The Interband Cascade Laser

Jerry R. Meyer 1,*, William W. Bewley 1, Chadwick L. Canedy 1, Chul Soo Kim 1, Mijin Kim 2, Charles D. Merritt 1 and Igor Vurgaftman 1

1 Code, 5613, Naval Research Laboratory, Washington, DC 20375, USA
2 Jacobs Corporation, Hanover, MD 21076, USA
* Correspondence: jerry.meyer@nrl.navy.mil; Tel.: +1-202-767-3276

Received: 23 August 2020; Accepted: 10 September 2020; Published: 15 September 2020

Abstract: We review the history, development, design principles, experimental operating characteristics, and specialized architectures of interband cascade lasers for the mid-wave infrared spectral region. We discuss the present understanding of the mechanisms limiting the ICL performance and provide a perspective on the potential for future improvements. Such device properties as the threshold current and power densities, continuous-wave output power, and wall-plug efficiency are compared with those of the quantum cascade laser. Newer device classes such as ICL frequency combs, interband cascade vertical-cavity surface-emitting lasers, interband cascade LEDs, interband cascade detectors, and integrated ICLs are reviewed for the first time.

Keywords: interband cascade laser; semiconductor laser; laser design; interband cascade detector; single-mode laser; vertical-cavity surface-emitting laser; interband cascade LED; optical frequency comb; photonic integrated circuit

1. Introduction

The type-II interband cascade laser (ICL) has recently gained acceptance as an important coherent optical source for the mid-wave infrared spectral band (mid-IR, defined here as spanning 3–6 µm). It combines a relatively long upper-level lifetime, characteristic of semiconductor interband transitions, with the voltage-efficient cascading scheme originally introduced for the quantum cascade laser (QCL, which employs intersubband transitions to produce light). Both electrons and holes are present in each stage of the ICL’s cascaded active region, even though the contacts inject and remove only electrons. The wealth of complex physics responsible for various aspects of the ICL operation have yet to be fully unraveled.

Given the volume of recent activity in the mid-IR spectral region, its technological importance requires little introduction. The most pervasive application involves the sensing of trace gases such as methane, carbon dioxide, carbon monoxide, formaldehyde, etc., in ambient air [1]. While this typically requires cw emission into a single spectral mode, low output powers on the order of 1 mW are generally sufficient. The leading military interest is for infrared countermeasure sources to jam heat-seeking missiles. Although high spectral purity is typically not needed for this application, substantial cw output powers are essential. Other potential applications of mid-IR sources include industrial process control [2], combustion diagnostics [3], clinical breath analysis [4], isotope differentiation [4], free space optical communications [5], IR scene projection [6], and the detection of explosives [7].
In spite of the mid-IR spectral band’s technological prominence, until 2002, no practical coherent semiconductor source offered continuous-wave (cw) operation at ambient temperature [8]. Until then, it seemed the most straightforward route to achieving that goal was to push the well-established multiple-quantum-well (MQW) diode laser, with type-I alignment of the conduction and valence bands, to longer wavelengths by employing more strain in the layers containing Sb [9,10]. However, challenges included the rapid wavelength scaling of Auger non-radiative decay [11] and carrier escape associated with a marginal valence-band offset (VBO) in the MQWs, not to mention the general immaturity of GaSb-based growth and processing technologies.

The invention of the QCL [12,13] loomed large against this backdrop, although in its first decade a high threshold power density precluded the room temperature (RT) cw operation that is required for nearly all practical applications. Besides its high threshold current density associated with a short carrier lifetime in the upper lasing subband (≈1 ps), the QCL requires a bias of at least ≈10 V because 30–40 stages are needed to supply sufficient gain. Nonetheless, a modest temperature sensitivity coupled with efficient heat dissipation eventually allowed the QCL to achieve not only RT cw operation, but also cw output power above 5 W into a near-diffraction-limited beam [14,15]. We will discuss similarities between the cascading interband and intersubband transitions in ICLs and QCLs, and also some important distinctions. We will focus on understanding the very different temperature performances of the two device classes. Even though the ICL was invented only shortly after the QCL, far fewer resources have been devoted to its development over of the intervening period. This may be attributed in part to the general paucity of expertise with GaSb-based materials.

In parallel, several research groups have continued to extend the emission wavelengths of more conventional diode lasers. GaSb-based type-I MQW diodes increasingly display excellent performance over the neighboring 2–3 µm spectral window [16], and RT cw operation has been reported for wavelengths as long as 3.44 µm [17,18]. Since the fundamental physical limitations of the type-I approach are not fully delineated, this work will offer only general remarks on the comparative characteristics of type-I vs. type-II-IR interband mid-IR lasers.

Although ICLs have been reviewed previously [19,20], our intention here is to provide a more in-depth and up-to-date description of both the physics of operation and the prospects for future improvement. We begin by historically overviewing the concept and its evolution. The ICL design principles are then discussed in some detail, with some ingredients identified that may be missing from the current picture. The experimental results for broad-area (pulsed) and narrow-ridge (cw) ICLs are also presented, along with newer topics that were not treated in the previous reviews, such as vertical-cavity ICLs, ICL frequency combs, and ICLs incorporated into photonic integrated circuits. Following a discussion of the mechanisms now limiting ICL performance at different mid-IR wavelengths, we offer some thoughts on the prospects for future improvements.

2. Evolution of the Interband Cascade Laser Concept

2.1. Initial Proposal

The first suggestion that interband transitions may be cascaded dates to a paper by Rui Yang, then at the University of Toronto, written (and reported at a conference) in 1994 and published the following year [21]. While most of the paper was devoted to a proposal that interband tunneling may be used to remove carriers from the lower lasing subband of a QCL-like device, the final figure in that work contained the first sketch of an interband cascade active core. In this first version, the composition and layer sequence of the electron injector remained unspecified, and there was no suggestion that a hole injector was needed. However, the first recognition that the QCL’s very short optical phonon relaxation channel could be eliminated by substituting an interband active core was notable.
This proposal focused extensively on the use of interband transitions to efficiently extract electrons from the lower lasing subband, a necessary element of any cascading architecture that incorporates interband active transitions. The initial use of the semimetallic (SM) overlap between neighboring layers of InAs and Ga(In)Sb to ensure rapid interband transfer dates back to the resonant interband tunneling diodes (RITDs) demonstrated by the McGill group at Caltech [22], although it also bears close kinship to some slightly earlier ideas discussed by Sweeny and Xu [23]. Prior to that work, interband transfer within a device was always accomplished by heavily doping both sides of the p-n junction in an Esaki diode.

2.2. Improvements

While the initial paper did not use the name “interband cascade laser”, it was introduced in short order by NRL researchers in collaboration with Rui Yang, who had by that time moved to the University of Houston [24]. This work introduced a single hole-injector QW into the design, noting that the active hole well by itself is unlikely to prevent excessive electron tunneling directly from the electron active well into the electron injector. Furthermore, the work extended the ICL concept to include the possibility of a type-I active region. It was pointed out that the number of wells in the electron and hole injectors could vary, as long as the total thickness was sufficient for a reasonable external field (say, 100 kV/cm) to drop a voltage of $\hbar \omega / q$, and efficient injector transport was assured.

At that time, it was unclear how strong Auger recombination would be for a type-II mid-IR QW structure, so the report limited itself to calculating the radiative current density only. A follow-up calculation by the NRL group [25] estimated a RT threshold current density ($J_{th}$), strongly dominated by the Auger process, of ~1 kA/cm². In fact, because the Auger coefficient turns out to be much lower than the value assumed in that paper, lasing thresholds in the 100–200 A/cm² range are now routinely attainable in the laboratory. Even so, that work [25] predicted that the device should be capable of RT operation. The specific design employed as many as 15 InAs QWs in the electron injector, which was too thick because the internal field was not properly taken into account. The proposed configuration had a single active InAs QW, and did not include a barrier between the electron and hole injectors.

A few months later, Vurgaftman et al. designed and analyzed a vertical-cavity surface-emitting laser with an interband cascade active region [26]. That design incorporated the so-called “W” structure with two active InAs electron wells on both sides of the GaInSb active hole well [27]. The motivation is to increase the optical matrix element and optical gain due to a stronger electron-hole wavefunction overlap. Nevertheless, since 2004 it has been a standard feature in nearly all ICL designs by the different research groups. In practice, the “W” ICL has been observed to outperform structures with a single active electron well, even after the modal gain per unit current density became much higher following other design improvements.

2.3. Early Experimental Realizations

The first successful experimental realization of an ICL was reported in early 1997 [28]. For short pulses, the device with 20 active stages operated at $T = 80–120$ K with a very high threshold current density of several kA/cm². The performance soon improved considerably, although only pulsed lasing was observed, and only at temperatures below 225 K [29,30]. It is clear in retrospect that the electron injectors in these initial ICL designs (e.g., Ref. [25]) were much too thick. Despite having 20 or more active stages, the external differential quantum efficiencies (EDQEs) per stage were low, e.g., only 6–7.5% at $T = 100$ K in Refs. [30] and [31]. These two working devices were the first to feature the “W” ICL configuration. Note also that by defining the EDQE as the efficiency per stage, here and below, we can directly compare the values measured for devices with a wide range of stage multiplicities ($M$). A slightly higher EDQE of nearly 10% at $T = 80$ K was soon reported [32].
Whereas the hole injectors in early ICLs had a single hole QW, a joint NRL-University of Houston publication [33] introduced the additional design improvement of incorporating a second QW, with the goal of further minimizing electron leakage straight from the active region into the electron injector. The number of QWs in the electron injector was also reduced, to eight, although the overall thickness was still high at ≈ 400 Å. The ICL with this design was the first to operate nearly to room temperature (\( T = 286 \) K) in pulsed mode. While most of the improvement probably resulted from the thinner electron injector rather than the additional well in the hole injector, the two-QW hole injector configuration subsequently became a standard feature of ICL designs.

Whereas the first structures were all grown at the University of Houston, in 1999 the Army Research Laboratory (ARL) began to grow ICLs as well. They soon improved the low-temperature EDQE per stage to ≈ 20% (> 4.5 photons emitted from all the stages for every electron injected into the device) at \( T = 150 \) K [34], although high-temperature operation was still elusive because the threshold power density remained high. Like the other early ICLs, these structures had 20–25 active stages that required a high threshold voltage approaching 10 V. Pulsed operation was observed at temperatures up to 217 K [35,36], and later 250 K [37]. Cw lasing was also achieved at temperatures up to 142 K [38,39].

The next significant milestone was the demonstration of pulsed operation at RT [40,41] by the ARL group, which by this time had spun off the company Maxion Technologies. Even though the RT threshold current density approached 7 kA/cm², this was tolerated by using wide ridges to manage current spreading. The devices still had 18 active stages, and the efficiency was quite low at higher temperatures. Subsequently, similar ICLs were operated up to \( T = 214 \) K in cw mode. A thick layer of electroplated Au provided heat sinking to the narrow ridge, which was etched through the active core [42]. Since the ARL/Maxion articles from this period supplied few details of the device structures, it is difficult to reconstruct the design factors that improved or limited their performance (in the context of what is now known about ICL operation).

Rui Yang’s move to the Jet Propulsion Laboratory (JPL) in 2002 marked the beginning of that group’s activity in the design and fabrication of ICLs. Using active regions based on the “W” configuration, they fairly soon reached RT pulsed operation and cw lasing to \( T = 200 \) K [43,44]. Their reduction of the RT \( J_{th} \) to \( \approx 1 \) kA/cm², in an ICL with 15 active stages and emitting at \( \lambda = 3.3 \) μm, represented a substantial improvement over previous thresholds. Distributed-feedback (DFB) ICLs were also demonstrated [45]. The performance became considerably worse at somewhat longer mid-IR wavelengths (\( \lambda = 4.3–5.6 \) μm), where the maximum pulsed operating temperatures (\( T_{max} \)) were well below RT [46,47]. An ICL structure grown on a GaAs substrate suffered only a moderate performance penalty resulting from the large lattice mismatch [48].

