1. Introduction

Traditional copper networking cables have been replaced with fiber optic cables for transcontinental distances down to meters through much...
more efficient and cost-effective light transmission. Highly integrated 100-Gbps silicon photonics transceivers are now being shipped in volume and 400G transceivers are expected to reach volume production in Q4 2019 (Jones et al., 2019). While the revolution of photonic integration was started on InP, and are boosted with hybrid wafer bonding, higher production volumes requires low-cost, large-area Si substrates containing high gain III-V epitaxial materials (Bowers et al., 2019). In contrast to conventional long-haul fiber optics, datacom applications focus on power efficiency, reliability, high temperature operation, volume manufacturability and overall system cost in order to be technically and economically competitive (Cheng et al., 2018). Many traditional CMOS foundries are now exploring epitaxial growth of III-V on Si as a “More-than-Moore” technology to combat the gate length scaling limits of transistors. The same technology is also being leveraged for photonics, with recent advances in using QDs to mitigate crystal defects during III/V-on-Si growth (Liu and Bowers, 2018). InAs QD lasers grown on Si substrates are already showing exceptional performance, exceeding what has been achieved through heterogeneous integration, and reliability is reaching a point of commercial viability (Norman et al., 2018). Formerly, the use of miscut Si substrates to avoid antiphase boundary (APB) formation would hamper full compatibility with CMOS processing, but recent breakthroughs with microelectronics-standard nominal (001) Si substrates have demonstrated comparable or even superior performance (Li and Lau, 2017). A number of breakthroughs, including sub-milliamp threshold (Wan et al., 2017a), 175 mW single-facet output powers (Jung et al., 2017a), 119°C continuous wave temperature (Liu et al., 2014), near zero linewidth enhancement factor (Zhang et al., 2018a), isolator-free stability at optical feedback levels of up to 90% (Duan et al., 2019), and >100,000h lifetimes at 35°C (Jung et al., 2017b), have now made the III/V-on-Si epitaxial growth an appealing platform for photonic integration, from both a performance and cost standpoint, warranting discussion for more advanced photonic integration (Wang et al., 2017).

For photonic integration, the forces that drove the evolution of discrete optical modules to compact elements on a single chip will continue to drive improvements in cost, size, weight, and power (cSWaP). By confining light to small volumes with resonant recirculation thus eliminating the need for gratings or Fabry-Perot (F-P) facets, microcavity lasers have generated great interest for compact on-chip light sources (No, 2019). While the best edge-emitting lasers have thresholds on the order of tens of mA, microcavity lasers
have achieved below-ten-mA thresholds by significantly improving the spontaneous emission efficiency and modifying the electromagnetic field distribution. Through light confinement by total internal reflection at a curved boundary, high-quality modes support light propagation close to the boundary of the cavity periphery and can be coupled to a bus waveguide for wavelength division multiplexing applications (Yao et al., 2018). Since the first demonstration of microcavity lasers with four cleaved facets forming a closed loop (Scifres et al., 1976), continuous research has been conducted with a target for low power consumption and small footprint. However, sidewall scattering and recombination hinder performance when the size of the resonator is reduced to tens of microns. For microring semiconductor lasers, to obtain minimal bending losses down to a radii of 5 μm or smaller, the waveguide must be etched at least 1 μm below the active region (De La Rue et al., 2014). The necessity of deeply etched waveguide structure, with the active region exposed to the sidewall, exacerbates the issue with nonradiative sidewall-recombination, resulting in higher thresholds and enhanced device degradation.

Combining the advantages of QDs with microcavity structures has opened a new size regime for microcavity lasers (Kryzhanovskaya et al., 2014). The same property that mitigates the sensitivity to dislocations during III/V-on-Si epitaxial growth, also reduces the sensitivity to nonradiative recombinations at sidewalls of the microresonator. In quantum well (QW) lasers, carrier in-plane diffusion length reaches several microns and can readily reach the device sidewalls. Consequently, radiative recombination efficiency drops drastically with the shrinkage of the cavity size. However, in QD lasers, the in-plane diffusion length is reduced to ~0.5 μm. This enhanced carrier localization gives rise to a weaker surface recombination effects and enables deeply etched process through the active layer without imposing an obvious penalty on the lasing threshold (Moore et al., 2006). This provides notable advantages in scaling to small dimensions, shrinking on-chip lasers with size as small as 1 μm in diameter (Kryzhanovskaya et al., 2015a,b; Wan et al., 2016a), and sub-milliamp thresholds (Wan et al., 2017b).

2. Fundamentals of semiconductor QDs and historical review of QD lasers on Si

First proposed in 1982 (Arakawa and Sakaki, 1982), QD lasers have made advances and outperformed bulk, QW and quantum wire lasers
QDs can be formed through a self-assembly process using InAs on (In,Ga,Al)(As,P) layers, and are promising to make lasers with low threshold current densities, temperature insensitive behavior, high power operation, isolator-free stability, suppressed sidewall recombination rates, and mitigated sensitivity to defects in epitaxial integration on Si (Norman et al., 2019). Each of these advantages can be attributed to the discrete density of states and three-dimensional carrier confinement (Bimberg et al., 2000).

2.1 Principle of operation: Quantization of the density of states

In conventional double heterostructure lasers, the active region is comprised of a low band-gap material with a thickness of 0.1–0.3 μm that is surrounded on either side with a larger band-gap material. This double heterostructure helps to effectively trap electrons and holes within the active region, thereby reducing the required threshold current density for lasing. The electron and the hole in the exciton are normally confined to an Exciton Bohr Radius. However, when the thickness of the active layer reduces to 50–100 Å, which is comparable to the de Broglie wavelength, the carriers are forced to exist in bound states with discrete allowed energies, as opposed to the continuous band of states associated with typical bulk semiconductors (Coldren et al., 2012). Modification of the quantum confinement results in redistribution of the allowed electronic states over energy and momentum, i.e., changing the shape of the density of states. This energy quantization effect was first utilized to increase the density of states near the band edge in QWs and effectively lower the lasing threshold (Dingle and Henry, 1976). When the active layer has quantum confinement in three dimensions, as in the case of QDs, a complete localization of electrons and holes ideally gives rise to a sequence of delta-function-like states possessing atom-like degeneracy. In practice, these delta-function-like states show a small finite line width from phonon scattering, as shown in Fig. 1.

2.2 The promise of QD lasers

For efficient lasing, it’s important that there exists a large density of carriers in both the electron and hole bands at energies close to the band-edge. In QD lasers, the atom-like density of states ensures that low densities of carriers are needed for population inversion. The short carrier lifetime of the ground state supports a high level of radiative optical transitions per unit volume. This mechanism contributes to a low threshold carrier density, and
subsequently low threshold current densities (Coleman et al., 2011). In bulk material, when the temperature rises, the charge carriers are distributed continuously to higher energy states. In QDs, each individual dot is separated to neighboring dots by large intersubband separations on the order of 60–85 meV and a total separation of a few hundred meV from the dot ground state. The much suppressed carrier thermalization contributes to temperature-insensitive operation (Sugawara and Usami, 2009). In addition, the gain bandwidth of QDs can be engineered from broad to narrow and to emit at a wide range of wavelengths (Feng et al., 2019). The stochastic process of dot formation results in a Gaussian gain distribution which enables ultralow linewidth enhancement factors and isolator-free stability (Otto, 2014). The carriers injected into the dots are confined by a lateral energy barrier with a few hundred meV from the dot ground state. The reduced carrier lateral diffusion from the device center to the sidewall, and to dislocations generated during epitaxial growth, is important for future ultra-dense photonic integration up to 450 mm diameter Si wafer.

2.3 GaAs-on-Si epitaxy

For laser growth and processing up to 450 mm Si wafers, high crystalline quality III/V-on-Si metamorphic buffers are a prerequisite, in which the main challenge relates to dislocations generated from the mismatch of lattice constant, thermal expansion coefficient, and polar/nonpolar incompatibility between the III/V and Si. Nowadays, dislocation densities in the III-V materials on Si has been reduced to the level of $10^6$ cm$^{-2}$ using techniques such as two-step growth (Bolkhovityanov and Pchelyakov, 2008),

![Fig. 1 Variation in density of states with degrees of quantum confinement.](image-url)
strained-layer superlattices (El-Masry et al., 1988), thermal cycle annealing (Lee et al., 1987), relaxed and graded SiGe layers (Groenert et al., 2003), selective area or patterned growth (Li et al., 2016), and GaP buffer layers (Yaung et al., 2016). To maintain full compatibility with CMOS processing, microelectronics-standard nominal (001) Si substrates are preferred. While conventional heteroepitaxy on planar Si wafers needs a 4°–6° miscut angle to prevent APB formation, recent breakthroughs with no miscut angle have demonstrated comparable or even better performance (Wan et al., 2017c). These approaches range from patterned growth on exposed \{111\} V-groove Si facets (Wan et al., 2015), 0.15° misorientation in the [110] direction (Alcotte et al., 2016), direct nucleation of a GaAs film with special Si wafer preparation (Chen et al., 2017), growing an Al0.3Ga0.7As seed layer (Kwoen et al., 2018), the use of a thin GaP buffer layer (Jung et al., 2017a), and U-shaped patterned Si (001) substrates with \{111\}-faceted-sawtooth Si hollow structures (Wei et al., 2018). A comprehensive discussion of the physics of the growth mechanisms can be found in Fang et al. (1990). An overview of defect control in III/V-on-Si heteroepitaxy can be found in Kunert et al. (2018) and Li and Lau (2017).

