. 2 [27]), whereas double-layer (or even-numbered) steps do not. The two III-V domains are separated by a plane of wrong bonds, either III-III or V-V bonds, which is referred to as an antiphase boundary (APB) or inversion domain boundary (IDB). APBs are electrically charged planar defects, acting as non-radiative recombination centers in optoelectronic devices and leakage paths in electronic devices. The impact of APBs on the optical properties of III-V layers is often characterized by photoluminescence quenching and spectral broadening [28,29], while their role on electrical properties is reflected in the significantly degraded electron mobilities [29,30]. With certain growth or etching conditions, APBs rising to the material surface could be visible under scanning electron microscopy (SEM) or atomic force microscopy (AFM). As an example, Fig. 3 displays an AFM image of as-grown GaAs on an exact (001) Si substrate by metal-organic chemical vapor deposition (MOCVD), showing irregular, curved boundaries.

Another fundamental problem originating from growing a polar semiconductor on a non-polar substrate is the lack of charge neutrality [31,32] at the III-V/Si interface, and the related interdiffusion leading to cross-doping of both materials. In GaAs/Si heteroepitaxy, as As forms strong bonds with Si, whereas Ga does not, the first atomic layer bonding to the Si substrate will be an As layer. Without atomic re-arrangement, the As-Si bonds would carry a charge density equivalent to $3 \times 10^{14} \text{V/cm}$ inside the growing GaAs layer [31]. Such a large field can lead to massive atomic re-arrangements, attempting to restore a neutral interface. Harrison et al. [32] have considered two idealized atomic arrangements, both of which involving a fraction of Si atoms being replaced by Ga atoms and re-incorporated into the first As plane. Under realistic finite-rate growth conditions, the Si atoms that are removed from the top Si layer have a good chance to be simply taken up by the growing GaAs as bulk dopant. As a consequence, the accumulated Si atom on the GaAs surface can easily diffuse across the heterointerface during high temperature annealing. Moreover, experimental observations revealed a significant enhancement of the Si diffusivity in the presence of a high level of crystalline disorder at the heterointerface [33]. This is even more severe in GaAs-Ge-Si structures. The outdiffusion of Si and Ge into the III-V layer implies a need for diffusion barrier layer to prevent large leakage current that makes device operation untenable [34].

3. Wafer-scale hetero-epitaxial growth of III-V thin films on Si

Since the 1980s, tremendous effort has been applied to the epitaxial growth of device-quality III-V thin films on lattice-mismatched Si wafers, mostly by MOCVD or molecular beam epitaxy (MBE). GaAs-on-Si has been the most heavily investigated since the problems associated with GaAs/Si hetero-epitaxy are representative of the difficulties associated with the growth of lattice-mismatched and polar-on-non-polar semiconductor hetero-structures [35]. The insight gained from studying this material could aid the conception and optimization of other hetero-epitaxial systems. The APB problem has been fairly successfully resolved by the use of misoriented Si substrates in conjunction with high-temperature pre-bake treatment [36–38]. Beginning with a 4–6° offcut Si substrate tilted toward the [110] direction, single atomic steps tend to reorganize into energetically more stable double steps under high-temperature annealing conditions [39]. As a result, the APB can be minimized or eliminated. In direct GaAs-on-Si epitaxy with an abrupt transition from the substrate to the overgrown GaAs, a two-step method [30] is widely adopted to improve the structural perfection of the hetero-epitaxial thin films. Post-growth thermal cycle annealing [40–42], by which an epilayer is subjected to large temperature oscillations and thus periodically switching between compressed and tensile states [43], has been found effective in reducing threading dislocation densities. Insertion of dislocation filter layers, such as a strained layer superlattice [44] or thin strained layers [45], can facilitate the annihilation of threading dislocations thereby minimize intrusion of dislocations in the active layers of interest. Growth of compositionally graded buffers (e.g., SiGe, GaAsP) [46–48] to bridge the lattice constant between Si and GaAs is another approach that has been heavily investigated. Achieving high strain relaxation while maintaining a low threading dislocation density are figures of merit of such metamorphic graded buffers. Over the past three decades, the above-mentioned techniques, often combined together, have been used to grow various compound semiconductors on Si, leading to different levels of success. The reader is referred to several previous reviews [35,49] for more comprehensive discussions.

More recently, some major advancements have been made in blanket hetero-epitaxy of III-V compound semiconductors on planar Si substrates, including scaling III-V integration up to 300-mm Si wafers, overcoming the APB problem on exact (001) oriented Si, and demonstrating high-mobility quantum-well (QW) transistors and low-threshold quantum-dot (QD) lasers.

