

High-Power Mid-Infrared ($\lambda \sim 3\text{-}6\ \mu\text{m}$) Quantum Cascade Lasers

Luke J. Mawst[✉], *Fellow, IEEE*, and Dan Botez[✉], *Life Fellow, IEEE*

(Invited Paper)

Abstract—The performances of mid-infrared (IR) quantum cascade lasers (QCLs) are now reaching a maturity level that enables a variety of applications which require compact laser sources capable of watt-range output powers with high beam quality. We review the fundamental design issues and current performance limitations, focusing on InGaAs/AlInAs/InP QCLs with emission in the 3–6 μm wavelength range. Metamorphic materials broaden the available compositions for accessing short emission wavelengths ($\lambda \leq 3.5\ \mu\text{m}$) or for integration with GaAs- and Si-photonics platforms. Conduction-band engineering through the use of varying compositions throughout the active-region structure has been utilized to achieve the highest performance levels to date. Interface roughness scattering plays a dominant role in determining both the lower-laser-level lifetime as well as the carrier-leakage current. Numerous approaches have been implemented in attempts to control, scale, and stabilize the spatial mode to high output powers. Of all approaches photonic-crystal structures with high built-in index contrast, thus capable of maintaining modal properties under strong self-heating, are the most promising device configuration for achieving single-spatial-mode, single-lobe reliable CW operation to multiwatt-range power levels. Such devices have demonstrated to date $>5\text{W}$ front-facet output powers with diffraction-limited beams in short-pulse operation.

Index Terms—Semiconductor lasers, quantum cascade lasers, mid-infrared lasers.

I. INTRODUCTION

HIGH-output-power lasers emitting in the mid-IR range (i.e., 3–6 μm) are currently of great interest for civilian- and defense-sector applications such as remote sensing of environmental gases, stand-off spectroscopic sensing of toxic-chemical agents, free-space optical communications, directed infrared countermeasures (DIRCM) and LIDAR. Quantum cascade lasers (QCLs), first demonstrated by Faist *et al.* in 1994 [1], have shown great potential for such applications. While there have been impressive advancements made in terms of maximum output power, the overall CW power-conversion efficiency at room temperature, $\eta_{\text{wp,CW}}$, remains well below that of near-infrared diode lasers; that is, $\eta_{\text{wp,CW}} \sim 21\%$ for QCLs

[2] vs. $>70\%$ for diode lasers [3]. In addition, there is still a lack of sufficiently powerful, compact sources of high beam quality, which has prevented the use of QCLs for many applications. High-output power ($>1\text{ W}$ CW) diode lasers have been realized [4] for wavelengths $< 2.5\ \mu\text{m}$. However, as the emission wavelength increases beyond $\sim 2.5\ \mu\text{m}$ the corresponding lower bandgap active-region materials lead to inherently higher non-radiative carrier losses via Auger processes and subsequently very low CW output powers. By contrast, intersubband-transition devices, such as QCLs, decouple the active-region material bandgap from the emission wavelength, and instead rely on quantum size effects to define the transition energy. Furthermore, for transitions in QCLs Auger effects are negligible. Unlike conventional (bipolar) diode lasers, QCLs are unipolar devices, which generally involve electron transitions in the conduction band. This fact provides great flexibility for conduction-band engineering to develop and optimize QCL designs, while ignoring the valence band, as discussed below in Section II. We focus this review on the $\text{In}_x\text{Ga}_{1-x}\text{As}/\text{Al}_z\text{In}_{1-z}\text{As}/\text{InP}$ material system, since the highest performance QCLs emitting in the 3–6 μm spectral region have been obtained by employing this material system. The quantum well/barrier compositions and strain values (on InP substrate) dictate the conduction-band energy offsets, which in turn determine the quantum-confined energy levels and the associated electronic wavefunctions. However, the design of such device structures is generally limited to a compositional space by strain-relaxation considerations for the superlattice (SL) materials comprising the active region, leading to a significant fraction of the compositional space being largely inaccessible, except for use employing very thin layers ($\sim 1\text{ nm}$). To help circumvent the strain-relaxation issue, the use of a “virtual substrate” can be employed to significantly broaden the pallet of materials accessible for QCL applications. This approach has been identified as particularly advantageous for accessing the short-wavelength portion ($\lambda \leq 3.5\ \mu\text{m}$) of the mid-IR spectral region. However, such “lattice-mismatched” materials still present many challenges for implementation into high performance devices, as discussed below in Section III.

While the $\text{In}_x\text{Ga}_{1-x}\text{As}/\text{Al}_z\text{In}_{1-z}\text{As}/\text{InP}$ material system is relatively mature and well characterized for application to diode lasers, many of the material properties relevant to QCLs are less well established and are subject to the subtleties of the crystal-growth methods employed. Being intersubband-transition devices relying on the quantum-mechanical tunneling of electrons,

Manuscript received October 17, 2021; revised November 27, 2021; accepted November 30, 2021. Date of publication December 2, 2021; date of current version January 13, 2022. This work was supported in part by AFOSR under Grant FA9550-19-1-0385, NSF ECCS 1806285, Navy contract N68335-19-C-0196, and ARO under contract W911NF16C0116. (Corresponding author: Luke J. Mawst.)

The authors are with the University of Wisconsin-Madison, Madison, WI 53706 USA (e-mail: ljmawst@wisc.edu; botez@engr.wisc.edu).

Digital Object Identifier 10.1109/JPHOT.2021.3132261

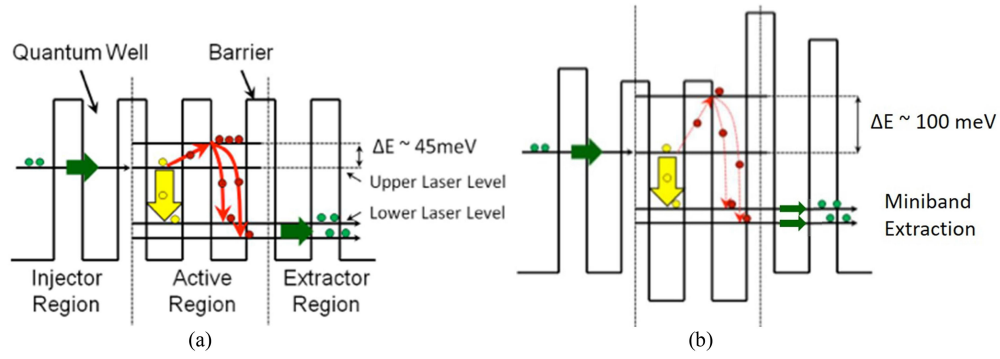


Fig. 1. Schematic representation of the conduction-band energy diagram for one stage of: (a) conventional QCL structure; (b) the step-tapered active-region (STA) QCL structure. Vertical arrows indicate the lasing transition.

QCLs are inherently sensitive to interfacial properties such as compositional grading and roughness. The electronic-states lifetimes determine the population inversion [5] and characterize the carrier-leakage mechanisms [6]. The carrier lifetimes themselves are determined by inelastic (i.e., LO-phonon scattering) and elastic scattering (i.e., alloy-disorder (AD) and interface-roughness [IFR] scattering [5], [6]). While LO-phonon and AD scattering determine the upper-laser (*ul*) level lifetime, IFR scattering generally determines the lower-laser (*ll*) level lifetime [6]. In addition, IFR scattering is key in triggering carrier leakage from the *ul*-level and injector-region states [6]. Finally, interfacial compositional grading leads to deviations of the emission wavelength from design targets based on “square well” approximations.

Conventional QCLs are devices that rely on electron relaxation between the quantized states in a superlattice of quantum wells (QWs) and barriers of *fixed* compositions. As a result, there is severe carrier leakage via electron thermal-excitation from the *ul* level (and injector-region states) to higher energy active-region (AR) levels *and* relaxation to low energy AR levels (see Fig 1(a)).

This leakage is a strong function of the energy difference between the *ul* level and next higher AR level (ΔE in Fig. 1): $\propto \exp(-\Delta E/kT)$. By contrast, the step-tapered active-region (STA) QCL [7], [8], schematically shown in Fig. 1(b), consists of stepwise tapering the barrier heights in the AR such that their conduction-band offsets increase in energy from the injection to the exit barriers. This causes more than doubling the ΔE value; thus, virtually suppressing carrier leakage. In turn the devices’ electro-optical characteristics become much less sensitive to temperature, thus allowing for significantly more powerful and efficient CW operation. Furthermore, since QCL-device failures are directly related to the device-core temperature rise [9], carrier-leakage suppression will also lead to dramatic improvements in reliability at high (≥ 1 W) CW powers. Such 4.6-5.0 μm -emitting STA QCLs have demonstrated significant advantages in device design offered by the flexibility to control, via metal-organic chemical vapor deposition (MOCVD) crystal growth, the composition (and strain) of each layer within a QCL structure. Such flexibility in crystal growth can be obtained with gas-source molecular beam epitaxy (GS-MBE) and has been used to design linearly-tapered active-region QCLs [2], so called tapered-active (TA) QCLs [7], [8] which, however, inherently

have lower ΔE values than STA-type device [7]; thus, higher carrier leakage [6].

Scaling the coherent power of mid-IR QCLs to the multi-watt range remains a significant and important objective for applications where the laser beam needs to travel through air to remote targets in real-world environments. Single-element, edge-emitting QCLs operating in the important 4.0-6.0 μm wavelength region generally necessitates a relatively narrow emitter width ($\leq 4.5 \mu\text{m}$) to maintain stable, single-spatial-mode CW operation up to watt-range output powers. Utilizing a single-stripe buried heterostructure QCL [10], the highest reported single-spatial-mode CW output power at a 4.5 μm emission wavelength is 1.4 W. Higher CW output powers (~ 5 W) have been achieved from wider aperture BH devices [2], although at the expense of multi-spatial-mode operation, generally characterized by beam steering with increasing drive level [11]. External cavity beam combining many single-emitter QCLs offers a path to scaling the coherent output power of a mid-IR emitting source of high brightness. However, in many applications monolithic approaches to power scaling are preferred, because monolithic sources are more compact, less expensive, more rugged and thus more reliable. Furthermore, any improvements in the single-spatial-mode output powers of individual monolithic QCLs can be directly utilized in externally beam-combined systems.

Unlike diode lasers, QCLs exhibit a maximum operating current density (J_{max}) which is dependent on the injector doping level, and is typically in the range of $3\text{-}4 \times$ the threshold-current density, J_{th} . Thus, since the power scales with the number of periods, N_p , the maximum output power at J_{max} is ultimately limited by the core-region volume, defined by the product of N_p , the period thickness, and the device area. Longer cavity length can be used to scale the area, although internal losses will generally limit the practical cavity lengths which can be used without incurring a significant reduction in slope efficiency, η_{sl} . The number of core-region periods can be increased for optical gain, but is constrained by thermal-conductance considerations. Increasing the emitter width is limited by the onset of multi-spatial-mode operation as well as the effectiveness of heat removal in CW operation. Self-heating under QCW/CW operation has been linked to QCL failures, which are quite distinct from those in diode lasers, for which optical absorption at the laser facet can induce catastrophic mirror damage. New QCL

architectures are being developed to increase the active volume of the device and scale the brightness over current state-of-the-art single-element, single-mode QCLs, as discussed in Section IV. These new architectures need to address strong self-heating in CW operation that results in thermal lensing which can trigger beam instabilities, like beam steering and multi-mode operation, as well as present challenges for watt-range reliable operation. Thus, new methods for stabilizing the optical mode in CW operation without introducing significant penalty in optical loss, thermal conductance, and reliability are needed.

Section II describes the QCL performance limitations under short-pulse and CW operation, as well as a review of state-of-the-art performance levels in the 3.8-6.0 μm wavelength region. The emphasis is on the roles of conduction-band and IFR-scattering engineering on maximizing device performance. QCL reliability and observed failure modes are also discussed in Section II. Section III reviews QCL material constraints imposed by strain limitations, and the use of lattice-mismatched “metamorphic” materials for expanding the accessible QCL compositional space. The use of a lattice-engineered growth platform, or “virtual substrate”, enables short-wavelength ($\lambda < 3.5\ \mu\text{m}$) QCL designs with potential for reaching the performance levels of their longer-wavelength counterparts. The application of such materials to realize QCLs integrated with Si- or GaAs-based photonics is also reviewed. Section IV reviews the limitations for high-brightness QCLs, focusing on device concepts used for scaling and maintaining single-spatial-mode operation to watt-range output powers. Finally, conclusions and future directions are discussed.

II. DEVICE CHARACTERISTICS AND PERFORMANCE OF 3.8-6.0 μm -EMITTING QCLs

A. Pulsed Operation

1) *Threshold-Current Density*: Assuming that the tunneling-injection efficiency $\eta_{\text{inj,tun}}$ is basically unity, which is generally the case, the threshold-current density J_{th} is given by [8]:

$$J_{\text{th}} = \frac{\alpha_m + \alpha_w}{\Gamma g} + J_{\text{bf}} + J_{\text{leak,tot}} = \frac{\alpha_m + \alpha_w + \alpha_{\text{bf}}}{\eta_p \Gamma g} \quad (1)$$

where α_m and α_w are the mirror and waveguide losses, respectively, J_{bf} is the current corresponding to thermal backfilling of the ll level [12], α_{bf} is a loss term that is used to represent backfilling, Γ is the optical-mode confinement factor, g is the differential gain as defined in [8], and η_p is called pumping efficiency which reflects the degree of carrier-leakage suppression [6]:

$$\eta_p = 1 - J_{\text{leak,tot}}/J_{\text{th}} \quad (2)$$

where $J_{\text{leak,tot}}$ is the sum of leakage-current densities triggered from the ul level and injector states, respectively, caused by LO-phonon and IFR scattering [6].

For conventional QCLs at room temperature (RT) η_p takes values of ~ 0.70 [13] and is in the 0.80-0.90 range for state-of-the-art 4.5-5.0 μm -emitting QCLs [6]. Therefore, η_p is a key factor in determining the J_{th} value, especially under CW operating conditions [7], since the leakage currents strongly increase with increasing temperature [6], [14]. As schematically

shown in Fig. 1, and further discussed below in Section II-B, for state-of-the-art QCLs carrier leakage mostly consists of scattering from the ul level and key injector states to the next higher energy AR state, level $ul+1$, followed by relaxation to lower AR energy states. That is, for most QCLs carrier leakage is primarily a shunt-type current through the next higher AR energy state, *not* leakage to the continuum. That is why η_p has to be considered as part of the total injection efficiency (i.e., $\eta_{\text{inj}} = \eta_{\text{inj,tun}}\eta_p$) [8], just like for interband-transition devices [15].

The J_{th} temperature dependence is defined by a characteristic temperature coefficient T_0 :

$$J_{\text{th}}(T_{\text{ref}} + \Delta T) = J_{\text{th}}(T_{\text{ref}}) \exp\left(\frac{\Delta T}{T_0}\right) \quad (3)$$

where $T_{\text{ref}} + \Delta T$ is the heatsink temperature, T_{ref} is the reference heatsink temperature, and ΔT is the range in temperature over which the J_{th} increases in an exponential fashion. Looking at (1), T_0 is determined by the terms that have a strong dependence on temperature: J_{bf} and $J_{\text{leak,tot}}$ [12], [6]. As far as $J_{\text{leak,tot}}$, it has been shown [6], [14] that as long as the energy difference between the $ul+1$ and ul levels, $E_{ul+1,ul}$ (i.e., ΔE in Fig. 1) ≥ 50 meV both LO-phonon- and IFR-scattering triggered carrier leakage primarily occur thorough the $ul+1$ level. Then, it can be shown [6], [16], [17] that:

$$J_{\text{leak,tot}} \propto \exp\left(-\frac{E_{ul+1,i}}{kT_{ei}}\right) \quad (4)$$

where i stands for either the ul level or an injector-region state, and T_{ei} is the electron temperature in the energy state i [6]. Thus, as the $E_{ul+1,ul}$ value increases the leakage-currents' relative increase with temperature sharply decreases and, in turn, T_0 increases. Moderately-high doped conventional 4.5-5.5 μm -emitting QCLs, as needed to reach 1 W CW power, suffered from severe carrier leakage due to small (~ 45 meV) $E_{ul+1,ul}$ values, and, in turn, that was reflected in low T_0 values: 130-150 K [12]. Increasing the $E_{ul+1,ul}$ value was achieved via conduction-band (CB) engineering: deep-well QCLs [18]; tapered active-region (TA) QCLs, for which the barrier heights increase linearly across the AR [19], [11], [20], and finally step-tapered AR (STA) QCLs, for which both the barrier heights and quantum-well depths increase stepwise across the AR [7], [8]. Those designs progressively led to high $E_{ul+1,ul}$ values, reaching the 100-120 mV range [8], [6] and to T_0 values, for moderately-high doped devices, in the 230-250 K range. As carrier leakage was suppressed, the temperature sensitivity of the slope efficiency also dramatically decreased (see next subsection), which led to maximum RT CW powers in the 2.6-5.1 W range [8], [12].

2) *External Differential Efficiency*: The external differential efficiency η_d , at and close to threshold, for a QCL having N_p periods is given by [5], [8]:

$$\eta_d = \eta_i \frac{\alpha_m}{\alpha_m + \alpha_w} N_p \quad (5a)$$

where η_i is the internal differential efficiency per period which can be well approximated by [8]:

$$\eta_i \cong \eta_{\text{inj}}^{\text{tun}} \eta_p \eta_{\text{tr}} = \eta_{\text{inj}} \eta_{\text{tr}} \quad (5b)$$

where the η_p term within η_{inj} stands for the differential pumping efficiency above threshold [21]; that is, $\eta_p = \Delta(J - J_{\text{leak,tot}})/\Delta J$,

where J is the total current density and $J - J_{\text{leak,tot}}$ is the ‘useful’ current density flowing through the upper laser level. Then, for relatively small variations in J above lasing threshold, the differential pumping efficiency is the same as the pumping efficiency at threshold (i.e., $1 - J_{\text{leak,tot}}/J_{\text{th}}$) [8], [12], just as is the case for interband-transition lasers [15]. As for η_{tr} , the differential lasing-transition efficiency [5], it is given by:

$$\eta_{\text{tr}} = \frac{\tau_{\text{up},g}}{\tau_{\text{up},g} + \tau_{\text{ll},g}}; \tau_{\text{up},g} = \tau_{\text{ul},g} \left(1 - \frac{\tau_{\text{ll},g}}{\tau_{\text{ul-ll},g}} \right); \quad (5c)$$

where $\tau_{\text{up},g}$ and $\tau_{\text{ll},g}$ are the global ‘effective’ ul -level lifetime and global ll -level lifetimes, respectively [7]; while $\tau_{\text{ul},g}$ and $\tau_{\text{ul-ll},g}$ are the global ul -level and lasing-transition lifetimes [7].

In order to achieve high output power and high wall-plug efficiency the η_d value has to be maximized. While the α_w value has been reduced to values as low as $\sim 0.5 \text{ cm}^{-1}$ at $\lambda = 4.6\text{--}4.9 \text{ }\mu\text{m}$ [12] and N_p was raised to 40 periods, the η_i values were found to be rather low for conventional-type QCLs: 50–60% in the $4.5\text{--}6.0 \text{ }\mu\text{m}$ wavelength range [12] with no clear explanation why that was the case.

On the one hand, in order to increase η_{inj} , carrier leakage was strongly suppressed by using CB engineering such as step-tapering the AR barrier heights and well bottoms [8] which has resulted in extracted η_{inj} values as high as 84.5% [6]. Linear tapering of the AR barrier heights in so-called shallow-well QCLs [11], which were shown [7] to be (linear) tapered-active (TA)-type QCLs, has resulted in an extracted η_{inj} value of 75% [6]. On the other hand, in order to increase η_{tr} , fast, effective carrier extraction from the AR, so-called miniband extraction, was employed [8] by extracting carriers from both the ll -level and one energy level below it (see Fig. 1(b)), called resonant extraction (RE) [8]. Taking into account elastic scattering, for $5 \text{ }\mu\text{m}$ -emitting STA-RE [8] and $4.9 \text{ }\mu\text{m}$ -emitting TA-RE QCLs [2], has resulted in enhancing η_{tr} by 9–10% to values in the 91–93% range [6]. In turn, for $\sim 5 \text{ }\mu\text{m}$ wavelength STA-RE QCLs η_i values as high as 77% were obtained [8]; that is, 30–50% higher than for conventional QCLs emitting in the $4.5\text{--}6.0 \text{ }\mu\text{m}$ wavelength range. Recently, a 45-period, $\sim 4.9 \text{ }\mu\text{m}$ -emitting TA-RE QCL [22] has demonstrated an η_i value of $\sim 76\%$. Even more recently, $4.05 \text{ }\mu\text{m}$ - and $4.6\text{--}4.7 \text{ }\mu\text{m}$ -emitting STA-RE QCLs have been designed [23], [6] with large $E_{\text{ul}+1,\text{ul}}$ values: 97 meV and 120 meV, respectively, which, in turn, resulted in projected η_i values of 79% and 86%, respectively. Therefore, by combining carrier-leakage suppression with highly effective carrier extraction, η_i values are steadily increasing towards upper limits of $\sim 90\%$ for mid-IR QCLs [6], [8].