### 2.4 Current status of QD lasers on Si

Despite the impressive progress of heteroepitaxy of III/V materials on Si, defect-free material or even material comparable to the state-of-the-art III-V wafers with defect level at $10^3–10^4$ cm$^{-2}$ remains elusive. This, to a large extent, deteriorates the reliability of conventional III-V devices on Si, particularly lasers. In fact, the most prolonged lifetime reported among GaAs-based QW lasers on Si after more than two decades of research is merely $\sim 200$ h (Kazi et al., 2001). On the contrary, for QD devices, with typical dislocation densities of $\sim 10^6–10^7$ cm$^{-2}$ in optimized III/V-on-Si buffers and typical dot densities of $\sim 6 \times 10^{10}$ cm$^{-2}$, the chance of charge carriers encountering a defect is far less than that of finding a dot and recombining radiatively. In addition, the strain field in QDs restricts the in-plane movement of dislocations. Evolution of dark line defects, which is the most critical mechanism responsible for rapid device failure (Ueda and Pearton, 2012), is thus effectively restrained. While the first QD laser grown on Si could only operate at 80 K under pulse biased conditions with a large threshold current density of 3.85 kA/cm$^2$ and a center wavelength at 1.038 $\mu$m (Linder et al., 1999), the performance has progressed rapidly in a short time frame. In 2011, 1.3 $\mu$m operation wavelength was reported (Wang et al., 2011). A threshold current density of 725 A/cm$^2$ with an
output power of \( \sim 26 \text{ mW} \) was achieved in broad-area lasers (3 mm \( \times \) 50 \( \mu \)m) at room temperature under pulsed injection. In this case, current was injected through the defective GaAs/Si interface with bottom contact placed on the Si substrate. Three years later, the pulsed lasing threshold for a 3 mm \( \times \) 25 \( \mu \)m broad-area laser was reduced to 200 A/cm\(^2\) (Chen et al., 2014). This was accomplished with an improved dislocation filtering by replacing In\(_{0.15}\)Ga\(_{0.85}\)As/GaAs with In\(_{0.15}\)Al\(_{0.85}\)As/GaAs strained layer superlattices, and an improved current injection scheme with top–top contact geometry to avoid current flow via the defective GaAs/Si interface. The maximum temperature under pulse injection was increased to be 111°C, with more than 100 mW of output power from a single cleaved facet. Using an intermediate Ge buffer nearly lattice matched to GaAs, the first CW lasing from InAs/GaAs QDs was reported with a maximum output power of 3.7 mW from both facets, a maximum operation temperature of 30°C, and a low threshold current density of 163 A/cm\(^2\) (Lee et al., 2012). Using a higher number of QD layers, narrow ridge waveguide structures, and high-reflection coatings, lasers grown on a similar Ge/Si substrates were demonstrated with low thresholds (16 mA), high output power (176 mW), high-temperature lasing (up to 119°C), and high characteristic temperature \( T_0 \) (>200K) (Liu et al., 2014). For the first time, over 2700 h of CW operation with no catastrophic failures at 30°C was reported, with extrapolated mean time to failures up to 4600 h (Liu et al., 2015a). This is much longer than any previous lifetime test of III–V lasers epitaxially grown on a Si substrate. Starting from 2016, efforts have been focused on developing CMOS compatible epitaxial platform on (001) Si, where III–V growth can be scaled up to 300-mm Si wafers (Alcotte et al., 2016). Exceptional laser performance has been realized with high single-side output power of 175 mW, low threshold currents below 1 mA, and extrapolated lifetime of more than 10,000,000 h at 35°C (Norman et al., 2018). Chronological evolution of InAs/GaAs QD lasers epitaxially grown on Si is shown in Fig. 2. Using QDs to achieve a commercially-viable monolithically integrated on-chip laser, with a push from the Si industry, is also boosted by the numerous performance advantages of QDs over QWs as described above.

3. Miniaturization of QD lasers using microcavities

Parallel to the success in engineering electronic confinement using QDs, optical confinement can be achieved with whispering-gallery-mode (WGM) resonators, enabling scalable, low-threshold, efficient light
sources that occupy a small volume on chips with compact sizes and high integration densities.

### 3.1 Whispering gallery mode microcavity lasers

The concept of WGMs was first proposed by Rayleigh in the study of the acoustic properties in St Paul’s Cathedral, where reflection of sound from walls forms high-quality resonators for acoustic waves to circulate around the wall (Rayleigh, 1910). In 1961, an optical analogue of the above architecture was realized by Garrett et al. (1961), where stimulated emission was achieved in a samarium-doped CaF$_2$ sphere. Since then, WGMs have been extensively and intensively investigated in a range of micro-sized systems used in a range of applications from sensing, frequency combs, optomechanics, quantum optics, to microcavity lasers (Nosich et al., 2007). For microcavity lasers, lasing can be achieved by the intrinsic nonlinearities of the resonator, or introducing active materials to the resonator (He et al., 2013). Resonators made from various liquid and solid materials such as water droplets, silica, two-dimensional materials, bulk semiconductors, organic semiconductors, epitaxial QDs, colloidal QDs, polymers, etc., have been reported. Various geometries including microdroplets, microspheres, photonic crystals, microtoroids, microdisks, microrings, microtubes, and microcylinders have been demonstrated (Song, 2019). In what follows, we will be interested primarily in QD microdisks and microrings where the resonator is formed by the total internal reflection along the circular perimeter, allowing both optical pumping and electrical injection, as well as output coupling to a waveguide.

![Figure 2: Chronological evolution of InAs/GaAs QD lasers epitaxially grown on Si.](image-url)
3.2 Microdisks and microrings: Major figure of merits

Microdisks and microrings are relatively thin (<1 μm) waveguide structures with a suspended disk or a surface ring. The refractive index contrast between the resonator material and the surrounding material contributes to strongly confined modes. The mode circulates the cavity in a closed-loop path, eliminating the necessity of mirror facets, gratings, or reflective coatings for optical feedback, which is beneficial for in-plane photonic integration. The two designs differ from one and another primarily in the schemes for vertical optical confinement. For microdisks with a suspended cavity, a relatively thin post was formed to connect the disk to the substrate by selectively removing the buffer layer consisting of InP in InGaAsP-based structures or of AlGaAs with high-Al compositions in GaAs-based structures. At least a half-wavelength of air gap was formed between the disk and substrate to avoid mode leakage into the substrate. For microrings with a circular mesa structure, the vertical optical confinement is obtained via a lower index semiconductor cladding at the bottom and a semiconductor-air refractive index contrast on the top surface. Generally, the suspended disks provide better optical confinement, and are more commonly adopted for optically pumped devices, while the microrings enable smaller thermal resistance and are more commonly used for electrically injected devices. In the following, we first define several important and key parameters of cavity-based light sources such as quality factor, spontaneous emission factor, threshold current, etc., and explain how critically the cavity design and carrier pumping/injection schemes are related.

3.2.1 Quality factor

Lasing threshold, wavelength, and spectral linewidth are the primary parameters to characterize a microcavity laser, and are all related to the volume \( V \) of the resonant mode and the quality factor \( Q \) of the laser cavity. While the \( V \) determines the level of light field confinement, \( Q \) defines the loss of the cavity. By examining the relation between the cavity loss and the power storage, the cavity \( Q \) can be deduced as the ratio of peak wavelength \( \lambda_c \) to the peak width \( \Delta \lambda \):

\[
Q = \frac{\lambda_c}{\Delta \lambda}
\]

In conventional “macroscale” optical resonators consisting of two or more mirrors, for example, a F-P cavity, the resonant condition implies that
there is an integer number of the wavelength of light in a round-trip optical path. Since the modes are reflected at near normal incidence, the $Q$ factor is generally below 100, and is given by

$$Q = \frac{2\pi n L}{\lambda (1 - R)}$$

where $L$ is the cavity length, $R$ is the reflectance, $n$ is the refractive index. From the equation, the $Q$ factor is inversely correlated to the mirror reflectivity, and proportional to the cavity length. To achieve a high light intensity confined in the resonator, high-reflectivity mirrors and small cavity size are required to reduce the optical loss as well as to form strong light confinement. This is challenging for mirror-based cavity constructions from both a technical and economical standpoint. On the contrary, microcavities can confine light to small volumes via resonant recirculation, with reflections at small grazing angles. While the $Q$-factor of microcavities is determined by a number of complicated mechanisms, including intrinsic loss from tunnel coupling of the confined mode to a continuum of propagating modes, and optical loss caused by absorption of the active layers and scattering mechanisms. This value is determined to a significant degree by the latter, and can be approximately as

$$Q = \frac{2\pi n}{\lambda \alpha}$$

where $\alpha$ is the optical loss (Kryzhanovskaya et al., 2014). In this case, by addressing the initial material quality and the microresonator fabrication process, cavities with $Q > 10^4$ can be fabricated using GaAs/AlGaAs materials (Kryzhanovskaya et al., 2015b). Characterizing the cavity $Q$-factor requires mode linewidth measured at transparency when stimulated emission exactly compensates the absorption. As an example, Fig. 3A shows emission spectra from a 4-µm diameter microdisk grown on Si with increasing pump power from a 532nm CW diode laser (Wan et al., 2016b). At a low pump intensity of 165µW, a weak cavity mode at 1308nm can be observed with a broad background emission. This sub-threshold emission can be fitted with bi-Lorentzian curves to extract the narrow cavity emission (Fig. 3B). The extracted cavity emission has a small full-width at half-maximum (FWHM) of 0.27nm (Fig. 3C), and $Q$ can be estimated to be $\lambda/\Delta\lambda = 4844$. It should be noted that this $Q$ is limited by the resolution limit of the spectrometer (0.07nm). Subject to a similar resolution limit of the instrument, a $Q$ factor of 3700 is obtained in a 1.2-µm diameter
GaN/InGaN microdisk (Tamboli et al., 2007). By conducting the measurement with a high spectral resolution of 30 pm in a cryostat environment at 78K, a much higher value of $C_20$ is obtained in a 7-μm diameter InAs QD microdisk (Kryzhanovskaya et al., 2015b).