The diameter of III-V substrates is often limited to 150 mm or less, while the maximum Si wafer diameter is 300 mm commercially, with 450 mm in development. Scaling III-V integration up to 300 mm Si to make use of state-of-the-art Si manufacturing foundries provides cost advantages. In the past few years, direct growth of III-V materials of interest on 200 mm and 300 mm Si wafers has been reported by a number of groups. Roesener et al. [50] proposed and investigated the growth of GaP nucleation and metamorphic Ga$_{1-x}$In$_x$P$_{1-y}$P buffers on 300 mm Si for solar cell integration. Today, 300 mm GaP/Si templates (NASP$_{300}$.GmbH [51]) grown by MOCVD are commercially available. Recent GaAsP metamorphic solar cells [52,53] epitaxially grown on GaP/Si substrates have demonstrated efficiencies above 11.5% by reducing the TDD values down to 4.0–4.6 $\times 10^6$ cm$^{-2}$, showing a promising path towards high-efficiency dual-junction GaAsP-on-Si cells. Motivated by the development of III-V channel MOSFETs for high-performance and low-power logic applications, Huang et al. [54] and Orzali et al. [55] reported epitaxy of smooth In$_{0.53}$Ga$_{0.47}$As layers on 300 mm Si wafers using metamorphic InP/GaAs buffers by MOCVD. The GaAs/InP metamorphic buffer was reduced to $\sim 860$ nm in thickness, among the thinnest in the literature. By defect counting or Ayers’ model [56], TDD in the low $10^7$ cm$^{-2}$ range was determined.

Having of III-V/Si epitaxy on 300 mm substrates also presents opportunities to overcome wafer size incompatibility in wafer bonding technology. By growing III-V layers on a Si donor wafer instead of on a lattice-matched III-V native donor wafer, III-V-on-insulator-on-Si substrates with any available wafer size can be realized [57,58]. Proof-of-principle demonstrations of 100 mm InGaAs-on-insulator [59] and 200 mm InGaAs-on-insulator wafers [60,61] were first reported. In the fabrication process flow described in Fig. 4, In$_{0.53}$Ga$_{0.47}$As was directly grown by MBE on a 200 mm Si substrate. A 2.5 $\mu$m Ge buffer, followed...
by 0.5 µm Ga(Al)As and 1.5–2 µm InGAs metamorphic buffer, was grown to accommodate the lattice mismatch between InGaAs and Si. The wafer was then planarized using chemical mechanical polishing (CMP). An ultrathin Al2O3 buried oxide (BOX) layer was deposited for direct wafer bonding. The donor wafer was subsequently removed by wet-etching, leaving a 200 nm InGaAs-on-insulator substrate. Very recently, a similar process but improved to allow donor wafer recycling was developed, leading to successful fabrication of 300 mm InGaAs-on-insulator substrates [62]. These processes open the way to very large-scale production of III-V-on-insulator hybrid substrates for future technology nodes. However, a major limitation still lies in the high density of crystal defects generated in blanket hetero-epitaxy on the donor Si substrates.

For III-V/Si integration to be compatible with CMOS processing and devices, microelectronics-standard nominal (001) Si substrates with a miscut angle less than 0.5° are preferred [5]. However, except for nearly lattice-matched GaP on Si [5], conventional hetero-epitaxy on planar Si wafers generally requires a 4°–6° offcut angle in order to form a prominent double-stepped Si surface that prevents APDs. Recently, there has been notable progress made to achieve APB-free III-V epitaxial layers on the so-called “exact” Si (001) substrates by careful treating of (001) Si with a slight misorientation (< 0.5°) prior to III-V nucelation [29]. Fig. 5(a) shows an AFM image of a 400 nm thick GaAs layer grown on un-optimized on-axis Si (001) substrates, showing a high density of randomly oriented APBs. In Fig. 5(b), a quasi Si (001) substrate with a 0.15° misorientation in the [110] direction is deoxidized in a SiConi™ chamber using NF3/NH3 remote plasma and then annealed (1 min–10 min) in an MOCVD reactor at high temperature (800 °C–950 °C) in H2 ambient. Such a surface preparation procedure promotes the structuring of the 2 × 1 surface and forms predominant double steps, as shown in Fig. 5(b). This observation appears to contradict theoretical thermodynamics, which suggests it is energetically unfavorable to form a double-stepped Si surface at a low miscut angle. Bruckner et al. [63] attributed this anomalous phenomenon to the interaction of the Si surface with the H2 ambient driving a dynamic step formation procedure governed by surface vacancy generation, diffusion, and annihilation at the step edges. With the double-stepped Si in Fig. 5(b), subsequent growth of a 150 nm GaAs overlayer shows APB-free surface, as shown in Fig. 5(c). These results coincide with the experimental observation from a number of groups that single-domain Ge and III-V layers can be deposited on nominal Si (001) substrates [54,55,64–67].
The last few years have witnessed a booming development in heterogeneous integration of III-V transistors and light emitters on Si. High-mobility III-V semiconductors, such as InGaAs/InP- and GaSb/InAs-based materials, are of great interest for digital logic device applications due to their potential to replace Si channel in emerging MOSFETs and their potential for developing innovative device structures like tunnel FETs [2,68]. Both MBE and MOCVD have demonstrated excellent InGaAs QW heterostructures on (001) Si substrates [64,66,69,70], with room-temperature Hall mobilities exceeding 10,000 cm²/V·s, matching the best results on native InP substrates. In these metamorphic device structures, compositionally graded InAlAs/GaAs or composite InP/GaAs buffers, a few micrometers thick, have been used to convert the lattice constant of Si to the desired one in the active layers, as shown by the cross-sectional transmission electron microscopy (TEM) images in Fig. 6. Different transistor structures, including QW field-effect transistors with Schottky gates [69,71], and QW channel MOSFETs with high-k dielectrics [70] and source/drain regrowth [72–76], have been reported. These exploratory transistors demonstrated record-high drive current and extrinsic transconductance at low operation voltages, suggesting potential for high-speed switching applications with reduced power consumption. Challenges remain, however, in their compatibility with CMOS processing, integration with p-channel transistors, and device uniformity and reliability issues arising from surface roughness and high dislocation density.