We show in Fig. 2 a comparison between experimentally obtained η_i values, for various QCL types, over the $4\text{--}6 \text{ }\mu\text{m}$ wavelength range. Devices with both carrier-leakage suppression and miniband extraction have significantly higher η_i values than conventional QCLs. That is, STA- and TA-RE QCLs have reached η_i values of $\sim 77\%$ at $\lambda = 5 \text{ }\mu\text{m}$; and $\sim 70\%$ and $\sim 76\%$ at $\lambda = 4.9 \text{ }\mu\text{m}$, respectively, compared to 50–60% for conventional QCLs. The $\sim 62\%$ value at $\lambda = 5.65 \text{ }\mu\text{m}$ was obtained from a two-QW design QCL [24] which, while having carrier-leakage suppression (i.e., $E_{\text{ul}+1,\text{ul}} = 88 \text{ meV}$), has an apparently low $\eta_{\text{inj,tun}}$ value due to parasitic injection from the injector ground level to the lower laser level as well as a relatively

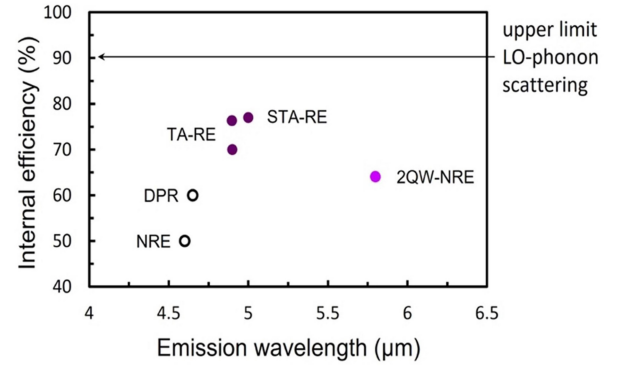


Fig. 2. Internal efficiency as a function of emitting wavelength over the $4\text{--}6 \text{ }\mu\text{m}$ range. The data points correspond to references given in [8] and to [22]. Dark purple disks and empty circles correspond, respectively, to QCLs with both carrier-leakage suppression and miniband extraction, and to conventional QCLs. The light purple disk indicates a QCL with only carrier-leakage suppression [24]. Adapted with permission from [8] © at The Optical Society.

low η_{tr} value, typical of nonresonant-extraction (NRE) QCLs [5]. The upper-limit η_i value of 90% is obtained by considering an ideal LO-phonon-relaxation case when the lasing transition occurs to the top of a single-subband miniband, instead to a discrete energy state, thus, providing very short ($\sim 0.1 \text{ ps}$) lower-level lifetimes [25], a generic 1.0 ps LO-phonon effective upper-state lifetime (i.e., $\eta_{\text{tr}} \sim 90\%$), no carrier leakage and unity tunneling-injection efficiency. As we shall see in next subsection, when considering both inelastic and elastic scattering, for devices of strong diagonal transition, η_i can indeed ultimately reach values as high as 90% over the $4.0\text{--}5.5 \text{ }\mu\text{m}$ range.

The η_d temperature dependence is defined by a characteristic temperature coefficient T_1 as such:

$$\eta_d(T_{\text{ref}} + \Delta T) = \eta_d(T_{\text{ref}}) \exp\left(-\frac{\Delta T}{T_1}\right) \quad (6)$$

where $T_{\text{ref}} + \Delta T$ is the heatsink temperature, T_{ref} is the reference heatsink temperature, and ΔT is the range in temperature over which the η_d decrease can be approximated by an exponential function.

The T_1 value, just as for interband-transition lasers [26], is a key signature of carrier leakage above RT, whose suppression is key to high CW power [7], [12], [26] and high CW wall-plug efficiency [7], [12]. Looking at (5a), T_1 is determined by the terms that have a strong dependence on temperature: η_p and α_w [12]. η_p reflects the degree of carrier-leakage suppression and since the carrier leakage is a strong function of temperature [see (4)], η_p sharply decreases with increasing heatsink temperature, in turn, accounting for a large part of the T_1 value. However, carrier leakage alone cannot account for the whole T_1 value [14], leaving an increase with temperature in α_w value as the most likely other factor that determines the T_1 value. That increase in α_w was shown [12] to most likely be due to the increase with temperature of the (nonresonant) intersubband (ISB) absorption term of the α_w value.

Suppressing carrier leakage via CB engineering of $3.8\text{--}5.6 \text{ }\mu\text{m}$ -emitting QCLs [7], [24], [27] has led to high T_1 values, in the $285\text{--}800 \text{ K}$ range [8], [12], [27]–[28]. $4.9\text{--}5.0 \text{ }\mu\text{m}$ -emitting STA- and TA-RE QCLs had their T_1 values drop from $\sim 650 \text{ K}$

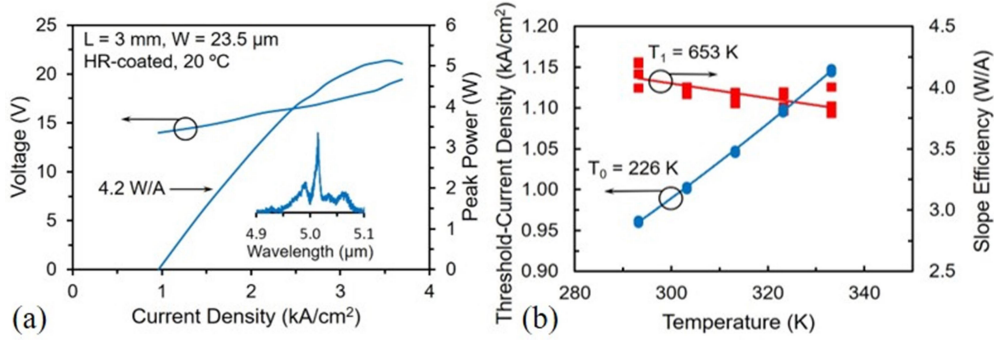


Fig. 3. 5.0 μm -emitting STA-RE QCL: (a) Light- and voltage-current characteristics, and spectrum; (b) Temperature dependence of the threshold-current density J_{th} and the slope efficiency. T_0 and T_1 are the characteristic temperatures for J_{th} and the slope efficiency, respectively. Reproduced with permission from [8] © at The Optical Society.

for low-doped devices [8], [11] to $\sim 400\text{ K}$ for moderately-high doped devices [8], [2], confirming that the α_w increase with temperature is in part responsible for the T_1 value. In sharp contrast, for conventional, moderately-high doped 4.5-5.0 μm -emitting QCLs the T_1 values are rather low: $\sim 140\text{ K}$ [12] as a consequence of severe carrier leakage.

We show in Fig. 3 the electro-optical characteristics typical of 5.0 μm -emitting STA-RE QCLs. The low J_{th} value (0.96 kA/cm^2) reflects a high η_{inj} value (84.5%) as a result of strong carrier-leakage suppression [6]. The high single-facet η_{sl} value (4.2 W/A) reflects both carrier-leakage suppression as well as a high η_{tr} value ($\sim 91\%$) due to a dramatic decrease in the ll -lifetime, τ_{3g} , via IFR scattering in STA-RE QCLs [6]. The high T_0 the T_1 values: 226 K and 653 K, respectively, shown in Fig. 3(b), are proof of both strong carrier-leakage suppression for devices of relatively low injector doping sheet density ($0.7 \times 10^{11}\text{ cm}^{-2}$). For STA-RE QCLs of same geometry and moderately-high injector doping ($\sim 10^{11}\text{ cm}^{-2}$) the T_0 and T_1 values were found [8] to be somewhat lower: 216 K and 400 K; reflecting the influence of increased backfilling on T_0 , and the increase in the ISB-absorption part of the α_w value on T_1 .

3) *Wall-Plug Efficiency*: In low-duty-cycle pulsed operation the electrical-to-optical power-conversion efficiency, so-called wall-plug efficiency, η_{wp} , is defined per pulse; that is, considering the peak-pulsed optical power and drive current. Although not of direct practical use, the maximum pulsed η_{wp} value is useful in that it provides the upper limit for the wall-plug efficiency in CW operation $\eta_{wp,CW}$. A good approximate expression for the pulsed maximum wall-plug efficiency $\eta_{wp,max}$ [28], slightly modified by considering unity $\eta_{inj,tun}$ value and the factor η_s , that accounts for the droop in the pulsed L - I curve at the $\eta_{wp,max}$ point [14], is given by:

$$\eta_{wp,max} \simeq \eta_s \eta_p \eta_{tr} \frac{\alpha_{m,opt}}{\alpha_{m,opt} + \alpha_w} \left(\frac{1}{1 + \Delta_{inj,res}/(h\nu)} \right) \times \left(1 - \frac{J_{th}}{J_{max}} \right) \quad (7)$$

where $\alpha_{m,opt}$ is the optimal mirror loss, $h\nu$ is the photon energy, and $\Delta_{inj,res}$ is the voltage defect at resonance (i.e., the energy difference between the ll level and the main injecting state,

in the next stage, at $J = J_{max}$). (7) is similar to the equation derived for the $\eta_{wp,max}$ value in [29], with the key difference that carrier leakage is taken into account via the η_s and η_p terms. The η_s term, which is likely due to carrier leakage high above threshold [12], [14], is generally found to be ~ 0.90 [12], while the η_p value, even for devices of the highest $\eta_{wp,max}$ values reported to date [2], can be as low as 0.80 at RT [6]. Besides carrier-leakage suppression, transition-efficiency maximization and α_w minimization, the other key condition for high $\eta_{wp,max}$ value is minimizing the $\Delta_{inj,res}$ term. The latter can be achieved by designing devices with injection into the ul level from an excited injector-region state; so-called pocket-injector design [30], [31], since that significantly lowers the applied field at threshold and at J_{max} . The pocket-injector design allows for both low voltage values as well as low thermal backfilling, as needed for efficient CW operation [12]. Furthermore, the pocket-injector design coupled with a strong diagonal lasing transition leads to photon-induced carrier transport (PICT) [31], [32]. More specifically strong coupling between the injecting state and the ul level leads to one degenerate energy state; that is, a large-spatial-extent ul level composed of the two states, such that the lasing transition *de facto* occurs from the degenerate state to the ll level, causing it to be highly diagonal. In turn, the lasing transition becomes primarily photon assisted [32] (i.e., not controlled by the electric field), thus the differential resistance above threshold significantly decreases. Furthermore, due to the strong coupling (6-10 meV) and no need for resonant tunneling, the transit time within the AR significantly shortens, thus the dynamic range increases [30], [31]. The net result is that the $\eta_{wp,max}$ value significantly increases, which explains why the current record for CW wall-plug efficiency at RT; i.e., 21% at $\lambda = 4.9\ \mu\text{m}$ [2], was obtained from devices found [33] to have strong PICT action.

Fig. 4 shows experimental $\eta_{wp,max}$ values reported at 298 K operation or adjusted for 298 K operation, by using their respective T_0 and T_1 values [12], for both-facets, uncoated and single-facet, HR-coated devices; and two curves, derived using (7), for the following conditions: (a) early upper limits [29] using the highest η_i value at the time (i.e., $\sim 67\%$) [28], assuming a linear L - I curve (i.e., $\eta_s = 1$), and $\Delta_{inj,res} = 150\text{ meV}$; and (b) upper limits using a projected η_i value of 90%, over the 4.0-5.5

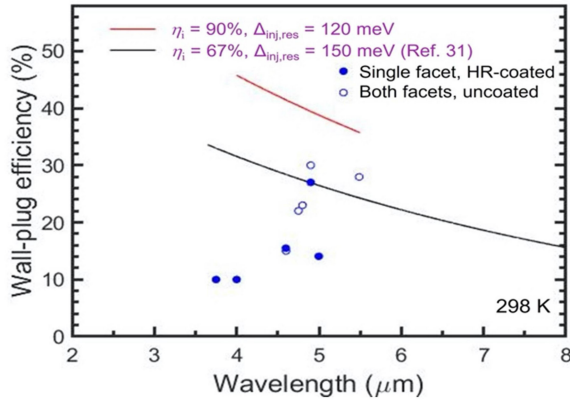


Fig. 4. Upper limits for the wall-plug efficiency of 3.8-6.0 μm -emitting QCLs as a function of emission wavelength, at 298 K heatsink temperature. The black curve is for a 'voltage defect' at resonance $\Delta_{\text{inj},\text{res}} = 150$ meV and a 70 ps dephasing time [29], and using a derived [28] internal efficiency, η_i , value of 67%. The red curve is for: $\Delta_{\text{inj},\text{res}} = 120$ meV, in agreement to values found for TA-RE QCLs of strong PICT action [5], [33] and $\eta_i = 90\%$, corresponding to when both IFR scattering is taken into account for TA- and STA-RE QCLs (i.e., $\eta_{\text{tr}} \sim 95\%$ [6], [23]) and there is strong carrier-leakage suppression (i.e., $\eta_{\text{inj}} \sim 95\%$ [6], [23]). The experimental data points are from Fig. 11 of [12] and from [34]–[38].

μm wavelength range, $\eta_s = 1$, and $\Delta_{\text{inj},\text{res}} = 120$ meV. The projected η_i value is derived by taking an average value of 95% for η_{tr} , obtained when considering elastic scattering for TA- and STA-type devices [6] and emitting in the 4-5 μm range [6], [23], and an average value of 95% for η_{inj} , obtained for STA-RE devices of strong carrier-leakage suppression due to a very large (120 meV) $E_{\text{ul}+1,\text{ul}}$ values [6], [23]. Strong carrier-leakage suppression also justifies taking $\eta_s = 1$. As for $\Delta_{\text{inj},\text{res}} = 120$ meV, it is justified by the fact that for 4.9 μm -emitting QCLs with PICT action [2] $\Delta_{\text{inj},\text{res}}$ has been found to be ~ 123 meV [5], [33]. The updated upper-limits curve results in projected $\eta_{\text{wp},\text{max}}$ values $\geq 40\%$ for $\lambda \leq 4.8$ μm , and $\eta_{\text{wp},\text{max}} \sim 42\%$ at $\lambda = 4.6$ μm . That is, $\eta_{\text{wp},\text{max}}$ values ≥ 1.5 times the current record (i.e., 27% at $\lambda = 4.9$ μm) look quite possible.

However, below $\lambda \sim 4.5$ μm the experimental, single-facet $\eta_{\text{wp},\text{max}}$ values decrease such that for $\lambda = 3.8$ -4.0 μm [34], [35] they are about 66% of the single-facet values for conventional 4.6-5.0 μm -emitting QCLs (i.e., $\sim 10\%$ vs. 14-16%). The main reason appears to be poorer well/barrier interfaces with increasing strain. That is, in order to accommodate the larger transition energies involved, higher strain (1.5%) wells and barriers have to be grown, that lead to significantly lower thermal-conductance G_{th} values than in conventional mid-IR InP-based QCLs [34], [39]. Those decreases in G_{th} have been attributed to interface-quality deterioration [39] which, in turn, results in wide electroluminescence spectra (e.g., 50 meV FWHM linewidth for the 3.8 μm -emitting devices [34]) and subsequently high threshold-current density J_{th} values [34], [35] (i.e., $\sim 50\%$ higher than for similar design 4.6 μm -emitting QCLs). That, together with the high threshold fields needed (~ 100 kV/cm), leads to higher leakage currents [6] and lower dynamic ranges both of which will lower the $\eta_{\text{wp},\text{max}}$ values (The increase in photon energy is compensated for by higher $\Delta_{\text{inj},\text{res}}$ values as result of higher applied fields). For instance, for 4.05 μm -emitting STA-RE

QCLs designed with excited-state injection and strong PICT action [23] the projected single-facet $\eta_{\text{wp},\text{max}}$ value was found to be $\sim 30\%$, mostly because an η_i value of only $\sim 70\%$. That is, for 3.8-4.1 μm -emitting QCLs the realistic upper limit for the pulsed $\eta_{\text{wp},\text{max}}$ value appears to be 30%.

Looking at the experimental data in Fig. 4, by far the highest single-facet $\eta_{\text{wp},\text{max}}$ value (27%) is for the $\lambda = 4.9$ μm device [2], as expected, since that device has both pocket injection and PICT action [33]. In order to approach the new upper-limit curve, further carrier-leakage suppression is needed, similar to what can be obtained from devices of STA-RE design [6], [23] (i.e., increase the η_i value from 70% to $\sim 90\%$, and the η_s value from 0.92 [12] to a value closer to unity).

We indicate results from both-facets, uncoated devices and single-facet, HR-coated devices for the following reasons. Both-facets, uncoated $\eta_{\text{wp},\text{max}}$ values are higher than single-facet, HR-coated values simply because of higher mirror-loss α_m values, especially for short-cavity (e.g., 3 mm-long) devices. Then to obtain single-facet operation, while maintaining the same α_m value, one can adjust the cavity length and the mirror-facets' reflectivities, but the J_{th} values remain relatively high compared to those for devices that have to start with an HR-coated back facet. In turn, the high J_{th} values severely impair CW performance [40]. Similarly, both-facets, uncoated devices of high J_{th} values (~ 2 kA/cm²) and possessing strong carrier leakage [36], [12], [37] will have their CW performance severely degraded since the core-temperature rise, ΔT_{act} , [12] is directly related to J_{th} and the degree of carrier leakage. For example, the 4.8 μm -emitting QCL of 23% pulsed $\eta_{\text{wp},\text{max}}$ value [37] reached only 6.7% in CW operation, in spite of being of the buried-heterostructure (BH) type (i.e., in spite of good thermal management). Finally, it is highly relevant that the two highest reported $\eta_{\text{wp},\text{max}}$ values (i.e., $\sim 30\%$ at $\lambda = 4.9$ μm , and $\sim 28.3\%$ at $\lambda \sim 5.65$ μm , in Fig. 4) were obtained from (uncoated-facets) devices unable to operate CW at room temperature. That is, optimization for high pulsed $\eta_{\text{wp},\text{max}}$ operation does not necessarily mean optimization for high CW $\eta_{\text{wp},\text{max}}$ operation.

B. Carrier Leakage Considering Both Inelastic and Elastic Scattering

1) *Background:* Early work on carrier leakage [41], [42] was performed on short-barrier QCLs (i.e., GaAs-based ones) and involved LO-phonon-triggered thermal excitation (i.e., inelastic scattering) of hot electrons from injector states to the continuum [see Fig. 1 in [12]]. Concurrently, a study of medium-tall-barrier, GaAs-based QCLs [43] revealed that the leakage to the continuum was dominated by leakage from the *ul* level. However, in GaAs-based devices the *ul* level is the highest energy state in the AR, thus leakage to the continuum was an obvious consequence. The picture dramatically changed with the advent of tall-barrier QCLs (i.e., InP-based ones) in which case hot-carrier leakage was found to mostly occur from the *ul* level through AR energy states above it [14], followed by relaxation to low-energy states [see Fig. 1 above, and Fig. 2 in [12]]; that is, the leakage current was found to be mainly an intra-AR shunt current, with carriers lost to the continuum being a negligible quantity [12].

The picture further changed with the findings that: (a) IFR-assisted, hot-carrier leakage (i.e., elastic scattering) through high-energy AR states is the only leakage mechanism at cryogenic temperatures [17]; and (b) hot-carrier leakage at RT is triggered by both LO-phonon and IFR scattering (i.e., both by inelastic and elastic scattering) from both the ul level and key injector states, with the IFR-scattering part being the dominant leakage component [6], [13]. The latter finding allowed bridging the gap between theoretical and experimental η_i values [6], thus providing a tool for further enhancing η_i via CB- and IFR-scattering engineering.

2) *Elastic Scattering*: The elastic-scattering mechanisms that primarily affect the performance of mid-IR QCLs are IFR scattering [6], [17], [44]–[47] and alloy-disorder (AD) scattering [6], [44], [47]. The impact of IFR and AD scattering on device characteristics is discussed mostly for diagonal-transition QCLs, since such devices have led to maximizing the wall-plug efficiency of high-power ($> 1\text{W}$) QCLs [6], [8], [12].

a) IFR-Scattering Rates

Downward Transitions: IFR scattering is currently mostly characterized, for sharp-interfaces devices, by two structural parameters: Δ , the average roughness height, and Λ , the average roughness correlation length along an interface plane, that characterize a standard Gaussian autocorrelation of the roughness [44], [45].

While the actual well/barrier interface nature is generally unknown, the Gaussian distribution function remains a useful tool for extracting Δ and Λ parameters that fit experimental data, thus can be used for high-performance device design. For example, for InGaAs/AlInAs QCLs the extracted Δ and Λ values have been found to be: 0.10-0.12 nm and 9 nm for MBE-grown QCLs [44], [48], and 0.11-0.125 nm and 10.2-10 nm for MOCVD-grown QCLs [6], [13]. Since recently a comprehensive IFR formalism that takes into account interface grading [49], with a Gaussian distribution for Δ and Λ , was successfully employed [33] for reproducing published experimental data of high-performance QCLs emitting at both $\lambda = 4.9\ \mu\text{m}$ [2] and $\lambda = 8.3\ \mu\text{m}$ [50], the Gaussian autocorrelation function appears to be a valid assumption for the Δ and Λ values characterizing mid-IR, InP-based QCLs. Then, comprehensive device design should be done by also using two additional IFR parameters: the interface width and axial correlation length [49].

For sharp-interfaces devices, the IFR scattering rate for the transition between a state m and a lower-energy state n (i.e., a downward transition), in QCL structures of varying barrier and well compositions, has the following dependencies [8]:

$$\frac{1}{\tau_{mn}^{IFR}} \propto \Delta^2 \Lambda^2 \sum_i m_{ci} \delta U_i^2 \varphi_m^2(z_i) \varphi_n^2(z_i) \exp\left(-\frac{\Lambda^2 m_{ci} E_{mn}}{2\hbar^2}\right) \quad (8)$$

where, m_{ci} is the effective mass at the i^{th} interface in the AR, δU_i is the CB offset at the i^{th} interface, $\varphi_m(z_i)$ and $\varphi_n(z_i)$ are the wavefunction amplitudes of the m and n energy states at the i^{th} interface, and E_{mn} is the energy difference between states. For transitions from the ul level, m is the ul -level state number and n is the ll -level state number or the state number of any of the rest

of low-energy AR and extractor states [8]. For transitions from the ll level(s), m is the ll -level state number and n is the state number of any of the lower-energy AR and extractor states [8].