2.4. Toward Room-Temperature cw Operation

In 2005, the JPL group demonstrated a 150 μm wide × 1.5 mm long ridge whose pulsed RT \( J_{th} \) was only 630 A/cm² [49]. This substantial reduction of the threshold allowed a 15 μm × 1.5 mm ridge to lase in cw mode up to \( T_{max} = 237 \) K at \( \lambda = 3.3 \) μm. These devices had a much thicker bottom optical cladding layer to prevent mode leakage into the high-index GaSb substrate, and featured only 12 stages even though the threshold voltage \( V_{th} \) still exceeded 6 V at RT. The addition of a thick layer of electroplated Au soon led to cw operation up to \( T_{max} = 264 \) K [50], as well as single-mode emission from a DFB ICL at a similar temperature [51]. These higher operating temperatures, in the thermoelectric cooler range, enabled JPL to qualify DFB ICLs emitting at \( \lambda = 3.27 \) μm for methane detection on the NASA Mars Curiosity Mission. Since 2013, the ICL-based spectrometer on Curiosity has confirmed the presence of methane on Mars, with rare bursts up to 7 parts per billion by volume [52].

The first ICLs designed, grown, processed, and characterized at NRL were reported in 2006 [53]. At that time, “W” diode lasers with similar active QWs were operating in pulsed mode at RT [54], albeit with threshold current densities well in excess of 10 kA/cm². The early NRL ICL designs employed “W” active regions, single-QW hole injectors, and relatively thick electron injectors with low doping in four of the wells. In fact, they were conceptually similar to the structures grown seven years
earlier at the University of Houston [33], except that only 10 active stages were employed and the hole injector comprised a single QW. The initial RT $J_{th}$ of 7 kA/cm$^2$ was quite high, being only a little lower than for the “W” diode lasers. NRL reported a cryogenic DFB ICL based on this structure [55].

In the next round of NRL designs, the number of stages was reduced to $M = 5$ [56] in order to lower the threshold power density. Simulations had determined that the gain per unit current density is actually quite high in an ICL, e.g., much higher than in a QCL that typically requires 30 or more stages to realize low $J_{th}$ at RT. Although the pulsed RT $J_{th}$ at the emission wavelength of $\lambda$ = 3.7 µm still exceeded 2 kA/cm$^2$, a 12 µm wide × 3.9 mm long Au-electroplated ridge fabricated from this material operated cw to $T_{\text{max}} = 257$ K, a little lower than the best JPL result at the time [57].

Subsequently, NRL reported a series of advances in such key performance figures of merit as $J_{th}$, $P_{th}$, and EDQE, at and above RT. While the design modifications responsible for these improvements did not appear in the journal publications from the period, they were later disclosed in a series of patent applications. One of the first changes was to employ a much thinner (200–250 Å) electron injector, which allowed operation at a higher external electric field. In conjunction with the double-QW hole injector originally attempted in 1998 [33], this enhanced the electron occupation of the active region while reducing somewhat the electron density in the electron injector [58]. The result was a lower RT threshold current density of 1.15 kA/cm$^2$ for a ten-stage structure. A narrow, Au-electroplated ridge fabricated from that wafer operated cw to $T_{\text{max}} = 269$ K [59]. A cavity-length study found that the internal loss increased rapidly with temperature, e.g., to $\alpha_i = 28$ cm$^{-1}$ at $T = 275$ K [60]. A gain per unit current density of 4 cm/kA/stage was determined at the same temperature.

In 2008, an NRL ICL emitting at $\lambda = 3.75$ µm achieved RT cw operation [61]. Several additional design improvements contributed to this milestone: (1) the number of stages was reduced to five, to further lower the threshold power density; (2) two $n$-doped GaSb separate-confinement layers (SCLs) were introduced on both sides of the active core; (3) the doping of the cladding regions immediately adjacent to the SCLs was reduced, to $1.5 \times 10^{17}$ cm$^{-3}$; (4) the AlSb barrier between the electron and hole injectors was thickened, to 20 Å, along with several other minor modifications. As a result, the pulsed $J_{th}$ at RT dropped to $\approx 400$ A/cm$^2$, the lowest value reported for an ICL up to that point. While some of the improvements were anticipated based on general design principles for semiconductor lasers, it was not obvious that an optimized 5-stage ICL could maintain approximately the same $J_{th}$ as a 10-stage device.

3. Design of Interband Cascade Lasers

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

3.1. Design and Doping of the Active Core

Since any mid-IR semiconductor laser is likely to need multiple QWs if it is to provide sufficient gain for cw operation well above the threshold at RT, we first consider how to best connect the QWs. For many years, electrically-pumped mid-IR lasers by default copied the standard MQW diode geometry, in which electrons and holes are injected from opposite sides of a $p$-$n$ heterojunction into QWs positioned adjacent to each other and separated by barrier layers.

One disadvantage of this arrangement is that if the electrons and holes starting at opposite ends of the line of QWs fail to populate each well uniformly, the gain generated by a given injected current density degrades. Furthermore, $J$ scales with the number of QWs $M$ because the injection must be sufficient to simultaneously invert the populations in all of them. The high current density characteristic of the conventional “parallel” QW configuration induces a large parasitic voltage drop $MJ\delta_i$, where $\delta_i$ is the series resistance-area product. At mid-IR wavelengths, the parasitic drop tends to be significant compared to the “useful” voltage of $\hbar \omega / q$, where $\omega$ is the photon frequency.
Cascading provides a practical alternative means for connecting the MQWs, namely in series with the same current flowing through every stage, and each carrier traversing every QW in turn [62]. The parasitic voltage drop then decreases to \( \rho_s \), as compared to the “useful” voltage of \( Mh\omega/q \). This may considerably reduce the threshold power density, provided the quasi-elastic recycling of carriers from valence to conduction band within each stage does not require much additional voltage or current [63, 64]. While the well-known Esaki or tunnel diode [65–67] may be used to implement carrier recycling, most ICLs alternatively take advantage of the unusual semimetallic band alignment between the narrow-gap semiconductors InAs and Ga(In)Sb [68, 69]. This approach avoids the appreciable mid-IR optical losses, especially due to free hole absorption, that occurs if every stage incorporates a heavily-doped tunnel junction.

Figure 1 illustrates the possibility of tuning the semimetallic overlap with an applied electric field. Since at RT the conduction band of InAs lies \( \approx 0.2 \text{ eV} \) below the valence of GaSb, a modest amount of quantum confinement is sufficient to open a small energy gap \( E_g \), with electrons and holes localized on opposite sides of the semimetallic interface as in Figure 1a. If we now apply an external field \( F \), the bottom conduction subband is lowered with respect to the top valence subband, as shown in Figure 1b. This produces a voltage-dependent band overlap of \( E_{SM}(V) = F(d_e + d_h) - E_g \), where \( d_e \) and \( d_h \) are the center-of-mass distances of the electron and hole probability densities from the semimetallic interface. At typical ICL operating voltages, this overlap is \( \approx 100 \text{ meV} \). Assuming parabolic bands and a common quasi-Fermi level across the semimetallic interface, the generated electron and hole densities are: \( n = p \approx m_e E_{SM}/\pi \hbar^2 \), where \( m_e = m_e m_h/(m_e + m_h) \) is the reduced effective mass.

![Figure 1](image.png)

**Figure 1.** Schematic of the energy alignment for adjacent InAs and GaSb QWs: (a) in equilibrium and (b) under bias, where the solid blue and red lines are the conduction and valence-band edges, respectively, in the bulk materials. In the absence of bias, quantum confinement leads to an energy gap \( E_g \) between the lowest conduction and highest valence-band states (blue and red dotted lines, respectively), whereas under bias, a semimetallic overlap \( E_{SM} \) is imposed that leads to the generation of equal electron and hole densities. The applied field also causes both carrier types to flow away from the interface (arrows), requiring that they be replaced on an ongoing basis to maintain quasi-thermal equilibrium populations. Adapted from Ref. [70].

The active core of the ICL is comprised of multiple repeated stages, each of which can be subdivided into the following regions: (1) the active QWs; (2) the hole injector; and (3) the electron injector. The active QWs can have a type-II or type-I band alignment. A single active QW suffices in the latter case, while in the former the design may have either a single (InAs/GaInSb) or two type-II interfaces (InAs/GaInSb/InAs, the “W” structure illustrated in Figures 2 and 3 below). While multiple QWs may be employed in each stage, there is no obvious advantage since the threshold current density would scale with the multiplicity. This is strictly true for type-I active QWs, and while strong coupling between the electron QWs may complicate the scaling for type-II structures, the same principle generally holds because the hole QWs tend to fully decouple.
Figure 2. Band diagram of 1½ stages of the ICL active core. Probability densities and zone-center energies (indicated by the wavefunction zero points) for some of the most important subbands are superimposed. The probability densities for the active electron (hole) subbands are indicated with blue (red) lines, while those for the injector-electron (hole) subbands are indicated with wine-colored (green) lines. The blue (red) arrows indicate the position of the quasi-Fermi levels in each stage. The layer structure of one period can be found in Ref. [70], from which this figure is adapted.

Figure 3. Calculated quasi-equilibrium electron (wine-colored) and hole (green) density distributions for ICL designs without carrier rebalancing (dashed), and with the carriers rebalanced (solid) by doping four of the InAs electron injector QWs with Si to $5 \times 10^{18}$ cm$^{-3}$. The blue and red lines indicate the conduction- and valence-band bulk band edges, respectively. Adapted from Ref. [70].
Figure 2 shows the conduction and valence band profiles for 1 1/2 stages of a typical type-II ICL with “W” active region designed for emission at $\lambda = 3.7$ $\mu$m. The semimetallic interface is seen to separate the hole injector, comprised of coupled GaSb/AlSb QWs, from the electron injector consisting of coupled InAs/AlSb QWs. By analogy with Figure 1, it is the field-dependent band overlap between states in the hole injector and active hole QW on the one hand, and states in the electron injector and active electron QW in the next stage on the other, that determines the carrier densities throughout the active core. We assume a common quasi-Fermi level (QFL) on both sides of the semimetallic interface, since the carrier transport times via direct, phonon-assisted, and other tunneling mechanisms are quite rapid. The much longer carrier lifetime in the active QWs (~1 ns) then assures that the QFL is discontinuous across the active region of each stage.

In the ideal design, the applied bias must separate the QFLs in successive stages (equal to the single-stage voltage drop) enough to simultaneously: (1) produce sufficient optical gain to compensate the photon loss in the cavity ($E_{Fi} - E_{Fi+1} \geq \hbar \omega \geq E_g$, with $E_{Fi}$ the electron QFL in stage $i$ being equal to the hole QFL in stage $i+1$), and (2) internally generate a quasi-equilibrium carrier density consistent with the voltage drop (accounting for any extrinsic doping). While the first condition is a property of the active QWs only, the second follows from the design and doping of all the electron and hole QWs on both sides of the semimetallic interface. If, for example, the QWs in the electron injector are thicker than optimal, excess carriers are generated that induce unnecessary free-carrier absorption. Conversely, if the injector QWs are too thin, the threshold voltage will exceed that needed to induce the ideal QFL separation of $E_{Fi} - E_{Fi+1} \approx \hbar \omega$ (or somewhat larger if the loss is very high).

Since many of the states on both sides of the semimetallic interface reside in the injectors rather than the active QWs, only a fraction of the carriers generated internally at that interface, or introduced by extrinsic doping, contribute to the optical gain. Therefore, a central design goal is to maximize the fraction of injected carriers populating the active states, while retaining sufficient carrier transport through the injectors to maintain a single QFL throughout the stage.

To a large extent, the success in meeting this goal is governed by the electron and hole injector designs, although the many constraints are not yet fully understood. Nonetheless, the broad picture that has emerged from experimental results corresponding to numerous diverse ICL designs is that the energy subbands in the two hole injector QWs should lie substantially below the topmost active hole subband. At the same time, the states of the thicker electron injector should be located in the vicinity of the semimetallic interface and energetically lower than the active electron QW state. In fact, designs with very thin electron injectors were unsuccessful, whereas thus far more typical injectors with about six QWs (as in Figure 2) have been optimal for the 3–4 $\mu$m spectral range. While this may be due in part to the large threshold electric field that accompanies a thin injector, most likely there are other factors. Furthermore, the requirement for a single QFL appears to dictate that the electron injector states not lie substantially above the active electron subband.