Integration of III-V lasers on Si has been under investigation for some time. More recent interest has been motivated by the fast-evolving Si photonics technology. Si is a good waveguiding material but poor in light emission. Despite progress in Si-based light modulation and detection technology, and low-cost Si optoelectronic integrated devices enabled by the mature CMOS technology [77,78], an efficient, reliable laser on a Si substrate remains the “holy grail” for Si photonics. Heterogeneous integration of III-V materials on silicon combines the superior gain characteristics of compound semiconductors with the passive waveguide characteristics of Si. Among various integration schemes, hybrid III-V-on-Si laser through wafer bonding technology is considered the most successful and being commercialized [79]. There have been many significant advances in device performance [80–85] using improved fabrication technologies compatible with silicon photonics. However, epitaxial growth of III-V lasers on Si is ultimately preferred as a monolithic and sustainable integration solution. Compared to the bonding approaches, it offers a scalable process which is desirable for large-scale commercial applications. Cost advantages follow as well when process integration schemes are well-defined.

Attempts to grow III-V QW lasers directly on Si date back to the 1980s. In 1987, room-temperature continuous-wave (CW) operation of large-area GaAs/AlGaAs hetero-structure lasers on (100) Si substrates [86] was obtained. The laser operated on four different occasions for a total of 4 min before degradation took place. In 1991, room-temperature CW operation of an InGaAs/InGaAsP multi-quantum-well (MQW) laser diode on a Si substrate was reported [87], and in 2003, GaAs/AlGaAs QW lasers were demonstrated on relaxed graded Ge/GeSi virtual substrates on Si [88]. In general, most of previous QW lasers on Si suffered from poor reliability and short lifetimes, impeding their practical application. The rapid degradation was attributed to the formation of dark line defects [89–95]. Fig. 7 illustrates the growth of a dislocation network in the active layer of a laser device. Non-radiative carrier recombination at dislocations initiated in non-lattice-matched heteroepitaxy leads to dislocation climbing into the active zone. This recombination-enhanced process [96] has been extensively investigated in the literature and has become widely accepted as the degradation
mechanism of III-V lasers and light emitting diodes. As a consequence, the achievement of a GaAs- or InP-based QW laser on Si that can simultaneously exhibit low threshold current, high temperature stability, and long lifetime has remained elusive [97]. To overcome this fundamental challenge, recent implementation of compound semiconductor lasers on Si has been focused on using QDs as the gain material. Because of their inherent property of carrier confinement/localization and reduced interaction with defects, QDs have proved to be less sensitive to defects than conventional bulk materials and QW structures [97–99]. As shown in Fig. 8, a threading dislocation can only ‘kill’ a very limited number of QDs, leaving the rest intact and able to provide optical gain. Moreover, the strong strain field of a QD array can bend conduction of threading dislocations [97] or propel threading dislocations away from the QDs [99]. For these reasons, III-V QD lasers grown on Si could potentially deliver superior reliability with relaxed tolerance to TDD. Pioneering work on self-organized InGaAs/GaAs QD lasers on Si incorporated multiple stacked InAs QD layers in the buffer as effective dislocation filters [97,100]. These QD lasers, with an emission wavelength close to 1 μm, exhibited relatively low threshold current (900 A/cm²) and very high characteristic temperature (T₀ = 278 K).

Significant progress has been made ever since to achieve room-temperature CW lasing at the telecommunication band of 1.3 μm. The first operation of InAs QD lasers epitaxially grown on a Ge substrate was demonstrated in 2011 [101]. Using a Ga prelayer growth technique, an APB-free GaAs buffer layer was grown on offcut Ge substrates by MBE. Broad-area laser devices with a five-layer InAs/InGaAs dot-in-a-well (DWELL) active structure were fabricated, realizing room-temperature CW lasing at 1305 nm with an output power of ~28 mW per facet and a very low threshold current density of 55.2 A cm⁻². Since Ge is widely used as the transitional buffer for GaAs-on-Si hetero-epitaxy, this result marks an important milestone in the monolithic integration of long-wavelength InAs/GaAs QD lasers on a Ge/Si substrate. Leveraging the well-established Ge-on-Si epitaxy technique, Liu et al. [102] were able to fabricate 1.3 μm InAs QD ridge lasers more in line with commercial telecom laser designs. Room-temperature low thresholds (16 mA), high output power (176 mW), and high T₀ (> 200 K) were reported. The device yield measurements showed repeatable performance across different dies and wafers. Reliability tests uncovered over 2700 h of continuous wave operation at 30 °C along with an extrapolated mean time to failures of up to 4600 h [103]. This remarkable progress suggests great potential for producing reliable and efficient QD lasers on Si by further increasing the ratio of lasing QDs to dislocations.