Since for high-power QCLs the lasing transition needs to be strongly diagonal (i.e., to obtain PICT action) the sum of probability products at interfaces is low, thus reducing the lasing-transition IFR scattering rate. In addition, the IFR scattering rate strongly decreases with increased transition energy. Thus, IFR scattering has a negligible impact on the ul -level lifetimes of 3-6 μm -emitting QCLs. Conversely, IFR scattering dominates the ll -level lifetime due to the low transition energies involved [6], [8]. Furthermore, the ll -level lifetime is significantly reduced in STA- and TA-RE QCLs since the CB offsets on the downstream side of the AR are much higher than those on the upstream side of the AR. The net effect is that the lasing-transition efficiency η_{tr} [see (5c)] reaches values in the 91-96% range for 4.0-5.0 μm -emitting QCLs [6], [8], [23].

Upward Transitions: For upward transitions, that is those considered in IFR-triggered carrier leakage [6], [17], the IFR scattering rate from the ul level to the next higher AR state, level $ul+1$, is characterized by the following scattering rate [6], [17]:

$$\frac{1}{\tau_{ul,ul+1}^{IFR}} = \frac{1}{\tau_{ul+1,ul}^{IFR}} I_{ul+1,ul} \left(\frac{E_{ul+1,ul}}{kT_{e,ul}} \right) \exp\left(-\frac{E_{ul+1,ul}}{kT_{e,ul}}\right) \quad (9)$$

where $1/\tau_{ul+1,ul}^{IFR}$ has the same form as (8) (i.e., for downward transitions); $I_{ul+1,ul}$ is a relatively weak function of temperature and energy [6], [17]; and $E_{ul+1,ul}$ and $T_{e,ul}$ are as defined for (4). The same equation is used for IFR-triggered carrier leakage from key injector states to the $ul+1$ level. As mentioned above, IFR-triggered leakage is found to dominate carrier leakage [6], [13].

Just as for LO-phonon triggered carrier leakage [14], the IFR upward-transition scattering rate decreases exponentially with increased $E_{ul+1,ul}$ value, thus, to suppress carrier leakage STA-RE-type devices are required. In addition, in STA-RE devices the wavefunction for the $ul+1$ level is ‘pushed’ to the AR downstream side [7], thus causing a low value for the sum of probability products at interfaces and, in turn, further reducing carrier leakage. Then, for STA-RE QCLs the relative carrier leakage, $J_{\text{leak}}/J_{\text{th}}$, reaches values in the 5-9% range compared to 20-30% for TA-RE [6] or conventional QCLs [13].

b) *AD-Scattering Rate*: In contrast to IFR scattering, AD scattering is not a function of the energy-dependent scattering vector because of the short-range nature of the scatterers. Its main role is to impact the ul -level carrier lifetimes, since it dominates the elastic part of the ul -level scattering rate [6], [8], [13], [44]. In turn, for high-power, 4-5 μm -emitting QCLs of strong diagonal transition the ul -level lifetime is controlled by LO-phonon and AD scattering [6], [23]. The AD scattering rate between an energy state m and a lower-energy state n , is given by [6]:

$$\frac{1}{\tau_{mn}^{AD}} = \frac{1}{8\hbar^3} \sum_i m_{ci} a_i^3 (V_i^{\text{alloy}})^2 x(1-x) \int_{\text{alloy}_i} \varphi_m^2(z) \varphi_n^2(z) dz \quad (10)$$

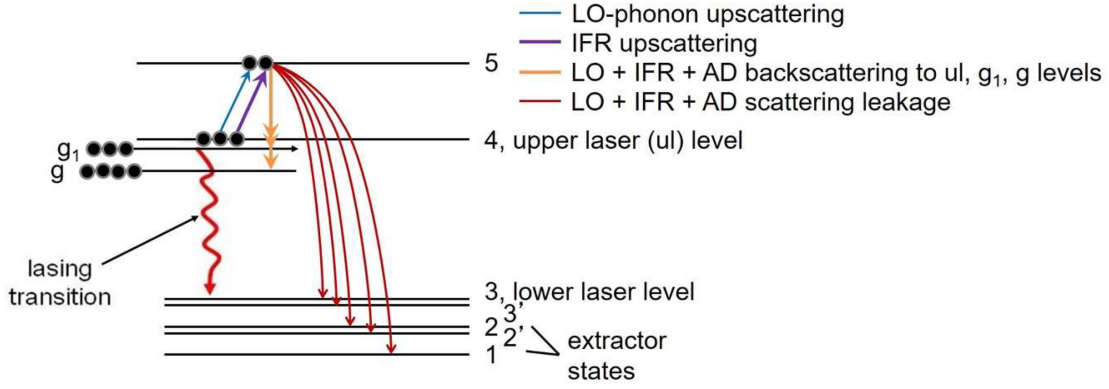


Fig. 5. Schematic representation of carriers excited from the ul level, state 4, to the $ul+1$ level, state 5; backscattering from state 5 to the ul , g_1 and g levels; and leakage from state 5 to low-energy AR and extractor states. Here, level g_1 , the first-excited injector state, is the main injecting state into the ul level. Reproduced from [6], with the permission of AIP Publishing.

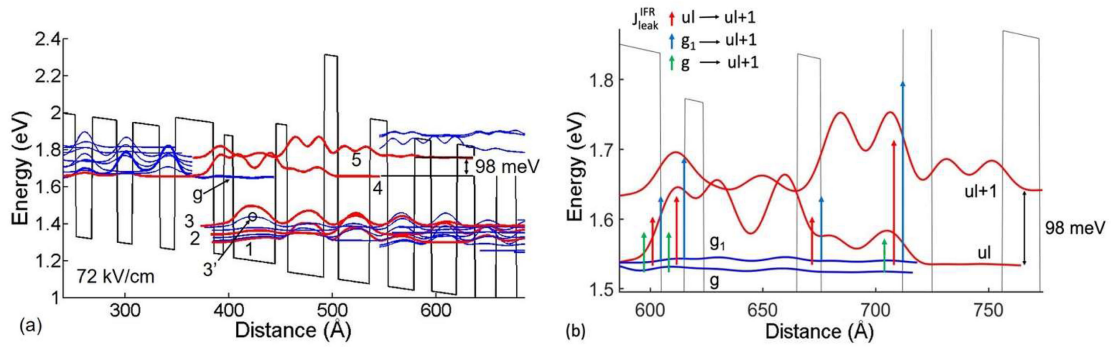


Fig. 6. STA-RE QCL: (a) conduction-band diagram and relevant wavefunctions. Reproduced with permission from [8] © at The Optical Society. (b) upper active-region band diagram, key-state wavefunctions, and arrows at interfaces indicating the main components of the IFR-triggered leakage. Reproduced from [6], with the permission of AIP Publishing.

where i is the alloy layer number in the sequence of well and barrier layers across which the wavefunctions $\varphi_m(z)$ and $\varphi_n(z)$ extend; m_{ci} and a_i are the effective mass and lattice constant, respectively, of the layer i ; V_i^{alloy} is difference between the CB minima of the well/barrier binary-alloy components of the layer i ; and x is the alloy fraction in the $\text{In}_{1-x}\text{Ga}_x\text{As}$ or $\text{In}_{1-x}\text{Al}_x\text{As}$ alloys. Since the AD scattering rate is proportional with the sum of integrals of probability products in each layer, it is strongly affected by the lasing-transition degree of diagonality. That is, in strong diagonal-transition devices, as required for PICT action, the AD scattering rate is minimized for transitions from the ul level, thus maximizing the global ul -level lifetime $\tau_{ul,g}$ [7], [8] and, in turn, helping minimize the J_{th} value, since $J_{th} \propto 1/\tau_{ul,g}$ [5], [8] and maximize the η_{tr} value [see (5c)].

3) *Overall Carrier-Leakage Picture*: The LO- and IFR-triggered leakage currents are interconnected [6] which leads to the carrier-leakage process to unfold as follows: 1) electrons from the ul level and key injector states (i.e., the injector ground and a couple of excited states) are excited to the $ul+1$ level via LO-phonon and IFR scattering; 2) because of short overall lifetimes (i.e., involving LO-phonon, IFR and AD scattering) from the $ul+1$ to the ul level and injector states, a large part of the excited electrons return to the ul level and injector states; and 3) the rest of the electrons in the $ul+1$ level relax via LO-phonon,

IFR and AD scattering to low-energy AR states and extractor states. As an example, Fig. 5 shows schematically the scattering mechanisms involved when considering electron excitation, in a four-QW AR, only from the ul level (state 4) [6]. Electrons in state 4 are excited to state 5 via LO-phonon and IFR scattering. Then, because of a relatively short (0.1-0.2 ps) backscattering lifetime, a large part of the electrons (85-90%) returns to the states 4, g_1 and g . The rest of state-5 electrons (10-15%) relax to low-energy AR states (states 3, 2 and 1) and low-energy extractor states (states 3' and 2') and constitute the intra-AR shunt leakage current.

4) *Examples of Carrier Leakage: STA-RE QCLs*: Both the STA-RE QCL ($\lambda \sim 5 \mu\text{m}$) as well as the TA-RE QCL ($\lambda \sim 4.9 \mu\text{m}$) were analyzed by using the comprehensive carrier-leakage formalism developed in [6]. Here we show results for the STA-RE QCL [8]. Fig. 6(a) shows the conduction-band diagram and relevant wavefunctions. The barrier heights in the AR increase stepwise: $x = 0.56, 0.63$, and 0.93 in $\text{Al}_x\text{In}_{1-x}\text{As}$, and the wells depths increase stepwise: $x = 0.57, 0.60, 0.70$ and 0.70 in $\text{In}_x\text{Ga}_{1-x}\text{As}$. Because of the resulting asymmetry and Stark-shift reduction [7] the energy difference between the ul level and the next higher level, E_{54} , increases to 98 meV from ~ 45 meV in conventional 4.5-5.0 μm -emitting QCLs [7]. Fig. 6(b) displays the top part of the AR and the states involved

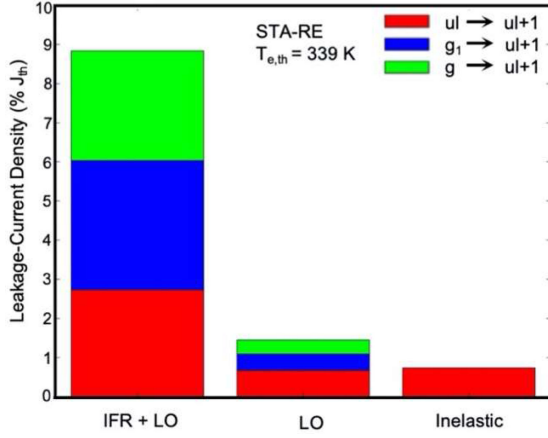


Fig. 7. Bar graph of components of the relative leakage-current density for $\sim 5.0\ \mu\text{m}$ -emitting STA-RE QCLs. LO and Inelastic stand for LO-phonon-triggered leakage in the presence and the absence of elastic scattering, respectively. Reproduced from [6], with the permission of AIP Publishing.

in triggering carrier leakage through the $ul+1$ level, as well as arrows corresponding to the main IFR-leakage paths. The device has conventional injection; that is, from the injector ground state, level g . Level g_1 , unlike in Fig. 5, is not an injecting state, but a parasitic injector state penetrating into the AR that plays a role in the carrier leakage.

The LO- and IFR-triggered leakage currents from the ul level and the g and g_1 levels are calculated as in [6]. The values, normalized to the J_{th} value, are shown in Fig. 7. The total relative leakage is 8.9% of which $\sim 37\%$ is from the parasitic state g_1 . For comparison, also shown are the LO-triggered leakage and the ‘classical’ case of leakage [14] (i.e., only LO-scattering triggered leakage and only from the ul level). The first comparison reveals that most of the leakage is IFR triggered (i.e., $\sim 83\%$). The second comparison reveals that the total leakage is 12 times higher than the conventionally calculated leakage (i.e., 0.74%) and explains the $\sim 13\%$ gap found [6] between theoretical and experimental η_i values. That is, by considering IFR-triggered carrier leakage the gap in η_i values was bridged, in that the recalculated value, with error analysis, was: 78-80.5%, which is basically the same as the experimental value: 77% [8]. Similar bridging between theoretical and experimental η_i values was found for the TA-RE QCL [6].

Since IFR-triggered leakage is the dominant leakage mechanism, it is interesting to see where it happens within the AR. As shown in Fig. 6(b), IFR scatterings between the ul and $ul+1$, and between g_1 and $ul+1$ levels are strongest at the upstream interface of the third AR barrier. That happens, due to both high overlap of the ul and g_1 wavefunctions with the $ul+1$ wavefunction at that interface, and the fact that interface corresponds to the highest CB offset in the AR [6], [8]

The finding that consideration of IFR leakage as well as the insertion of elastic scattering in the leakage process covers the gap between theoretical and experimental η_i values offered not only a complete understanding of the internal efficiency in QCLs, but also a valuable tool for achieving η_i values well in excess of 80% via IFR scattering-rate engineering; thus, leading

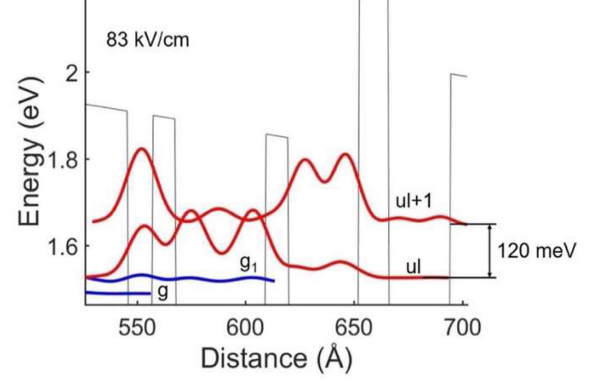


Fig. 8. Upper AR band diagram and key states' wavefunctions for optimized STA-RE QCL. Reproduced from [6], with the permission of AIP Publishing.

to substantial enhancements in the $\eta_{wp,max}$ value. With this new tool we reached [6] a preliminary design for an optimized $4.6\ \mu\text{m}$ -emitting, STA-RE QCL with three key new features: (a) elimination of parasitic injector states penetrating the AR; (b) pocket injection from the first excited injector state; and (c) an increased $E_{ul+1,ul}$ value (i.e., from 98 meV to 120 meV). The upper-AR band diagram and key energy states are shown in Fig. 8. By considering same doping and extracted IFR parameters as for GSMBE-grown TA-RE QCLs [6] the relative leakage decreases to 4.7% which, in turn, results in an η_i value of 86%. Further considering that the new design, just like the TA-RE design, is of the pocket-injector type and that the calculated J_{th} ($1.29\ \text{kA/cm}^2$) is basically the same value as for the TA-RE device [2], we extrapolated the $\eta_{wp,max}$ value, primarily based on differences in the η_i value, and obtained a value of 36%. Furthermore, the double-facet equivalent value (37.2%) approaches the upper-limit, double-facet value of 42% (see Fig. 4).

C. CW Operation

1) *Device Self-Heating*: For a given heatsink temperature T_h , the device-core temperature rises by the quantity ΔT_{act} [12]:

$$\begin{aligned} \Delta T_{act} &= T_l - T_h = R_{th} (P_{el} - P_{opt}) \\ &= R_{th} P_{el} (1 - \eta_{wp,cw}) \end{aligned} \quad (11)$$

where T_l is the lattice temperature, R_{th} is the thermal resistance, P_{el} is CW input electrical power (i.e., $P_{el} = A J V$, where A is the pumped area), P_{opt} is CW output optical power and $\eta_{wp,cw}$ is the wall-plug efficiency in CW operation. At threshold (i.e., $P_{opt} = 0$) ΔT_{act} is given by $\Delta T_{act,th}$:

$$\Delta T_{act,th} = \frac{J_{th,cw} V_{th}}{G_{th}} = J_{th}(T_h) \exp\left(\frac{\Delta T_{act,th}}{T_o}\right) \frac{V_{th}}{G_{th}} \quad (12)$$

where $J_{th,cw}$ is the CW threshold, V_{th} is the voltage at threshold and G_{th} is the thermal conductance. Then, the keys to minimizing self-heating at threshold are maximizing the T_o and G_{th} values and minimizing the V_{th} value. As pointed above, the T_o value can be maximized by suppressing carrier leakage, while keeping moderate-high values for the injector doping as needed

for high-power devices [12]. For a given number of stages, V_{th} can be significantly reduced by using excited-state injection [30], [31]. As for G_{th} , it can be maximized by considering the buried-ridge width, w , and optimizing the chip-packaging configuration. For instance, from published data on BH 4.6–4.8 μm -emitting conventional, 40-period QCLs mounted episode-down on diamond, it has been found [12] that, as the buried-ridge width w varies from 8.6 μm to 11.6 μm , G_{th} is approximately inversely proportional with $\sqrt{w}$, in good agreement with theory [5]. There are also data indicating that the $G_{th} \propto 1/\sqrt{w}$ dependence holds down to $w = 5 \mu\text{m}$ for episode-down mounted BH devices [5]. This is expected, since a significant portion of heat removal occurs laterally, away from the buried core region. Therefore, w needs to be lowered. However, as w decreases so does the output power. If the emitted-beam quality is not of great concern, for high CW power a good compromise for the w value has been 8 μm , as used for record-high CW power and wall-plug efficiency at RT [2]. However, as pointed out in Section IV-A, stable, single-lateral-mode operation for 4.6 μm -emitting BH devices requires a w value in the 4.0–4.5 μm range [51]. That can be compensated though by using long-cavity (5–10 mm), low- α_w devices in order to achieve stable, single-mode QCL operation to watt-range CW power levels.

2) *Maximum Wall-Plug Efficiency*: The maximum, single-facet CW wall-plug efficiency can be expressed as shown in (13a), at the top of the next page, [7], [12], where $\eta_d(T_h)$ is the external differential efficiency for an optimal mirror loss $\alpha_{m,opt}$, and $\Delta T_{act,wpm}$ [(13b)] is the core temperature rise at the $\eta_{wp,max}$ point, where $T_{l,wpm}$ is the lattice temperature at the CW $\eta_{wp,max}$ point and $P_{el,wpm}$ is the electrical CW power dissipated at the $\eta_{wp,max}$ point (i.e., $P_{el,wpm} = A J_{wp,max} V_{wpm}$, where A is the pumped area). As clear from the equations, both the CW $\eta_{wp,max}$ and $\Delta T_{act,wpm}$ values are strong functions of the T_1 and G_{th} values. Thus it becomes apparent that in order to maximize $\eta_{wp,max}$ and minimize $\Delta T_{act,wpm}$ values, carrier leakage has to be suppressed (to achieve high T_0 and T_1 values), the G_{th} value needs to be increased as much as possible, and the operating voltage V_{wpm} needs to be minimized by using excited-state injection and especially designing devices with strong PICT action.

The device of highest published single-facet, CW $\eta_{wp,max}$ values at 298 K heatsink temperature (i.e., 21%) is a 4.9 μm -emitting TA-RE QCL [2]. That is expected given its high η_i value (i.e., 70%) and strong PICT action [33] [e.g., the differential resistance above threshold is only $\sim 1.6 \Omega$ compared to $\sim 3.6 \Omega$ for conventional BH devices of same pumped area (i.e., 0.04 mm^2)]. However, the R_{th} value ($\sim 3.4 \text{ K/W}$, as extracted from CW-curve fitting [12]), is about twice higher than that extracted [12] from conventional 5 mm-long devices of similar buried-ridge width and using diamond submounts (i.e., $\sim 1.6 \text{ K/W}$). This is most likely due to the insertion of a multitude of tall barriers (AlAs) per period. More specifically, the G_{th} value is about half that for conventional, 8 μm -wide-ridge BH devices (i.e., $\sim 735 \text{ W/cm}^2 \text{ K}$ vs. $1450 \text{ W/cm}^2 \text{ K}$ [12]). The virtual halving of the G_{th} value appears to be primarily due to both significantly more interfaces per period [52] (i.e., 34 vs. 22 interfaces) with the rest likely due to the multitude of highly strained layers [39]. The

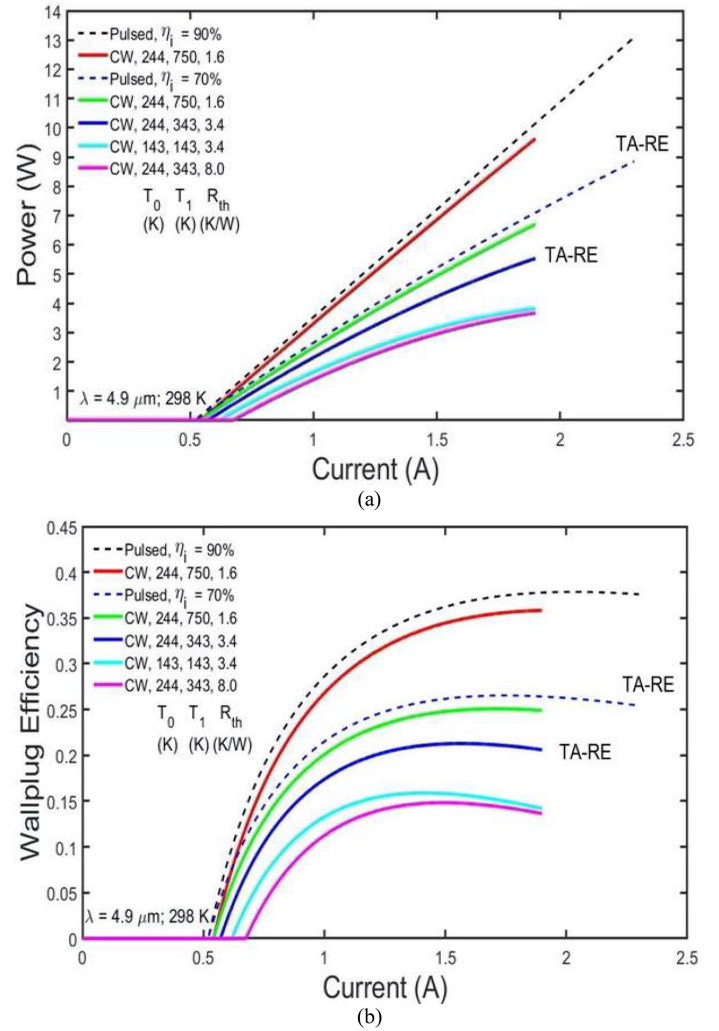


Fig. 9. Characteristics of 4.9 μm -emitting TA-RE QCL [2] and for QCLs of different T_0 , T_1 and R_{th} values, and for a η_i value of 90%: (a) CW and pulsed power; (b) CW and pulsed wall-plug efficiency. Adapted from [12].

relatively high R_{th} value results in a relatively high $\Delta T_{act,wpm}$ value (i.e., $\sim 54 \text{ K}$). By using (13a) and the experimental, pulsed slope-efficiency value (5.7 W/A), the calculated CW $\eta_{wp,max}$ value is 20.7%, that is quite close to the experimental value [2], confirming that (13a) is a good approximation.