The empirical constraints described above imply that most of the electrons populating a given stage reside in the injector states rather than the active region, whereas nearly all of the holes transfer successfully to the active hole QW. This is illustrated in Figure 3, which plots the electron (wine-colored) and hole (green) densities at the threshold for the ICL structure of Figure 2. The consequence is that if no extrinsic doping is introduced (dashed curves) to supplement the carriers generated at the semimetallic interface, the active hole density will substantially exceed the active electron density at the lasing threshold ($p_{th} >> n_{th}$) [70]. Unless $mp$ (multi-electron) Auger processes strongly dominate the carrier lifetime, the large hole/electron density ratio will tend to reduce the gain per unit current density, and also increase the internal loss because holes usually have a larger free carrier absorption-cross section. To investigate this potential issue, ICL structures were grown with a series of heavy $n$-doping levels in the electron injectors of devices emitting at $\lambda = 3.6$–3.9 $\mu$m [70]. While many of the additional electrons remained behind in the injector, NRL simulations predicted that a sufficient fraction would be transferred to the active QWs to alter the hole/electron density ratio there (solid curves in Figure 3).

The total sheet doping, $n_s$, was varied from $4.8 \times 10^{11}$ to $7.4 \times 10^{12}$ cm$^{-2}$, with 2–4 of the injector QWs
receiving the additional doping. This study found that $J_{th}$ at RT was minimized at $n_s \approx 5 \times 10^{12} \text{ cm}^{-2}$, which according to the simulations corresponded to $n_{th}/p_{th}$ only slightly larger than unity. Note that the required sheet doping density substantially exceeds $n_{th}$ and $p_{th}$, due to the inefficient electron transfer from the doped injector. Another conclusion, from the dependence on doping level, is that the $nnp$ and $ppn$ Auger coefficients most likely have comparable values.

We next examine other aspects of the hole and electron injector designs. For example, whereas empirical studies showed it advantageous to employ two QWs (rather than one) in the hole injector, nominally to prevent electron tunneling leakage from the active region to the electron injector, adding still more QWs is unlikely to prove beneficial. Since the thicknesses of the two hole QWs are adjusted so as to roughly align their topmost subbands at the threshold field, the QWs become strongly coupled with two closely-spaced subbands. The QW thicknesses are also chosen so as to place these subbands $\approx 80–100 \text{ meV}$ below the maximum in the active hole QW. This empirical rule should ensure very low occupation of the hole injector states in quasi-equilibrium, as illustrated in Figure 3. The thicknesses of the AlSb barriers separating the active and hole-injector QWs are typically $\approx 10–12 \text{ Å}$, in order to assure unencumbered hole transport. However, somewhat thicker barriers degrade the ICL performance only gradually.

The QW layer thicknesses in the electron injector should be adjusted so as to produce the required threshold carrier densities, $n_{th}$ and $p_{th}$, when the heavy extrinsic doping is taken into account. In practice, this occurs when the first electron injector QW has a thickness between 40 and 50 Å. The electric field in the active core is then $\approx 70–90 \text{ kV/cm}$ at the lasing threshold. The thickness of each subsequent well is reduced (chirped) so as to maintain coupling between the adjacent QWs. However, the electron injector subbands do not actually form a miniband with similar occupation probabilities in all of the QWs. Figures 2 and 3 indicate that the lower-energy subbands concentrated near the semimetallic interface are more heavily populated than the higher-energy subbands adjacent to the active region of the next stage. This arrangement reduces the density of states (DOS) at lower energies, while still allowing sufficient electron transport from the semimetallic interface to the active electron QWs.

Like the barriers in the hole injector, the AlSb barriers in the electron injector are relatively thin, e.g., 12–14 Å. However, two of the barriers in each stage are much thicker, at 25–30 Å. One separates the electron and hole injectors at the semimetallic interface, in order to minimize parasitic interband absorption across that interface. The other separates the electron injector from the first active electron QW, so as to isolate the active electron subband from the injector states and prevent significant hybridization while still permitting sufficient electron tunneling. Empirically, ICLs with AlSb barrier thickness near the top of the range specified above have shown slightly improved performance characteristics over those with thinner barriers. However, the difference is not significant enough to warrant a further increase in the AlSb thicknesses.

We finally consider the design of the active QWs. It was mentioned above that state-of-the-art ICLs employ the “W” active region, in which two InAs electron wells sandwich a single GaInSb hole well [27]. The typical Ga$_{1-x}$In$_x$Sb alloy composition in an ICL of $x \approx 0.35$ the induces strain near the maximum of what is allowable for coherent layer growth. While hole well thicknesses ranging from 20 to 30 Å have been employed, no significant trends attributable to a strong variation of the Auger coefficient were observed in the experimental laser thresholds or slope efficiencies. This finding contrasts the early theoretical predictions of final-state optimization from the literature [71,72]. Possible reasons are discussed below.

The InAs QW thicknesses are chosen so as to dial in the laser emission wavelength via quantum confinement while maintaining roughly equal electron wavefunction probabilities in the two wells. The “W” structure displays a stronger electron-hole wavefunction overlap than an InAs/GaInSb QW with a single type-II interface [27]. For example, when the emission wavelength is 3.2 µm and the GaInSb thickness is 20 Å in both structures, the square of the overlap peaks at 42% for the “W” structure and only 17% for the single-interface structure. However, it is not immediately obvious that this is
beneficial, insofar as both the optical gain and the recombination rates should scale similarly with overlap. This will be discussed further in what follows.

3.2. Design of the Laser Waveguide

The goal in designing the optical waveguide is to simultaneously minimize both the material gain required to reach transparency and the internal loss. While the ICL’s active core is unique among semiconductor lasers, in most ways its waveguide is analogous to those of other mid-IR emitters. The primary difference is that since all the holes are generated internally, there is no need for $p$-doped cladding or separate confinement layers. This is advantageous since the free-carrier absorption cross-section is usually much larger for holes than electrons.

While the ICL’s active region can in principle comprise the entire waveguide core, this would require more stages than are generally optimal from other standpoints that will be discussed below [64,73]. Therefore, it is useful to view the waveguide as constructed from the following basic building blocks: (1) the active core, (2) $n$-doped optical claddings, most commonly comprised of InAs/AlSb short-period superlattices (SLs), although bulk alloys such as AlGaAsSb provide an alternative [74,75], and $n$-InAs may be employed when the ICL is grown on an InAs substrate [76]; (3) lightly $n$-doped GaSb separate-confinement layers (SCLs); and (4) transition SLs that separate the other three regions from each other, and from the GaSb substrate buffer and $n^+$-InAs(Sb) top contact. The transition SLs reduce parasitic voltage drops that would occur at the otherwise abrupt heterointerfaces between adjacent regions with very different conduction-band offsets (CBOs). Insofar as the SLs employed in the transition layers are quite similar to those in the cladding regions, it is sometimes convenient to lump their optical mode overlaps together.

If low drive power is the overriding figure of merit for a given application, a relatively small number of active stages is advantageous since the bias needed to reach threshold scales with the stage multiplicity $M$. For many years, $M = 5$ was the NRL norm, although now that high-temperature cw operation may be taken for granted, our simulations indicate that $M = 2–4$ may ultimately be ideal. On the other hand, the same simulations find that a larger stage multiplicity is likely to be preferable if maximized cw output power and brightness, rather than minimized drive power, is the primary objective. Nominally, the slope efficiency scales with $M$, although this is accompanied by a decrease of the current at which heat accumulation in the active region causes the slope to roll over.

The optimal number of stages depends to a large degree on how $M$ influences the internal loss. The net internal loss combines contributions from up to four primary sources: (1) the active gain region comprising the $M$ stages layered in series, (2) the two $n$-GaSb SCLs that surround the active core, (3) the top and bottom optical cladding layers that surround the SCLs, and (4) absorption (e.g., in the dielectric coatings) and scattering (e.g., by sidewall corrugations or unintentional non-uniformities) in the narrow ridge waveguide. Since the fourth is nominally insensitive to the epitaxial layering design, it will be taken as fixed in this discussion. We also note that the various transition superlattices separating the cladding, SCL, and active regions may add a non-negligible contribution to the loss. It will be seen below that the optical confinement factors for the active and cladding layers are governed largely by the SCL thicknesses. Relatively thick SCLs are nominally advantageous, since lightly $n$-doped GaSb should exhibit low material loss. However, the active-region optical confinement factor must also remain high enough to assure adequate gain for overcoming the net losses. Whereas empirical studies of 5-stage ICLs found an SCL thickness of $\approx 500$ nm to be optimal [64], later work found thicker SCLs to be advantageous in devices with 7 and 10 stages.

In particular, increasing the stage multiplicity (at fixed SCL thickness) provides additional gain. For example, NRL simulations showed that even though a typical 7-stage design with thick SCLs had 6% less gain than a 5-stage structure with thin SCLs, it was nonetheless beneficial that the mode fraction in the cladding layers was reduced by 40%, with most of the difference being transferred into the SCLs [77]. For example, the experimental slope efficiencies for 7-stage devices were found to be higher than those for 5-stage devices by more than a factor of $7/5$, indicating that the net loss had
decreased. The experimental results discussed in the next section confirm this advantage. Section 5.3 will further discuss how the stage multiplicity affects performance.

Figure 3 of Ref. [20] showed the waveguide refractive index profile and optical intensity profile for the fundamental TE mode in a typical 7-stage ICL with two 800-nm-thick SCLs (λ = 3.7 μm). For that case, the optical mode overlap with the SCLs was ≈ 77%. In most NRL structures, the center of each SCL is n-doped at N_i = 5 × 10^{15} cm^{-3}, whereas 50-nm-thick layers at the edge of each SCL (containing Γ_h ≈ 9% of the mode) are doped at N_o = 1 × 10^{17} cm^{-3} to reduce the voltage at the heterointerfaces. The clad doping is also non-uniform, in that the inner regions closest to the SCL are doped to N_i = 7.5 × 10^{16} cm^{-3}, while the outer regions with less mode overlap are doped more heavily, to N_o = 5 × 10^{17} cm^{-3}. The objective is to trade minimization of the series resistance against minimization of the free-carrier absorption loss. While the estimated total mode overlap with the claddings is typically ≈ 6%, the overlap with the more heavily doped outer regions nearly vanishes. The confinement factor for the active region (including the electron and hole injectors as well as the active QWs) is Γ_a ≈ 14%. The remainder of the mode (≈ 3%) resides in the InAs/AlSb SL transition regions, which are typically doped to N_i = 2 × 10^{17} cm^{-3}.

Figure 4 of Ref. [20] plotted the modal overlap factors for the active core, InAs/AlSb SL cladding layers, separate-confinement layers, and transition regions vs. the total thickness of two identical SCLs on each side of the active core. Whereas the clads have the greatest confinement factor when each SCL is only 100 nm thick, the SCLs become dominant at thickness beyond ≈ 200 nm each, with the active core having the second-highest concentration. Were the free-carrier absorption cross-sections for the claddings and SCLs well known, we could estimate the net loss using the nominal doping densities. However, the literature values for n-GaSb display some spread [78,79], with typical mid-IR values on the order of 10^{-17} cm^{-2}. Little information is available concerning free carrier absorption in the cladding materials, although the contributions of multiple, potentially-rough interfaces in the InAs/AlSb SLs, and alloy scattering in bulk Al(Ga)AsSb, implies that the free carrier absorption cross-sections are surely larger than in n-GaSb. Since the mode-weighted doping of the SCLs is typically an order of magnitude lower than N_i, relatively thick SCLs that gather a lion’s share of the optical mode are preferred. However, thicker SCLs also reduce the confinement in the active core, resulting in lower optical gain. Furthermore, the SCLs can become so thick that the TE_2 mode with two nodes along the growth direction (and an antinode at the active core) can have greater confinement than the TE_0 mode. In state-of-the-art ICLs emitting at λ ≈ 4 μm, the optimal SCL thickness determined by these trade-offs is t_s ≈ 800 nm.