The results of high-performance QD lasers on both Ge-on-Si and Ge substrates are encouraging. However, it would be more attractive to realize a laser that does not require an intermediate Ge layer, both because the requirement of the Ge layer restricts the range of Si circuits to which it can be applied and because it is difficult to couple light through this layer to a Si waveguide due to the large optical absorption coefficient of Ge at telecommunications wavelengths [99]. Combining a thin nucleation layer made of AlAs, a multi-temperature growth method and InGaAs/GaAs superlattices filters, Chen et al. [99] reported InAs/GaAs QD lasers directly grown on Si substrates with a record-low threshold current density of 62.5 cm⁻². Broad-area lasers were fabricated with as-cleaved facets, realizing CW lasing up to 75 °C, a high output power exceeding 105 mW at room temperature, and a long extrapolated lifetime of over 100,158 h. In all the above-mentioned QD lasers on Si, offcut (001) substrates were used to suppress the anti-phase disorder arising from the polar-on-non-polar hetero-epitaxy.

Most recently, extensive research has been performed on quantum dot lasers monolithically grown on “exact” (001) Si substrates that are standard in microelectronics fabrication. Leveraging an APB-free GaP-on-Si (001) template, Liu et al. [104] demonstrated the first electrically pumped CW InAs QD lasers epitaxially grown on (001) Si without offcut. By growing APB-free GaAs thin films on V-groove patterned (001) Si, Norman et al. [105] reported electrically injected, Fabry–Perot QD ridge lasers, showing CW operation up to 80 °C with threshold currents as low as 37 mA. Chen et al. [106] reported electrically pumped 1.3 μm InAs/GaAs QD lasers directly grown on microelectronics-standard planar (001) Si substrates using a sophisticated Si surface preparation step. These results demonstrate the compatibility of high-performance monolithic III–V light sources with on-axis Si substrates and their potential for CMOS foundry integration.

4. Selective-area hetero-epitaxy from micrometer to nanometer length scale

Selective-area hetero-epitaxy, which restricts epilayer growth in pre-defined regions by substrate patterning, offers additional control over the strain relaxation process and brings benefits like the defect necking effect. Over the past few decades, selective-area hetero-epitaxy has been extensively investigated in Ge/Si and III–V/Si material systems, spanning the micro- to nanoscales. One unique growth method derived from selective hetero-epitaxy is referred to as the “aspect ratio trapping” (ART) technique. By this approach, III-V materials are selectively grown in high-aspect-ratio holes or trenches formed by a patterned dielectric (typically SiO₂) on Si. The threading dislocations originating from the III–V/Si hetero-interface are guided to the oxide sidewalls should the aspect ratio allow, resulting in dislocation-free regions above a critical thickness. The “trapping” of threading segments (i.e., the epitaxial necking effect) in the ART technique is attributed to the inherent crystallographic geometry. In the (111)/<110> cubic slip system, misfit dislocations lie along the (110) directions in the (100) growth plane, while the threading segments rise up on the (111) planes in the (110) directions [107]. As shown in Fig. 9(a), for an [110] direction aligned trench with a trench width of w and an oxide sidewall...
height of $h$, <110>-oriented threading segments would all be inclined to the [100] oxide sidewalls, and their projections onto the (110) plane would form a $\sim 55^\circ$ angle with the [110] direction. In principle, when the aspect ratio ($AR = h/w$) is larger than 1.4, all the threading dislocations will project into the oxide wall and terminate there. This also implies an interesting fact, that if the trenches/holes are small enough, the dislocation elimination can be accomplished within a very thin buffer layer, preventing stress accumulation caused by the thermal expansion coefficient mismatch. The ART concept (see Fig. 9(b)) was firstly revealed in Ge/Si hetero-epitaxy [108], and then applied to III-V/Si epitaxy. The trapping capability of the ART structure for dislocations and planar defects is further explained schematically in Fig. 9(c). For 60° threading dislocations generated during growth, they tend to glide along the $\langle 111 \rangle$ planes. These dislocations will eventually hit an oxide wall if the trench aspect ratio is sufficiently high. In the same way, planar defects on $\langle 111 \rangle$ planes parallel to the trenches are trapped. Only planar defects lying in planes perpendicular to the trenches will never be blocked [109]. Experimental studies show that those dislocations in selective hetero-epitaxy can undergo significant redirection, so as to follow the normal to the newly formed crystal facets as growth proceeds [107], but the defect trapping effect still holds.