### 3.2.2 Characteristic curves

To characterize the onset of lasing, L-L curve (laser output power versus pump power) and a L-I-V (light-current-voltage) plots are typically used, in optically pumped and electrically injected devices, respectively. Fig. 4 shows two typical L-L curves from a 4-μm diameter microdisk grown on Si (Wan et al., 2016c): a linear plot in (a) and a log-log plot in (b). Three regimes of lasing operation can be observed. The first regime of operation is dominated by spontaneous emission, in which the gain medium is excited well below the lasing threshold. In this region, the characteristic curve in the
linear plot doesn’t differ from that of a light emitting diode, and has a constant slope in the log-log plot. With an increase in pump power, the rate of stimulated emission becomes comparable to that of the spontaneous emission. This point denotes the transition to the second regime of operation, named the amplified spontaneous emission regime. In this region, the overall photons emitted are incoherent, with random phases and a significant spread in wavelengths due to the large number of spontaneous emission events involved. Both plots have an exponential increase in photon population, denoting a kink in the linear plot and a turning point of a “S” shape in the log-log plot. As the pump power is further increased beyond the threshold, the rate of the stimulated emission dominates over the spontaneous emission, with photons emitted coherently in phase and wavelength. Both plots show a linear dependence between the output power and pump power. Lasing threshold can then be determined by either extracting the kink in the linear plot, or the turning point of the “S” shape characteristics in the log-log plot, which is $175\,\text{μW}$ in this case.

### 3.2.3 Spontaneous emission factor

At this point, it’s important to introduce the spontaneous emission factor ($\beta$), which characterizes the amount of spontaneous emission into a single mode (Coldren et al., 2012). $\beta$ can be extracted by calculating theoretical values
and fitting them to experimental data. In the log-log plot of L-L curve in Fig. 4B, the rate equation model solutions for various values of $\beta$ are calculated by solving equations using a coupled rate equation model (Wang et al., 2015) for the carrier density $N$ and the photon density $P$:

$$\frac{dN}{dt} = \eta \cdot \frac{P_{in}}{\hbar \omega \cdot V} - V_g \cdot g \cdot P - B \cdot N^2 - C \cdot N^3$$

$$\frac{dP}{dt} = \Gamma \cdot V_g \cdot g \cdot P + \Gamma \cdot \beta \cdot B \cdot N^2 - \frac{P}{\tau_p}$$

Here, $\eta$ is the internal quantum efficiency, $P_{in}$ is the pump power, $V_g$ is the group velocity ($7.8 \times 10^9$ cm/s), $B$ is the bimolecular recombination coefficient ($2 \times 10^{-10}$ cm$^3$/s), $C$ is the Auger coefficient ($8 \times 10^{-17}$ cm$^6$/s), $\hbar \omega$ is the pump photon energy, $V$ is the volume of the active region, $\Gamma$ is the optical confinement factor (0.21), and $\tau_p$ is the photon lifetime (4.1 ps). Assuming a gain function of $g = g_0 \cdot \frac{N - N_t}{1 + \varepsilon \cdot P}$, where $g_0$ is the differential gain ($3 \times 10^{-16}$ cm$^2$), $N_t$ is the transparency carrier density ($1.23 \times 10^{18}$ cm$^{-3}$), $\varepsilon$ is the gain compression factor ($1 \times 10^{-17}$ cm$^3$), $\beta$ was extracted to be 0.1.

In conventional semiconductor F-P lasers or for any gas or solid-state laser, typical values of $\beta$ are in the range of $10^{-5}$ to $10^{-4}$. In such lasers, there is a distinct increase in output intensity with the onset of stimulated emission, similar to the scenario described above. It’s noted that $\beta$ is inversely proportional to the laser cavity size, and can be theoretically close to 1 when the cavity volume becomes so small, such that all the spontaneous emission occurs in a single optical mode. This case is termed a zero-threshold laser, where the L-L curve is linear without a clear distinction between the spontaneous and the stimulated emission region. As a matter of fact, for ultrasmall microdisk lasers where the diameter of the disk is smaller than the lasing wavelength (termed subwavelength), the lasing kink is gradually replaced with an increasingly smoother feature when $\beta$ increases. In Fig. 5A and B, laser output power and mode linewidth are plotted as a function of pump power for a subwavelength microdisk with a diameter of 1 $\mu$m grown on Si (Wan et al., 2016a). From the emission spectra (Fig. 5C), a weak cavity mode with a broad background emission appears at a low pump intensity of 16.5 $\mu$W (Fig. 5D). Upon surpassing the threshold, the cavity mode increases sharply in intensity, with a clear saturation of the background QD emission and a measurably larger signal, compared to the background optical emission. In Fig. 5E, emission below threshold (8.25 $\mu$W) was amplified by 100 times and overlaid with the emission well above threshold (660 $\mu$W),
Fig. 5 Laser characteristics for a 1-μm diameter microdisk grown on silicon: (A) output power and (B) mode linewidth against pump power; (C) Power dependent emission spectra; (D) emission spectrum below threshold (taken at a pump power of 16.5 μW). The black dots are measured data, the blue line is a fit to the broad photoluminescence background of the QDs; (E) Emission spectrum measured below (light blue, 8.25 μW) and above (red, 660 μW) the threshold. Black dots are measured data and blue lines are Gaussian fits. The spectrum below the threshold (light blue, 8.25 μW) was amplified 100 times to be visible (Wan et al., 2016a).
To highlight the optical transition positions in the gain medium. While lasing is evidenced with the pronounced narrowing of the lasing mode and the strong saturation of spontaneous emission with pumping, no well-defined transition can be observed from the L-L curve (Fig. 5A).

The well-established laser rate equations used for the 4-μm diameter microdisk described above in Fig. 4 is not sufficient to determine device behavior here (Strauf et al., 2006). Instead, β of the subwavelength microdisk laser can be extracted using a semiconductor cavity quantum electrodynamics (CQED) model (Gies et al., 2007). In Fig. 6A, L-L curve were fitted with the computed input/output curves for various values of β that bound the experimental data in the log-log scale (Wan et al., 2016a). In the model, for a given carrier injection rate into the QD states, equations of motion for carrier, photon populations, and photon correlations are solved numerically using the light-matter coupling coefficient and the dephasing rate as inputs to the model. The light-matter coupling coefficient can be computed from the active region and the optical resonator configuration, represented by a Rabi frequency of $6 \times 10^{10}$ s$^{-1}$. The dephasing rate was assumed with a value of 1 ps$^{-1}$ at 10 K. By analyzing both the onset of linewidth reduction and the bends in the “S” shape of the input/output intensity in the log-log plot (Fig. 6), β was extracted to be 0.3. This high β contributes to a low lasing threshold of 35 μW determined by a combination of characteristic L-L curve, linewidth narrowing profile, and strong saturation of spontaneous emission. While a truly zero-threshold laser hasn’t been demonstrated due to a number of practical problems not including fabrication issues, high-β can be obtained synthesizing a small modal

![Fig. 6](image)

**Fig. 6** Calculation and experiment for (A) input/output intensity and (B) the linewidth versus pump intensity log-log plot (Wan et al., 2016a).
volume, a high-quality factor of the mode, and an enlarged overlap of the single mode with the gain medium (Chow et al., 2014). To date, a record high $\beta$ of 0.95 has been reported in nanoscale coaxial cavities (Khajavikhan et al., 2012). A value of 0.85 has been reported on photonic crystal lasers sustained by only 2–4 QDs as the active gain material (Strauf et al., 2006). More information of the recent experimental progress in the control of $\beta$ by manipulating optical modes with nanocavities can be found in Noda et al. (2007).