![Figure 4](image-url)

**Figure 4.** Pulsed threshold current densities vs. emission wavelength (at 300 K) for 85 different broad-area (150 μm × 2 mm) ICLs with 5,7, and 10 stages that were grown and processed at NRL when operated at 300 K (blue) and 375 K (red). The longer-wavelength devices indicated by stars employed a hybrid bottom cladding, as explained in the text, while the points at the top of the graph represent devices that did not lase at 375 K.
The internal loss produced by free carriers in the active core is also unknown. Two significant contributions are due to: (1) electrons concentrated in the electron injectors; and (2) holes in the GaInSb active hole QWs. While the intervalence absorption cross-section is likely much larger, at the threshold, there are many more electrons than holes in each stage of the ICL. The high injector electron density is also likely to blur any intervalence resonances that can in principle occur in the active hole QW. In state-of-the-art ICLs with fewer than 10 active stages, the small confinement factor $\Gamma_a$ reduces the internal loss in the core and also the optical gain available to overcome losses elsewhere in the waveguide.

The top cladding layer must be thick enough to minimize mode overlap with the lossy top-contact metallization. The typical top cladding thickness of 1.5 $\mu$m for NRL structures emitting at $\lambda = 4$ $\mu$m (and assuming thick SCLs) is designed to reduce the nominal loss to $< 0.1$ cm$^{-1}$. Additional loss also occurs if the bottom cladding is too thin, since the mode then leaks into the high-index GaSb substrate [54]. The maximum value of this spectrally-modulated loss is: $\exp(-2k_m t_a)/d_m$, where $k_m = 2\pi\sqrt{(n_m^2 - n_c^2)/\lambda}$, $n_m$ is the modal index, and $d_m$ is the modal extent along the growth direction [80,81]. While there are many ways to define $d_m$ for a complicated multilayer waveguide, fortunately, the loss tends to be insensitive to the precise definition. We take $d_m = (2t_s + t_a)/(\Gamma_h + \Gamma + \Gamma_a)$, where $t_a = Md$ is the total thickness of the active core and $d$ is the thickness of one stage.

3.3. Assessment of the Current Level of Understanding

The previous subsections outlined some primary factors influencing the design of state-of-the-art ICLs. Numerous implementations have established that ICL wafers grown in adherence to these guidelines generally yield high performance over a wide range of wavelengths. In particular, the threshold current densities, threshold voltages, and slope efficiencies, and their dependences on temperature tend to be predictable to within $\approx 20\%$. It does not follow, however, that the ICL operation is so well understood that any given performance issue can be addressed straightforwardly, via careful re-engineering of the laser’s active and waveguide regions. Whereas numerous variations on the baseline ICL designs at NRL and elsewhere, particularly at U. Würzburg [82,83], have produced laser characteristics following the expected trends (most notably, the effect of carrier rebalancing on $J_{th}$), in other cases to be described below the observations continue to defy a straightforward interpretation in accordance with known principles.

First, we briefly overview the design aspects that seem in greatest need of attention. For example, it remains unclear why the experimental threshold current density (nominally proportional to the Auger coefficient) and slope efficiency (which is sensitive to the free hole absorption cross-section) show no obvious dependences on the active region layering details such as the thickness and composition of the hole well. Possibly resonance phenomena do not emerge because, for any given well width and composition, a wide variety of final states are available to the various intervalence Auger and absorption transitions. Furthermore, whereas the optical matrix element at the zone center is slightly larger for a thinner active hole QW, the calculated gain per unit current density is nearly constant for any thickness between 20 and 30 Å as long as the electron/hole density ratio is held fixed. The group at Wroclaw University of Technology, in collaboration with U. Würzburg, has proposed and simulated a number of other design variations, for example, employing GaInAsSb or GaAsSb for the active hole QW rather than GaInSb [84–86].

In a second example, while the “W” active QW configuration is projected to produce more than twice the gain of a single type-II interface, it would seem that the gain per unit current density should not vary significantly since the non-radiative recombination rates should scale similarly. In fact, $J_{th}$ may actually be lower in the single-interface structure, since a smaller wavefunction overlap lowers the threshold current density. Instead, the experimental results to date yield a much higher $J_{th}$, conceivably due to design flaws in the particular structures, but also possibly indicating that the non-radiative recombination does not have a simple dependence on wavefunction overlap.
While the projection that very thick injector QWs inhibit electron transfer to the active region has been validated experimentally, the same simulations imply that even thinner QWs may allow a substantial reduction of the heavy doping in the electron injector. The experiments are ambiguous, however, in that the thinner injector QWs failed to provide any advantage, and possibly a slight degradation.

In fact, the very need for an electron injector becomes questionable in light of the carrier-rebalancing results [70]. With no injector, the large density of electrons residing there, and potentially contributing to the internal loss, would disappear. While the injector serves to reduce the applied field at the threshold, the field may nonetheless remain tolerable if the injector is removed or thinned to only one or two QWs. However, such designs have thus far failed to operate at RT even in pulsed mode. Nonetheless, this strategy should not be discounted completely until the electron transport throughout each stage is better understood.

Similarly, there is no clear theoretical rationale for incorporating a second QW into the hole injector. Even though it further suppresses electron tunneling from the active region to the electron injector, in principle the single QW should quite effectively eliminate this leakage path. Experimentally, ICLs containing a single hole injector QW have displayed disappointing performance to date, albeit based on a small sampling. Furthermore, although the barrier thickness at the semimetallic interface has been optimized systematically, there has been little exploration of the other barrier thicknesses. While increasing the barrier thicknesses throughout the active core appears to improve the performance slightly, it seems unlikely that most of those barriers (e.g., internal to the injectors) play any critical role.

As will be discussed further below, even the latest optimized generation of ICLs is challenged by an unexplained efficiency droop that occurs with increasing current density [87]. Its magnitude is anomalous because, in the absence of significant heating, the carrier density in an ideal laser becomes pinned as soon as sufficient gain is generated to overcome the cavity loss at the threshold. Instead, limited measurements of the spontaneous emission above the lasing threshold [88] appear to indicate a continuing increase of the carrier density in the active QWs, although the data do not preclude growing differences in the electron and hole densities. Furthermore, the cavity-length data to be discussed in Section 5.2 imply that most of the efficiency droop results from an increasing internal loss rather decreasing internal efficiency. The higher loss may be connected with the increasing carrier density above the lasing threshold, while the lower internal efficiency may derive from carrier heating in addition to the same mechanism. Unfortunately, no design modification attempted to date has appreciably mitigated this efficiency droop phenomenon. Curiously, the carrier densities in type-I mid-IR diode lasers have shown a similar reluctance to pin at injection currents above the lasing threshold [89,90], although the conditions under which the carrier density in an ICL pins above the threshold remains unresolved [91].

Finally, the same cavity-length studies yield experimental gains per unit current density that are only 60–65% of the predicted values, a discrepancy that substantially exceeds the expected experimental and theoretical uncertainties. This may ultimately be linked to the lack of carrier pinning as well, although the details remain to be clarified.

4. Experimental Performance

4.1. Broad-Area Devices in Pulsed Mode

Intrinsic limits to the device performance are best evaluated from measurements in pulsed mode (with a low duty cycle to preclude significant lattice heating) on ridges wide enough to minimize current leakage and optical scattering losses at the sidewalls. The ridge should be etched to a depth stopping below the entire active core, since significant lateral current spreading otherwise occurs due to the core’s high impedance to vertical transport at current densities well below the lasing threshold [92]. Unless otherwise specified, the broad-area lasers discussed in this work are 150 µm wide and 2 mm
long, with uncoated facets. Sidewall corrugations are introduced to suppress lasing in “dark” modes that would otherwise bounce between the sidewalls and output facet. The measurements typically employed 250 ns pulses with a repetition rate of 3 kHz.

Figure 4 plots the pulsed threshold current densities as a function of emission wavelength for 85 different 5-, 7-, and 10-stage ICLs operating at 300 K (blue) and 375 K (red) [93]. All of the devices were grown, processed, and characterized at NRL, employing designs with carrier rebalancing to minimize Jth as discussed above. The 5-stage devices had n'-GaSb SCLs of thickness 500 nm on each side of the active stages, while the SCL thicknesses in the 7-stage structures varied from 600 to 1250 nm as the wavelength increased from 2.9 to 6.1 µm, so as to maintain approximately constant overlap of the optical mode with the SCLs. Because the confinements of the active cores were similar, the threshold current densities were not expected to display any pronounced variation with stage multiplicity. While the 5-stage ICLs had lower threshold power densities, the 7- and 10-stage devices exhibited higher slope efficiencies, and hence the potential for higher cw output powers.

For λ = 3.7 µm, the RT threshold current densities measured at NRL have been as low as 134 A/cm². The group at U. Würzburg [83] observed an even lower Jth for an ICL with 10 stages, namely 98 A/cm² at T = 20 °C (which translates to Jth = 114 A/cm² at 300 K, using their stated T0). While minimizing the RT Jth has never been an overriding objective on its own, it is noteworthy that the best ICL thresholds for λ = 3.4–3.9 µm are now only moderately larger than the typical values for mature diode lasers emitting in the near-IR spectral band. The temperature variation of Jth may be expressed in terms of a characteristic temperature T0, even though the observed dependence above RT is usually somewhat faster than exponential [94]. Between 300 and 350 K, the T0 values for lasers emitting at λ < 4 µm typically range from 45 to 55 K, with a few lasers falling outside these bounds.

Below a minimum of around 3.3 µm, the thresholds begin to increase and exceed 300 A/cm² when λ < 2.95 µm. This could be due in part to the use of thicker SCLs that minimize the internal loss at the expense of optical gain, and also the stronger interface roughness scattering that accompanies the thinner InAs QWs required to induce greater quantum confinement. However, neither of these mechanisms seems sufficient to fully explain the higher thresholds, which occur in U. Würzburg devices as well as those grown and processed at NRL [20]. Nominally, these results imply a somewhat larger Auger coefficient, as will be discussed below. Furthermore, the slope efficiencies of shorter-wavelength ICLs imply higher internal loss, and the anomalous “droop” of the slope at currents somewhat above the threshold becomes more significant with decreasing wavelength [87]. These observations are not explained by current models for the device operation.

The higher Jth at wavelengths longer than the minimum is less surprising since we expect both the Auger coefficient and the internal loss due to free carrier absorption to increase with increasing λ. When a conventional n-InAs-AlSb superlattice was employed for both the top and bottom cladding layer (blue circles), the room-temperature thresholds at still longer wavelengths increased more rapidly, to ≈400 A/cm² near λ = 4.7 µm and 650 A/cm² at λ = 5.5 µm [95]. However, the thresholds at λ > 5 µm decreased substantially when a hybrid cladding was employed on the bottom (blue stars). The hybrid cladding consisted of 1.5 µm of the usual InAs-AlSb SL, followed by an n'+-InAs0.91Sb0.09 layer doped at 4 × 10¹⁸ cm⁻³ to reduce the refractive index via the plasma effect. The active stages (8 in the NRL devices designed for λ < 5 µm and 8 or 10 when λ > 5 µm) were sandwiched between low-doped GaSb SCLs of thickness 850–1200 nm. While typical T0 values do not vary appreciably with wavelengths in the 3–4 µm window, they decrease to ≈40–45 K at the longer wavelengths.