Fig. 10(a) displays a tilted-view SEM image of GaAs selectively grown on a sub-micron-stripe-patterned (001) Si substrate by MOCVD. We can observe $\langle 111 \rangle$ GaAs faceting at the growth front with an excellent growth uniformity. In Fig. 10(b), the propagation of crystalline defects revealed by cross-sectional annular dark field scanning transmission electron microscopy (STEM) agrees well with the ART concept, although the dielectric stripes are not tall enough to provide a sufficiently large aspect ratio for complete defect necking. Development of the ART growth of III-V materials on Si has gone from micro- to nanoscale lengths [110–113]. In Fig. 11(a), the aspect ratio of the trenches is large enough so that all the dislocations are essentially trapped under the dash line. However, hard-to-control asymmetry of GaAs facets remains a problem. It should be noted that the top facets of GaAs can be manipulated by growth conditions. Fig. 11(b) shows a flat GaAs(001) top surface which is considered more amenable for device fabrication.

Fig. 9. (a) Schematic showing the crystallography of mismatch relaxation-related glissile threading dislocations and their projections onto the (110) plane ($w$ is the width of the trench, while $h$ is the height of the sidewall) [107]. (b) Cross-section diagram demonstrating the principles of epitaxial necking, showing zero threading dislocations at the upper surface [108]. (c) The trapping capability of the ART technique for threading dislocations and planar defects [109].

Within the ART approach, engineering the Si surface at the bottom of the trenches is critical. Realizing an oxide-free III-V/Si interface and minimizing the bottom trench roughness are effective in improving morphology uniformity and reducing twin defects. However, it is still difficult to control APD generation on a flat (001) Si surface. The
density of defects such as twin planes travelling along the trench direction is fairly high. By etching Si in alkaline solutions (such as potassium hydroxide, tetramethylammonium hydroxide or ammonium hydroxide), Si V-grooves are formed as a result of the lowest Si etch rates in (111) planes. Growing III-V materials on V-grooved (111) Si surfaces can greatly enhance the quality of epitaxial III-V materials in the ART process [112,113]. Fig. 11(c)–(e) display bright field STEM images of GaAs-on-V-grooved-Si in directions both perpendicular and parallel to the trenches. All threading dislocations (meandering lines) are found annihilated on the oxide walls and confined at the trench bottom. Very few (111) planar defects can be identified and none of them reach the surface, suggesting the upper part of the inspected GaAs portion is free of defects.

Integration of InP and associated alloys on Si by the ART method is of great interest for device applications ranging from high mobility transistors to emitters at telecom wavelengths. However, the 8% lattice mismatch between InP on Si poses a big challenge. With a high surface mobility of the Indium adatom, InP/Si nucleation in nanosized regions can have significant spatial non-uniformity [114]. Fig. 12 summarizes four growth schemes that have been investigated in the past few years. In Fig. 12(a), InP growth is performed on rounded Ge surfaces [115,116]. The surface steps of the Si and Ge are carefully engineered to avoid APDs. A high-density dislocation network appears at the InP/Ge hetero-interfaces. According to results of FinFETs fabricated on such a material platform, out diffusion of Ge into the InP layer generates severe buffer leakage current, compromising transistor on-off ratios.

Fig. 12(b) shows direct growth of InP on V-grooved Si using a low nucleation temperature and an extremely high V/III ratio of larger than 2000 [114]. TEM and x-ray diffraction (XRD) characterizations suggest greatly enhanced crystalline quality as compared to the InP-Ge-Si structure. By mimicking metamorphic InP/GaAs buffer on planar Si [55,66], Fig. 12(c) illustrates the use of a GaAs intermediate layer [117,118]. It eases the difficulty in subsequent InP epitaxy because of the halved lattice mismatch. This GaAs layer filling up the entire V-shaped Si pockets, however, decreases effective aspect ratio for defect trapping. It was found that a thick GaAs transitional buffer might not be necessary. In the design of Fig. 12(d), a low-temperature GaAs nucleation layer which is only a few-nanometer thick is deposited prior to InP epitaxy [119,120]. Compared to the structure in Fig. 12(b), similar crystalline quality is achieved under much more relaxed growth conditions.

The use of (111) Si V-grooves in the ART growth process clearly demonstrates its unique advantages. First, the crystallographic alignment between the Si and III-V materials in the V-grooves avoids the introduction of APDs. Fig. 13(a) shows a III-V lattice in the V-shape of Si with (111) facets along the [110] direction [121]. Crystallography analysis indicates that the III-V semiconductors on the two (111) facets of the “V-shape” have the same polarity. In principle, the Si (111) surface can also have surface steps, as in the case of Si (001). However, a single step on the Si (111) surface has the height of one Si (111) double-layer (0.31 nm) [122,123]. As shown in Fig. 13(b), such steps will not lead to the formation of APBs. Secondly, III-V nucleation on Si (111) generates less defects as compared to nucleation on Si (001). For InP epitaxy on Si (111), under proper growth conditions, the large lattice mismatch can be mostly absorbed by a thin layer of one-dimensional twins that are parallel to the growth surface [114,119,121]. This is in sharp contrast to the hetero-epitaxy on the Si (001) surface in which stress is mainly relaxed by misfits and 60° threading dislocations.