3.2.4 Modal field distribution
WGMs are the eigenstates of a microcavity system and can be solved through the Maxwell equations for a given field $\{E, H\}$ and a wave vector $k$, given below:

$$\nabla^2 E + k^2 E = 0$$
$$\nabla^2 H + k^2 H = 0$$
$$\nabla \cdot E = 0$$
$$\nabla \cdot H = 0$$

Using the cylindrical coordinate system $(\rho, \phi, z)$ and arbitrarily choosing the longitudinal component $E_Z$ and $H_Z$ as the generating field, the wave function can be written as:

$$\psi(\rho, \phi, z) = P(\rho)\Theta(\phi)Z(z)$$

where $P(\rho)$ is the $m$th order Bessel function, $\Theta(\phi)$ is a Harmonic function of $\phi$, $Z(z)$ is an exponential decay outside the cavity. The wave function can thus be separated into three equations:

$$\frac{1}{\rho} \frac{d}{d\rho} \left( \rho \frac{d}{d\rho} P \right) + \left( (k_\rho \rho)^2 - m^2 \right) P = 0$$
$$\frac{1}{\Theta} \frac{d^2 \Theta}{d\phi^2} = -m^2$$
$$\frac{1}{Z} \frac{d^2 Z}{dz^2} = \beta_\perp^2$$

where $m$ and $\beta_\perp$ (mode index and propagation constant) are constants to solve the respective differential equations. By considering the boundary condition where $E_Z$ and $H_Z$ are finite at the origin and periodic with a period of $2\pi$ with respect to $\phi$ inside the cavity, $E_Z$ and $H_Z$ can be written in the following form inside the cavity.
$E_{Z1} = AJ_m(k_{\rho 1}\rho) \left( \cos(m\phi) \sin(m\phi) \right) e^{\beta z}$

$H_{Z1} = BJ_m(k_{\rho 1}\rho) \left( \sin(m\phi) \cos(m\phi) \right) e^{\beta z}$

$k_{\rho 1}^2 = k^2 + \beta^2$

and can be written as the following form outside the cavity

$E_{Z2} = CH_m(k_{\rho 2}\rho) \left( \cos(m\phi) \sin(m\phi) \right) e^{\beta z}$

$H_{Z2} = DH_m(k_{\rho 2}\rho) \left( \sin(m\phi) \cos(m\phi) \right) e^{\beta z}$

$-k_{\rho 2}^2 = k^2 + \beta^2$

In this way, the field is related to the azimuth angle $\phi$ by $\exp(\pm im\phi)$. The radial distribution within the cavity takes on the form of a Bessel function ($J_m$ of the first kind and the $m_{th}$ order), while the evanescent radial field outside the disk is described by the Hankel function ($H_m$ of the first kind and the $m_{th}$ order). WGMs can then be characterized by $(n, m)$ pairs where $n$ and $m$ are the radial and azimuthal mode numbers that describe the spatial field distribution in the radial direction and the azimuthal direction. While there are no analytical solutions to determine the eigenvalues of the mode, finite difference time domain method can be used to numerically solve the equations and study the WGM emission patterns. In a microdisk cavity, high-radial order modes generally have low quality factors and are easily lost via the underlying pedestal. Since gain material gives more TE gain, only the transverse electric (TE)-polarized modes with the first radial mode number will be considered. A computed 1st order WGM mode profile of the 1-\(\mu\text{m}\) diameter microdisk discussed above is shown in Fig. 7. Only one maximum in the radial direction is observed. Here, The TE$_{1,5}$ mode has been identified with a resonant wavelength at 1.194\(\mu\text{m}\), which matches well with the measured lasing wavelength at 1.197\(\mu\text{m}\) from the spectrum in Fig. 5C.

### 3.2.5 Free spectral range (FSR)

The azimuthal solution discussed above also offers a method to calculate the FSR of the WGM modes. To form a standing wave in the azimuthal direction, the wave must repeat itself upon one round trip around the circle. In this regard, the resonant condition in a disk of radius $R$ can be
approximated when the field equals to zero on the disk surface (Slusher et al., 1993), given by

\[ 2\pi R_{\text{neff}} = m\lambda \]

The resonant condition is achieved when there is an integer number \( m \) of waves with wavelength of \( \lambda/n_{\text{eff}} \) along the circumference of a waveguide with a cavity length of \( 2\pi R \). This can be used to derive FSR as a function of the circumference of the cavity, as

\[ \text{FSR} = \frac{\lambda^2}{2\pi R_{\text{neff}}} \]

FSR increases with the decrease of \( R \). This value can reach several tens of nanometers in micron-sized resonators. This is important for low-threshold lasers as reduced number of modes increases the spontaneous emission factor. Besides the cavity size, the FSR also decreases with the decrease of the mode wavelength, and an increase of \( n_{\text{eff}} \). The former implies more internal bounces within the cavity, and the latter is mostly influenced by the variation of temperature and pumping power for a given material and structure. In Fig. 8A and B, emission spectra from a 4-μm diameter microdisk were measured at both 10 and 300K (Wan et al., 2016c). To denote the optical transition positions in the gain medium, emission below threshold (black dots) was magnified 100 × and plotted together with that well above the threshold (red lines). At 10 K, four lasing peaks were observed. The sub-peak (1188 nm) close to the main peak (1191 nm) comes from
out-of-plane polarization while the other three well-separated modes (1162, 1191, 1220 nm) are WGMs with different azimuthal orders. At 10 K, FSR is extracted to be around 29 nm, corresponding to a $n_{\text{eff}}$ of 3.95. At room temperature, FSR increases to 35 nm, corresponding to a lower $n_{\text{eff}}$ of 3.84. The large FSR (compared to the FWHM of the gain media) results in well-separated neighboring resonances that favor single mode operation, which is the case at room temperature in Fig. 8B. At lower temperature in Fig. 8A, WGMs of various radial orders form their own sequences of resonances, which may overlap and complicate the optical spectrum with simultaneous emission of multiple modes. Let’s consider the modes as photon states: If the modes have high $Q$, then once a photon occupies a particular mode, it has a reasonably long life-time in that state. The gain spectrum emitted by QDs has a relatively broad luminescence peak because of the variation in size of the dots. Suppose the gain material emits a photon
at the exact energy where there is a mode: the higher the Q and the longer the lifetime of the mode, the more likely the photons will be seen at the mode energy when optically interrogating the microdisk. The role of the cavity (and its modes) is thus to selectively filter out photons of a certain energy. That is an important first step to lasing: using the cavity to selectively filter out a certain optical transition, and avoiding pumping energy into all other (ultimately non-lasing) transitions. In both cases of high injection and low temperature, there is more efficient production of photons which can interact with the higher order modes, but because these modes have very low Q, with a relatively high photon loss, the effect is still minimal on the spectrum and tends to disappear at room temperature. At room temperature, a histogram of the dominant lasing mode for a series of 4-µm diameter microdisks on Si is presented in Fig. 8C. The histograms are plotted together with a normalized photoluminescence spectrum from the as-grown material. The broad photoluminescence spectrum corresponds to optical emissions from the entire range of QDs, and the dominant modes denote optical transitions that are most efficiently coupled to the cavity, due to advantages in spatial overlap with a mode, radiative emission rate, or resonance frequency.

3.2.6 Characteristic temperature

Two figures of merit are primarily used to describe the laser performance at elevated temperatures: the maximum temperature, and the characteristic temperature, \( T_0 \). \( T_0 \) measures how much a laser’s threshold power \( (P_{th}) \) or threshold current \( (I_{th}) \), changes with temperature, and is defined by \( P_{th} = P_0 e^{T/T_0} \) and \( I_{th} = I_0 e^{T/T_0} \) for an optically pumped device and an electrically injected device, respectively. While temperature invariant operation \( (T_0 = \infty) \) in a range of 5–70°C (Fathpour et al., 2004) and maximum operating temperature up to 220°C (Kageyama et al., 2011) have been demonstrated for QD lasers with “macroscale” resonators on native substrates, the high aspect ratio of sidewall/active region volume deteriorates the temperature performance of microcavity lasers to some extent. Despite the fact that thermal conductance and injection pumping are difficult to be achieved in suspended disks, it is in this configuration that QD WGMs were first observed (Gayral et al., 1999). Among the limited research carried out on the temperature characteristics of InAs QD microdisk lasers on a native GaAs substrate, Ide et al. reported CW pumped micordisk lasers with a \( T_0 \) of 64K in the temperature range of 130–230K and of 36K above 230K (Ide et al., 2005), while Yang et al. demonstrated a \( T_0 \) of 31K above room temperature under pulsed excitation (Yang et al., 2007). A record-high \( T_0 \) of 105K was reported in a 4-µm diameter microdisk epitaxially grown on Si through
continuous optical pumping (Wan et al., 2016c). This was achieved by optimizing the lateral undercut for high light confinement within the disk while maintaining sufficient heat sinking via the underlying pedestal. Normalized lasing spectra of this microdisk laser were compared from 10 to 300 K in Fig. 9A, together with the corresponding L–L curves in Fig. 9B. In the spectra, narrow lines of high-\(Q\) WGMs shift to longer wavelengths with the increase of temperature, with a small temperature coefficient of less than 0.04 nm/K from 10 to 80 K, and from 220 to 280 K (Fig. 9C). The lasing line is also determined by the temperature shrinkage of an active region bandgap. Mode hopping toward longer wavelengths at 80 and 280 K was observed due to thermal redshifting of the gain spectrum. The overall trend follows the theoretical shrinkage line of InAs band-gap, given below (Heitz et al., 1999).

\[
\Delta E = A \cdot \frac{T^2}{T + B}
\]

assuming \(A = 0.00042 \text{eV/K}^2\) and \(B = 199 \text{K}\). In Fig. 9D, the threshold pump power increases by a factor of \(\sim 2\) as temperature increases from 10 to 300 K, giving a \(T_0\) of 105 K.