In order to reach ultimate limits in CW $\eta_{wp,max}$ values for 4.5–5.0 μm -emitting QCLs one needs to consider devices possessing strong PICT action, high T_1 values (e.g., 750 K, as obtained from deep-well, TA-type devices [20]), low R_{th} values (e.g., the 1.6 K/W value mentioned above), and an η_i value of 90% (see Fig. 4). For example, by using as reference the η_{wp} vs. I CW curve of the 4.9 μm -emitting TA-RE QCL [2], a 36.3% $\eta_{wp,max}$ value is projected [see Fig. 9(b)], that is about 1.7 times higher than that reported in [2]. Extrapolating for an emission wavelength of 4.6 μm , an $\sim 8\%$ improvement is anticipated [12] which leads to a projected CW $\eta_{wp,max}$ value of $\sim 39\%$. Maximum CW wall-plug efficiency values close to 40% would be quite beneficial for many QCL-related applications since thermal-load management drives the packaged laser system's

size, weight and overall power consumption.

$$\eta_{wp,maxCW} \approx \eta_s \eta_d (T_h) \exp\left(-\frac{\Delta T_{act,wpm}}{T_1}\right) \times \left[1 - \frac{J_{th}(T_h)}{J_{wpm}(T_h)} \exp\left(\frac{\Delta T_{act,wpm}}{T_0}\right)\right] \frac{h\nu}{qV_{wpm}} \quad (13a)$$

$$\Delta T_{act,wpm} = T_{l,wpm} - T_h = R_{th} P_{e,wpm} (1 - \eta_{wp,max}) \\ = \frac{J_{wpm} V_{wpm}}{G_{th}} (1 - \eta_{wp,max}) \quad (13b)$$

As for devices emitting in the $3.8\text{-}4.1\ \mu\text{m}$ range, besides the fact that, as mentioned above, the upper limit for the η_i value is expected to be only $\sim 70\%$, the voltages increase with decreased emission wavelength and the G_{th} value has been found [34], [39] to be $\sim 70\%$ the value of $4.5\text{-}5.0\ \mu\text{m}$ -emitting QCLs of same buried-ridge width which, in turn, lead to strong self-heating. The best published single-facet result in that range is 6% at $\lambda \sim 3.8\ \mu\text{m}$ [34]. Using STA-type device designs for significant carrier-leakage suppression, excited-state injection and strong PICT action we have found [23] that RT CW $\eta_{wp,max}$ values as high as $\sim 18\%$ become possible for $\lambda \sim 4\ \mu\text{m}$.

3) *Optical Output Power*: The expression for the CW optical power [12] is shown in (14a), with ΔT_{act} given by (14b) where the CW wall-plug efficiency, $\eta_{wp,cw} = P_{opt}/AJV$, follows from (11).

Just as for interband-transition devices [26], the key parameters for high-CW-power operation are high values for T_1 and low values for R_{th} . The CW L-I curve from [2] (i.e., for $4.9\ \mu\text{m}$ -emitting TA-RE QCL) was matched [12], by using the experimental characteristic temperature coefficients: $T_0 = 244\ \text{K}$, $T_1 = 343\ \text{K}$; and curves fitting the experimental pulsed L-I and V-I curves. The best fit (i.e., $5\ \text{W}$ CW output power at $1.73\ \text{A}$ drive current) was obtained for $R_{th} = 3.4\ \text{K/W}$ [Fig. 9(a)]. Then, several cases were considered for comparison. If $R_{th} = 8\ \text{K/W}$, as may be the case for episcide-up mounting, the CW power at $I = 1.73\ \text{A}$ drops from $5\ \text{W}$ to $3.4\ \text{W}$, and the maximum $\eta_{wp,cw}$ value drops from 21% to 14.5% . If the T_0 and T_1 values are lowered to $143\ \text{K}$, typical of conventional QCLs [12], while keeping the same R_{th} value, the CW power at $I = 1.73\ \text{A}$ drops from $5\ \text{W}$ to $3.6\ \text{W}$, and the maximum $\eta_{wp,cw}$ value drops from 21% to 15% . Therefore, low T_1 values have quite the same deleterious effect on $\eta_{wp,cw}$ as high R_{th} values. Then, we considered devices with $T_0 = 244\ \text{K}$, $T_1 = 750\ \text{K}$ and $R_{th} = 1.6\ \text{K/W}$, as may be the case for $4.9\ \mu\text{m}$ -emitting STA-RE devices; that is, devices of virtually complete carrier-leakage suppression [6] at no price in R_{th} value, since there is no need for AlAs barrier inserts throughout the injector region. The CW power increases from $5\ \text{W}$ to $6\ \text{W}$, and the maximum $\eta_{wp,cw}$ value increases from 21% to 25.2% ; that is, quite close to the maximum reported pulsed wall-plug efficiency value (i.e., 27%).

Finally, the η_i value experimentally obtained in [2] (i.e., 70%) is short of the above-deduced upper-limit value of 90% for $4.5\text{-}5.5\ \mu\text{m}$ -emitting QCLs of strong carrier-leakage suppression (see Fig. 4). The CW L-I curve for the case: $T_0 = 244\ \text{K}$, $T_1 = 750\ \text{K}$, $R_{th} = 1.6\ \text{K/W}$ and $\eta_i = 90\%$, is plotted as the limiting

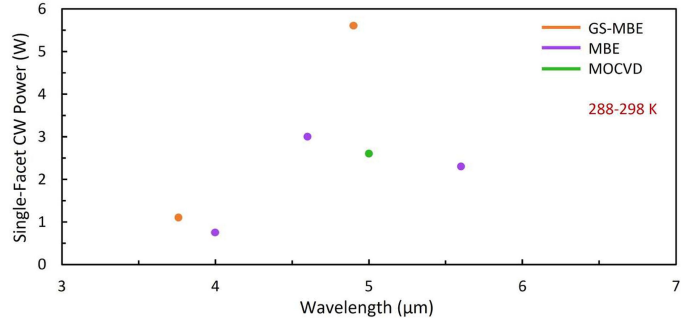


Fig. 10. Single-facet-emitted, RT (288-298 K) maximum CW powers vs. emission wavelength for QCLs grown by different crystal growth methods. The data are taken from Refs. 8, 34, 35, 58, 59, and 60.

case for $4.9\ \mu\text{m}$ -emitting QCLs. Then the maximum projected CW power is $9.7\ \text{W}$; that is, almost twice the highest CW value obtained at $298\ \text{K}$ heatsink temperature [2].

$$P_{opt}(T_l) = A \frac{h\nu}{q} \eta_{d,CW}(T_l) [J - J_{th,CW}(T_l)] \\ = A \frac{h\nu}{q} \eta_d(T_h) \exp\left(-\frac{\Delta T_{act}}{T_1}\right) \\ \times \left[J - J_{th}(T_h) \exp\left(\frac{\Delta T_{act}}{T_0}\right) \right] \quad (14a)$$

$$\Delta T_{act} = T_l - T_h = R_{th} P_{opt} [(1/\eta_{wp,cw}) - 1] \quad (14b)$$

Given CW L-I and η_{wp} vs. I curves, one can compare ΔT_{act} values. While $\Delta T_{act} = 54\ \text{K}$ for the TA-RE QCL at the $\eta_{wp,max}$ point (i.e., 21% at $4.25\ \text{W}$ output power), for the case $T_0 = 244\ \text{K}$, $T_1 = 750\ \text{K}$, $R_{th} = 1.6\ \text{K/W}$, $\eta_i = 90\%$ ΔT_{act} is only $16.6\ \text{K}$ at $4.25\ \text{W}$ output power. Such a relatively low value is close to the $\sim 15\ \text{K}$ value for the low-CW-power ($\sim 0.2\ \text{W}$) $4.6\ \mu\text{m}$ -emitting QCLs that have demonstrated long-term reliability [53]. Therefore, we conclude that optimized QCLs hold the potential to have long-term reliability at multi-watt CW output power levels.

It was recently proposed [54] that instead of scaling the power by increasing the number of periods, N_p , to scale it laterally by using structures with $N_p = 10\text{-}15$ for broad-area (i.e., $20\text{-}30\ \mu\text{m}$ -wide) ridge-waveguide devices. The basic idea is that $\Delta T_{act,th}$ [see (12)] decreases with decreasing N_p [54], [55] since the threshold-power density decreases with decreasing N_p . From 10 -stage, 30 -wide ridge-waveguide, $4.6\ \mu\text{m}$ -emitting devices CW operation was obtained to above $100\ ^\circ\text{C}$ [54], something that only BH devices have been able to achieve [2], [56], [57]. However, the maximum CW P_{opt} and η_{wp} values were only $0.9\ \text{W}$ and 5% , respectively [55] since both P_{opt} and η_{wp} are strong function of N_p [12]. A similar scaling approach was applied to BH QCLs [58] by using 15 -stage, broad-area BH devices, with the best result being $2.3\ \text{W}$ single-facet CW power from $21\ \mu\text{m}$ -wide ridges at $\lambda = 5.7\ \mu\text{m}$. As expected, P_{opt} was limited since it is dependent on N_p .

Fig. 10 shows the maximum single-facet (i.e., useable) CW powers achieved at or near RT (i.e., in the $288\text{-}298\ \text{K}$ heatsink

temperature range) over the 3–7 μm wavelength range, and for different crystal-growth methods. The highest CW power (i.e., 5.6 W) was obtained [59] at $\lambda = 4.9 \mu\text{m}$ by using a GSMBE-grown TA-RE QCL with strong PICT action. The $\sim 10\%$ improvement over the best prior value [2] was claimed to reflect effective heat removal from devices with regrowth planarization, although it was obtained a heatsink temperature 5 K lower than the prior result. The highest values for each growth method [8], [59], [60] occur in the 4.5–5.0 μm wavelength range because of low α_w values and/or relatively high thermal-conductance values and/or strong carrier-leakage suppression [12]. MBE-grown QCLs of NRE, vertical-transition design [58], [60] have yielded lower maximum P_{opt} values, in spite of similarly low α_w values as GSMBE-grown devices, due both to relatively low η_i values [8], [24], [61] and inherent lack of PICT action. MOCVD-grown QCLs currently have higher α_w values than GSMBE- and MBE-grown QCLs, which for STA-RE QCLs are in large part compensated for by very high η_i values (i.e., 77% at $\lambda = 5 \mu\text{m}$) [8]. The best result in the 3.8–4.0 μm range (i.e., 1.1 W) is from a GSMBE-grown device [34] due both to relatively low carrier leakage as a result of an AR design of relatively high $E_{\text{ul}+1,\text{ul}}$ value [12], [34] and to miniband-type (resonant) extraction [12].

D. Reliability and Failures Modes

While remarkable progress has been made in understanding the design constraints of QCLs and their optimization for high CW output-power operation, there is a lack of understanding regarding the degradation and failure mechanisms under high CW-drive conditions. Furthermore, there is still a need for long-term reliability studies to be carried out at high CW output powers, and/or under environmentally stressed conditions. QCLs are expected to have different degradation and failure modes than diode lasers, because nonradiative recombination and subsequent optical absorption at the facets is not an issue. QCL lifetests have been reported which were carried out at relatively low output powers ($\sim 200 \text{ mW}$) and revealed activation energies as high as 1.2 eV, with the primary failure mechanism being oxidation of the front uncoated facet [53]. Higher output power ($\sim 1 \text{ W CW}$), constant-power, lifetest studies of QCLs emitting at $\lambda \sim 5.0 \mu\text{m}$ have shown that stable operation can be achieved over 2500 hours at room temperature, as shown in Fig. 11 [62]. Interestingly, some devices exhibit a slight improvement initially with aging and stabilize after an initial burn-in, as shown in Fig. 11(b), which has also been reported previously [53]. The mechanism responsible for such behavior needs further study. The estimated active-region temperature is $\sim 70^\circ\text{C}$, based on correlating thermal simulations with thermal-reflectance measurements on BH lasers [63]. In [62], to mitigate the failure mechanism previously observed at lower output powers in [53] as well as to improve device output, both facets have coatings: a high-reflectivity (HR) back-facet coating and a 14% low-reflectivity (LR) front-facet coating. The devices were mounted episcide-down on copper with indium and tested under constant-power operation in a controlled environment. The active-region width is nominally $10 \mu\text{m}$ with a cavity length of 5 mm. Such QCLs have demonstrated room-temperature

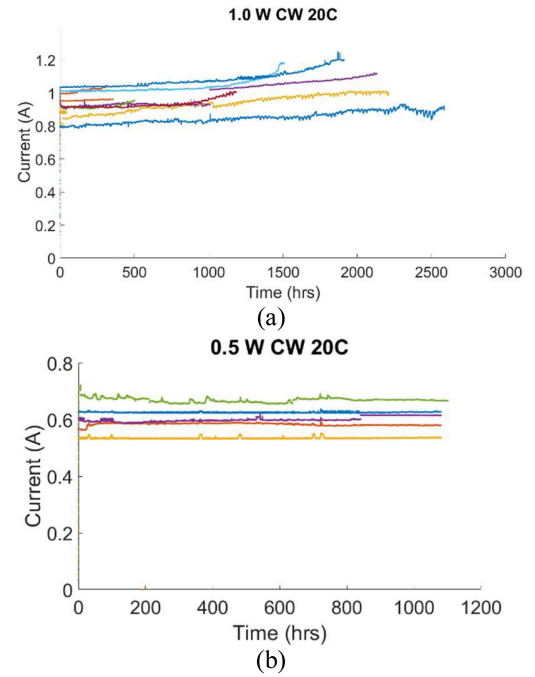


Fig. 11. Single emitter BHs operating at a submount temperature of: (a) 20°C at $\sim 1 \text{ W CW}$ for devices with HR/LR coatings. © (b) 20°C at 0.5 W CW for devices with HR-coated back facet and uncoated front facet.

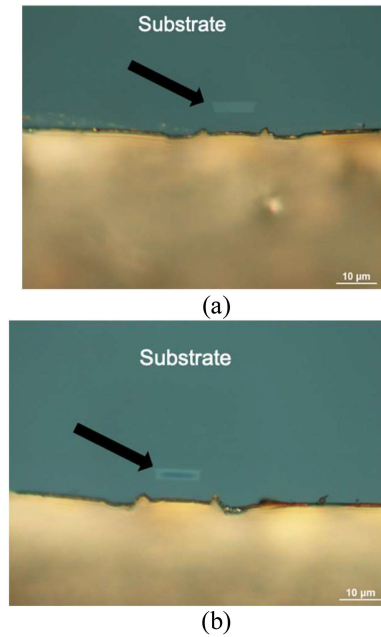


Fig. 12. Optical microscope images of the LR-coated front facet of BH QCL: (a) prior to burn-in, (b) after observing gradual degradation over 1000 hrs at 1 W CW output power, as shown in Fig. 11(a).

operation to front-facet CW output powers as high as 2.6 W [8].

All devices operating at $\sim 1 \text{ W CW}$ in Fig. 11(a) show a relative increase in the current, attributed to front-facet coating degradation. Fig. 12 shows the observed facet coating degradation before and after lifetesting. No gradual degradation was observed for devices operated at 0.5 W , although without LR

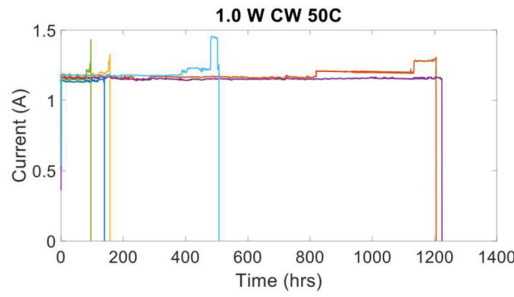


Fig. 13. QCL lifestest data at 1W CW at room temperature and 50°C submount temperature. Optical inspection after failure confirmed indium shorting as the cause of failures.

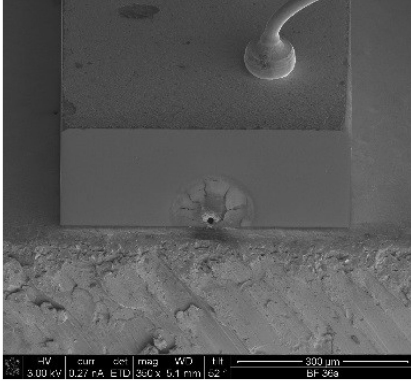


Fig. 14. SEM image showing the damaged area on the QCL front facet after catastrophic mirror damage [67].

front-facet coatings. Indium-mounting failures, which are more prevalent while operating at elevated temperatures and under pulsed conditions, were also observed. The use of indium solder is not optimal for long-term reliability, since the electromigration of indium is a known issue [64]. This fact prevents higher temperature reliability studies, as shown in Fig. 13, since indium creep leads to device shorting.

A few studies of high power QCL lasers indicate that the reliability limitations are associated with Catastrophic Mirror Damage (CMD) occurring at the facets [9], [65]–[67]. Facet failures in QCL appear to be directly related to thermal shear stress [9], resulting from the large temperature gradient between the core region and the heatsink, at the laser facet, when operated at high CW drive currents or when heat management is inadequate. In contrast to what is observed in strained QW lasers, where the CMD-affected area typically displays a blister on the front facet, the damaged area of QCLs is significantly extended into the InP substrate [9], [67]. An SEM image after CMD is shown in Fig. 14, indicated a large void is present [67]. TEM failure analysis on devices that exhibited CMD, reveal that dislocations are not confined within the active layer, but propagate into the substrate [67]. This observation is in marked contrast to what is observed in degraded diode lasers.

Note that no correlation has been established between these facet failures and optical-power density at the facet, except for longer wavelength ($\lambda \sim 8\ \mu\text{m}$) devices with uncoated facets [68] for which an oxide, absorbing at $\sim 8\ \mu\text{m}$ wavelength, formed on the emitting facet. At high power output ($>1\ \text{W}$) operating conditions, a large amount of heat is generated in the core region of the devices. For a typical QCL with a power-conversion

efficiency $\eta_{\text{wp,CW}} \sim 10\%$, at 1 W output power 9 W of heat must be extracted to avoid excessive device heating. At maximum output powers, typically this heat load increases to 25–30 W. By contrast, for highly efficient near-infrared diode lasers ($\eta_{\text{wp,CW}} > 50\%$), 1W of heat must be removed for watt-range output power devices. Thus, the maximum reliable QCL output power is not necessarily limited by the structure, but by the efficiency with which heat generated at the semiconductor (active) core region (during laser operation) is removed by the cooling scheme employed for heat removal. Since the device performance and lifetime are exponentially related to the operating temperature, controlling the heat removal from the (active) core region is critical to achieving reliable performance of high-power lasers.

III. METAMORPHIC MATERIALS FOR QCLS

The direct growth of InGaAs/AlInAs/InP QCLs on substrates other than InP opens up exciting possibilities for achieving either higher performance levels or exploiting well established integration technologies on GaAs or Si optoelectronic platforms. Monolithic integration of lasers by direct growth on Si has generated much interest as a means to take advantage of the rapidly expanding capabilities of photonic integrated circuit manufacturing offered by large-area, full wafer (silicon IC manufacturing is up to 18"-diameter wafers at present) Si. The direct integration of optical sources, such as quantum cascade lasers and detectors with Si-based photonics (waveguides, modulators, diffraction gratings, etc.) could enable compact integrated lab-on-chip gas sensing systems [69], [70]. In recent years the functionality of Si-photonics has exploded, although the realization of direct heteroepitaxy for III/V-based devices still poses many challenges including: a large (8%) lattice mismatch between InP and Si, a 50% mismatch in the thermal coefficient of expansion, and the formation of antiphase domains originating from the growth of polar III-V (InP) compounds on non-polar (Si) substrates [71]. To overcome these issues, specially designed metamorphic buffer layers (MBLs) are required to provide a low-defect-density growth surface for the subsequent device (in this case QCL) growth. While such MBLs have proven successful for the integration of interband-transition quantum-dot active region diode lasers onto Si [72], only a few studies have been reported on InGaAs/AlInAs/InP QCLs grown on Si [73] or other highly-lattice-mismatched substrates such as GaAs [74]–[76].

A compositionally graded buffer layer (MBL) is the most common approach to accommodate materials with lattice constants that differ from the substrate. These buffer layers consist of intermediate layers of material, which are grown in a compositionally-graded manner to shift from the substrate lattice constant to the target value. In many materials systems, threading defects tend to annihilate as films grow thicker, leading to a roughly logarithmic decrease in threading dislocation density with film thickness. As a result, the achievement of an acceptably low defect density MBL often requires the growth of a rather thick buffer layer (several microns thick). There are many reports of compositionally-graded MBLs across a wide variety of material systems and growth techniques [77], [78].

The strain relaxation, which occurs within the MBL, gives rise to a surface with a cross-hatched morphology that negatively impacts the performance of device structures grown atop MBLs. While threading-dislocation densities in the layers grown on top of the MBL can be sufficiently low (typically in the 10^5 – 10^6 cm $^{-2}$ range) for many device applications, the underlying surface morphology of the MBL will impact the interfacial structure and compositional uniformity of device structures grown atop the MBL. Chemical mechanical polishing (CMP) was found to be effective to planarize the MBL surface prior to regrowth of the device structure, although procedures must be established to prepare an epi-ready surface for regrowth [79].