Growing III-V crystals in nano-sized cavities by the ART technique offers a viable path to integrate III-V high-mobility channel materials.
into advanced CMOS technology. A replacement fin process (See Fig. 14) was recently developed to fabricate InGaAs FinFETs on 300 mm Si substrates [124]. A sequence of epitaxial growth and CMP steps were used to form the InGaAs channel. After the shallow trench isolation (STI) template formation, the dummy Si was removed and MOCVD growth was performed to fill up the STI trench with InP buffer. The first CMP step was then applied to get a flat InP surface. This was followed by recess etching which ultimately determines the height of the InGaAs fin. The InGaAs channel was then grown and the Indium content target is 53% to be lattice matched to the InP. The InGaAs was planarized by a second CMP step and the fin was revealed by recessing of the STI oxide. This process, where defects are trapped at the STI sidewall, led to the first realization of III-V FinFETs (see Fig. 15) with well-behaved on-off characteristics on large scale Si wafers in a fully VLSI compatible flow.

It was later applied to fabricate various other transistor structures incorporating III-V high mobility channels. By removing the leaky InP buffer layer from underneath the InGaAs channel, InGaAs gate-all-around (GAA) and nanowire transistors [125] have been demonstrated on the same platform, showing greatly enhanced performance.

In addition to GaAs/Si and InP/Si integration, the application of the ART technique extends to ternary alloys, hetero-structures and 6.1 Å family heteroepitaxial layers. The cross-sectional TEM image in Fig. 16(a) shows the structure of a new type of GaAs-InGaAs-GaAs fin-array Esaki tunnel diode fabricated on (001) Si substrates using the ART integration process [126]. A room-temperature peak-to-valley current ratio (PVCR) of 5.4 is obtained and negative differential resistance characteristics remain up to 200 °C. Such fin-array tunnel diodes, serving as the building blocks, are used to form inverters [127] and fin-array tunneling triggers with tunable hysteresis [128], demonstrating potential logic applications. GaSb and InAs are materials of interest for...
high-mobility p-channel and n-channel transistors, respectively. The broken-gap band alignments formed by their hetero-structures could be exploited for novel tunnel FETs with steep subthreshold swings. Fig. 16(b) and (c) present TEM images of GaSb and InAs fins grown on top of GaAs fins on Si [129]. These results highlight the flexibility of the ART process in heterogenous materials integration and its potential in co-integration of n-channel and p-channel FinFETs based on high-mobility 6H-1 family compound semiconductors.

The aforementioned nano-scale selective epitaxy shows promise for next generation logic transistors. Substantial efforts are being made to leverage the same technique for integrated photonic devices requiring a large material volume for a sufficient modal gain. In one approach, III-V materials are grown out of tiny trenches to form bigger structures. Fig. 17(a) and (b) show SEM images of GaAs box-shaped ridges by this method [109]. The straight and uniform ridges with flat surfaces in all directions are promising for realizing waveguides with low light scattering losses. Compressively strained InGaAs/GaAs MQWs are deposited on top of the fully relaxed GaAs ridges, exhibiting pronounced QW photoluminescence at room-temperature. Fig. 17(c) illustrates a similar epitaxial scheme, but to grow tilted InP nanowires out of nanotrenches on a patterned SiO2-on-Si substrate [130]. The nanowires, with a hexagonal top, consist of a mixture of wurtzite and zincblende crystal phases, forming type II hetero-structures that promote lasing over a wide wavelength range. An alternative path to integrate III-V structures on Si with sufficient volume for laser cavities is to use a relatively large trench size. Fig. 17(d) presents highly ordered InP ridges selectively grown on sub-micron V-groove-patterned Si [119]. Quantum wires and ridge QW [120] structures emitting at telecom wavelengths have been demonstrated. Fig. 17(e) schematically depicts the fabrication of an InP distributed feedback laser array on Si using the selective-area growth technique in confined regions and top-down lithographic patterning [131]. Compared to other integration schemes, this result presents a highly scalable monolithic solution compatible with standard high-volume and low-cost CMOS manufacturing processes.

Recently, an alternative to the ART technique known as the confined epitaxial lateral overgrowth (CELO) technique [132] has been developed. This method begins with the formation of a 3D cavity in an oxide containing a crystalline seed. The 3D cavity is then selectively filled by the epitaxial material, during which the growth direction is forced to turn 90° (from vertical to lateral) with respect to the original seed. For traditional ART, planar defects traveling along the trench cannot be filtered geometrically, whereas defect trapping in CELO occurs in two directions and all defects are presumably confined in the seed region by geometry (see Fig. 18(a)). Compared to classical epitaxial lateral overgrowth, growth by CELO is constrained in a cavity that is pre-defined by a sacrificial layer. As a result, the epilayer thickness and morphology are well controlled. A smooth surface can be achieved without the need for CMP. Fig. 18(b)–(f) show typical SEM and TEM images of InGaAs CELO for scalable integration of CMOS-compatible InGaAs-on-insulator MOSFETs on large-area Si substrates.