Ring resonators provide much improved thermal conductivity and are better suited to electrical injection than microdisks. A similar structure using ring geometry achieved a record-high lasing temperature of 107 °C and a \(T_0\) of 77 K under continuous optical pumping with a 6-\(\mu\)m outer diameter and a 2-\(\mu\)m inner diameter (Kryzhanovskaya et al., 2014). To achieve electrical injection in a QD microlaser, a microdisk was the first configuration where WGMs were observed at a temperature of 5 K, with a complicated air bridge structure (Zhang and Hu, 2003). However, the highest temperature to date was 300 K, achieved by burying the disk in benzocyclobutene (BCB) cladding to alleviate the thermal and fragility problem (Mao et al., 2011). On the contrary, microring QD lasers show much superior temperature performance, with the highest CW temperature up to 100 °C, and \(T_0\) of 197 K (Fig. 10) (Wan et al., 2017a).

In addition, the temperature characteristics have been improved to a great deal, attributed to special techniques such as tunnel injection of electrons and modulation \(p\)-doping of the active region (Bhattacharya and Mi, 2007). In tunnel injection, cold carriers (electrons) are tunneling injected into the QD lasing states so that other carriers are not heated. This leads to an enhanced \(T_0\) by minimizing carrier occupation in the states from wetting layer and barrier (Bhattacharya et al., 2003). \(P\)-modulation doping of the QDs is more generally utilized to reduce the sensitivity to temperature.
Fig. 9 Temperature dependent characteristics of a 4-μm diameter microdisk epitaxially grown on Si: (A) normalized lasing spectra from 10 to 300 K taken at three times the threshold. Inset: SEM image of the disk; (B) L-L curves of the lasing peak from 10 to 300 K. (C) Lasing wavelength versus temperature and theoretical InAs band-gap shrinkage; (D) Threshold power versus temperature (Wan et al., 2016c).
Typically, the separation between the ground state and first excited state band is up to 80 meV in the conduction, but is only around 10 meV in the valence band (Kageyama et al., 2011). At room temperature, while the electrons can be well confined in the dots, the holes can easily thermalize and escape, as the valence band offset is well below $kT$. This effect of thermalization can be suppressed by adding extra holes via $p$-type modulation doping, leading to increased temperature stability. As a direct comparison, two microring lasers were tested at various heatsink temperatures and their $L-I$ characteristics were analyzed in Fig. 11A and B, respectively (Wan et al., 2018a). For the laser in Fig. 11B, the GaAs barriers separating the QDs was modulation $p$-doped using beryllium with a hole concentration of $5 \times 10^{17}$ cm$^{-3}$, otherwise the two structures are nominally the same. In Fig. 11A, CW lasing was sustained up to 40°C for the undoped laser with a $T_0 \sim 22$ K. On the contrary, for the same laser structure grown on a separate wafer but with modulation $p$-doped active regions, CW lasing temperature was elevated up to 80°C, with a $T_0$ of $\sim 103$ K near room temperature (20–40°C) (Fig. 11B).

### 3.2.7 3dB bandwidth

The two techniques described above to improve temperature characteristics also increased the differential gain, leading to an enhanced modulation...
bandwidth. The small signal modulation bandwidth describes the laser frequency response and is proportional to the square root of the differential gain. The small-signal response, $S_{21}$ of ring lasers can be measured by directly probing the device using a signal/ground (SG) RF probe. The experimental set up is schematically shown in Fig. 12A.

Fig. 11 Temperature characteristics of two QD microring lasers grown on Si with an outer ring radius of 15 μm and ring waveguide width of 4 μm: L-I curves with varied heatsink temperature for (A) QDs with an intrinsic active region and (B) QDs with a modulation $p$-doped active region. (C) Threshold current versus heatsink temperature (Wan et al., 2018a).
**Fig. 12** (A) Measurement setup for small-signal modulation. (B) Small-signal modulation responses of the QD ring laser with an outer ring radius of 50 μm and ring waveguide width of 4 μm. Inset: $L$-$I$-$V$ characteristics (Wan et al., 2018a).
Fig. 12B shows the measured $S_{21}$ of a microring laser with an outer ring radius of 50 $\mu$m and ring waveguide width of 4 $\mu$m, under three injection currents (21, 54, and 86 mA). A 3dB bandwidth of 6.5 GHz can be attained at a bias current of 86 mA. Generally, a three-pole fitting function $H(f)$ is used to fit the $S_{21}$ curves to extract the damping rate $\gamma$ and relaxation oscillation frequency $f_r$ (Nagarajan et al., 1992).

$$H(f) = \frac{1}{\left(1 + \left(\frac{2\pi f \tau_p}{f_r^2}\right)^2\right)\left(f_r^2 - f^2\right)^2 + \left(\frac{\gamma f}{2\pi}\right)^2}$$

where $\tau_p$ is the RC or carrier transport delay. The extracted $f_r$, together with the measured $f_{3dB}$ can be used to extract the modulation efficiencies by plotting the resonance frequency versus $(I_b - I_{th})^{1/2}$. In Fig. 12B, modulation efficiencies of 0.38 GHz/mA$^{1/2}$ for $f_{3dB}$ and 0.34 GHz/mA$^{1/2}$ for $f_r$ were extracted by linear fitting, using the data points below $(I_b - I_{th})^{1/2} = 8$ mA$^{1/2}$ (Wan et al., 2018a). An analysis of the small signal response shows that the highest modulation rate attainable at the transition of a laser diode to conditions of the pronounced damping of relaxation oscillations is defined in terms of the so-called $K$ factor (Spencer et al., 1997).

$$f_{3dB,\text{max}} = 2\sqrt{2\pi/K}$$

Theoretically, the $K$ factor is the summation of contributions from the photon lifetime ($\tau_{ph}$) and gain compression ($\xi$), and can be represented as a function of optical loss ($\alpha$) (Zhukov et al., 2012).

$$K = \left\{ \begin{array}{l} K_{ph} = 4\pi^2 \tau_{ph} \\ K_{\xi} = 4\pi^2 \frac{\xi}{\nu_{gr} G_{\text{sat}}} \end{array} \right\} = \frac{4\pi^2}{\nu_{gr} G_{\text{sat}}} \left[ \frac{1}{\alpha} + \frac{\xi}{g_{\text{max}} \left(1 - \frac{\alpha}{G_{\text{sat}}}\right)} \right]$$

where $\nu_{gr}$ is the group velocity of photons, $g_{\text{max}}$ is the largest value of differential gain (at zero losses), $G_{\text{sat}}$ is the saturated gain. Due to the unequal carrier scattering rates of electrons and holes within the QD lasers, the gain compression in QDs is $>10 \times$ higher than that in QW lasers. This results in strong damping of relaxation oscillations, a high $K$ factor value, and a limited maximum (small signal) direct modulation bandwidth of $\sim 10$–15 GHz, as is illustrated in Fig. 13.
Experimentally, the $K$-factor can be derived by linear fitting the damping rate $\gamma$ as a function of squared $f_r$ using the following equation (Inoue et al., 2018a)

$$\gamma = K \cdot f_r^2 + \gamma_0$$

where $\gamma_0$ represents damping offset. In Fig. 14, the $K$-factor and a maximum $K$-factor limited $f_{3dB,\text{max}}$ was calculated to be 0.91 ns and 9.7 GHz, respectively.

Fig. 13 (A) Dependence of the $K$ factor (the solid line) and its components caused by the lifetime of photons (dashed line) and nonlinear gain saturation (dotted line) on losses for $G_{\text{sat}} = 50 \text{ cm}^{-1}$ and (B) the dependence of the highest modulation frequency on losses for various values of the saturated gain. The values $\varepsilon / g_{\text{max}} = 0.1 \text{ cm}$ and $v_{gr} = 0.9 \times 10^{10} \text{ cm/s}$ were used in the calculations (Zhukov et al., 2012).

Fig. 14 3dB bandwidth $f_{3dB}$ and relaxation oscillation frequency $f_r$ versus square-root of the bias current above threshold. Inset: Damping rate $\gamma$ versus squared relaxation oscillation frequency $f_r$ (Wan et al., 2018a).
The modulation bandwidth is at the same time affected by the device design and extrinsic electrical properties. To analyze the discrepancy between the theoretical $f_{3dB, \text{max}}$ and the measured $f_{3dB}$ reflection coefficient $S_{11}$ is generally used to extract the impedance and $RC$ cutoff frequency. A general equivalent circuit model (Bowers et al., 1986) consists of an inductance $L$, a total capacitance $C$, a device resistance $R$, a voltage source $V_s$, and a characteristic impedance of the transmission line (50 Ohm) $Z_0$ (Fig. 15). The inductance $L$ can be neglected if the electrical contact was performed directly by a RF probe without any wire bonding. The pad capacitance was calculated to be 0.33 pF by approximating the $p$-probe pad area with 80-μm width and 120-μm length over a 1-μm-thick SiO₂ layer (RF dielectric constant $\varepsilon = 3.9$). The total capacitance was extracted to be 0.74 pF from the 50-Ohm normalized Smith chart. $RC$ cutoff frequency is calculated to be 10.3 GHz using the circuit parameters ($C = 0.74$ pF, $R = 20.9$ Ohm). The extrinsic electrical properties can be improved by reducing the pad capacitance to tens of fF using thick dielectric passivation and dielectric embedded channel ridge structure (Kobayashi et al., 2013).