Other effective approaches to reduce dislocation densities involve the use of the Aspect Ratio Trapping (ART) technique [71], [80]–[82] and strained defect filtering layers [83]. In some cases, these techniques have been combined to sequentially lower the dislocation density [71]. Note that since QCLs are unipolar devices, dislocation assisted electron-hole recombination is not a concern. Nevertheless, the residual dislocations may lead to electron scattering as well as poor tunneling due to poor quality well/barrier interfaces, reduced state lifetimes, and increased device internal losses and/or carrier leakage.

A. 4.6 μ m-Emitting QCLs Employing InP Buffer Layers on GaAs or Si

QCLs grown on an InP buffer layer, which has a lattice-mismatch to the underlying substrate (Si or GaAs), allows for a direct comparison with laser performance on actual InP substrates. Such studies provide insights into the impact of threading dislocations and non-ideal surface morphology on the growth and characteristics of QCLs, in addition to allowing for the integration of QCLs with lower-cost Si or GaAs photonic platforms.

InGaAs/AlInAs/InP QCLs grown on mismatched substrates such as GaAs [74] or Si [73] have been demonstrated by employing molecular beam epitaxy (MBE). On a GaAs substrate, a compositionally graded AlInAs MBL was employed for the subsequent MBE growth of a QCL emitting near $\lambda \sim 4.6$ μ m. Room-temperature laser operation under short-pulse current-drive conditions was demonstrated, although with threshold-current densities significantly higher than when employing an InP substrate (i.e., 4.1 kA/cm 2 vs 1.3 kA/cm 2). Presumably, the moderately high density of threading dislocations (10^6 – 10^7 cm $^{-2}$), originating from a large lattice mismatch, is a contributing factor limiting the device performance on the MBL.

InGaAs/AlInAs/InP $\lambda \sim 4.6$ μ m QCLs have also been grown by MBE on Si substrates with 6° miscut towards [111], employing a Ge buffer layer in order to shift the lattice constant from Si towards the GaAs one, followed by a compositionally graded AlInAs MBL terminated with an InP buffer layer [73]. The growth surface exhibited a relatively high residual threading-dislocation density ($\sim 5 \times 10^8$ cm $^{-2}$), although with a relatively low RMS roughness value (0.7 nm). In this case, QCL performance was

limited to low-temperature (170 K) short-pulse current operation. It is possible that the higher threading-dislocation densities on Si, compared to those of the MBL on GaAs substrate [74], are severely impacting device performance. Note that the large Si-substrate miscut, may lead to step-bunching growth for the strained InGaAs/AlInAs SL materials comprising the QCL active core region, which may also be a factor in the poorer device performance on the Si substrate. Studies have also been reported for the growth of QCLs on (100) Si substrates using MOCVD [84]. In this case the MBL consisted of multiple layers of InAs/InP quantum dot (QD) insertions with an InP top surface, and chemical polishing was utilized to create a smooth surface for the subsequent MOCVD growth of 5-period and 40-period QCLs employing a step-tapered active-region (STA) design for emission near 4.8 μ m. The estimated dislocation density of the MBL was 3.2×10^8 cm $^{-2}$ and the RMS roughness was measured to be 2.56 nm after polishing. Cross-sectional TEM images indicated that in regions that were free of dislocations, planar QCL structures are observed for 5-period QCLs. However, when a dislocation was present, the planarity of the layer growth was strongly disrupted by the defect.

Since the X-ray diffraction (XRD) beam samples relatively large regions of the wafer containing many dislocations, the diffraction fringes were severely broadened with respect to those from a QCL on a planar InP-substrate growth [84]. Nevertheless, (224) asymmetrical XRD RSM scans indicated that the active-core region layers remain fully-strained. Electroluminescence measurements from 40-period full QCLs on the InP MBL had a measured FWHM 48.6 meV value, at 77 K, for the primary lasing transition near 5 μ m, but no lasing was observed.

These initial demonstrations of QCL structures grown on MBLs appear to indicate that the growth of a MBL with a low residual threading-dislocation density (TDD) is essential for realizing high-performance devices grown on mismatched substrates. However, understanding the underlying mechanisms behind exactly how TDs affect the performance of unipolar devices such as QCLs needs further investigation. It is interesting to note that longer wavelength ($\lambda \sim 8$ μ m) QCLs employing binary (InAs/AlSb) active regions, grown on a GaSb buffer layer on a Si substrate, exhibit performance levels only marginally degraded compared to devices grown on native InAs substrates, even though the dislocation densities are estimated to be in the 10^8 cm $^{-2}$ range [85]. Such material systems appear to be more tolerant to residual threading dislocations than ternary, InGaAs/AlInAs-based QCL structures on InP MBLs grown on Si substrates.

B. Short-Wavelength QCLs Employing MBLs

While the highest performance QCLs with multi-watt-range output powers have been reported in the 4.5–5.0 μ m wavelength region, as discussed in Section II, significant challenges remain to realizing the same high performance at other emission wavelengths, particularly wavelengths shorter than about 3.5 μ m. There are currently no CW high-power (>0.5 W) QCLs operating in the 3.0–3.5 μ m wavelength region. These performance shortcomings at shorter emission wavelengths are largely

dictated by material constraints, and the limited accessibility to alloy compositions imposed by strain-relaxation considerations. Current state-of-the-art QCLs employ superlattice (SL) materials, such as strained-layer InGaAs/AlInAs grown pseudomorphic to InP substrates. Through strain engineering, employing MBLs which act as a virtual substrate with an optimal designer-chosen lattice-constant, the QCL design space can be significantly expanded to enable device performance at short wavelengths comparable to that now achieved in the well-established $4.5\text{--}5.0\ \mu\text{m}$ wavelength region. To accommodate the larger electron-transition energy, deeper wells and taller barriers (i.e., higher strain) are necessary to prevent excessive active-region carrier leakage. However, the barrier and well compositions that can be accessed are limited by strain-thickness considerations, in order to avoid strain relaxation. Note that the critical thickness at which strain relaxation occurs is dependent on the growth conditions, since relaxation can be controlled to some extent by kinetics. InP-based $\sim 3.5\ \mu\text{m}$ -emitting QCLs [86], [10] have in some cases demonstrated high CW output power (0.5 W [86]), although such devices utilize InGaAs/AlInAs superlattice (SL) active regions of highly-strained ($\Delta a/a \sim 1.5\text{--}2\%$) QWs and barriers, and as a result the thermal conductance is expected to degrade relative to lower strain active-core regions [39]. Furthermore, the impact of such highly-strained materials on device reliability is largely unknown. Even higher strain ($\Delta a/a \sim 3\%$) barriers have been used to achieve emission wavelengths as short as $3.0\ \mu\text{m}$ [87], although, as expected, high thermal resistance as well as low T_0 and T_1 values limited the CW output power to quite low values ($\sim 2.5\ \text{mW}$). Such degraded performance relative to longer wavelength QCLs may be due in-part to strong carrier leakage to satellite valleys (L, X) for $\sim 3.05\ \mu\text{m}$ -emitting devices grown on InP, since the quantum wells for such devices require very high indium contents [88], [89]. Another approach to mitigate the high strain is to employ composite barriers, allowing pulsed lasing at $\sim 3.3\ \mu\text{m}$, although leakage to satellite valleys is still an issue as evidenced by low T_1 values [90].

InGaAs/AlInAs active-core region QCLs which utilize a MBL grown on a GaAs substrate, offer an approach which significantly broadens the compositional and layer-thickness range that can be accessed for QCL-device design [75], [91]–[93]. Such designs enable short-wavelength emission (i.e., $\lambda \leq 3.5\ \mu\text{m}$), while maintaining similar strain values as those used for $4.5\text{--}5.0\ \mu\text{m}$ -emitting QCLs grown on InP. Furthermore, such designs utilize significantly *lower* indium contents in the quantum wells compare to those employed for conventional $3.0\text{--}3.5\ \mu\text{m}$ -emitting QCLs, so carrier leakage to (L, X) satellite valleys can be eliminated. That is, the employment of an adjustable lattice-constant or *virtual substrate* (i.e., MBL) may enable QCLs with the potential for significantly improved performance at shorter wavelengths than currently possible. Of course, to take advantage of such designs, appropriate MBLs with smooth surface morphology and ultra-low dislocation densities must be available.

To illustrate the benefit of choosing an optimal substrate lattice constant for QCL design, the simulated conduction band offset, ΔE_c , between compressively-strained (CS) In_xGa_{1-x}As QWs

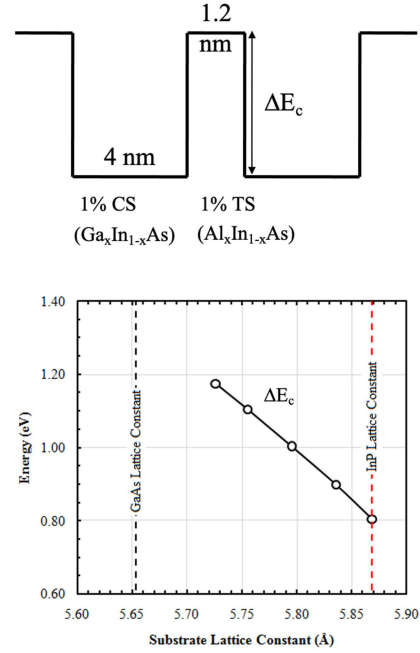


Fig. 15. Simulated conduction-band offset, ΔE_c , between an In_xGa_{1-x}As quantum well and an Al_xIn_{1-x}As barrier as a function of the virtual-substrate lattice constant, when the compressive and tensile strain in QWs and barriers is 1%. Adapted from [75] with permission from Elsevier Publishing.

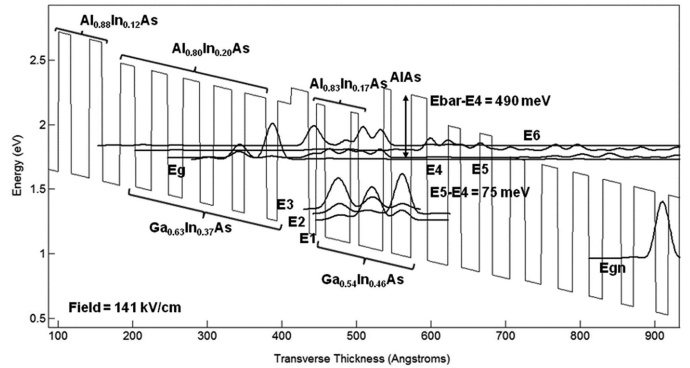


Fig. 16. Conduction band diagram under bias for a STA QCL stage design for emission near $3.1\ \mu\text{m}$. The lattice constant of the MBL is $0.577\ \text{nm}$, which corresponds to relaxed In_{0.30}Ga_{0.70}As (representing the top surface of the MBL). Reproduced with permission from [91]. Copyright 2014 IET.

and tensilely-strained (TS) Al_xIn_{1-x}As barriers as a function of the virtual-substrate lattice constant is shown in Fig. 15.

Fig. 15 clearly shows that significantly deeper QWs can be achieved by choosing an MBL lattice-constant value between those for GaAs and InP. Note that below a lattice constant of $5.72\ \text{\AA}$ it becomes difficult to maintain the strain*thickness product values without utilizing excessively tall (i.e., AlAs) barrier thickness. A large ΔE_c value is required for the design of efficient short-wavelength ($\lambda < 3.5\ \mu\text{m}$) QCLs, since it allows for the implementation of carrier-leakage-suppression AR designs. One such design, shown in Fig. 16, illustrates the conduction-band-engineered step-taper active (STA) region for emission at $3.1\ \mu\text{m}$ [91]. The advantages of the STA design over the conventional active region QCL design are addressed in Section II. In addition, carrier leakage to satellite valleys (L, X) is

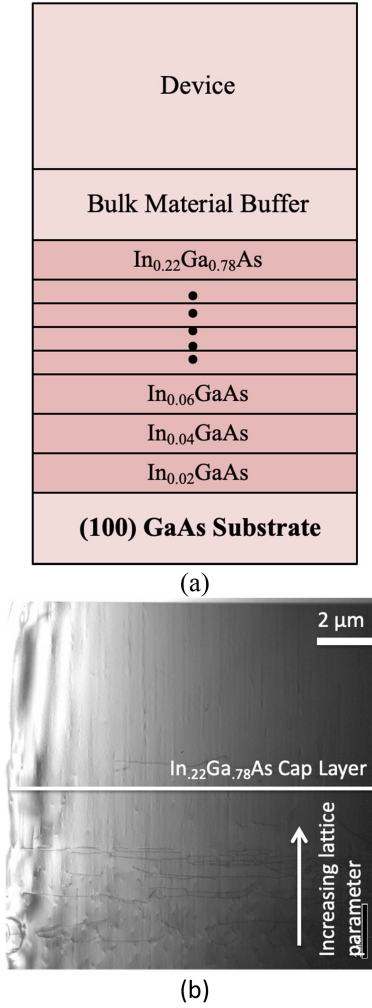


Fig. 17. (a) Schematic diagram of a compositionally graded InGaAs MBL as a growth platform for subsequent device growth; (b) HVPE-grown metamorphic buffer layer consisting of $\sim 2 \mu\text{m}$ thick layers with stepped-changes in the In concentration. The dark line features are dislocations introduced to accommodate the lattice mismatch between layers of different composition and lattice constant and the GaAs substrate.

also essentially eliminated owing to the relatively low In content in QWs (i.e., 46%) for this QCL on MBL when compared to conventional QCLs grown on InP substrates for this wavelength range.

The strain $\times$ thickness products for the active-region wells and barriers, $-0.330 \text{ \AA}$ and $0.093 \text{ \AA}$ respectively, are comparable to those for high-performance QCLs on InP emitting near $4.8 \mu\text{m}$ [20]. These values are also significantly lower than that reported for highly-strained conventional QCLs emitting near $3 \mu\text{m}$, $-0.570 \text{ \AA}$ (well) and $0.614 \text{ \AA}$ (barrier) [87].

To achieve a lattice constant in the range 5.72 to $5.77 \text{ \AA}$, a compositionally graded $\text{In}_x\text{Ga}_{1-x}\text{As}$ MBL can be utilized on a GaAs substrate with a final indium content $x \sim 0.22$ to 0.3 , as shown schematically in Fig. 17(a). Note that since the final constant composition of the MBL is generally not fully relaxed (typically $>90\%$ relaxation) the residual strain must be considered when choosing the final layer indium composition in order to target a specific lattice constant.

One attractive approach for realizing a suitable MBL involves ultra-thick compositionally graded $\text{In}_x\text{Ga}_{1-x}\text{As}$ buffer layers grown by hydride vapor phase epitaxy (HVPE), taking advantage of the extremely high growth rates [94], [95]. For example, the step-compositionally graded $\text{In}_x\text{Ga}_{1-x}\text{As}$ MBL shown in Fig. 17 is grown using hydride-vapor-phase-epitaxy (HVPE) at very high growth rates ($\sim 40 \mu\text{m/hr}$), allowing for the growth of ultra-thick MBLs with relatively short growth times. In addition, constant composition capping layers on top of the MBL can be easily grown with thickness of $10\text{--}20 \mu\text{m}$, resulting in very low residual threading dislocation densities $\sim 10^5 \text{ cm}^{-2}$ [94]. Such InGaAs MBLs are attractive for realizing short wavelength ($\lambda < 3.5 \mu\text{m}$) QCLs with low strain.

The first report of short-wavelength ($\lambda < 4 \mu\text{m}$) electroluminescence (EL) was reported for a QCL on HVPE-grown MBL employing a single active-core region stage consisting of $\text{In}_{0.37}\text{Ga}_{0.63}\text{As}$ QWs and $\text{Al}_{0.80}\text{In}_{0.20}\text{As}$ barriers [91]. The measured emission wavelength from half-mesa devices (i.e., no Fabry-Perot cavity) at 80 K was $\lambda \sim 3.6 \mu\text{m}$, which deviated from the simulated value of $3.1 \mu\text{m}$ for this active-region device.

Full-QCL (30-period) active-core regions were also fabricated using MOCVD regrowth on HVPE-grown MBLs, and structural characterization of the active region was reported [76], [96]. Prior to the growth of the device structure, a wet-etching procedure was utilized to remove residual damage from the CMP process used to planarize the surface cross-hatching. This pre-regrowth etching process was found to be critical to avoid indium segregation at the top surface of the ternary MBL [76]. Conventional, single-phonon-resonance active-region structures in conjunction with resonant-tunneling extraction from the lower laser level were employed [92], [76], [93]. The MBL top layer had a composition of $\text{In}_{0.22}\text{Ga}_{0.78}\text{As}$, with an active core region consisting of AlAs barriers and $\text{In}_{0.45}\text{Ga}_{0.55}\text{As}$ wells, and designed to emit at $3.39 \mu\text{m}$.

Experimental studies on the properties of QCL active-region materials grown on compositionally graded $\text{In}_x\text{Ga}_{1-x}\text{As}$ MBLs indicate that high structural integrity full-QCLs be achieved, although surface morphology and residual dislocation density of the underlying MBL are important considerations. Furthermore, while the active region materials can consist of strain-balanced SL materials, the choice of the optical waveguide cladding layer materials are restricted to ternary materials (i.e., $\text{Al}_x\text{In}_{1-x}\text{As}$ or $\text{In}_x\text{Ga}_{1-x}\text{P}$) which can be lattice-matched to the top of the MBL. This presents a challenge for achieving both optical confinement and low thermal resistance, since the ternary materials generally have poorer thermal conductivity compared with that of InP. Another important issue to be addressed is the doping level within device structures grown on MBLs. Doping levels in the laser's optical waveguide cladding layers need to be optimized to minimize free-carrier absorption, which can significantly impact the device internal losses and slope efficiency. In addition, doping in the SL injector regions of the device controls the dynamic operational current range and also impacts threshold-current density and carrier backfilling (i.e., the T_0 value) [12]. Since all calibration layers must be grown on the MBL, the use of a semi-insulating substrate is not possible. This precludes the use

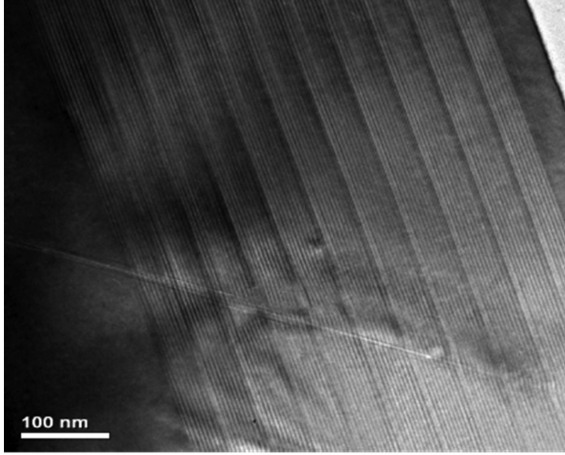


Fig. 18. Cross sectional STEM images of the 10-stage QCL structure MOCVD regrowth on an HVPE-grown MBL. SL layers near the defect site are deformed in thickness.

of Hall measurements for doping calibrations, making it more challenging to characterize doping levels within the device.

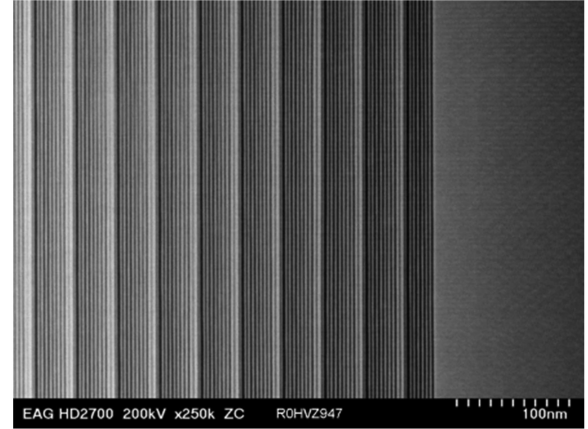
10-period [76] and 30-period [96] QCL structures grown by MOCVD with lattice-matched AlInAs cladding layers on top of polished HVPE-grown MBLs were used to analyze the structural characteristics of the layers using TEM imaging and XRD analysis. Cross-section TEM images (Figs. 18 and 19) were used to evaluate layer thickness and uniformity. Certain areas of the image (Fig. 18) show defects penetrating into the active region and deforming the SL layers near the defect. It is likely that the compositions of the layers near the defect are also highly nonuniform. However, other areas which are defect free, show uniform QCL stages, although some slow modulation of the layers is observed under high magnification, as shown in Fig. 19.

XRD data, for 30-stage devices show diffraction peaks which are in good agreement with the XRD simulation, although some degree of fringe broadening is evident compared with the XRD simulation [101]. Doping was not yet calibrated, so current injection was not possible for these initial devices.

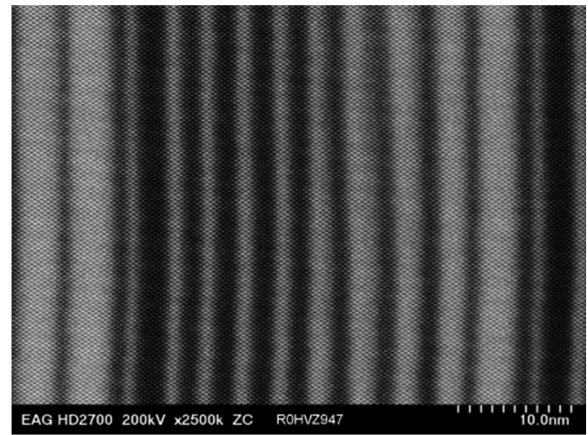
Remaining challenges for realizing high performance In-GaAs/AlInAs active-region QCLs on MBLs may require an improved understanding of how the residual dislocations impact both composition and layer thickness, as well as the sensitivity of the device performance to these factors. The use of three-dimensional compositional mapping using atom probe tomography may help elucidate the structural properties of the active region in the vicinity of the defects. New MBL structures with reduced threading dislocation densities are also needed to take full advantage of the potential performance advantages offered by the MBL “virtual substrate” platform.