Combining the ART technique with epitaxial lateral overgrowth (ELOG) is able to provide large regions of device materials. A new monolithic integration platform based on InP ELOG technology has been proposed to integrate III-V lasers on Si, as shown in Fig. 19. A layer of InP is first deposited on planar Si wafers and then polished by CMP. Microsized SiO2 stripe patterns are defined for the subsequent InP ELOG. The method exploits the SiO2 mask to filter the dislocations at the same time as the waveguide layer, enabling the integration of a monolithic evanescently coupled silicon laser (MECSL) [133]. High-quality QWs have been successfully grown on the ELOG InP layers, exhibiting comparable light emission intensities (> 85%) to those on a
Selective-area hetero-epitaxy in conjunction with the ART technique has been used to realize various heterogeneous structures on Si. Yet large-area planar thin films are more amenable in some device applications. Towards this goal, III-V materials are epitaxially grown out of closely spaced trenches and continued laterally on top of the oxide stripes until materials from adjacent trenches merge together. The resultant III-V on Si templates can be treated as virtual substrates (or compliant substrates) for subsequent epi-structure growth and device fabrication. Initial investigations of continuous GaAs films on SiO2-patterned Si substrates discovered coalescence defects appearing in the form of threading dislocations, twin defects and stacking faults above the SiO2 (Fig. 20(a) and (b)). This was accompanied by a rough crystallographic surface. A two-temperature step growth optimization was therefore implemented to reduce the coalescence defects, yielding an XRD $\omega$-rocking curve full-width-at-half-maximum (FWHM) of 190 arcsec for 1.5-µm GaAs on Si. However, the APD issue is not addressed and surface undulations are still visible (see Fig. 20(c)). CMP is generally required to establish a flat surface for device integration.

An improved method to grow coalesced GaAs thin films on patterned Si was demonstrated recently to produce GaAs-on-V-grooved Si (GoVS) templates. It starts with an on-axis Si (001) substrate covered by a layer of SiO2. The SiO2 is patterned by dry etching into [110] oriented stripes with a line opening of 90 nm and an oxide spacing of 40 nm. Then a 70 °C heated KOH solution (45%) is used to etch Si into diamond-shaped trenches bounded by {111} facets. After a pre-baking step in a MOCVD chamber, GaAs nanowires are formed by selective area hetero-epitaxy. The SEM image in Fig. 21(a) presents the resultant highly ordered array of GaAs nanowires free of APDs. According to analysis of x-ray diffraction $\omega$-rocking curves, these nanowires with a thickness of 150 nm yield an FWHM comparable to that of 1 µm-thick GaAs thin films on planar off-cut Si. Fig. 21(b) displays a cross-sectional TEM image taken along the [110] zone axis. Only a few (111) plane stacking faults are observed in the bulk of the GaAs. The zoomed-in TEM image at the GaAs/Si hetero-interface indicates that the 4% lattice mismatch is accommodated by a few-nanometer-thick stacking disordered layer in parallel with the (111) growth surface. With this ultrathin GaAs stress relaxing layer, the GaAs bulk layer has a very high crystalline quality.

Looking more closely at the glide of the few-layer stacking faults/twins at the GaAs-Si hetero-interface reveals that the disordered layers are stopped by a “tiara”-like structure beneath the SiO2 walls (see Fig. 21(c)).