The group at U. Oklahoma had previously used n'-InAs layers to provide plasma confinement of the lasing mode in ICLs grown on InAs substrates. At λ = 10.4 µm [96], cw lasing was observed up to 166 K. Devices emitting at 4.6 µm and 5.2 µm exhibited threshold current densities of 247 A/cm² and 340 A/cm², respectively, when operated in pulsed mode at 300 K [97,98]. The thresholds for subsequent devices grown at NRC Canada remained low at even longer wavelengths, namely 333 A/cm² at λ = 6.0 µm and 395 A/cm² at λ = 6.3 µm, for pulsed operation at 300 K [76,99]. The team of U. Würzburg and Nanoplus also demonstrated longer wavelength ICLs grown on InAs substrates [100,101].
They reported $J_{\text{th}} = 800 \text{ A/cm}^2$ at $\lambda = 7.0 \mu\text{m}$ for pulsed operation at room temperature, and cw operation to $-10^\circ\text{C}$ at $\lambda = 5.9 \mu\text{m}$.

For portable IR systems requiring long battery lifetimes, a more meaningful figure of merit is the threshold power density $P_{\text{th}}$. It is significant because $J_{\text{th}}$ can in principle be reduced to its “transparency” value by increasing the number of stages $M$, whereas $P_{\text{th}}$ displays a minimum as a function of $M$. Figure 5 shows the RT $P_{\text{th}}$ as a function of $\lambda$ corresponding to the same ICLs with 5 stages (blue points) and 7 or more stages (red points) whose $J_{\text{th}}$ are given in Figure 4. A power density threshold as low as 350 W/cm$^2$ has been observed, and most of the values for 5-stage devices emitting at wavelengths between 3.5 $\mu\text{m}$ and 4.1 $\mu\text{m}$ are $<500$ W/cm$^2$. For both the 5- and $\geq 7$-stage designs to date, $P_{\text{th}}$ increases at both shorter and longer wavelengths. In the next section, we show that $P_{\text{th}}$ may reach its minimum for $M < 5$.

For applications requiring high mid-IR output powers, it is essential to maximize the external differential quantum efficiency per stage (EDQE), i.e., the fraction of electrons injected into each stage which is converted to emitted photons beyond the lasing threshold. Figure 6 plots the wavelength dependence of the EDQE just above the threshold at $T = 300 \text{ K}$, for the same ICLs considered in Figures 4 and 5. Again, the blue points represent devices with 5 stages (SCL thickness $= 500 \text{ nm}$), while the red points are devices with 7 or more stages (with SCLs $\geq 700$ nm thick). The plot of internal loss vs. wavelength in Figure 7 confirms that the higher EDQE for structures with more stages correlates with lower internal loss, which results because sufficient gain is generated even though a larger portion of the optical mode is shifted out of the active stages and into the lower-doped SCLs. The internal loss in each structure was estimated by analyzing the slope efficiencies observed when broad-area devices were operated in pulsed mode. For devices emitting at wavelengths up to $\approx 5 \mu\text{m}$, stage multiplicities between 7 and 10 provide similar EDQEs and internal losses. However, at longer wavelengths, it may be preferable to employ at least 10 stages. The highest RT EDQE to date is 56%, while more typical values fall in the 40–50% range for emission in the 3–4 $\mu\text{m}$ window and 25–30% for $\lambda = 4.5–5.0 \mu\text{m}$. For $\lambda = 3–4 \mu\text{m}$, the EDQE at $T = 350 \text{ K}$ is typically $\approx 60–70\%$ of the value at RT.

Figure 5. Pulsed threshold power densities at $T = 300 \text{ K}$ as a function of wavelength for the same NRL broad-area ICLs as in Figure 4. Devices with 5 stages are represented by blue points and those with 7 or more stages by red points.
**Figure 6.** Pulsed external differential quantum efficiency above the lasing Table 300. K as a function of emission wavelength, for the same NRL broad-area ICLs as in Figures 4 and 5. Again, blue points represent 5-stage devices, while red points indicate devices with \( \geq 7 \) stages.

**Figure 7.** Internal loss vs. wavelength at 300 K as estimated from the pulsed slope efficiencies, for NRL broad area ICLs with 5 (blue), 7 (red), and \( \geq 8 \) (green) stages, assuming 80% internal efficiency per stage.

Pulsed cavity-length studies will be discussed below in Section 5.2, in the context of limitations on device performance.
4.2. Narrow-Ridge Lasers in CW Mode

Narrow ridges are required for most applications in order to ensure lasing (mostly) in a single lateral mode, and also to provide lateral heat dissipation for cw operation. The ridges are typically defined by photolithography and reactive-ion etching. Although both methane-based and BCl₃-based inductively coupled plasma (ICP) processes have been employed [102], most NRL devices to date have used the CH₄-based ICP. The dry etching, which is designed to stop just below the active core of the device in the bottom GaSb SCL, is usually followed by cleaning with a phosphoric-acid-based dry etch. In the standard NRL fabrication procedure, a 250-nm-thick Si₃N₄ layer is deposited by plasma-enhanced chemical vapor deposition, after which a top contact window is etched back using SF₆-based ICP. Approximately 100 nm of SiO₂ is also deposited by sputtering, to block occasional pinholes in the Si₃N₄.

The ridge is next metallized for electrical contact, and then electro-plated with ≈ 5 µm of Au. The electro-plating is patterned so as to leave non-plated gaps of ≈50 µm for cleaving into laser cavities. The standard metallization thickness before electro-plating of 200 nm for 5-stage ICLs is increased to ≈ 1 µm for 7-stage devices [77]. This is needed to prevent damage to the non-plated portions of the ridge due to excessive heating during cw operation. The thicker initial metallization does not interfere with the cleave quality.

Each device is then bonded to a C-mount attached to a thermoelectric cooler. The mounting is either epitaxial side up, relying on lateral heat flow in the electro-plated Au layers for heat dissipation, or epitaxial-side-down, using a high-yield method that was developed at NRL [103]. In order to maximize both the cw output power and the wallplug efficiency, it is useful to coat the two facets for high reflection (HR) and low reflection (anti-reflection, AR). The reflectivity of the HR coating on the back facet, comprising 200 nm of Al₂O₃ topped by 100 nm Au, is estimated to be >95%. The AR coating, consisting of a λ/4 layer of Al₂O₃ in the front facet, may have a minimum reflectivity of as low as 1–2%.

The first 5-stage ICL narrow ridges to incorporate carrier rebalancing lased to a maximum cw operating temperature (T∗ maxcw) of 107 °C when mounted epitaxial side up [70]. A 0.5-mm-long cavity with HR-coated back facet and uncoated front facet displayed a threshold power of only 29 mW, confirming the potential for much longer battery lifetimes than QCLs in applications that do not require high output powers. Section 5.3 below provides a more detailed comparison of the potential merits of both mid-IR device classes.

The advantages of epi-down vs. epi-up mounting may be quantified by analyzing the thermal impedance-area product RgA derived from the light-current-voltage (L-I-V) characteristics of each device, and comparing with numerical simulations, e.g., using the COMSOL finite-element package. The RgA values ranged from 3.3 to 5.2 K/(kW/cm²) for epi-down ridges with widths varying from 7.7 to 15.7 µm, as compared to RgA = 5.1–8.6 K/(kW/cm²) for epi-up ridges with the same dimensions. These results are consistent with the numerical simulations if we take the InAs/AlSb SL cladding to have an anisotropic thermal conductivity with vertical and in-plane values of σ⊥ = 2 W/(m-K) and σ∥ = 10 W/(m-K). Note that the simulations are much more sensitive to the vertical heat conduction, which represents by far the largest contribution to the net device thermal resistance. The fitted value of σ⊥ is a little lower than the 2.7–3.3 W/(m-K) range measured for InAs/AlSb SLs with a somewhat larger period. We also note that the extracted thermal conductivity is consistent with independent measurements based on the 3ω method [104–106].

Epi-down-mounting led to considerable improvements in the cw output power, wall-plug efficiency (WPE), and maximum operating temperature. Here we present narrow-ridge results, while scaling of the cw power with ridge width will be discussed in the next subsection. The emission spectra shown in Figure 8, for a 5-stage device with 7.7 µm width, 4 mm cavity length, and HR/AR coated facets [94], are typical of narrow-ridge ICLs operating in cw mode when wavelength selectivity is not imposed by a distributed feedback or external gratings. The centroid wavelength for this device tuned from ≈ 3.6 µm to ≈ 3.9 µm as the temperature varied from 20 °C to 115 °C. While each spectrum
contains many longitudinal modes, analysis of the far-field patterns described below indicates that nearly all the emission is into a single lateral mode.

![Normalized CW Spectral Density](image)

**Figure 8.** Cw emission spectra for a 7.7-μm-wide 4-mm-long narrow-ridge 5-stage ICL mounted epi-side-down, with an HR-coated back facet and an AR-coated front facet, at temperatures between 20 °C and 115 °C. Adapted from Ref. [94].

### 4.3. Interband Cascade Lasers with High Output Power and Brightness

Since the ICL’s threshold power density is relatively low and the thermal impedance-area product (RtA) decreases with ridge width, a wider ridge will generally produce higher cw output power, although a maximum is reached if the threshold power becomes large enough to cause appreciable heating. However, most practical applications requiring high power also require high brightness, meaning that the output must not depart too far from the diffraction limit ($M^2 = 1$, where $M$ is the beam quality factor). Increasing the ridge width should therefore be accompanied by some means for maintaining acceptable beam quality.

Although tapering the laser cavity to increase the net volume will be discussed below, we first consider patterning the sidewalls of the laser ridge with corrugations [107,108]. The rationale is to suppress higher-order optical modes that are preferentially scattered at the sidewalls, as compared to the fundamental mode which is more localized in the center of the ridge. However, the net brightness increases only if this mode selection outweighs the accompanying decrease in efficiency that results from increased internal loss due to scattering of the fundamental mode. A comparison of the $L-I-V$ characteristics for ridges processed from the same wafer material with both straight and corrugated sidewalls found 25% higher brightness for the devices with sidewall corrugations, even though the WPE was 10% lower [94]. A study of the corrugation parameters found that a relatively long period of ≈ 10 μm provides stronger suppression of the higher-order modes than a shorter period of 2–4 μm. A corrugation amplitude of 2 μm provided roughly the same higher-order-mode suppression as a larger amplitude of 3.5 μm, while imposing less loss due to scattering.

Figure 9 shows: (a) $L-I-V$ characteristics and (b) far-field profiles at 25 °C for a 10-stage ICL with 22 μm ridge width, 4.5 mm cavity length, and HR/AR facet coatings, which was processed with corrugated sidewalls [109]. While a conventional straight-sidewall ridge of this width would normally exhibit poor beam quality, we find $M^2 \leq 2.5$ up to the injection current that produces 500 mW of cw output.
Whereas most of the early ICL narrow ridges that achieved RT cw operation employed 5 stages, we saw in the previous Section that the loss is generally lower and the EDQE higher when a few more stages in the 7–10 range are employed. It follows that higher maximum power and maximum wallplug efficiency are possible with the larger number of stages. Figure 10 compares the wallplug efficiency at 25 °C as a function of cw injection current for two ICL narrow ridges with corrugated sidewalls, but with either 5 (red) or 7 (blue) active gain stages. Clearly, the WPE advantage of the 7-stage ICL goes well beyond that attributable to its slightly shorter cavity (3 mm vs. 4 mm). The values of 13.2% at the WPE maximum and 10.1% at the highest current compare with 8.7% and 6.5% for the 5-stage design. These results represent the consistent trend that ICL narrow ridges emitting in the 3–4 μm spectral range display much higher maximum cw output powers and maximum WPEs for 7–10 gain stages.

Figure 9. (a) L-I characteristics and wallplug efficiency in cw mode at $T = 25 \, ^\circ\text{C}$ for a 10-stage ICL ridge with corrugated sidewalls. The ridge width is 22 μm, the cavity length is 4.5 mm, and the facets are HR/AR coated. (b). Far-field emission profiles at several currents, along with extracted values of the beam quality factor [109].
Figure 10. Wallplug efficiency at $T = 25 \, ^\circ\text{C}$ as a function of the cw injection current for a 7-stage ICL narrow ridge with 32.4 $\mu$m width and 3 mm length ICL (blue), along with the corresponding dependence for a 5-stage narrow ridge with corrugated sidewalls and roughly similar dimensions of 25.2 $\mu$m $\times$ 4 mm (red). Both devices had corrugated sidewalls and were mounted epi-side-down with HR/AR facets [77].