IV. BEAM CONTROL AND STABILITY

Many applications for QCLs, such as local sensing and spectroscopy, do not necessitate high output powers. However, scaling the coherent power of mid-IR-emitting QCLs to the multi-watt range remains an important objective for a variety of applications such as remote sensing of pollutants and explosives,



(a)



(b)

Fig. 19. Cross sectional STEM images of the full 30-stage QCL structure MOCVD regrowth on an HVPE-grown MBL, shown at (a) low magnification and (b) high magnification.

free-space communications as well as infrared countermeasures. Many approaches for achieving beam control and coherent-power scaling have been based on those originally developed for near-infrared diode lasers. In this review, we focus on edge-emitting lasers, although there are also ongoing parallel developments of surface-emitting devices with promising results.

A. Single-Element Narrow-Aperture Devices

Single-element, edge-emitting QCLs operating in the atmospheric-transmission wavelength window of $4.5\text{-}5.0\ \mu\text{m}$ generally require relatively narrow element widths ($4.0\text{-}4.5\ \mu\text{m}$) in order to maintain stable, single-spatial-mode CW operation up to the >1 watt-range CW output power. The highest reported single-mode CW output power at $4.5\ \mu\text{m}$ is $1.4\ \text{W}$ from a buried-heterostructure (BH) QCL [10], although commercially available devices are specified up to $2.5\ \text{W}$ CW [97]. Higher CW output powers ($3\text{-}5\ \text{W}$) have been achieved from wider width BH-ridge widths, although such devices operate multi-mode [2], [11] and exhibit beam steering with increasing drive level [11], [98]. External cavity beam-combining many single-element QCLs offers a path to scale the coherent output power of high-brightness mid-IR-emitting sources. However, in many applications, monolithic approaches to coherent-power scaling

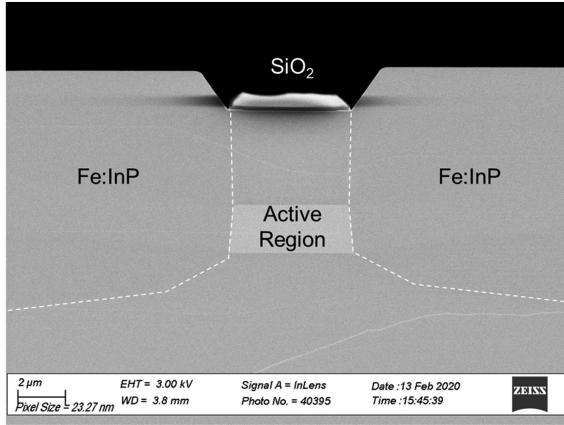


Fig. 20. SEM cross-section image of narrow ridge BH QCL after HVPE regrowth of Fe:InP. Reproduced with permission from [51] © The Optical Society.

are preferred. In addition, any improvements in the single-mode output power of individual QCLs can be directly exploited in externally beam-combined systems. Beam stability is also important for applications such as remote sensing or free-space communications, which require a high degree of beam-pointing accuracy. Studies have shown [51], [98], [99] that multi-mode operation leads to beam instability and wandering. Even for devices which exhibit low M^2 values beam instabilities may be present, thus making them unsuitable for applications requiring high pointing accuracy.

Studies on narrow-ridge-width BH QCLs indicate that the beam properties and mode stability can be highly sensitive, under QCW/CW operating conditions, to the dimensions of the width of the buried ridge as well as the BH-regrowth morphology [100]. Accurate control of the ridge width and Fe:InP regrowth planarization is facilitated by employing ICP dry etching and hydride vapor phase epitaxy (HVPE), as shown in Fig. 20. The high growth rates afforded by HVPE [51], [100] are attractive, since the Fe:InP regrowth necessary is generally very thick (10–15 μm) when forming the BH. While MOCVD regrowth of the Fe:InP is also possible, it is generally more challenging to achieve highly planar surface morphology for the selective growth process [101].

As shown in Figs. 21 and 22, nominal ridge widths, less than $\sim 4.5 \mu\text{m}$ -wide, result in high lateral beam quality ($M^2 < 1.1$) and beam stability without any measurable beam wander, up to QCW output powers of 0.55W [51]. By contrast, larger (i.e., $\sim 5 \mu\text{m}$ -wide) ridge widths provide higher QCW output powers ($\sim 1\text{W}$) although they exhibit a small degree of lateral beam wander with increasing output power ($< 0.125 \text{ mrad}$), even though M^2 maintains a value less than 1.25. Note that the transverse (perpendicular to the epi-layers) beam is stable with drive current, due to the fact that the transverse waveguide is generally designed to support only a single spatial mode. Since higher-order lateral modes are not cut-off, even for a 4 μm -wide ridge width, differences in the lateral optical-confinement factors and metal losses originating from overlap of the optical fields with the top metallization play a role in establishing a threshold-gain difference between lateral modes. While the

small amount of lateral-beam instability shown on Fig. 22(a) may be acceptable for some applications, it corresponds to a beam-pointing inaccuracy of $\sim 1.25 \text{ cm}$ over a distance of 100 m.

Since QCLs rely on the tunneling injection of carriers into the upper laser level of the active region, they exhibit a maximum operating current density (J_{max}) which is strongly dependent on the injector doping level and the electronic-wavefunction coupling strength between the upper laser level and the appropriate injector energy state, but is typically in the range of 3–4 times the threshold-current density, J_{th} . Thus, the peak output power at J_{max} under short-pulse current operation (i.e., no heating) is ultimately limited by the active-core region volume, defined by the number of periods and the *device area*. Under QCW/CW operation, self-heating degrades the QCL performance and leads to thermal roll-over of the power-current (P-I) characteristics at lower current densities than J_{max} under short-pulse operation [2]. Note that this J_{max} limitation is very different from the behavior of diode lasers, where in the absence of self-heating (i.e., short current pulse operation) the maximum output power is limited by either catastrophic optical mirror damage (COMD) or P-I rollover due to strong carrier leakage.

As discussed in Section II, self-heating under QCW/CW operation leads to an increase in J_{th} , lower slope efficiencies, and under extreme heating triggers device failures [9], [65]. To predict the CW performance and optimize the device geometry for a given active-region design, semi-empirical models of the QCL are valuable. These models utilize measured device parameters under short-pulse operation, such as internal losses, T_0 , T_1 , and the I-V curve as model inputs. Thermal diffusion and full-wave optical models of the device structure are used to calculate the optical confinement factor, Γ , and average core-region temperature rise as a function of the device geometry and dissipated power. Since optical-mode competition is not included in such a model, the device structure is assumed to support only the fundamental TM_{00} mode. Nevertheless, such semi-empirical models are very useful to accurately predict QCW/CW performance as a function of device geometry and active-region properties.

To illustrate the utility of such a model, the pulsed and CW output characteristics of a single-stripe (8.5 μm wide, 5 mm long, 40-period core, HR-coated back facet, LR front-facet coating with a nominally 14% reflectivity) BH STA-RE-type QCL emitting at $\sim 5.1 \mu\text{m}$ has been correlated with experimental data [102]. The simulated output power vs current (P-I) under both short pulse (i.e., no heating) and CW operation are both in excellent agreement with the measured data for this device, as shown in Fig. 23.

Furthermore, the model can be used to optimize the device for a specific device performance metric, such as CW output power or CW wall-plug efficiency. For example, varying the number of core-region stages, for a given active-region material, serves to illustrate important design tradeoffs between device width and the number of core-region periods. A larger number of core-region periods improves the slope efficiency, but leads to large operating voltages. While a narrow stripe width allows for better heat extraction, it also reduces the active volume.

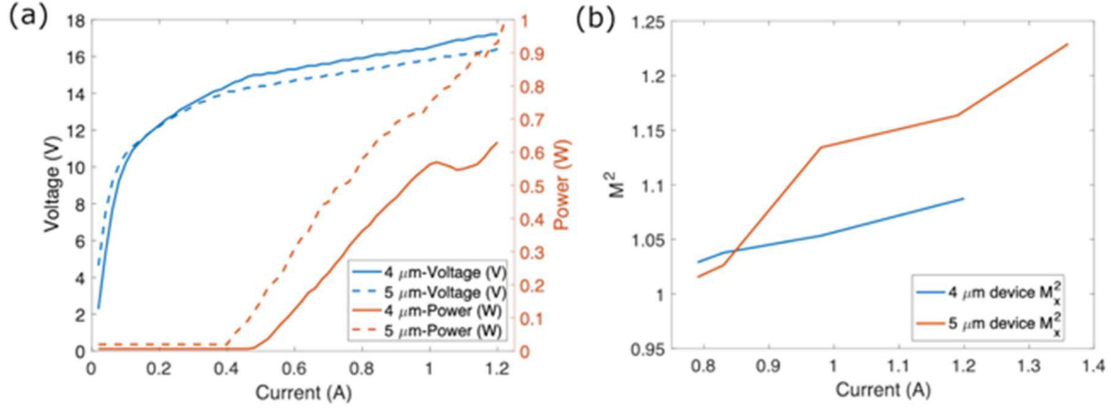


Fig. 21. (a) Measured L-I-V characteristics under QCW operation and (b) M^2 along x-direction (lateral) under QCW operation for the 4- and 5- μm -wide BH devices after beam collimation. Reproduced with permission from [51] © at The Optical Society.

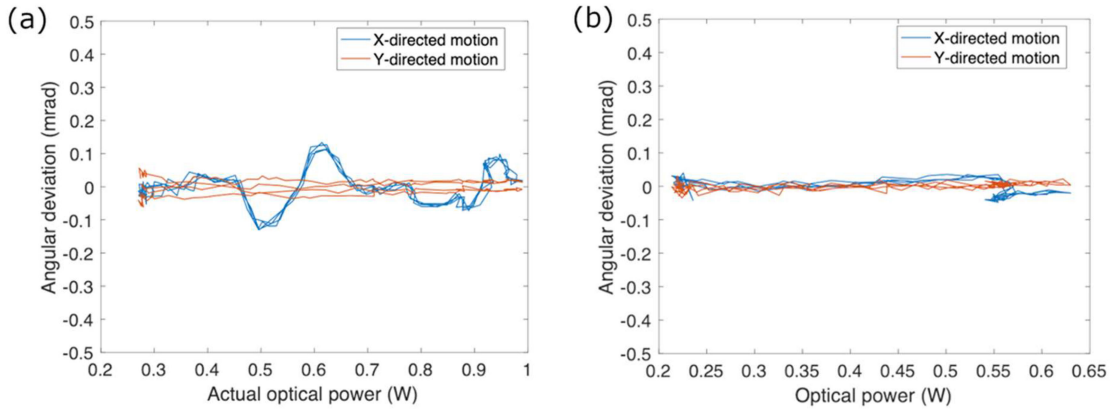


Fig. 22. Measured angular deviation of the collimated-beam far-field centroid position along x- (lateral) and y- (transverse) directions, as a function of output power under QCW operation for (a) $\sim 5\ \mu\text{m}$ -wide ridge device and (b) $\sim 4\ \mu\text{m}$ -wide ridge device. Reproduced with permission from [51] © at The Optical Society.

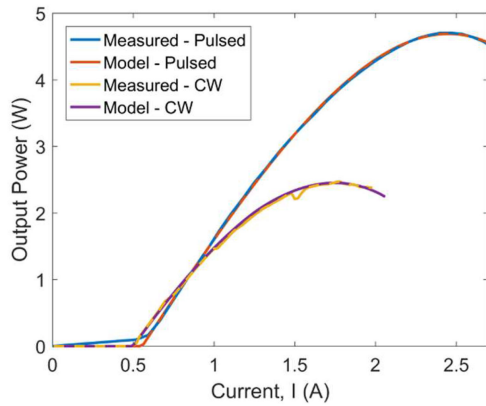


Fig. 23. Simulated and measured P-I characteristics for a STA-RE QCL emitting at $\sim 5.1\ \mu\text{m}$ wavelength [102].

The result of such a simulation is shown in Fig. 24, indicating the optimal number of stages and device width which would maximize the CW output power. In this case, 20 periods and a 22 μm -wide ridge width are projected to increase the maximum CW output power from $\sim 2.5\ \text{W}$ (for 40 periods, 8.5 μm ridge width) to $\sim 3.1\ \text{W}$. However, note that penalties will occur for

other device metrics as a result of such an optimization for CW output power. A fewer number of periods leads to a higher J_{th} value and a lower wall-plug efficiency, while the wide ridge will support many lateral modes and result in degraded beam quality.

B. Scaling the Aperture Width: Edge-Emitting Lasers

Many approaches have been reported in the literature aimed at coherently scaling the QCL output power beyond that possible for a narrow-ridge BH device, while maintaining high beam quality and stability. Many of these device concepts were originally developed for near-infrared diode lasers and then implemented to QCLs. These approaches can be broadly classified in three categories: 1) broad-area (weakly-guided) structures; 2) strongly-index-guided structures with weak overall interelement coupling (i.e., series coupling) [103]; and 3) strongly-index-guided (or antiguided) structures with strong overall interelement coupling (i.e., global coupling) [103], including High-Index Contrast Photonic Crystal (HC-PC) approaches.

Weakly-guided devices, such as broad area lasers with lateral-mode filtering [104], [105], tapered lasers [106], [107], Master

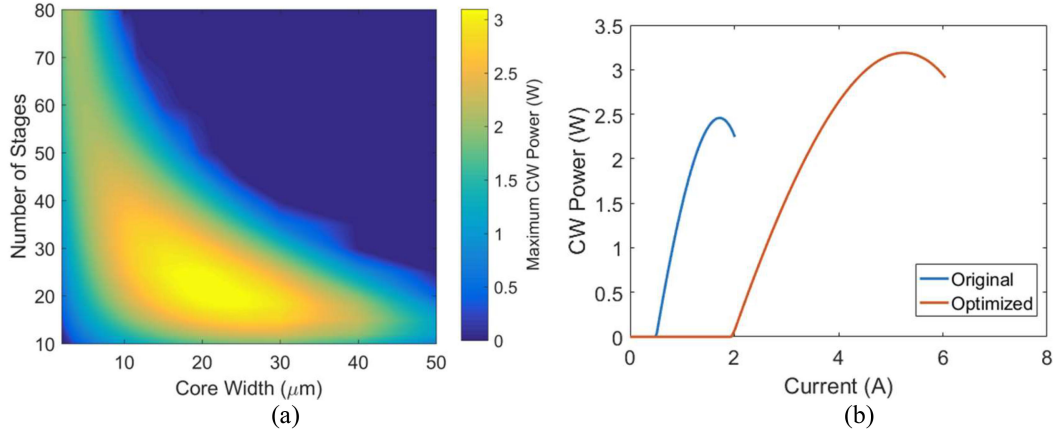


Fig. 24. (a) Simulated maximum CW output power from a parametric sweep of current density, width, and number of stages, (b) simulated L-I curves for a 8.5 μm -wide (40-period) BH QCL and a 22 μm -wide (20-period) device optimized for CW output power [102].

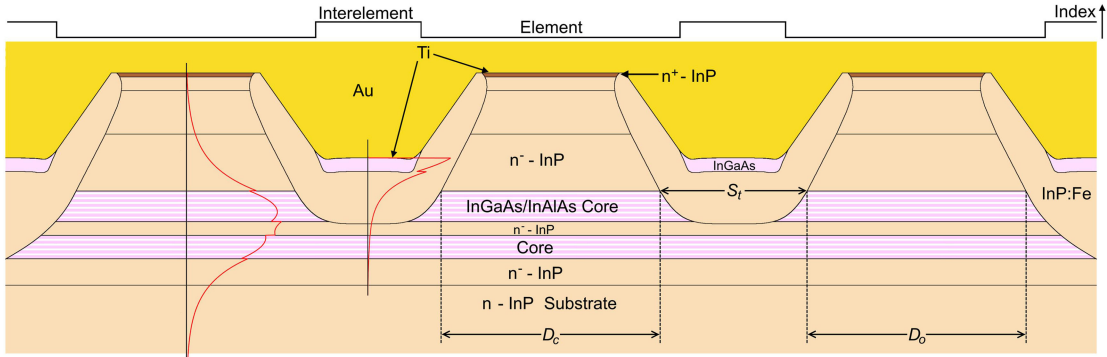


Fig. 25. Schematic representation of the cross-sectional view of a 3-element HC-PC QCL structure © 2019 IEEE. Reprinted with permission from [122].

Oscillator Power Amplifiers (MOPAs) [108], [109], angled-facet waveguide devices [110], [111] and PC lasers with weakly index-guided structures such as diffraction gratings (i.e., PC-DFB lasers) [112], [113] have all been used to provide a higher degree of spatial coherence compared with multi-mode, wide-ridge lasers under low-duty cycle pulsed-current operation. However, as device self-heating becomes significant under QCW/CW operation, thermally induced index perturbations can destabilize the optical mode in those weakly-guided structures. Strongly index-guided structures with weak overall interelement coupling include evanescently coupled phase-locked laser arrays [114], [115], and diffraction-coupled phase-locked arrays with Talbot-type cavities or intracavity filters [116]–[118]. For such devices the interelement coupling is of the nearest-neighbor type [103], [119] which creates nonuniform array-mode intensity profiles and subsequently leads to easy multimoding above threshold via gain spatial hole burning at the array level. A summary of reported beam properties for selected devices is summarized in Table I.

In general, a built-in strongly-index-guided architecture is preferred to help suppress the impact of the unavoidable thermally-induced-index variations across the device during QCW/CW operation. In a weakly index-guided device, such thermal lensing can lead to multi-mode operation and/or mode instability as heating increases with increasing output powers or device degradation. However, as pointed above, even strongly-index-guided structures can become multimode and/or

temporally unstable if the overall interelement coupling is weak (i.e., of the nearest-neighbor type).

Strongly-index guided monolithic approaches to scale the device area, and thereby the maximum output power, while maintaining single-mode operation, include the use of closely-spaced phase-locked antiguided arrays of QCLs [120]–[122], tree arrays of QCLs employing either Y-junctions [123], [124] or multi-mode interference (MMI) coupling sections [125], [126], and BH HC-PC structures [127]. The beam properties for strongly-index-guided devices with global interelement coupling are also summarized in Table I. Below we discuss the only two approaches which have produced devices capable of pure D.L. beams to multi-watt range output powers.

The array employing MMI coupling is the only approach to date that has enabled scaling the CW output power while maintaining a near-diffraction-limited beam [126]. This approach combines the attributes of a large built-in index step with strong (non-evanescent) optical coupling between array elements for mode stability. The inherent large separation between emitters reduces thermal cross-talk between array elements, enabling high CW output powers while mitigating thermal lensing. However, the large emitter spacing (i.e., low array fill factor) leads to a highly multi-lobe beam pattern with relatively low power ($\sim 5\%$) contained within the central lobe.

An approach with both high array fill factor as well as potential for high QCW/CW output powers is to employ phase-locking of multiple lasers using resonant leaky-wave coupling, which

TABLE I
BEAM PROPERTIES OF WIDE APERTURE EDGE-EMITTING QCLS

Weakly-Index Guided:	
Broad Area with mode filters	$\lambda \sim 4.8\ \mu\text{m}$, 6.0 W pulsed, $2.3 \times$ D.L. [104, 105]
Tapered Lasers	$\lambda \sim 8.1\ \mu\text{m}$, 4.3 W pulsed, $M^2 \sim 1.6$ [106] $\lambda \sim 4.6\ \mu\text{m}$, 4.5 W CW @ 283K, multi-mode, single-lobed beam [107]
MOPA	$\lambda \sim 9.2\text{-}9.8\ \mu\text{m}$, 3.9 W pulsed, near-D.L. [108] $\lambda \sim 7.26\ \mu\text{m}$, 1.5 W pulsed, $< 2 \times$ D.L. [109]
PC-DFB	$\lambda \sim 4.7$, 0.5 W per facet pulsed, D.L. beam [112] $\lambda \sim 4.36$, 34 W, pulsed power, $M^2 \sim 2.5$ near threshold, $M^2 \sim 9$ at high drive current [113]
Angled-stripe QCL	$\lambda \sim 8\ \mu\text{m}$, 1.2 W pulsed, $M^2 \sim 2.0$ [110] $\lambda \sim 4.8\ \mu\text{m}$, 203 W pulsed, $M^2 \sim 3.0\text{-}5.0$ at high drive currents [111]
Strongly index-guided structures with weak overall interelement coupling:	
Evanescently-coupled arrays	$\lambda \sim 8.4\ \mu\text{m}$, 0.23 W pulsed, double-lobed, near-D.L. far-field [114] $\lambda \sim 4.6\ \mu\text{m}$, multi-mode, $> 4.2 \times$ D.L. [115]
Talbot cavity /filter	$\lambda \sim 4.8\ \mu\text{m}$, 4 W pulsed, two array modes, $1.4 \times$ D.L. [116] $\lambda \sim 4.6\ \mu\text{m}$, 1.2 W pulsed, $\sim 1.8 \times$ D.L. [117]
Strongly-index-guided /antiguide structures with strong overall interelement coupling:	
BH HC-PC Structure	$\lambda \sim 8.5\ \mu\text{m}$ 0.88W CW, first lateral mode D.L. [127]
Y-Junction Tree Array	$\lambda \sim 4.6\ \mu\text{m}$ 1.5W CW, multi-mode [124]
MMI Tree Array	$\lambda \sim 4.8\ \mu\text{m}$, 15 W pulsed, 30-lobes D.L. beam [125] $\lambda \sim 8\ \mu\text{m}$, 7 W CW, 35-lobes near-D.L. beam [126]
Resonant Leaky-Wave Coupled Arrays	$\lambda \sim 8.4\ \mu\text{m}$ 5.5 W pulsed, $1.6 \times$ D.L. beam [120] $\lambda \sim 5.3\ \mu\text{m}$ 5.3W pulsed, D.L. beam [122]

has been previously successful in the near-IR spectral region [128]–[131]. Such devices rely on the formation of interelement regions which create a high-index-contrast ($\Delta n = 0.08\text{-}0.10$) [131] defining the array structure. These devices are essentially 1D high-index-contrast photonic-crystal (HC-PC) structures that allow global coupling between all array elements and operate in a single *lateral* mode corresponding to a photonic-crystal, band-edge state [132]. In the effective-index approximation, the interelement regions can be designed to contain an integer number of half-lateral projected wavelengths of the radiation leakage in the interelement regions, leading to global coupling across the array and a nearly uniform near-field intensity profile [103], [128]. In practice, achieving the resonant coupling condition is accomplished by using an array structure shown schematically in Fig. 25, where the interelement regions are formed by MOCVD regrowth. The placement and thickness of the InGaAs layer within the interelement regions, as shown in Fig. 25, allows for control over the lateral index-step and optical loss due to plasmonic coupling to the metal. Such structures are fabricated using a two-step MOCVD process as described in [120], [121]

In the mid-IR, resonant leaky-wave coupling of QCLs emitting up to 5.5 W peak pulsed power at $8.36\ \mu\text{m}$ in a near-diffraction-limited (D.L.) ($1.65 \times$ D.L.) beam pattern [120], as

well as 5.3 W peak pulsed power from resonant leaky-wave coupled QCLs emitting near $5.3\ \mu\text{m}$ with a pure diffraction-limited beam with nearly uniform near-field intensity across the array, Fig. 26 and Fig. 27 [122].