Fig. 18. (a) Comparison of the CELO approach with traditional ART. By forcing the growth direction to turn 90° with respect to the original seed, defect trapping occurs in two directions. For traditional ART, defects generated along the trench cannot be filtered geometrically. (b) Top-view SEM for In$_0.7$Ga$_{0.3}$As epitaxy into CELO structures. (c) TEM cross-sectional view of the corresponding InGaAs CELO structure highlighting the long and thin InGaAs region on a 25 nm buried oxide and the Si seed region. (d) Plan-view and (e,f) cross-sectional HR-TEM images confirming that crystalline defects are geometrically confined in the seed region [132]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 19. Schematics of (a) epitaxial lateral overgrowth and (b) a monolithic evanescently coupled silicon laser (MECSL) structure based on the proposed integration platform [133].
Fig. 21(c). This “tiara”-like structure, made of Si, was formed together with the diamond-shaped Si pockets in the Si recessing process (the combination of dry etching and potassium hydroxide wet etching). This represents a slight modification from conventional simple Si V-grooves bounded by two {111} facets, but results in a unique defect trapping effect. Recall that in the conventional method [136] that combines ART growth within trenches and ELOG over SiO2, the SiO2 stripes lead to a high density of coalescence defects and a high surface roughness. Here, with most of the hetero-interface defects stopped by the “tiara” structure, it is possible to remove the SiO2 without jeopardizing the defect necking effect in the coalescence process. As shown in Fig. 22(a) and (b), the GaAs nanowires without SiO2 in between merged into a 300 nm planar GaAs film with a smooth surface. Fig. 22(c) schematically illustrates the defect trapping mechanism in the GaAs-on-V-grooved Si templates. The cross-sectional TEM image in Fig. 22(d) taken near the GaAs/Si interface confirms that most of the defects are confined within the diamond-shaped trenches. This growth method eliminates the necessity to use misoriented Si substrates or graded SiGe/Ge buffers, and erases the difficulties in achieving coalescence over dielectric patterns. A flat planar film can be obtained in a simpler way without an additional CMP planarization process. It was found later that, with the insertion of AlGaAs/GaAs or InGaAs/GaAs superlattices in the GaAs, the...
The defect densities of the GoVS templates have been assessed through electron channeling contrast imaging (ECCI) [105], plan-view TEM and x-ray rocking curves. ECCI uses a scanning electron microscope in backscatter mode at a given diffraction condition to image deviations from the Bragg condition caused by strain fields, such as those around dislocations and local phase shifts of the electron wave caused by the non-integer lattice translations across stacking faults. It allows for rapid acquisition of large-area scans without any material preparation. For a 2 µm GoVS, a threading dislocation density of \(7 \times 10^7\) cm\(^{-2}\) and a stacking fault density of \(2 \times 10^7\) cm\(^{-2}\) are obtained from the ECCI image in Fig. 23(a). A plan-view TEM image of a typical GoVS template in Fig. 23(b) suggests a dislocation density in the order of \(10^7\) cm\(^{-2}\), which is consistent with ECCI characterization. The GoVS templates provide opportunities for integration of electronic, photonic, or photovoltaic devices on Si wafers. InAs/GaAs QD structures grown on the GoVS template show superior optical properties as compared to those grown on planar offcut Si [140], while optically injected, Fabry–Perot, QD ridge lasers fabricated on the GoVS templates exhibit promisingly low threshold lasing characteristics [141–143]. Electrically injected, Fabry–Perot, QD ridge lasers on the GoVS templates were also reported, showing CW operation up to 80 °C, with threshold currents as low as 37 mA [105]. Applying a similar growth technique to develop InP and other III-V on Si virtual substrates is of great interest for future studies.

5. Summary

Monolithic III-V/Si integration can leverage the best physical properties of III-V devices with the advanced manufacturing process associated with silicon. Emerging technologies and applications provide new driving forces to transform this concept into an industrial solution. Notable progress has been made recently in developing novel epitaxial schemes as well as in overcoming the persistent challenges in material mismatch. Proof-of-principle demonstrations from exploratory transistors to on-chip quantum-dot lasers suggest good performance gain and integration potential. Efforts should continue to be devoted to resolving the compatibility of integration techniques with large-volume CMOS manufacturing, addressing device reliability issues and designing integrated systems.

Acknowledgements

This work was supported in part by an Initiation Grant (IGN15EG01) from HKUST and the Innovation Technology Fund of Hong Kong (No. ITS/273/16FP). We want to acknowledge Bei Shi, Yu Han and Yating Wan for discussions of results.

Reference


[33] R. Alcotte, M. Martin, J. Moeyaert, R. Cipro, A. David, F. Bassani, F. Ducrotte, Y. Bogumilowicz, E. Sanchez, Z. Ye, X. Bao, J. Jin, T. Baron, Epitaxial growth of antiphase boundary free GaAs layer on 300 mm Si(0 0 1) substrate by metalorganic chemical vapour deposition, ML. Mater. 4 (2016) 043503.


X. Zhou, Q. Li, C. W. Tang, K. M. Lau, 30 nm inverted In0.53Ga0.47As e-MOSFETs on Si substrate grown by MOCVD exhibiting high transconductance and low on-resistance, Int. Electron Devices Meet. 32 (2012) 5.


Qiang Li is currently a research assistant professor with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST). She received her B.S. degree in microelectronics from Peking University in 2009 and Ph.D. degree in electronic and computer engineering from HKUST in 2014. Her current research focuses on lattice-mismatched epitaxy of III–V compound semiconductors on silicon for electronic-photonic heterogeneous integration.

Kei May Lau is Fang Professor of Engineering at the Hong Kong University of Science and Technology (HKUST). She received the B.S. and M.S. degrees in physics from the University of Minnesota, Minneapolis, and the Ph.D. degree in Electrical Engineering from Rice University, Houston, Texas. She was an ECE faculty at the University of Massachusetts/Aberdeen and initiated MOCVD, compound semiconductor materials and devices programs. Since the fall of 2000, she has been with the ECE Department at HKUST. She established the Photonics Technology Center (PTC) at HKUST in 2001. She was on the ECE faculty at the University of California, Berkeley, and the ECE faculty at the University of Washington, Seattle, until 2017. She is a recipient of the US National Science Foundation (NSF) Faculty Awards for Women (FAW) Scientists and Engineers (1991), Hong Kong Croucher Senior Research Fellowship (2008) and IEEE Photonics Society Aron Kressel Award (2017). She is an Editor of the IEEE EDL and Associate Editor of Applied Physics Letters.