The WPE is maximized in relatively short cavities that optimize the balance between mirror loss and internal loss. Figure 11 shows the $L$-$I$ characteristics (red) and WPEs (blue) of four narrow-ridge ICLs with 1 mm cavity lengths and HR/AR facet coatings, when operated in cw mode at 25 $^\circ\text{C}$ [109]. Both ridge widths for both 10-stage (a) and 7-stage (b) devices reach maximum WPEs of 18%, which is the highest ever reported for an ICL. This compares to 22% for the record WPE of a QCL operating in cw mode at 20 $^\circ\text{C}$ [15].

Besides corrugating the narrow ridge sidewalls, another approach to scaling the brightness of a semiconductor laser is to employ a tapered laser or a tapered amplifier geometry [110,111]. The defining feature is that a ridge waveguide section, narrow enough to support only the fundamental lasing mode, adjoins a tapered section, usually gain-guided, with a typical half-angle of a few degrees. Gain guiding is difficult to implement for ICLs and QCLs because of the extensive current spreading that takes place below the lasing threshold. However, index guiding was used to realize a 5-stage tapered ICL with the geometry illustrated in Figure 12 [112]. The HR/AR-coated ridge consisted of a single 4-mm-long section with a back-facet aperture width of 5.5 $\mu$m, taper half-angle of 0.42$^\circ$, front-facet aperture width of 63 $\mu$m, and sidewall corrugations applied to the wider portion of the ridge. The net active (pumped) area of the device was 0.00139 $\text{cm}^2$, and index guiding was assured by etching the ridge through the ICL active core (as in the other ridges discussed above).
The maximum cw output power of 403 mW, limited by thermal rollover, was obtained at a current density of 210 A/cm² and maximum slope efficiency of 320 mW/A (EDQE = 19%). The threshold current density of 210 A/cm² consisted of a single 4-mm-long section with a back-facet aperture width of 5.5 µm and internal loss.

Figure 11. L-I characteristics (red) and wallplug efficiencies (blue) for narrow ridges with 1 mm cavity length and HR/AR facet coatings in cw mode at T = 25 °C: (a) Two 10-stage ICLs with 12 µm (dashed) and 15 µm (solid) ridge widths; (b) two 7-stage ICLs with 12 µm (dashed) and 15 µm (solid) ridge widths. All four lasers exhibit WPEs up to 18%, the highest ever reported for an ICL [109].

Figure 12. Schematic of the tapered ridge structure for the 5-stage ICL (left), also showing the position of the corrugations. The (right) panel is a micrograph of the processed device’s output facet. Reproduced from Ref. [112], with the permission of AIP Publishing.

Figure 13a shows L-I-V characteristics for the 5-stage tapered ICL, which indicate a cw threshold current density of 210 A/cm² and maximum slope efficiency of 320 mW/A (EDQE = 19%). The maximum cw output power of 403 mW, limited by thermal rollover, was obtained at I = 2A. Figure 13b shows the far-field intensity profiles measured at several cw injection currents, all normalized to the same area. The resulting M² ranged from 1.4 at I = 0.5A to 2.3 at I = 2A. The corresponding brightness figure of merit was ≈ 28% higher than the best earlier result for a 5-stage ICL (a 25-µm-wide...
corrugated ridge). Although the tapered ridge geometry has not been applied to ICLs with 7–10 stages, the 5-stage results indicate it may provide a promising approach to maximizing the brightness.

![Graph showing L-I-V characteristics for the tapered-ridge device at T = 25 °C.](image)

**Figure 13.** (a) CW L-I-V characteristics for the tapered-ridge device of Figure 12 at T = 25 °C. (b) Corresponding far-field emission profiles at a series of injection currents (all normalized to the same area). The extracted $M^2$ values range from 1.4 at $I = 0.5$A to 2.3 at $I = 2$A. Reproduced from Ref. [112], with the permission of AIP Publishing.

Other approaches to improving the ICL brightness include the broad-area photonic-crystal distributed-feedback concept [113]. While such geometries may be explored further in the future, it will be challenging to exceed the cw brightnesses demonstrated already, since very wide devices have high thermal impedance-area ratios, and also because ICLs are sensitive to any additional optical losses introduced into the structure.

### 4.4. Single-Spectral-Mode Emission

Mid-IR ICLs producing a single spectral mode have already found wide use in chemical sensing instruments, especially for platforms where compactness and/or long battery lifetimes are required [1]. The most popular technique for constraining a semiconductor laser to emit a single mode is to incorporate a distributed-feedback (DFB) line grating into a ridge narrow enough to support only the
fundamental lateral mode. Unfortunately, the conventional approach of etching a grating into the top optical cladding layer, and then overgrowing a conducting material with contrasting refractive index, is inapplicable to GaSb-based ICLs because no suitable overgrowth technology exists to date. Several alternative approaches have been explored instead.

For example, ridges have been patterned with periodic sidewall corrugations similar to those discussed in the previous subsection for enhanced brightness, except that the corrugation period was adjusted so as to form a fourth-order DFB \[107\]. One such structure operated cw up to 40 °C and produced 55 mW of single-mode output at \(T = 25\) °C \[73\], the highest reported to date for an interband mid-IR DFB laser.

In another approach, a thin germanium layer was deposited on top of the epilayer, and then periodically patterned to modulate the refractive index and also expose the \(n\)-type top cladding material for electrical contact \[114\]. For a 5-stage ICL emitting at \(\lambda = 3.7–3.8\) \(\mu\)m, reducing the top cladding thickness from its typical value of 1.4 \(\mu\)m to 0.7 \(\mu\)m provided sufficient coupling to the grating. At \(T = 45\) °C, single-mode lasing was observed for all the periods between 519 and 534 nm, with a maximum current tuning range of 10 nm. Temperature tuning of up to 21 nm was observed with the tuning rate \(d\lambda/dT = 0.4\) nm/°C. The maximum single-mode output was 27 mW at 40 °C, 22 mW at 60 °C, and > 1 mW at 80 °C. A slight shift of the grating period would have produced single-mode operation at \(T = 25\) °C, most likely with cw output >30 mW.

Nanoplus now offers DFB ICL products that span the wavelength range 3–6 \(\mu\)m \[115\]. As reviewed in Refs. \[116\] and \[117\], these typically employ gratings etched on both sides of the narrow ridge. Cw output powers at room temperature for \(\lambda = 3–4\) \(\mu\)m are typically > 5 mW and often exceed 10 mW, with temperature tuning of 20 nm. Up to 5 mW at \(\lambda = 4.5\) \(\mu\)m (\(T = 22\) °C) was also reported \[117\]. At the shorter wavelength of \(\lambda = 2.8\) \(\mu\)m, Nanoplus with U. Würzburg observed nearly 1 mW of cw output and 7 nm total tuning range \[118\]. The same team reported cw single-mode lasing to 0 °C at \(\lambda = 5.2\) \(\mu\)m \[119\], and later to 6.0 \(\mu\)m by growing on an InAs substrate \[101\]. A two-segment grating provided single-mode emission spanning 158 nm, effectively covering the entire ICL gain spectrum, although the tuning was not continuous \[120\].

Elsewhere, JPL demonstrated DFB ICLs with lateral gratings that produced up to 20 mW of cw power in a single mode (\(\lambda = 3.37\) \(\mu\)m) at 20 °C, with a sidemode suppression ratio (SMSR) of over 25 dB \[121\]. Earlier JPL/NRL devices had generated 18 mW at 3.6 \(\mu\)m when operated at 40 °C \[92\].

More recently, Thorlabs reported DFB ICLs emitting at 3.3 \(\mu\)m that generated up to 42 mW of cw power with 2% WPE in a single spectral mode at 25 °C, with 30 dB SMSR \[122\]. They accomplished this by longitudinally segmenting the narrow ridge into alternating sections that were patterned with either a top grating for wavelength selectivity or corrugated sidewalls to suppress higher-order lateral modes.

Since there is no mature overgrowth technology for ICLs, all of the DFB ICLs discussed above employed either gratings at the sides of the ridge, which limit the ridge width that can maintain sufficient coupling to the fundamental lasing mode, or a top grating with thinner top clad that induces higher loss associated with overlap of the lasing mode with the contact metallization. However, NRL recently demonstrated a strategy which strongly mitigates the additional loss \[123\]. This structure, shown schematically in Figure 14, employs the same grating and waveguide designs, but with metallization covering a small fraction of the top surface. This is possible because the highly anisotropic electrical conductivity in an ICL will spread the current \[92\] into the regions not lying directly underneath a contact.
A preliminary study showed that reducing the contact fill factor cycle toward 10% in 19-μm-wide Fabry-Perot ICL ridges decreased the RT cw threshold current density by \( \approx 50\% \), and increased the slope efficiency by a factor of 2. The DFB devices were processed from a 5-stage, ICL wafer with top cladding thickness reduced to \( \approx 0.5 \) μm and designed for emission at \( \lambda = 3.3 \) μm. The top grating was etched into a 250-nm-thick layer of InAs. Since this layer was not lattice-matched to the GaSb substrate, it relaxed and exhibited a morphology with considerable cross-hatching, although it was not obvious that this degraded the performance. The gratings etched into the InAs had depths comparable to its layer thickness and pitches ranging from 558 to 568 nm. The low-fill-factor top contacts were defined as openings in the insulating film deposited on top of the grating layer. The width of the contacts was held fixed at 10 μm, while the fill factor cycle was varied from 14% to 100%. The devices were mounted epitaxial side up with uncoated facets and gold electroplating for better heat dissipation. Unfortunately, the devices exhibited considerably more current leakage at the ridge sidewalls than usually observed.

Figure 15 shows the cw light-current characteristics measured for DFB ICL ridges from this lot with a series of contact fill factors varying from 14 to 100%. These devices all lased in a single spectral mode at some range of temperatures and currents, although some only at temperatures above 25 °C. Note that the slope efficiency was highest for the devices with 20% and 33% fill factors, and that both the threshold current and differential slope efficiency were much poorer for the device with a 100% fill factor. The tendency to lase in a single spectral mode showed no apparent dependence on the fill factor. The highest cw power emitted into a single spectral mode was 6.8 mW for the DFB with 33% fill factor.
At a VCSEL! via a tunnel junction between the stages (with each containing 5 active QWs) [128], Photonics ≈ tuned over 300 nm around a center wavelength = and SMSR µtuned continuously over 200 nm, from 3.05 tuning [134]. With Brookhaven National Lab, SUNY demonstrated an external cavity type-I ICL that type-I DFB ICLs (λe development of type-I mid-IR lasers with cascaded gain stages [131]. Broad-area devices emitting advantageous over type-II devices at wavelengths near or below about 3 µm to the origin of the ICL [24], although only a few early attempts were made to demonstrate such devices experimentally [127]. Type-I ICLs were demonstrated more recently at 2.3 µm by U. Montpellier (a VCSEL!), via a tunnel junction between the stages (with each containing 5 active QWs) [128], by SUNY Stony Brook at 1.9–3.3 µm, using a graded AlGaAsSb layer for the hole injector (typically 2 QWs per stage) [129], and by U. Oklahoma and NRC Canada at 3.2 µm, using a design much more analogous to the type-II structure illustrated in Figure 2 [130]. At this time, type-I ICLs are generally advantageous over type-II devices at wavelengths near or below about 3 µm.

Another approach to lasing in a single spectral mode is to incorporate an edge-emitting ICL into an external cavity that couples to a grating, which can be rotated for wavelength tuning. Using this approach, Caffey et al. achieved cw operation at 7 = 15 °C, with 110 nm of continuous tuning around a center wavelength of 3.2 µm [124]. The drive power was <1 W, and >1 mW of output was produced over the entire tuning range. At about the same time, Tsai et al. reported an external cavity ICL that generated slightly less power, but tuned by 150 nm around a center wavelength of 3.15 µm [125].