The emitted far-field beam pattern in Fig. 26 indicates operation primarily in the in-phase array mode up to rollover. The measured central lobe is diffraction limited at 6 A, just above threshold. At 10.1 A and 5.3 W front-facet output power, close to rollover, the measured central-lobe far-field beam's FWHM is $1.1 \times$ the diffraction limit, with relatively small sidelobes. The threshold current density is $\sim 3\ \text{kA/cm}^2$, pulsed slope efficiency of 1.6 W/A, and a pulsed wall-plug efficiency of up to 3.5%. Design studies of the three-element array structures were carried out using COMSOL to find the expected threshold currents and fabrication tolerances for a given design [121]. The simulations accurately predict that the observed array geometry should maintain sole in-phase-mode operation to high drive levels, in good agreement with the experimental results.

To elucidate the potential for such devices under CW operating conditions, an optimization simulation involving sweeping a 7-D parameter space of current, element width, number of periods, length, front facet reflectivity, injector doping, and pulse duty cycle was carried out [102], employing the semi-empirical model discussed above for single-element QCLs. The optimized design for a five-element array under CW operation has 22 periods, $12\ \mu\text{m}$ -wide elements, a length of 6 mm, a front facet reflectivity of 6.5%, an injector sheet doping of $1.6 \times 10^{11}\ \text{cm}^{-2}$. The predicted maximum CW output power for this device is 8 W at a current of 20.5 A and voltage of 11.0 V, for a wall-plug efficiency of 3.5%, while the maximum power in pulsed operation (with short pulses and low duty) is 20.3 W at a current of 26.7 A. Using devices of PICT-action design, as outlined in Section II, should significantly raise the CW wall-plug efficiency and output powers. The projected CW P-I curves for 3-element and 5-element arrays, compared to that for an optimized 40-periods ($5\ \mu\text{m}$ -wide, 6 mm-long) BH, are shown in Fig. 28. This simulation indicates that scaling the coherent CW output power by a factor of $\sim 2\text{-}2.5$ over the single-element BH is possible by optimizing five-element array structures. Further scaling using a larger number of elements may be possible, although at diminishing returns due to heat-extraction limitations.

V. CONCLUSIONS AND FUTURE PROSPECTS

Performance levels of quantum cascade lasers (QCLs) have reached watt-range output powers with high beam quality. For $3.8\text{-}6.0\ \mu\text{m}$ -emitting QCLs carrier leakage is a shunt-type leakage current within the active region (AR), which is thermally triggered by IFR and LO-phonon scattering, from the upper laser level and injector states to high-energy AR states, followed by relaxation to low-energy AR and extractor states. Suppression of the carrier leakage via conduction-band engineering and IFR engineering is one of the means for achieving high CW output power and wall-plug efficiency values. The other key means to maximize the CW power and wall-plug efficiency are: transition-efficiency maximization in diagonal-transition

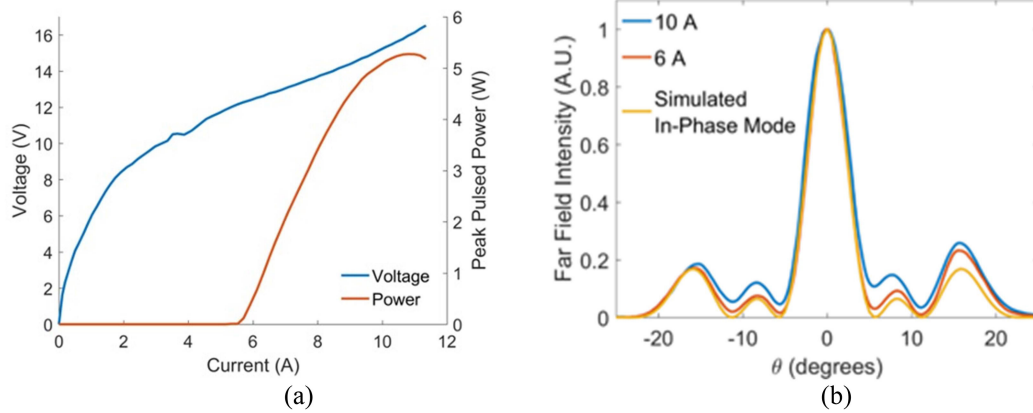


Fig. 26 Measured (a) P-I-V from 3-element resonant array, (b) lateral far-field pattern at two drive currents compared with simulated beam profile © 2019 IEEE. Reprinted with permission from © 2019 IEEE. Reprinted with permission from [122].

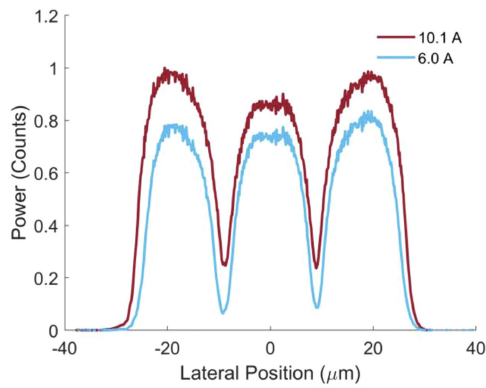


Fig. 27 Measured near-field intensity profile at two different drive currents for 3-element resonant array under low-duty cycle pulsed current operation © 2019 IEEE. Reprinted with permission from [122].

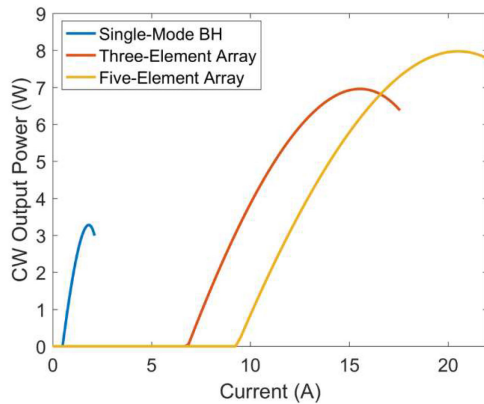


Fig. 28 Simulated CW L-I curves of three-element HC-PC, five-element HC-PC, and BH QCLs, all optimized to maximize the CW output power while maintaining single-spatial-mode operation [102].

devices via IFR-scattering engineering, and minimizing above-threshold voltages by designing devices with strong photon-induced carrier transport (PICT). The net effect will not only be CW-performance optimization, but also enabling long-term reliable operation at watt-range CW power levels

While current studies show promise for establishing long-term QCL reliability, larger studies on accelerated lifetesting

at high CW output powers are necessary. Furthermore, limited data is available on the underlying mechanisms that trigger QCL failures, and on the necessary preventive measures.

Emerging metamorphic materials hold potential to enable high performance levels for wavelengths less than $3.5 \mu\text{m}$, although lower threading-dislocation buffer layers with low surface roughness are needed. Such materials have also shown potential for integration of InP-based QCLs with more mature GaAs- and/or Si-photonics platforms, but much more work remains to be done.

Challenges remain for scaling the output power to multi-watt CW levels while maintaining stable, single-mode operation. Strong self-heating makes mode control difficult under high CW driving conditions. Photonic-crystal (PC) structures with high built-in index profiles appear as the most promising approach for minimizing the impact of thermal lensing under CW operation while scaling the coherent power.

For the future, several developments are in store. First, the combination of virtual carrier-leakage suppression with PICT action is expected to result in significantly higher CW output powers as well as $\sim 40\%$ CW wall-plug efficiency for $4.5\text{-}5.0 \mu\text{m}$ -emitting QCLs, and close to 20% for $\sim 4.0 \mu\text{m}$ -emitting QCLs.

Second, the development of high-quality QCL structures grown by using metamorphic-buffer materials should lead to both $\leq 3.5 \mu\text{m}$ -emitting QCLs with performances comparable to those from $4.0\text{-}5.0 \mu\text{m}$ -emitting QCLs as well as the realization of practical integration on GaAs and InP substrates.

Finally, the coherent power should be able to be scaled in two dimensions [133] by using efficient grating-coupled surface emission [134] from the elements of high-index contrast PC structures [120], [122]. CW powers, surface-emitted in a single-lobe, diffraction-limited beam, as high as 15 W are projected for 7-element, $4.6 \mu\text{m}$ -emitting devices.

ACKNOWLEDGMENT

The authors gratefully acknowledge valuable discussions with Jerome Faist.

REFERENCES

- [1] J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, A. L. Hutchinson, and A. Y. Cho, "Quantum cascade laser," *Science*, vol. 264, 1994, Art. no. 553.
- [2] Y. Bai, N. Bandyopadhyay, S. Tsao, S. Slivken, and M. Razeghi, "Room temperature quantum cascade lasers with 27% wall plug efficiency," *Appl. Phys. Lett.*, vol. 98, 2011, Art. no. 181102.
- [3] M. Kamskar, D. Botez, T. Earles, T. Goodnough, L. J. Mawst, and E. Stiers, "73% CW power conversion efficiency at 50 W from 970 nm diode laser bar," *Electron. Lett.*, vol. 41, pp. 245–247, 2005.
- [4] J. Chen *et al.*, "Type-I GaSb based diode lasers operating at room temperature in 2 to 3.5 μm spectral region," in *Proc. SPIE 7686, Laser Technol. for Defense Secur. VI*, May 2010, Art. no. 76860S.
- [5] J. Faist, *Quantum Cascade Lasers*, Oxford, U.K.: Oxford Univ. Press, 2013.
- [6] C. Boyle, K. M. Oresick, J. D. Kirch, Y. V. Flores, L. J. Mawst, and D. Botez, "Carrier leakage via interface-roughness scattering bridges gap between theoretical and experimental internal efficiencies of quantum cascade lasers," *Appl. Phys. Lett.*, vol. 117, no. 5, 2020, Art. no. 051101.
- [7] D. Botez *et al.*, "Multidimensional conduction-band engineering for maximizing the continuous-wave (CW) wallplug efficiencies of mid-infrared quantum cascade lasers," *IEEE J. Sel. Top. Quantum Electron.*, vol. 19, no. 4, pp. 1200312, 2013.
- [8] D. Botez *et al.*, "High-efficiency, high-power mid-infrared quantum cascade lasers," *Opt. Mater. Exp.*, vol. 8, no. 5, pp. 1378–1398, 2018.
- [9] Q. Zhang *et al.*, "Thermal induced facet destructive feature of quantum cascade lasers," *Appl. Phys. Lett.*, vol. 96, no. 14, 2010, Art. no. 141117.
- [10] F. Xie *et al.*, "Room temperature CW operation of short wavelength quantum cascade lasers made of strain balanced GaIn_{1-x}As/Al_yIn_{1-y}As material on InP substrates," *IEEE J. Sel. Topics Quantum Electron.*, vol. 17, no. 5, pp. 1445–1452, Sep./Oct. 2011.
- [11] Y. Bai, N. Bandyopadhyay, S. Tsao, E. Selcuk, S. Slivken, and M. Razeghi, "Highly temperature insensitive quantum cascade lasers," *Appl. Phys. Lett.*, vol. 97, no. 25, 2010, Art. no. 251104.
- [12] D. Botez, C.-C. Chang, and L. J. Mawst, "Temperature sensitivity of the electro-optical characteristics for mid-infrared ($\lambda = 3\text{--}16\ \mu\text{m}$)-emitting quantum cascade lasers," *J. Phys. D: Appl. Phys.*, vol. 49, no. 4, 2016, Art. no. 043001.
- [13] K. M. Oresick, J. D. Kirch, L. J. Mawst, and D. Botez, "Highly efficient long wavelength infrared, step-tapered quantum cascade lasers," in *Proc. SPIE*, vol. 11705, 2021, Art. no. 1170515.
- [14] D. Botez, S. Kumar, J. C. Shin, L. J. Mawst, I. Vurgaftman, and J. R. Meyer, "Temperature dependence of the key electro-optical characteristics for midinfrared emitting quantum cascade lasers," *Appl. Phys. Lett.*, vol. 97, no. 7, 2010, Art. no. 071101.
- [15] P. M. Smowton and P. Blood, "The differential efficiency of quantum well lasers," *IEEE J. Sel. Top. Quantum Electron.*, vol. 3, no. 2, pp. 491–498, Apr. 1997.
- [16] D. Botez, "Comment on 'Highly temperature insensitive quantum cascade lasers' [Appl. Phys. Lett. 97, 251104, (2010)]," *Appl. Phys. Lett.*, vol. 98, no. 26, 2011, Art. no. 216101.
- [17] Y. V. Flores, S. S. Kurllov, M. Elagin, M. P. Semtsiv, and W. T. Masselink, "Leakage current in quantum-cascade lasers through interface roughness scattering," *Appl. Phys. Lett.*, vol. 103, 2013, Art. no. 161102.
- [18] J. C. Shin *et al.*, "Highly temperature insensitive, deep-well 4.8 μm emitting quantum cascade semiconductor lasers," *Appl. Phys. Lett.*, vol. 94, no. 20, 2009, Art. no. 201103.
- [19] J. C. Shin, "Tapered active-region quantum cascade laser," Ph.D. dissertation, Univ. of Wisconsin-Madison, Madison WI, Ch. 6, pp. 92–103, 2010. [Online]. Available: <http://digital.library.wisc.edu/1793/52493>
- [20] J. D. Kirch, J. C. Shin, C.-C. Chang, L. J. Mawst, D. Botez, and T. Earles, "Tapered active-region quantum cascade lasers ($\lambda = 4.8\text{ m}$) for virtual suppression of carrier-leakage currents," *Electron. Lett.*, vol. 48, no. 4, pp. 234–235, 2012.
- [21] D. Botez *et al.*, "The temperature dependence of key electro-optical characteristics for midinfrared emitting quantum cascade lasers," in *Proc. SPIE*, vol. 7953, 2011, Art. no. 79530N.
- [22] F. Wang, S. Slivken, D. Wu, and M. Razeghi, "Room temperature quantum cascade lasers with 22% wallplug efficiency in continuous-wave operation," *Opt. Exp.*, vol. 28, no. 12, pp. 17532–17538, 2020.
- [23] S. Suri, B. B. Knipfer, L. Mawst, and D. Botez, unpublished work on pocket-injector 4 μm - and 4.6 μm -emitting STA-type QCL designs of strong diagonal lasing transition.
- [24] A. Lyakh, M. Suttinger, R. Go, P. Figueiredo, and A. Todi, "5.6 μm quantum cascade lasers based on a two-material active region composition with a room temperature wall-plug efficiency exceeding 28%," *Appl. Phys. Lett.*, vol. 109, no. 12, 2016, Art. no. 121109.
- [25] A. Tredicucci, F. Capasso, C. Gmachl, D. L. Sivco, A. L. Hutchinson, and A. Y. Cho, "High performance interminiband quantum cascade lasers with graded superlattices," *Appl. Phys. Lett.*, vol. 73, no. 15, pp. 2101–2103, 1998.
- [26] D. Botez, "Design considerations and analytical approximations for high continuous-wave power, broad-waveguide diode lasers," *Appl. Phys. Lett.*, vol. 74, no. 21, pp. 3102–3104, 1999.
- [27] Y. V. Flores *et al.*, "Thermally activated leakage current in high performance short-wavelength quantum cascade lasers," *J. Appl. Phys.*, vol. 113, no. 13, 2013, Art. no. 134506.
- [28] J. D. Kirch *et al.*, "86% Internal differential efficiency from 8-9 μm -emitting, step-taper active-region quantum cascade lasers," *Opt. Exp.*, vol. 24, no. 21, pp. 24483–24494, 2016.
- [29] J. Faist, "Wallplug efficiency of quantum cascade lasers: Critical parameters and fundamental limits," *Appl. Phys. Lett.*, vol. 90, no. 25, 2007, Art. no. 253512.
- [30] A. Bismuto, R. Terazzi, B. Hinkov, M. Beck, and J. Faist, "Fully automated quantum cascade laser design by genetic optimization," *Appl. Phys. Lett.*, vol. 101, no. 2, 2012, Art. no. 021103.
- [31] J. M. Wolf, "Quantum cascade laser: From 3 to 26 μm ," Ph.D. dissertation, ETH no. 24571, Zurich, Dept. Physics, ETH Zurich, 2017.
- [32] H. Choi *et al.*, "Gain recovery dynamics and photon-driven transport in quantum cascade lasers," *Phys. Rev. Lett.*, vol. 100, no. 16, 2008, Art. no. 167401.
- [33] S. Suri, B. B. Knipfer, L. J. Mawst, and D. Botez, unpublished work.
- [34] N. Bandyopadhyay *et al.*, "Watt level performance of quantum cascade lasers in room temperature continuous wave operation at $\lambda \sim 3.76\ \mu\text{m}$," *Appl. Phys. Lett.*, vol. 97, 2010, Art. no. 131117.
- [35] A. Lyakh, R. Maulini *et al.*, "High-performance continuous-wave room temperature 4.0- μm quantum cascade lasers with single-facet optical emission exceeding 2 W," *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 44, pp. 18799–18802, 2010.
- [36] Q. Yang *et al.*, "High-peak-power strain-compensated GaInAs/AlInAs quantum cascade lasers ($\lambda \sim 4.6\ \mu\text{m}$) based on a slightly diagonal active region design," *Appl. Phys. Lett.*, vol. 93, 2008, Art. no. 251110.
- [37] Y. Yao, X. Wang, J.-Y. Fan, and C. F. Gmachl, "High performance 'continuum-to-continuum' quantum cascade lasers with a broad gain bandwidth of over 400 cm^{-1} ," *Appl. Phys. Lett.*, vol. 97, no. 8, 2010, Art. no. 081115.
- [38] F. Wang, S. Slivken, D. H. Wu, and M. Razeghi, "Room temperature quantum cascade laser with 31% wall-plug efficiency," *AIP Adv.*, vol. 10, no. 7, 2020, Art. no. 075012.
- [39] H. K. Lee and J. S. Yu, "Thermal analysis of short wavelength InGaAs/InAlAs quantum cascade lasers," *Solid-State Electron.*, vol. 54, pp. 769–776, 2010.
- [40] R. Maulini, A. Lyakh, A. Tsekoun, and C. K. N. Patel, " $\lambda \sim 7.1\ \mu\text{m}$ quantum cascade lasers with 19% wall-plug efficiency at room temperature," *Opt. Exp.*, vol. 19, no. 18, pp. 17203–17211, 2011.
- [41] D. Indjin, P. Harrison, R. W. Kelsall, and Z. Ikonić, "Influence of leakage current on temperature performance of GaAs/AlGaAs quantum cascade lasers," *Appl. Phys. Lett.*, vol. 81, no. 3, 2002, Art. no. 400.
- [42] X. Gao, D. Botez, and I. Knezevic, "X-valley leakage in GaAs/AlGaAs quantum cascade lasers," *Appl. Phys. Lett.*, vol. 89, no. 19, 2006, Art. no. 191119.
- [43] S. R. Jin *et al.*, "Spectroscopy of GaAs/AlGaAs quantum-cascade lasers using hydrostatic pressure," *Appl. Phys. Lett.*, vol. 89, no. 22, 2006, Art. no. 221105.
- [44] A. Vasanelli *et al.*, "Role of elastic scattering mechanisms in GaInAs/AlInAs quantum cascade lasers," *Appl. Phys. Lett.*, vol. 89, no. 17, 2006, Art. no. 172120.
- [45] Y. T. Chiu, Y. Dikmelik, P. Q. Liu, N. L. Aung, J. B. Khurgin, and C. F. Gmachl, "Importance of interface roughness induced intersubband scattering in mid-infrared quantum cascade lasers," *Appl. Phys. Lett.*, vol. 101, no. 17, 2012, Art. no. 171117.
- [46] M. Franckie *et al.*, "Impact of interface roughness distributions on the operation of quantum cascade lasers," *Opt. Exp.*, vol. 23, no. 4, pp. 5201–5212, 2015.
- [47] R. Terazzi, "Transport in quantum cascade lasers," Ph.D. dissertation, Dept. Physics, ETH Zurich, 2011.
- [48] M. Lindskog *et al.*, "Comparative analysis of quantum cascade laser modeling based on density matrices and non-equilibrium Green's functions," *Appl. Phys. Lett.*, vol. 105, no. 10, 2014, Art. no. 103106.