While 25 Gbit/s direct modulation has been reported for high-mesa waveguide $F$-$P$ lasers on native substrates (Tanaka et al., 2010), 12.5 Gbit/s NRZ operation and a 25 Gbit/s PAM-4 operation have been reported for a similar “macroscale” resonator epitaxially grown on a Si substrate (Inoue et al., 2018b). For microcavity lasers, the relatively low fiber-coupling output power makes it difficult to perform an eye-diagram characterization.

![Fig. 15](image_url) (A) A general equivalent circuit model. (B) Measured and fitted curves of $S_{11}$ under reverse (−3V) and forward (50 mA) biased condition from 0.14 to 5 GHz (Wan et al., 2018a).
Still, with the design of an asymmetric 2-tap equalizer (Fan et al., 2019), a recent report from Hewlett Packard Labs achieved 12 Gbit/s operation from a hybrid-integrated directly-modulated QD microring laser.

4. Designs of WGM resonators based on III-V materials

4.1 Optically pumped microdisks

A typical cross-sectional transmission electron microscope (TEM) image of a microdisk laser structure is shown in Fig. 16A. In the disk, the principal emission direction is radially in-plane. To avoid mode leakage into the substrate, the gap between the microdisk and substrate should be at least half of a wavelength in air, which is formed from a 600-nm Al$_{0.7}$Ga$_{0.3}$As sacrificial layer. The thickness of the disk was chosen so that it’s thick enough to incorporate an adequate number of QD layers to achieve enough gain and thin enough to be below the cutoff thickness of the second order planar waveguide mode. A five-layer InAs/InGaAs dot-in-a-well (DWELL) structure was adopted as active region, giving a total thickness of 500 nm (Wan et al., 2016a). At this thickness, the gain for the modes with polarization perpendicular to the disk is highly suppressed, so that only modes with polarization parallel to the plane of the disk will lase. To acquire high density of uniform QDs (typical density $\sim 6.5 \times 10^{10}$ cm$^{-2}$), the Stranski-Krastanov growth mode is the most common approach where self-organized islands

Fig. 16 (A) Cross-sectional TEM of a microdisk grown on Si; (B) Schematic of the DWELL; (C) High-resolution TEM image of the QD; (D) QD photoluminescence spectra at different pump power (Wan et al., 2016a).
are formed after a few monolayers of layer-by-layer growth (Wu and Jin, 2015). The DWELL structure (Fig. 16B), in which the dots are inserted in a QW so that the well captures a high density of carriers and keeps them localized around the dots, is widely used to promote effective capture of electrons and holes in the relatively small and spatially separated QDs (Stintz et al., 2000). For typical growth parameters used in molecular beam epitaxial (MBE) or metal–organic vapor phase epitaxy (MOVPE), the dots have lateral size of 10–40 nm and heights of 5–8 nm. An example of a TEM image of a single dot is shown in Fig. 16C, where the dot has a size of \( \sim 21 \text{ nm} \) in diameter and \( \sim 6 \text{ nm} \) in height. A useful feature of the QD size distribution is a reduction in the photoluminescence spectral linewidth. In the PL characterization from the as-grown QDs in Fig. 16D, the PL FWHM is as narrow as 28 meV centered at the 1.3 \( \mu \text{m} \) telecommunication band. Similar results have been reported by other groups, with a linewidth of 28 meV (at 300 K) in Guimard et al. (2010), a linewidth of 29 meV (300 K) in Chen et al. (2016), and a record-low linewidth of 17.5 meV (10 K) in Bhattacharya and Mi (2007), providing evidence of high-quality defect-free QDs, which are needed for the fabrication of devices.

Fig. 17A schematically shows one process for the fabrication of a QD microdisk resonator. A dilute suspension of SiO\(_2\) spheres in isopropyl alcohol (IPA) was first dispersed onto the sample surface by pipette and quickly dried on a hot plate at 110 °C. The diameters of the SiO\(_2\) spheres are varied from 1 to 5 \( \mu \text{m} \) within ±5% precision. Using the isolated sphere-masks, the cylinder mesa was etched to a depth of \( \sim 1.1 \mu \text{m} \) into the full structure using Inductively coupled plasma etching. It’s essential to acquire a clean and smooth circumference sidewall surface, a vertical profile, and a low degree of plasma-induced damage. In refining the etching, dominant parameters include the gas ratio, RF power, temperature, pressure, and selection of the etching mask. Among the different plasma chemistries used to etch GaAs-based compound semiconductor films, the preference for chlorine-based plasmas to etch Ga containing III–V materials is due to the high volatility of the Ga and group–V chlorides. Mixed BCl\(_3\) and Ar gases were used with flow rates of 20 and 6.5 sccm, respectively. The physical nature of BCl\(_3\) plasma helps to achieve a moderate etch rate due to the generation of less reactive Cl compared to Cl\(_2\), while Ar was added as an etching buffer which performs physical anisotropic etching to carry away the etching by-product as well as to keep the cavity plasma stable. Substrate temperature controls the volatility of the etching species, and was adjusted to be 10 °C. Low pressure causes the plasma to diffuse from the generation region and drift to the
substrate at relatively low ion energy and is therefore maintained at 5 mTorr. Coil and platen power are set as 500 and 100 W, respectively, to obtain anisotropic profiles. After the cylinder mesa formation, the remaining SiO₂ microspheres were removed using acetone in an ultrasonic bath, and the underlying pedestal was defined by a 5% HF solution, which has an extremely high selectivity in etching aluminum-rich AlGaAs in preference to the GaAs in the disk area. By controlling the time of wet etching, the lateral undercut of the disk from the outer periphery was optimized to maintain effective light confinement within the disk, as well as low mode leakage into the substrate. A fabricated SEM images of a 4-μm microdisk can be seen in Fig. 17B and C, revealing a smooth sidewall, and a circularity of the disk cavity. Devices can be measured through a microphotoluminescence (μPL) system in a surface-normal pump/collection configuration (Fig. 17D). Excitation from a 532-nm laser spot can be focused to approximately the size of the disk by a high-NA microscope objective, and the pumping intensity can be varied using a polarizer and monitored using a power meter. The collected signal can be dispersed by a monochromator and recorded by a liquid-nitrogen-cooled InGaAs detector array. The sample stage contains a helium gas flow cryostat, which can control the
temperature from 10 to 300 K. Fig. 17E summarizes the threshold powers obtained from a series of microdisk lasers on Si with different diameters. An apparent monotonic decrease of threshold power can be seen as the disk diameter is scaled from 5 to 1 μm (Wan et al., 2016a). A similar result of microdisk lasers grown on native substrate indicated that the threshold power density remains nearly constant when decreasing the diameter of the disk down to 1.7 μm (Kryzhanovskaya et al., 2015a).

4.2 Electrically injected microrings

While optical pumping is a critical step for demonstrating lasing and is a useful method for studying cavity and material properties, electrical injection is a necessity for commercial applications, especially for integrated photonics. The main fabrication challenge associated with electrically injected microrings is the requirement for metal contacts, which, if in direct contact with the cavity, represent a major source of optical loss. To address this issue, it’s important to acquire small metal contacts that maintain a low resistance and thermal impedance while not significantly perturbing the optical cavity. Contrary to the case of micropillars where vertical confinement requires complicated distributed Bragg reflectors (Feng et al., 2019), or the case of suspended microdisks where an airbridge contact contains a “double-disk” structure that separates the contact layer and the active layer (Zhang and Hu, 2003), the epitaxial structure of electrically injected QD microrings presented in the following sections do not differ from a typical design of GaAs/Al$_x$Ga$_{1-x}$As graded-index separate-confined heterostructure lasers.

Fig. 18 shows the critical processing steps in fabricating electrically injected microrings (Wan et al., 2017a). Using oxide hard mask, the ring structures were patterned and etched all the way to the n-contact layer. The sidewall of the ring was covered with 12-nm conformal Al$_2$O$_3$ by atomic-layer deposition, which provides excellent surface passivation.

![Fig. 18 Electrically injected microrings: typical process flow (Wan et al., 2017a).](image-url)
(Zhou et al., 2018), widely used to improve photoluminescence efficiency (Black et al., 2017) and reduce scattering loss (Ikku et al., 2012). $p$- and $n$-contacts were formed using Pd/Ti/Pd/Au and Pd/Ge/Au layer stacks, respectively. A 1 $\mu$m-thick SiO$_2$ layer was sputtered to further passivate the sidewall and to isolate the optical modes from the metal contacts. After via opening, $n$-contact pads were displaced laterally around the ring mesa and $p$-contact were arranged on top of the ring mesa for efficient current injection and spreading.