Quite recently, the novel V-coupling approach realized single-mode emission, at cryogenic temperatures initially [126]. ICLs that emit vertically a single spectral mode will be discussed in Section 4.6 below.

4.5. Type-I Interband Cascade Lasers

The suggestion that type-I rather than type-II active transitions may be employed dates nearly to the origin of the ICL [24], although only a few early attempts were made to demonstrate such devices experimentally [127]. Type-I ICLs were demonstrated more recently at 2.3 µm by U. Montpellier (a VCSEL!), via a tunnel junction between the stages (with each containing 5 active QWs) [128], by SUNY Stony Brook at 1.9–3.3 µm, using a graded AlGaAsSb layer for the hole injector (typically 2 QWs per stage) [129], and by U. Oklahoma and NRC Canada at 3.2 µm, using a design much more analogous to the type-II structure illustrated in Figure 2 [130]. At this time, type-I ICLs are generally advantageous over type-II devices at wavelengths near or below about 3 µm.

The group at SUNY Stony Brook, with various collaborators, has carried out the most extensive development of type-I mid-IR lasers with cascaded gain stages [131]. Broad-area devices emitting around λ = 3.0 µm operated with pulsed threshold current densities as low as 100 A/cm² and wallplug efficiencies as high as 16% at room temperature, while narrow ridges emitted >100 mW cw [132]. At λ = 2.0 µm, the WPE was as high as 20% [133]. With JPL, SUNY reported laterally-coupled type-I DFB ICLs (λ = 2.9 µm) that emitted up to 13 mW cw at 20 °C, with 10 nm of temperature tuning [134]. With Brookhaven National Lab, SUNY demonstrated an external cavity type-I ICL that tuned continuously over 200 nm, from 3.05 µm to 3.25, with cw power > 1 mW over the entire range and SMSR = 25 dB [135]. They later demonstrated an external cavity laser with Aston University that tuned over 300 nm around a center wavelength ≈3.2 µm, with cw output power up to 8 mW [136].

A type-I ICL frequency comb will be discussed in Section 4.7 below.
4.6. Vertically-Emitting Interband Cascade Lasers

The vertical cavity surface emitting laser (VCSEL) geometry may be quite attractive in mid-IR chemical sensing applications, because a mesa device with diameter on the order of the wavelength could operate with ultra-low drive power, and potentially be manufactured in high volume at much lower cost. The first experimental VCSELs to derive gain from type-II antimonide QWs employed optical pumping and operated only to 160 K for cw operation and 280 K in pulsed mode [137]. All of the mid-IR HgCdTe [138] and lead-salt [139,140] VCSELs reported to date have also required optical pumping. Type-I GaSb-based [128,141], and type-II InP-based [142] quantum-well VCSELs have been demonstrated at \(\lambda = 2.3-2.6 \, \mu m\), and recently up to \(\lambda \approx 3.0 \, \mu m\) [143]. However, as with edge-emitting mid-IR lasers, the performance characteristics tend to fall off dramatically at wavelengths much beyond 3 \(\mu m\). In particular, the longest-wavelength GaSb-based VCSEL (\(\lambda \approx 3.0 \, \mu m\)) had a threshold current density of \(\approx 5 \, kA/cm^2\) in pulsed mode at \(T = 25 \, ^\circ C\), and a cw output power of 10 \(\mu W\) at 5 \(^\circ C\) [143].

While an ICVCSEL was originally designed and its performance predicted over two decades ago [26], the first experimental demonstration was not accomplished until 2016 by NRL [144]. The structure illustrated in Figure 16 had 15 active stages that were divided into three groups positioned at antinodes of the resonant optical field. The bottom Bragg mirror was formed by 22.5 repeats of \(n\)-GaSb and \(n\)-AlAs0.08Sb0.92 quarter-wavelength layers that were doped with Te for electrical conduction. The top mirror consisted of four quarter-wavelength Ge layers. The layer thicknesses were controlled to within \(< 1\%\) by in-situ near-IR laser interferometry, which was simulated to provide a reflectivity of 99.5\%. The total epilayer thickness was \(\approx 13 \, \mu m\). Mesas of diameter 30–60 \(\mu m\) were dry-etched to a depth below the active stages, and an annular top contact was deposited to leave an emission aperture of diameter 10–20 \(\mu m\) smaller than the mesa. The top mirror consisted of four quarter-wavelength periods of Ge and Al2O3, which had an estimated reflectivity of 99.3\%. Because the devices were mounted epitaxial-side-up, the thermal dissipation was too unfavorable for cw operation.

Figure 16. (a) Schematic of the epitaxial-side-up mounted ICVCSEL architecture employed by NRL; (b) schematic of the layered design, showing the GaSb substrate, grown GaSb/AlAsSb bottom mirror, active stages (grouped at antinodes of the optical field), deposited Ge spacer layer, and the deposited dielectric top mirror.
When operated in pulsed mode, the ICVCSELs emitted a circularly-symmetric beam at temperatures up to 70 °C. The threshold current density at \( T = 25 ^\circ C \) was as low as 390 A/cm\(^2\), although the slope efficiency was degraded by loss in the mirrors, overlap of the optical mode with the metal contact, and current injection into areas below the annular contacts where lasing could not occur. Despite having an emission area well beyond the wavelength scale, the smallest device with an aperture diameter of 20 \( \mu m \) operated in a single spectral mode. The wavelength of \( \lambda = 3.400 \mu m \) tuned with temperature at the rate 0.3 nm/\(^\circ C\). It should be possible to improve this performance substantially by growing an additional partial mirror stack above the active stages, and then mounting the mesa epitaxial-side-down on a low-loss, high-reflectivity metal contact. That approach would: (1) provide far superior thermal dissipation for cw operation, (2) avoid the annular contact that induces loss, restricts the emission area, and limits the minimum mesa diameter, and (3) avoid the dielectric mirror which was damaged due to the adjacent deposition of the annular contact on the same top surface. Once optimized, such a configuration can potentially provide single-mode emission from mesas \( \leq 10 \mu m \) in diameter, which could operate with only a few mW of drive power due to the extremely small active volume.

Subsequently, the group at Walter Schottky Institute demonstrated an electrically-pumped VCSEL emitting at \( \lambda = 4.0 \mu m \) that operated cw up to −7 °C and in pulsed mode up to 45 °C [145]. The gain region for that device comprised eight adjacent type-II W quantum wells, rather than cascaded stages. Using optical pumping of a type-I gain region, Praevium and Thorlabs demonstrated a VCSEL whose single-mode center wavelength of 3.35 \( \mu m \) could be tuned by 97 nm via positioning of an external top mirror with a micro-electromechanical system (MEMS) [146]. Cw operation was observed up to 20 °C. The same team quite recently demonstrated an electrically-pumped ICVCSEL (\( \lambda = 3.35 \mu m \)) that operated in cw mode to 26 °C. At \( T = 16 ^\circ C \), the threshold current was 6 mA, and 70 \( \mu W \) of single-mode cw output was produced [146]. The threshold input power was only 40 mW for RT cw operation. The longer-term objective is to produce MEMS-tunable ICVCSELs.

Vertical emission can also be obtained using a second-order grating to diffract a beam propagating in the plane, an approach applied to both QCLs [147] and ICLs. In 2006, NRL demonstrated an optically-pumped type-II “W” photonic crystal distributed feedback laser that emitted from the top surface at \( \lambda = 3.7 \mu m \) [148], with cw operation at 81 K and pulsed operation above 200 K. More recently, TU Wien and U. Würzburg demonstrated ICLs fabricated into rings and patterned with metallized second-order DFB gratings to provide vertical emission via the substrate [149]. The pulsed threshold current density at room temperature was 750 A/cm\(^2\) for a single-mode wavelength of 3.73 \( \mu m \), although a second mode appeared at temperatures above 20 °C. The latest ring ICLs were mounted epitaxial-side-down and lased in cw mode to 38 °C [150]. A device with 6 stages and ring diameter 800 \( \mu m \) operated at a threshold current density of 600 A/cm\(^2\) at 20 °C, and produced > 6 mW of cw power with WPE \( \approx 1 \% \). With higher-order modes suppressed by a narrow waveguide width of 4 \( \mu m \), the vertical emission occurred in a single spectral mode at \( \lambda = 3.8 \mu m \), which tuned continuously with current and temperature at rates of 0.06 nm/mA and 0.37 nm/\(^\circ C\), respectively.

4.7. Interband Cascade Laser Frequency Combs

An optical frequency comb emits a broad spectrum of coherent narrow lines that are evenly-spaced with fixed relation to a stable absolute reference frequency. The recent advent of frequency combs operating in the visible and near-IR has revolutionized precision spectroscopy [151,152], and they have become the preferred standards for measuring frequency [153], time [154], and other metrics. The extension of frequency combs into the mid-IR, for example, using combs based on frequency down-conversion [155], difference frequency generation (DFG) [156] and QCLs [157], is proving to be especially powerful for spectroscopy.

A common way to generate a frequency comb is to passively mode-lock a laser that emits into a single lateral mode, since the resulting pulse train has exactly the required spectral characteristics produced by the phase locking of individual modes. A phase-coherent train of very short pulses
(<1 ps) is generated by maintaining a well-defined phase relationship among the longitudinal modes in the laser cavity. This can be accomplished by placing a saturable absorber (SA) within the cavity, an intensity-dependent loss element that absorbs photons until a very high intensity (at the peak of the short pulse) saturates the transition. Conventional diode lasers can incorporate the SA monolithically, by forming individual top contacts to separated gain (forward-biased) and SA (typically, reverse biased) sections of the ridge waveguide, as shown schematically in Figure 17. For AM-like passive mode-locking, the lifetime in the upper lasing level must be longer than the round-trip time for photons in the laser cavity (typically 50–100 ps for a cavity length of 2–4 mm), which in turn must be longer than the recovery time for loss in the SA. It follows that the QCL gain medium is unsuitable for passive mode-locking, since the upper lasing subband has a very short (∼1 ps) relaxation rate due to phonon scattering. Instead, QCL frequency combs display a parabolic phase profile (referred to as FM) and do not require a SA, taking advantage of Kerr frequency mixing instead. On the other hand, the carrier lifetime of 300–500 ps in an ICL at threshold appears much more favorable.

In 2018, JPL and NRL demonstrated the first ICL frequency combs, which were the first electrically-pumped combs to operate in the λ = 3–4 μm range [158]. Using the monolithic geometry illustrated schematically in Figure 17, the gain section (3.7 mm long) and SA section (200 μm) were separated by a gap that prevented electrical cross-talk (100 μm). The SA section was ion bombarded, and sometimes also reverse-biased, to assure that the SA recovery time (the time to extract carriers from that section) was much shorter than the lifetime in the gain section. The comb bandwidth of 35 cm⁻¹ was centered on an emission wavelength of 3.6 μm. For cw operation at 15 °C, a very narrow rf beat-note of < 1 kHz was observed. The beating of two combs with slightly different repetition frequencies was also demonstrated, although in this early demonstration they were not applied to dual-comb spectroscopy. However, the interferometric autocorrelation data lacked the 8:1 peak-to-background ratio expected for complete mode-locking and emission into a single short pulse. It is therefore possible that the frequency comb operated with the FM phase relationship between neighboring modes rather than the AM one characteristic of passive mode-locking [159]. Nonetheless, it should be possible to realize both types of combs in ICLs by controlling group velocity dispersion and nonlinearity (via the use of SAs). TU Wien more recently induced active mode-locking of an ICL frequency comb by injecting 1 W of rf power [160].