- [49] T. Grange *et al.*, "Atomic-scale insights into semiconductor heterostructures: From experimental three-dimensional analysis of the interface to a generalized theory of inter-facial roughness scattering," *Phys. Rev. Appl.*, vol. 13, no. 4, 2020, Art. no. 044062.
- [50] W. Zhou, Q.-Y. Lu, D.-H. Wu, S. Slivken, and M. Razeghi, "High-power, continuous-wave, phase-locked quantum cascade laser arrays emitting at 8 μm ," *Opt. Exp.*, vol. 27, no. 11, pp. 15776–15785, 2019.
- [51] J. Ha Ryu *et al.*, "Beam stability of buried-heterostructure quantum cascade lasers employing HVPE regrowth," *Opt. Exp.*, vol. 29, pp. 2819–2826, 2021.
- [52] M. S. Vitiello *et al.*, "Influence of InAs δ layers on the optical, electronic, and thermal characteristics of strain-compensated GaInAs quantum-cascade lasers," *Appl. Phys. Lett.*, vol. 91, no. 16, 2007, Art. no. 161111.
- [53] F. Xie *et al.*, "Long term reliability study and life time model of quantum cascade lasers," *Appl. Phys. Lett.*, vol. 109, no. 12, 2016, Art. no. 121111.
- [54] M. P. Semtsiv and W. T. Masselink, "Above room temperature continuous wave operation of a broad-area quantum-cascade laser," *Appl. Phys. Lett.*, vol. 109, no. 20, 2016, Art. no. 203502.
- [55] A. Aleksandrova, Y. V. Flores, S. S. Kurlov, M. P. Semtsiv, and W. T. Masselink, "Impact of cascade number on the thermal properties of broad-area quantum cascade lasers," *Phys. Status Solidi A*, vol. 215, no. 8, 2017, Art. no. 1700441.
- [56] A. Wittmann, Y. Bonetti, M. Fischer, J. Faist, S. Blaser, and E. Gini, "Distributed-feedback quantum-cascade lasers at 9 μm operating in continuous wave up to 423K," *IEEE Photon. Tech. Lett.*, vol. 21, no. 12, pp. 814–886, 2009.
- [57] T. Kato, H. Mori, H. Yoshinaga, and S. Souma, "High-temperature operation of a quantum cascade laser with definite parity of the wave functions," *IEEE Photon. Technol. Lett.*, vol. 33, no. 10, pp. 507–510, 2021.
- [58] M. Suttinger, R. Go, P. Figueiredo, A. Todi, H. Shu, J. Leshin, and A. Lyakh, "Power scaling and experimentally fitted model for broad area quantum cascade lasers in continuous wave operation," *Opt. Eng.*, vol. 57, no. 1, 2018, Art. no. 011011.
- [59] F. Wang, S. Slivken, D. H. Wu, Q. Y. Lu, and M. Razeghi, "Continuous wave quantum cascade laser with 5.6 W output power at room temperature and 41% wall-plug efficiency in cryogenic operation," *AIP Adv.*, vol. 10, no. 5, 2020, Art. no. 055120.
- [60] A. Lyakh *et al.*, "3 W continuous wave room temperature single-facet emission from quantum cascade lasers based on nonresonant extraction design approach," *Appl. Phys. Lett.*, vol. 95, no. 14, 2009, Art. no. 141113.
- [61] R. Maulini *et al.*, "High power thermoelectrically cooled and uncooled quantum cascade lasers with optimized reflectivity facet coatings," *Appl. Phys. Lett.*, vol. 95, no. 15, 2009, Art. no. 151112.
- [62] B. Knipfer *et al.*, "Failure analysis of high-power (One-Watt) room-temperature continuous wave MOCVD quantum cascade lasers," in *Proc. IEEE Int. Semicond. Laser Conf.*, 2018, pp. 1–2, doi: 10.1109/ISLC.2018.8516151.
- [63] N. Becher *et al.*, "Thermal imaging of buried heterostructure quantum cascade lasers (QCLs) and QCL arrays using CCD-based thermoreflectance microscopy," *J. Appl. Phys.*, vol. 125, 2019, Art. no. 033102.
- [64] X. Liu, R. W. Davis, L. C. Hughes, M. H. Rasmussen, R. Bhat, and C.-E. Zah, "A study on the reliability of indium solder die bonding of high power semiconductor lasers," *J. Appl. Phys.*, vol. 100, 2006, Art. no. 013104.
- [65] Y. Sin *et al.*, "Destructive physical analysis of degraded quantum cascade lasers," in *Proc. SPIE 9382, Novel -Plane Semicond. Lasers XIV*, Mar. 2015, Art. no. 93821P.
- [66] D. Pierścińska, K. Pierściński, M. Pluska, G. Sobczak, P. G. A. Kuźmicz, and M. Bugajski, "Temperature induced degradation mechanisms of $\text{AlInAs}/\text{InGaAs}/\text{InP}$ quantum cascade lasers," *Mater. Res. Exp.*, vol. 5, 2018, Art. no. 016204.
- [67] Y. Sin *et al.*, "Catastrophic degradation in high-power buried heterostructure quantum cascade lasers," in *Proc. Conf. Lasers Electro-Opt.*, 2019, Art. no. SW3N.3.
- [68] D. Hathaway *et al.*, "Output facet heating mechanism for uncoated high power long wave infrared quantum cascade lasers," *AIP Adv.*, vol. 10, 2020, Art. no. 085104.
- [69] T. Hu *et al.*, "Silicon photonic platforms for mid-infrared applications," *Photon. Res.*, vol. 5, no. 5, pp. 417–430, 2017.
- [70] B. Schwarz *et al.*, "Monolithically integrated mid-infrared lab-on-a-chip using plasmonics and quantum cascade structures," *Nat. Commun.*, vol. 5, 2014, Art. no. 4085.
- [71] Y. Han and K. M. Lau, "III–V lasers selectively grown on (001) silicon," *J. Appl. Phys.*, vol. 128, 2020, Art. no. 200901.
- [72] J. C. Norman *et al.*, "A review of high-performance quantum dot lasers on silicon," *IEEE J. Quantum Electron.*, vol. 55, no. 2, Apr. 2019, Art. no. 2000511.
- [73] R. Go *et al.*, "InP-based quantum cascade lasers monolithically integrated onto silicon," *Opt. Exp.*, vol. 26, pp. 22389–22393, 2018.
- [74] R. Go *et al.*, "Room temperature operation of quantum cascade lasers monolithically integrated onto a lattice-mismatched substrate," *Appl. Phys. Lett.*, vol. 112, no. 031103, 2018, pp. 2–6.
- [75] L. J. Mawst *et al.*, "InGaAs/AlInAs strain-compensated superlattices grown on metamorphic buffer layers for low-strain, 3.6 μm -Emitting quantum-cascade-laser active regions," *J. Cryst. Growth*, vol. 370, pp. 230–235, 2013.
- [76] A. Rajeev *et al.*, "Regrowth of quantum cascade laser active regions on metamorphic buffer layers," *J. Cryst. Growth*, vol. 452, pp. 268–271, 2016.
- [77] E. A. Fitzgerald *et al.*, "The science and applications of relaxed semiconductor alloys on conventional substrates," in *Proc. 3rd World Conf. Photovoltaic Energy Convers.*, May 2003, pp. 587–592.
- [78] T. F. Kuech *et al.*, "Metamorphic and non-conventional 'Buffer' layers," in *Proc. Compound Semicond. Week 23rd Int. Conf. Indium Phosphide Related Mater.*, May 2011, pp. 1–4.
- [79] B. Zutter *et al.*, "Surface crystalline defects in $\text{In}_{1-x}\text{Ga}_x\text{As}$ introduced by polishing methods," in *Proc. TMS Electron. Mater. Conf.*, Santa Barbara, CA, USA, Jun. 2014.
- [80] J. Z. Li *et al.*, "Defect reduction of GaAs epitaxy on Si (001) using selective aspect ratio trapping," *Appl. Phys. Lett.*, vol. 91, 2007, Art. no. 021114.
- [81] S. Jha *et al.*, "Defect reduction in epitaxial GaAs grown on nanopatterned GaAs substrates using full wafer block copolymer lithography," *Appl. Phys. Lett.*, vol. 95, no. 6, 2009, Art. no. 062104.
- [82] Y. Shi *et al.*, "Optical pumped InGaAs/GaAs nano-ridge laser epitaxially grown on a standard 300-mm Si wafer," *Optica*, vol. 4, pp. 1468–1473, 2017.
- [83] B. Shi, Q. Li and K. M. Lau, "Self-organized InAs/InAlGaAs quantum dots as dislocation filters for InP films on (001) Si," *J. Cryst. Growth*, vol. 464, no. September 2016, pp. 28–32, 2017.
- [84] A. Rajeev *et al.*, "III–V superlattices on InP/Si metamorphic buffer layers for $\lambda \approx 4.8 \mu\text{m}$ quantum cascade lasers," *Phys. Status Solidi A*, vol. 216, 2019, Art. no. 1800493.
- [85] Z. Loghmari *et al.*, "InAs-based quantum cascade lasers grown on on-axis (001) silicon substrate," *APL Photon.*, vol. 5, 2020, Art. no. 041302.
- [86] N. Bandyopadhyay, S. Slivken, Y. Bai, and M. Razeghi, "High power, continuous wave, room temperature operation of $\lambda \sim 3.4 \mu\text{m}$ and $\lambda \sim 3.55 \mu\text{m}$ InP-based quantum cascade lasers," *Appl. Phys. Lett.*, vol. 100, 2012, Art. no. 212104.
- [87] N. Bandyopadhyay, Y. Bai, S. Tsao, S. Nida, S. Slivken, and M. Razeghi, "Room temperature continuous wave operation of $\lambda \sim 3\text{--}3.2 \mu\text{m}$ quantum cascade lasers," *Appl. Phys. Lett.*, vol. 101, 2012, Art. no. 241110.
- [88] M. P. Semtsiv, M. Wienold, S. Dressler, and W. T. Masselink, "Short-wavelength ($\lambda \approx 3.05 \mu\text{m}$) InP based strain-compensated quantum-cascade laser," *Appl. Phys. Lett.*, vol. 90, 2007, Art. no. 90051111.
- [89] A. Aldukhayel *et al.*, "Investigations of carrier scattering into L-valley in $\lambda = 3.5 \mu\text{m}$ InGaAs/AlAs (Sb) quantum cascade lasers using high hydrostatic pressure," *Phys. Status Solidi B*, vol. 250, pp. 693–697, 2013.
- [90] A. Bismuto, M. Beck, and J. Faist, "High power Sb-free quantum cascade laser emitting at 3.3 μm above 350 K," *Appl. Phys. Lett.*, vol. 98, 2011, Art. no. 191104.
- [91] L. J. Mawst *et al.*, "Low-strain, quantum-cascade-laser active regions grown on metamorphic buffer layers for emission in the 3.0–4.0 μm wavelength region," *IEEE Photon. Technol. Lett.*, vol. 8, no. 2, pp. 25–32, 2014.
- [92] L. J. Mawst *et al.*, "Quantum-cascade-laser active regions on metamorphic buffer layers," *Proc. SPIE–Int. Soc. for Opt. Eng.*, vol. 9370, 2015, Art. no. 93700E.
- [93] A. Rajeev, C. Sigler, T. Earles, Y. V. Flores, L. J. Mawst, and D. Botez, "Design considerations for $\lambda \sim 3.0\text{--}3.5 \mu\text{m}$ -emitting quantum cascade lasers on metamorphic buffer layers," *Opt. Eng.*, vol. 57, no. 1, 2017, Art. no. 011017.
- [94] K. Schulte *et al.*, "Metalorganic vapor phase growth of quantum well structures on thick metamorphic buffer layers grown by hydride vapor phase epitaxy," *J. Cryst. Growth*, vol. 370, pp. 293–298, 2013.
- [95] K. L. Schulte *et al.*, "Compositionally graded $\text{Ga}_{1-x}\text{In}_x\text{P}$ buffers grown by static and dynamic hydride vapor phase epitaxy at rates up to 1 $\mu\text{m}/\text{min}$," *Appl. Phys. Lett.*, vol. 118, 2021, Art. no. 052106.

- [96] A. Rajeev, "Mid-Infrared-Emitting quantum cascade lasers on metamorphic buffer layers," Order No. 13815085 The University of Wisconsin - Madison, 2019. Ann Arbor: *ProQuest*. Web. Dept. Elect. Comput. Eng., 22 Jul. 2021.
- [97] THORLABS Quantum Cascade Lasers product data sheet, [Online]. Available: https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=6826
- [98] N. Yu *et al.*, "Coherent coupling of multiple transverse modes in quantum cascade lasers," *Phys. Rev. Lett.*, vol. 102, no. 1, 2009, Art. no. 013901.
- [99] W. W. Bewley *et al.*, "Beam steering in high-power CW quantum-cascade lasers," *IEEE J. Quantum Electron.*, vol. 41, no. 6, pp. 833–841, Jun. 2005.
- [100] W. Metaferia *et al.*, "Demonstration of a quick process to achieve buried heterostructure quantum cascade laser leading to high power and wall plug efficiency," *Opt. Eng.*, vol. 53, no. 8, 2014, Art. no. 087104.
- [101] L. Cheng *et al.*, "Analysis of InP regrowth on deep-etched mesas and structural characterization for buried-heterostructure quantum cascade lasers," *J. Electron. Mater.*, vol. 41, no. 3, pp. 506–513, 2012.
- [102] C. A. Sigler "2-D coherent power scaling of mid-infrared quantum cascade lasers." Order No. 10933067 The University of Wisconsin - Madison, 2018. Ann Arbor: *ProQuest*. Web. Dept. Elect. Comput. Eng., 14 Aug. 2021.
- [103] D. Botez, "Monolithic phase-locked semiconductor laser arrays," in *Diode Laser Arrays*, D. Botez and D. R. Scifres, Eds., Cambridge, U.K.: Cambridge Univ. Press, 1994, pp. 1–71.
- [104] R. Kaspi, S. Luong, C. Yang, C. Lu, T. C. Newell, and T. Bate, "Extracting fundamental transverse mode operation in broad area quantum cascade lasers," *Appl. Phys. Lett.*, vol. 109, no. 21, 2016, Art. no. 211102.
- [105] R. Kaspi, S. Luong, T. Bate, C. Lu, T. C. Newell, and C. Yang, "Distributed loss method to suppress high order modes in broad area quantum cascade lasers," *Appl. Phys. Lett.*, vol. 111, no. 20, 2017, Art. no. 201109.
- [106] R. Blanchard *et al.*, "High-power low-divergence tapered quantum cascade lasers with plasmonic collimators," *Appl. Phys. Lett.*, vol. 102, no. 19, 2013, Art. no. 191114.
- [107] A. Lyakh, R. Maulini, A. Tsekoun, R. Go, and C. K. N. K. N. Patel, "Tapered 4.7 μm quantum cascade lasers with highly strained active region composition delivering over 4.5 watts of continuous wave optical power," *Opt. Exp.*, vol. 20, no. 4, 2012, Art. no. 4382.
- [108] P. Rauter *et al.*, "Master-oscillator power-amplifier quantum cascade laser array," *Appl. Phys. Lett.*, vol. 101, no. 26, 2012, Art. no. 261117.
- [109] S. Menzel *et al.*, "Quantum cascade laser master-oscillator power-amplifier with 1.5 w output power at 300 K," *Opt. Exp.*, vol. 19, no. 17, pp. 16229–16235, 2011.
- [110] S. Ahn *et al.*, "High-power, low-lateral divergence broad area quantum cascade lasers with a tilted front facet," *Appl. Phys. Lett.*, vol. 104, no. 5, Feb. 2014, Art. no. 051101.
- [111] D. Heydari, Y. Bai, N. Bandyopadhyay, S. Slivken, and M. Razeghi, "High brightness angled cavity quantum cascade lasers," *Appl. Phys. Lett.*, vol. 106, no. 9, 2015, Art. no. 091105.
- [112] Y. Bai, B. Gokden, S. R. Darvish, S. Slivken, and M. Razeghi, "Photonic crystal distributed feedback quantum cascade lasers with 12 W output power," *Appl. Phys. Lett.*, vol. 95, no. 3, 2009, Art. no. 031105.
- [113] B. Gökden, Y. Bai, N. Bandyopadhyay, S. Slivken, and M. Razeghi, "Broad area photonic crystal distributed feedback quantum cascade lasers emitting 34 W at $\lambda \sim 4.36\ \mu\text{m}$," *Appl. Phys. Lett.*, vol. 97, no. 13, 2010, Art. no. 131112.
- [114] G. M. de Naurois, M. Carras, B. Simozrag, O. Patard, F. Alexandre, and X. Marcadet, "Coherent quantum cascade laser micro-stripe arrays," *AIP Adv.*, vol. 1, no. 3, Sep. 2011, Art. no. 032165.
- [115] F.-L. Yan *et al.*, "High-power phase-locked quantum cascade laser array emitting at $\lambda \sim 4.6\ \mu\text{m}$," *AIP Adv.*, vol. 6, no. 3, Mar. 2016, Art. no. 035022.
- [116] L. Wang *et al.*, "Phase-locked array of quantum cascade lasers with an integrated talbot cavity," *Opt. Exp.*, vol. 24, no. 26, pp. 30275–30281, Dec. 2016.
- [117] B. Meng, B. Qiang, E. Rodriguez, X. N. Hu, G. Liang, and Q. J. Wang, "Coherent emission from integrated Talbot-cavity quantum cascade lasers," *Opt. Exp.*, vol. 25, no. 4, pp. 3077–3082, 2017.
- [118] Z. Jia *et al.*, "Phase-locked array of quantum cascade lasers with an intracavity spatial filter," *Appl. Phys. Lett.*, vol. 111, no. 6, 2017, Art. no. 061108.
- [119] D. Botez, "Comment on 'Phase-locked array of quantum cascade lasers with an intracavity spatial filter' [Appl. Phys. Lett. 111, 061108 (2017)]," *Appl. Phys. Lett.*, vol. 111, no. 25, 2017, Art. no. 256101.
- [120] J. D. Kirch *et al.*, "5.5W near-diffraction-limited power from resonant leaky-wave coupled phase-locked arrays of quantum cascade lasers," *Appl. Phys. Lett.*, vol. 106, no. 6, Feb. 2015, Art. no. 061113.
- [121] C. Sigler, C.-C. Chang, J. D. Kirch, L. J. Mawst, D. Botez, and T. Earles, "Design of resonant leaky-wave coupled phase-locked arrays of Mid-IR quantum cascade lasers," *IEEE J. Sel. Top. Quantum Electron.*, vol. 21, no. 6, Nov. 2015, Art. no. 1200810.
- [122] C. Sigler *et al.*, "5.3 μm -emitting diffraction-limited leaky-wave-coupled quantum cascade laser phase-locked array," *IEEE J. Sel. Topics Quantum Electron.*, vol. 25, no. 6, Nov./Dec. 2019, Art. no. 1200509.
- [123] L. K. Hoffmann *et al.*, "Tree array quantum cascade laser," *Opt. Exp.*, vol. 17, no. 2, pp. 649–657, 2009.
- [124] A. Lyakh, R. Maulini, A. Tsekoun, R. Go, and C. K. N. K. N. Patel, "Continuous wave operation of buried heterostructure 4.6 μm quantum cascade laser Y-junctions and tree arrays," *Opt. Exp.*, vol. 22, no. 1, pp. 1203–1208, Jan. 2014.
- [125] W. Zhou, S. Slivken, and M. Razeghi, "Phase-locked, high power, mid-infrared quantum cascade laser arrays," *Appl. Phys. Lett.*, vol. 112, no. 18, 2018, Art. no. 181106.
- [126] W. Zhou, Q.-Y. Lu, D.-H. Wu, S. Slivken, and M. Razeghi, "High-power, continuous-wave, phase-locked quantum cascade laser arrays emitting at 8 μm ," *Opt. Exp.*, vol. 27, pp. 15776–15785, 2019.
- [127] R. Peretti *et al.*, "Room temperature operation of a deep etched buried heterostructure photonic crystal quantum cascade laser," *Laser Photon. Rev.*, vol. 10, no. 5, pp. 843–848, 2016.
- [128] D. Botez, L. J. Mawst, G. L. Peterson, and T. J. Roth, "Phase-locked arrays of antiguides: Modal content and discrimination," *IEEE J. Quantum Electron.*, vol. 26, no. 3, pp. 482–495, Mar. 1990.
- [129] D. Botez, L. J. Mawst, and G. Peterson, "Resonant leaky-wave coupling in linear arrays of antiguides," *Electron. Lett.*, vol. 24, no. 21, pp. 1328–1330, 1988.
- [130] H. Yang, L. J. Mawst, M. Nesnidar, J. Lopez, A. Bhattacharya, and D. Botez, "10 W near- diffraction-limited pulsed power from 0.98 μm -emitting, Al-free phase-locked antiguided arrays," *Electron Lett.*, vol. 33, no. 22, pp. 136–138, 1997.
- [131] H. Yang, L. J. Mawst, and D. Botez, "1.6 W continuous-wave coherent power from large-index-step ($\Delta n \cong 0.1$) near-resonant, antiguided diode laser arrays," *Appl. Phys. Lett.*, vol. 76, no. 10, pp. 1219–1221, 2000.
- [132] C. A. Zmudzinski, D. Botez, and L. J. Mawst, "Simple description of laterally resonant, distributed-feedback-like modes of arrays of antiguides," *Appl. Phys. Lett.*, vol. 60, no. 9, pp. 1049–1051, 1992.
- [133] S. Li and D. Botez, "Analysis of 2-D surface-emitting ROW-DFB semiconductor lasers for high- power, single-mode operation," *IEEE J. Quantum Electron.*, vol. 43, no. 8, pp. 655–668, 2007.
- [134] C. Sigler, J. D. Kirch, T. Earles, L. J. Mawst, Z. Yu, and D. Botez, "Design for high-power, single-lobe, grating-surface-emitting quantum cascade lasers enabled by plasmon-enhanced absorption of antisymmetric modes," *Appl. Phys. Lett.*, vol. 104, no. 13, 2014, Art. no. 131108.