4.2.1 Typical characteristics

Typical L–I–V characteristics and spectra of a QD microring laser under CW injection are shown in Fig. 19. The cavity has a radius of 15 $\mu$m and a mesa width of 4 $\mu$m. The lasing threshold is around 3 mA. The primary lasing mode centered at $\sim$1.3 $\mu$m, with an extinction ratio over 30 dB and a side-mode suppression ratio of 16 dB in the injection range of 20–25 mA (Wan et al., 2018a). Despite the high ratio of sidewall to active region volume of the ring structure, a record low threshold current density of 306 A/cm$^2$ has been reported with a radius of 80 $\mu$m and a mesa width of 4 $\mu$m (Wan et al., 2018b). The optical power of the microring is generally

Fig. 19 (A) L-I, (B) I-V characteristics, and (C) emission spectra for a microring laser with a 15-$\mu$m radius and a 4-$\mu$m ring width (Wan et al., 2018a).
collected by capturing the radiation out-coupling at one side. While the angular directivity pattern of radiation brings about only a fraction of total power that can be measured, no output saturation was observed when the output power reached 1.2 mW at an injection current of 100 mA with a radius of 50 μm and a mesa width of 4 μm (Wan et al., 2017a). This power level is large enough for optical interconnects where sensitive photodetectors exist. With the same size QD microring laser, a maximum 3 dB bandwidth of 6.5 GHz has been reported with a modulation efficiencies of 0.38 GHz/mA^{1/2} (Wan et al., 2018a). A maximum CW temperature of 100°C has been achieved, with T_0 of 197 K between 10 °C and 50 °C and of 55 K between 60 °C and 100 °C (Wan et al., 2017a). Probably the most important advantage of using epitaxially grown QDs over heterogeneous QWs in a microcavity laser is their high temperature stability. Si is one of the most thermally conductive semiconductors, which is beneficial for dissipating heating. However, in the hybrid Si platform, a 7.5-μm diameter microdisk was produced from an InP-based QW epitaxial structure and then molecularly bonded to the SOI waveguide wafer. Fast thermal rollover is obvious even at a small injection current of 1.5 mA (Van Campenhout et al., 2007). An InP-based QW microring that sits on a silicon microdisk with the same diameter was reported through a low temperature O_2 plasma-assisted wafer bonding process (Liang et al., 2009). While the hybrid ring has a similar size (a radius of 50 μm and a mesa width of 4 μm) as the epitaxial QD microring described above (Wan et al., 2017a), lasing can only sustain up to 65 °C, and CW L-I measurement typically stop at 25 mA in order to avoid excessive device heating. The primary cause of serious device heating and inefficient heat dissipation in QW hybrid microcavity lasers is due to the thick buried oxide layer (typically 1–3 μm, with a 100 times smaller thermal conductivity compared to Si), and the inferior thermal tolerance of QWs, as compared to their QD counterparts.

Another distinguishing characteristic of QDs used by microcavity lasers is the reduced sensitivity to recombination at device sidewalls that enables ultra-dense photonic integration. Fig. 20 plot the threshold currents of the microring lasers as a function of the outer-ring radii. The threshold current decreases monotonically with the reduction of the ring radius. For the smallest dimension with a radius of 5 μm and a ring width of 3 μm, an ultra-low threshold of ~0.5 mA is obtained with single mode operation centered at ~1.29 μm.
5. Future outlook

The field of epitaxially integrated QD microcavity lasers is evolving quickly. However, compared to conventional linear cavity lasers, the total output power of the ring cavity is relatively low. Much research is needed to achieve single mode operation, wavelength-division multiplexing (WDM) sources, and other emission wavelengths. A systematic study of reliability and reproducibility at the device level is required to warrant further large-scale on-chip integration. In terms of integration via epitaxy, majority of research still focus on individual devices, particularly lasers. The performance of amplifiers, photodetectors, and modulators using the QD materials need to be demonstrated. Low-loss active-passive coupling schemes need to be developed to integrate these QD-based components with waveguides, preferably, in Si photonics. With that achieved, one can envisage the leverage of photonics for post-Moore performance scaling of electronic systems, and the intimate integration of photonics with electronics into a package. In the sections below, some possible avenues of future research to further advance the field will be outlined.
5.1 Increasing the output power

Generally, the output power of microcavity lasers is isotropic and can’t be extracted efficiently due to their rotational symmetry. These challenges are less acute for applications requiring low output powers. For example, in Stock et al. (2013), on-chip quantum optics experiments can be conducted using an electrically pumped micropost WGM microlaser to optically excite the nearby microresonators, despite a low lateral free-space outcoupling efficiency. In Wan et al. (2017d), proof-of-principle demonstrations of on-chip data transmission has been realized from a microring laser to an on-chip photodetector sharing the same InAs/GaAs QD active region. In Munnely et al. (2017), tunable emission energy is achieved on a QD micropillar-based single-photon source triggered by a monolithically integrated on-chip QD WGM microlaser. However, for more efficient data transmission, coupling to external waveguides with directional couplers or multimode interferometers (MMI) is necessary. Advanced device implementation methods need to be utilized, such as vertical coupling into a silicon-on-insulator waveguide via bonding (Van Campenhout et al., 2007), deformed microcavities (Harayama and Shinohara, 2011), lateral evanescent coupling to a straight waveguide (Lv et al., 2013), waveguide-coupled microspirals (Yang et al., 2014), and nano-antennas (Moiseev et al., 2017).

5.2 Single mode operation

In addition to output power, the resonant wavelength of a microcavity is sensitive to ambient temperature changes and fabrication variations. While shrinking the cavity size to a microscale volume helps to get single mode operation as the resonant frequencies are spaced by more than the gain bandwidth. Mode competition can be complicated at increased temperature or pumping, with simultaneous emission from multiple modes, as described previously. Actively tuning and locking the ring resonances to the desired wavelengths is important, and can be acquired with harnessing notions from parity-time symmetry (Hodaei et al., 2014), integrated second order gratings on the ring waveguide (Arbabi et al., 2015), and light confinement in subwavelength structures using metallic cavities (Feng et al., 2014).

5.3 WDM source

Microring cavities can also be very versatile because they can both serve as on-chip light sources with active regions based on QDs, and as
wavelength-selective modulators of the QD laser output based on passive silicon microrings. It’s widely recognized that WDM may be the only practical choice for very short reach inter-processor links in the future, and a single broad-spectrum light emitter replacing a group of DFB lasers appears to be the most competitive solution. The ultrafast gain recovery and reduced feedback sensitivity of QD-based gain media promise a lower jitter than expected compared to their QW counterparts (Rafailov et al., 2007). Broad, flat gain spectrum can be obtained via the inhomogeneous broadening from different size of QDs, and the deliberate broadening due to the use of several layers of QDs differing in center wavelengths. These features of QDs make them naturally suited for broadband emission sources in WDM applications. One example is QD mode lock lasers, where 4.1 Tb/s transmission using a single mode lock laser grown on Si was demonstrated (Liu et al., 2019). Alternatively, WDM transmitters can be built by cascading a simple QD F-P laser diode with microring modulator arrays, which simultaneously provide multiple functionalities including wavelength multiplexing/de-multiplexing and modulation. Experimentally, a DWDM transmitter based on a single QD comb laser and an array of microring modulators has been reported (Chen et al., 2015; Wojcik et al., 2009).

5.4 Emission at 1.55 μm

While the emission wavelengths of most of the QD lasers grown on Si have been limited to ~1.3 μm, developing InP-based QD lasers on Si with 1550 nm telecom wavelength is of immense technological interest. 1.55 μm emission from InAs quantum dashes (QDashes) grown on InP substrates can be achieved with InAs/InP material systems (Khan et al., 2014). Since the first demonstration of InAs QDash lasers grown on a InP substrate (001) (Wang et al., 2001), device characteristics have been quickly improved (Sadeev et al., 2016), with lowest threshold current density being 730 A/cm² (46 A/cm² per layer) (Hein et al., 2009). By exploiting ultra-broad gain profile of QDashes, various broadband optical components have been made (Ooi et al., 2008), such as optical amplifiers, mode-locked lasers, and superluminescent diodes (Khan et al., 2014). On native substrates, large ring lasers with millimeter-length have been reported with a threshold current density of 6.4 kA/cm² at near room temperature (283K) (Hill et al., 2008). CW lasing up to 55 °C has been reported with a threshold current density of 528 A/cm² in a 15-μm radius ring (Wan et al., 2019). Epitaxially grown on Si substrates, only optically pumped QD microdisk lasers have been
demonstrated with a threshold power of 1.6 mW (Shi et al., 2017b), and no electrical injected embodiment has been reported yet. While Shockley-Read-Hall recombination in InP related compounds is less severe than in GaAs so that even longer operating lifetimes are expected for InP-based QD lasers on Si, significant research is still needed to catch up with the state-of-the-art GaAs-based devices epitaxially grown on Si.

5.5 Other components employing ensembles of QDs

From the perspective of using QDs for improving other devices: QD solar cells offer high power conversion efficiency, where the partially occupied intermediate band of the confined QD levels results in additional sub-bandgap absorption in a single-junction cell (Popescu et al., 2008). QD modulators are shown to have enhanced electro-optic coefficients and steep absorption edges due to the quantum confined Stark effect (Akca et al., 2008). Experimentally, InAs QD cores in GaAs/AlGaAs epitaxial layers grown on Si have demonstrated 1.76 V-cm modulation efficiency, 2.2 V $V_\pi$, and less than 3 dB/cm propagation loss in a Mach-Zehnder modulator with 8 mm long electrodes (Bhasker et al., 2018). QD semiconductor optical amplifiers (SOAs) possess distinct superiorities compared to the QW counterparts at higher current levels due to the fast carrier dynamics and the decoupling of gain and phase dynamics. This property is manifested through comparative measurements with a commercially available QW SOA (Meuer et al., 2009), where an almost threefold increase of cross–gain modulation bandwidth is measured in QD SOAs (which can be tuned well beyond 40 GHz). The QD heterostructure can be engineered for the detection of photons in the midinfrared, far-infrared, and terahertz frequencies, with low dark current, intrinsic sensitivity to normal incidence light, and much longer lifetime of photoexcited electrons due to the reduced electron–phonon scattering (Bhattacharya et al., 2007). Sharing the same material with the QD–based lasers epitaxially grown on Si, monolithically integrated waveguide photodetectors (Fig. 21) operating in the O-band have been reported with an internal responsivity of 0.9 A/W, 10 Gbit/s eye opening with non–return–to–zero signals, and most notably, an ultra–low dark current density of 0.13 mA/cm$^2$ (Inoue et al., 2018c), which is over two orders of magnitude better than the state–of–art dark Ge–on–Si detectors (Michel et al., 2010). The waveguide QD photodetector can detect free–space coupled light from a nearby microring laser, both heteroepitaxially integrated on the same chip (Wan et al., 2018c). To more efficiently extract
the laser emission, in-plane coupling waveguide structures were designed, where the light can be evanescently coupled to the bus waveguide, and detected at the integrated photodiodes at both ends (Shang et al., 2018).

5.6 On-chip integration with epitaxial III-V on Si

Epitaxial integration can be achieved by coupling light generated in the QD section to either the Si waveguide in an SOI substrate, deposited foreign waveguides, or within the epitaxially deposited III-V layers.

5.6.1 Couple light to Si waveguide of the SOI substrate

From the perspective of co-integration of QD lasers with other Si photonic components that are defined in the top Si layer of a SOI platform: Liu et al. proposed that the optical mode can be butt-coupled to a Si rib waveguide by selective area growth of III-V cavity onto a prepatterned SOI substrate, and the optical coupling can be maximized by aligning the height of the active QD region with the Si waveguide layer (Liu et al., 2015b). In this way, the Si is used as both a substrate carrier and a low-loss wave-guiding material. However, the height alignment is challenging with a thick buffer, efficient electrical injection is also difficult as the material volume obtained is very limited. Buffer-less selective growth of III-V on Si is expected to simplify the coupling as the III-V active region is close to the underlying Si waveguide (Wang et al., 2015). Furthermore, the shape of the material growing out of the trench can be engineered with a box shape, where the lateral
width of the box is on the order of a micrometer, comparable to the typical width of a “classical” III-V laser (Shi et al., 2017a). Electrical injection is more feasible as the material volume obtained is larger. Using the integration approach discussed above, QW nano-ridge lasers have been demonstrated under pulsed optical pumping, covering the emission wavelengths of 1050 nm (Shi et al., 2017a) and telecom bands (Han et al., 2019). Incorporating QDs inside the cavity can further widen the range of achievable emission wavelengths, as well as improve the thermal performance to demonstrate CW lasing. However, for both selective area growth approaches discussed above, either through butt coupling of the QD layer to a SOI waveguide, or buffer-less approach where the active layers are within reasonable coupling distances to the underlying Si, the key obstacle is electrical injection. This requires doping of both the Si and GaAs layers, which need to go through highly defective III-V/Si heterointerfaces that are of large resistance. Furthermore, selective area growth relies on MOCVD, which has lagged substantially behind MBE in QD laser performance due to the inherent temperature limitations in the growth technique. To reduce the electrical and coupling issues, the III-V layers can be either directly grown on SOI wafers, or be bonded to SOI wafers after its growth on Si. Light coupling from the QD region to the Si part in SOI wafers can be achieved by either tapers, or photonic wire bonding. In the former, emitted light from a laser at 54.6° with a low divergence angle of 1.7° have been effectively directed onto a vertical coupler in the Si substrate for waveguide integration (Zhang et al., 2018b). In the latter, insertion loss down to 0.4 dB has been obtained between the InP-based horizontal-cavity surface-emitting lasers and the passive silicon photonic circuits with a direct-write two-photon lithography (Billah et al., 2018).

5.6.2 Couple light to deposited foreign waveguides
The light coupling scheme described above, using either butt coupling (Liu et al., 2015b), inclined coupling (Zhang et al., 2018b), or photonic wire coupling (Billah et al., 2018), can be used to couple light from the QD region to foreign waveguides, like crystalline Si, amorphous Si (a:Si–H), and SiN. It has been demonstrated that crystalline Si membranes, with reasonably low loss of ≤0.5 dB/cm, can be transferred onto foreign substrates using a liftoff process (Yuan and Ma, 2006), and be readily extended for the monolithic integration of single crystalline Si waveguides with III–V lasers and modulators (Mi et al., 2009). Alternatively, a:Si–H waveguides can be deposited by plasma-enhanced chemical-vapor
deposition with low temperature processing, tunable refractive index (De Dood et al., 2002), and relatively low losses of 2.0 and 5 dB/cm at 1.55 and 1.3 μm, respectively (Harke et al., 2005). With a:Si-H waveguides, on-chip integration of edge-emitting QD lasers has been outlined by evanescent coupling (Wan et al., 2017e) and has been experimentally demonstrated by butt coupling (Yang and Bhattacharya, 2008). Compared to Si, SiN has even lower propagation loss and temperature sensitivity that enables advanced wavelength multiplexing functionality through the high waveguide thickness and the low index contrast. Monolithic co-integration of QD-based InP on SiN has been proposed using interposer-based coupling scheme (Vyrsokinos et al., 2019). Using a novel Si-Rich Nitride (SRN) material with low stress and high refractive index (n > 3.16), light from the InP-based emitter can be transferred to the SRN intermediate waveguide and then guided to the Si$_3$N$_4$ photonic circuitry.

5.6.3 Couple light to epitaxially grown III-V waveguides
A final approach is to follow the established methods in monolithic integration, similar to photonic integration on native III-V substrate (Coldren et al., 2011), but with a major difference being that the III-V layers are grown on a Si substrate. While the relatively small index contrast provided by III-V waveguides compared to that of the SOI material system reduces optical confinement and dense integration prospects to some extent, just the application of Si as a cheap, robust, and large substrate is very advantageous, as the fabrication could benefit from a well-developed massive manufacturing ecosystem that is capable of producing tens of millions of wafers per year at the 300-mm wafer scale with near perfect yield in all fabricated components.

5.6.4 CMOS compatibility
From the perspective of future co-integration of the epitaxially grown QD lasers with Si CMOS: the III-V thermal budget is a challenge. One approach is to begin with the front-end-of-line of Si CMOS part, then cover the Si CMOS part and expose the area where III-V needs to be processed. Both the mask sets and the process window need to be well designed. Recently, there is a lot of effort from both electronic and photonic communities for III-V integration on Si. IMEC is investing heavily to III-V high mobility logic transistors on Si using aspect ratio trapping process, where the process modules are developed in a Si-III/V hybrid 300mm R&D pilot line, compatible for future CMOS high-volume manufacturing. Intel is currently
using a 300mm heterogeneous integration Si photonics foundry line for volume production of optical transceivers that include III-V semiconductor gain regions bonded onto a single large Si wafer (Jones et al., 2019). A similar optimization has already been carried out when introducing Ge-detectors in a standard CMOS-flow, as was demonstrated in industry (Schow et al., 2007). Introducing epitaxially grown III-V on Si into the CMOS line should be of a similar exercise in engineering, and some of the lessons there can be leveraged. As the field matures and prices come down, on-chip integration with epitaxial III-V on Si may become ubiquitous and affordable for mass-production.

6. Conclusions

The field of QD microresonators is advancing rapidly and promises to complement the rise of Si photonics by populating these chips with ultralow-threshold and small-footprint light sources. As the field matures and prices reduce, future epitaxial integration of these devices into optoelectronic circuits may become ubiquitous and affordable for mass-production.

Acknowledgements

The authors thank Advanced Research Projects Agency-Energy (ARPA-E) (DE-AR0000672), American Institute for Manufacturing (AIM) Integrated Photonics, Research Grants Council of Hong Kong (RGC) (No 16212115), Innovation Technology Fund of Hong Kong (ITS/320/14), for funding this research. Important contributors to the epitaxial growth on Si research described here include Arthur C. Gossard, Roger Helkey, Evelyn L. Hu, Kei May Lau, Alan Y. Liu, Daehwan Jung, Qiang Li, Chen Shang, Daisuke Inoue, M. J. Kennedy, Di Liang, Chong Zhang, Songtao Liu, Paolo Pintus, Zeyu Zhang, Kaiyin Feng, Noelle Collins, and Mario Dumont.

References

Alcotte, R., Martin, M., Moeyaert, J., Cipros, R., David, S., Bassani, F., Ducroquet, F., Bogumilowicz, Y., Sanchez, E., Ye, Z., Bao, X.Y., 2016. Epitaxial growth of antiphase boundary free GaAs layer on 300 mm Si (001) substrate by metalorganic chemical vapour deposition with high mobility. APL Mater. 4 (4), 46101. AIP Publishing.


Further reading