![Figure 17. Schematic of an ICL frequency comb, with gain section, saturable absorber section, and uncontacted spacer layer to provide electrical isolation.](image-url)
In dual-comb spectroscopy, the signals from two combs with slightly different mode spacings are superimposed to map a mid-IR spectrum into the rf where it is detected using a sufficiently fast detector. If one of the beams passes through a chemical analyte of interest, absorption features appear in the rf spectrum. Princeton, JPL, and NRL applied dual-comb spectroscopy to sensing methane and HCl [161]. Beating the two ICL combs with center wavelengths \( \approx 3.61 \, \mu m \) provided a net spectral coverage of \( 33 \, cm^{-1} \), with \( 0.32 \, cm^{-1} \) frequency sampling interval. Each of the devices with 4 mm length and 4 \( \mu m \) width emitted 8 mW of average optical power and required \( < 1 \) W of combined drive power. These characteristics are favorable for compact, broadband, and energy-efficient chemical sensing systems. Princeton U. and NRL had previously applied the same rf beating effect that makes dual-comb spectroscopy possible to multi-heterodyne sensing of ethylene and methane [162]. That experiment superimposed the slightly-mismatched longitudinal cavity modes emitted by two narrow-ridge Fabry Perot ICLs.

JPL and NRL subsequently reported the simultaneous emission of mid-IR and near-IR frequency combs (centered on 3.6 \( \mu m \) and 1.8 \( \mu m \)), as a result of second harmonic generation (SHG) of the mid-IR beam by nonlinearities inherent to the ICL gain stages [163]. Most recently, 1,1-difluoroethane was sensed with a much higher signal-to-noise ratio by using dual-comb spectroscopy spanning \( \approx 20 \, cm^{-1} \) centered around 3.62 \( \mu m \), as shown in Figure 18 [164]. The introduction of an optical isolator, which prevented feedback of reflected signals back into the laser cavities, substantially improved the temporal and spectral stability of the dual-comb signal. Injection locking of the optical beat note frequency with an rf input was also demonstrated [165]. Detection was achieved by a fast interband cascade detector (ICD, discussed below) that was processed from the same bi-functional wafer material as the ICLs, although not monolithically on the same chip. Figure 18d shows that a very fast frequency response (> 10 GHz), which is required to detect the rf beat note at 9.7 GHz, was obtained by reducing the detector area to minimize its capacitance. Since both lasers consumed \( < 1 \) W of power, this clearly shows the potential for low-power and extremely compact spectrometers that combine both ICL frequency combs, passive waveguides for sensing, and ICDs processed on a single monolithic platform [166].

**Figure 18.** Dual comb spectroscopy with two ICL combs: (a) Schematic of the experimental setup; (b) superimposed emission spectra of the two ICL combs; (c) experimental (points) and simulated (curve) dual-comb spectrum of 1,1-difluoroethane at atmospheric pressure; (d) frequency response of ICDs with three different lateral dimensions from microwave rectification characterization. Reproduced from Ref. [164], with the permission of AIP Publishing.
TU Wien, with other collaborators, has independently reported the injection locking of ICL frequency combs [160]. Although their laser cavities were also divided into two regions, those functioned as a gain section and an rf injection/extraction section (rather than an SA section). When corrected for the capacitance of the rf section, the frequency response was flat up to 20 GHz. To unequivocally prove frequency comb operation, they performed shifted wave intermode beat Fourier transform spectroscopy (SWIFTS) [167], which measured the coherence and phase between each pair of comb lines. The amplitudes of the SWIFTS spectrum were found to match the beating amplitudes extracted from the intensity spectrum, proving comb operation over the entire emission spectrum. On the other hand, the minimum of the SWIFTS interferograms at zero path difference indicated that amplitude modulations were suppressed. They believe the phase-locking was instead an FM process governed by the combined effects of a Kerr nonlinearity and the group velocity dispersion. Fast interband cascade detectors fabricated on the same chip demonstrated 5 GHz frequency response. Subsequently, an ICL was actively mode-locked to produce pulses of length 3.2 ps arriving at $\approx 100$ ps intervals, with 2.7 mW average power [167].

Passive mode-locking has, however, been realized by SUNY Stony Brook using 3-stage type-I ICLs. The frequency combs emitting at $\lambda = 3.25 \mu m$ were processed as epitaxial-side-up narrow ridges with the top contact split into gain and SA sections as in the JPL/NRL design [168]. With a reverse bias on the SA section, the devices generated $\approx 10$ ps pulses with average power $> 1$ mW. The smooth, bell-shaped frequency comb spanned a spectral bandwidth of $\approx 20$ nm. Second-order interferometric autocorrelation studies revealed a strong spectral chirp in the pulse. Mapping the rf spectral dependence vs. biases applied to the gain and SA sections revealed switching between two bistable mode-locking regimes.

4.8. Interband Cascade LEDs

Mid-IR LEDs were recently reviewed by Krier et al. [169]. The same active stages that produce optical gain in an interband cascade laser can be used without feedback to produce spontaneous emission in an interband cascade light-emitting device (ICLED). Here we use the term “device” because the interband cascade structure is not, in fact, a diode, despite the presence of both electrons and holes. As illustrated schematically in Figure 19, a typical ICLED uses the same vertical emission architecture as an ICVCSEL, but with no top and bottom mirrors to form a cavity. For some applications that do not need high power, a broadband incoherent mid-IR source with continuous, rather than abrupt, L-I response (i.e., no threshold) is preferred. These include certain spectroscopic techniques for chemical detection [169–171], particularly multi-point sensors that require low unit cost, and dynamic IR scene projection [172]. Although mid-IR LEDs have been investigated for several decades [173,174], the maximum continuous wave output powers for packaged commercial devices operating at room temperature are still $\leq 300 \mu W$. Recent ICLEDs have considerably advanced this state of the art.

![Figure 19](image-url)

**Figure 19.** (a) Typical ICLED layering configuration (without grouped stages), where the active stages may be identical to those of an ICL; (b) schematic of the epitaxial-side-down mounted device, which emits through an annular bottom contact on the substrate side.
U. Houston, with Sandia and NRC Canada, demonstrated the first ICLEDs in 1996–1997, which at 77 K emitted 700 (50) nW at a peak wavelength of 5.8 (11) µm, and continued to operate up to room temperature where the peak wavelength was 7.4 (14) µm [175–177]. Several years later, Naresh Das at the Army Research Laboratory investigated ICLEDs emitting at a variety of mid-IR and LWIR wavelengths as a potential source for IR scene projection arrays [178–181]. He demonstrated that various treatments of the output surface could substantially enhance emission efficiency and power. The out-coupling from an LED is normally very inefficient due to total internal reflection, which redirects all but a small cone of photons striking the output surface back into the device which at 77 K emitted 700 (50) nW at a peak wavelength of 5.8 (11) µm, and continued to operate up to room temperature where the peak wavelength was 7.4 (14) µm [175–177]. Several years later, Naresh Das at the Army Research Laboratory investigated ICLEDs emitting at a variety of mid-IR and LWIR wavelengths as a potential source for IR scene projection arrays [178–181]. He demonstrated that various treatments of the output surface could substantially enhance emission efficiency and power. The out-coupling from an LED is normally very inefficient due to total internal reflection, which redirects all but a small cone of photons striking the output surface back into the device material (e.g., the out-coupling from an unpatterned GaSb substrate surface is <4%). He found that the maximum pulsed peak power emitted at room temperature by an array of ICLEDs with peak wavelength 3.8 µm, dimensions 100 µm × 100 µm, and bump-bonded to a Si readout circuit, could be increased from 40 µW per pixel for emission from an untreated back surface to nearly 400 µW (corresponding to an emission intensity of 4 W/cm²) when the substrate was thinned to 20 µm and a grating patterned on the output surface [179]. For 100 µm square ICLEDs with peak wavelength 8 µm, the peak power was > 7 µW (0.07 W/cm²) when the substrate was thinned to 25 µm and no grating was patterned [180]. He also found a substantial enhancement of the out-coupling when 20 nm of Au was deposited on the output surface to provide a localized surface plasmon resonance, although the observed powers were reported only in arbitrary units [181].

In 2008, U. Iowa demonstrated 3.8 µm LEDs in which 1, 4, 8, or 16 stages of a type-II InAs-GaSb superlattice active region were connected by tunnel junctions (which included semimetallic interfaces) to convert holes into electrons [182]. At 77 K, a 120 µm × 120 µm mesa with 16 stages emitted up to 960 µW (6.7 W/cm²) of cw power, with a maximum wallplug efficiency of 0.6%. For 50% duty cycle, later devices with 4.1 µm peak wavelength and 520 µm × 520 µm area emitted 0.8 mW (0.30 W/cm²) at room temperature (11 K, 4.1 W/cm², with maximum WPE = 1% at 77 K) [183], and > 600 µW (0.022 W/cm²), WPE = 0.036%, at 77 K for devices with 8.6 µm peak wavelength [184]. U. Iowa [185] broadened the ICLED emission spectrum by using tunnel junctions to connect active regions with different bandgaps. U. Glasgow also combined different mid-IR bandgaps, but with separate contacts to each emission region [186]. A distinct advantage of ICLEDs, whether the stages are joined by semimetallic interfaces or tunnel junctions, is that each stage can be independently designed to have a different bandgap. The reduction in the wall-plug efficiency due to parasitic voltage dropping at the series resistance can also be minimized.

In 2014, NRL reported 15-stage ICLEDs with 400 µm diameter and 3.3 µm peak wavelength that emitted up to 1.6 mW of cw power and up to 1.7 W/cm² emission intensity at 25 °C, via an annular contact on the substrate side as in Figure 19b, with a maximum wallplug efficiency of 0.15% [187]. Subsequently, NRL nearly doubled the maximum power by splitting an ICLED’s 22 active stages into four groups positioned at antinodes of the optical field [188]. As illustrated in Figure 20a, the IR radiation then interferes constructively with itself when reflected at small angles with respect to the norm from the top metal contact of the epitaxial-side-down mounted device. With this approach, an ICLED with peak wavelength 3.1 µm and 400 mesa diameter emitted up to 2.9 mW of cw power and 2.6 W/cm² emission intensity at 25 °C, with a maximum wallplug efficiency of 0.4%. The L-I-V curves shown in Figure 20b indicate a very gradual degradation with temperature, so that 1.7 mW cw is emitted even when the device is operated at 105 °C. One of these devices was used by Rice University to detect methane [170]. Most recently, NRL reported ICLEDs with peak wavelength 4.2 µm emitting 1.4 mW (1.1 W/cm²) and peak wavelength 4.7 µm emitting 0.5 mW (0.4 W/cm²) [189]. Nanoplus adopted the grouping of stages to obtain 5 mW of cw output at peak wavelength 3.7 µm from a larger mesa of a diameter of 640 µm [190]. While the emission intensity of 1.6 W/cm² was lower than the best NRL result, the maximum WPE of 0.7% is the highest ever reported for room temperature operation to date. We note that the WPE of any mid-IR LED decreases markedly with increasing drive current and output power, because the Auger lifetime that dominates at room temperature decreases rapidly (nominally as 1/t²) as more carriers are injected, leading to lower radiative efficiency.
Figure 20. (a) Schematic of an epi-side-down mounted ICLED mesa with 22 active stages split into four groups positioned at antinodes of the optical field reflected from the metal contact; (b) $C_w$ $L-I-V$ characteristics for an epi-down-mounted ICLED with mesa diameter 400 $\mu$m at six different temperatures ranging from 10 $^\circ$C to 105 $^\circ$C [188].

SUNY Stony Brook has recently demonstrated type-I ICLEDs with 5 or 10 stages connected by tunnel junctions [191]. At 20 $^\circ$C with 5 $\mu$s pulses at 50% duty cycle, the 10-stage devices with 3 active QWs per stage with diameter 500 $\mu$m emitted up to 2.3 mW (1.2 W/cm$^2$) at a peak wavelength of 3.1 $\mu$m, which increased to 6.8 mW (3.5 W/cm$^2$) at 77 K. The maximum WPE at room temperature was 0.19%.

The most obvious route to further improving the ICLED performance is to enhance the out-coupling efficiency by texturing the output surface [192], as was already demonstrated more than 10 years ago by ARL [179]. Another promising direction is to add top and bottom mirrors as in a VCSEL, but operate below the gain threshold to form a resonant cavity LED. This has the effect of enhancing the emission rate due to a higher optical field at the antinodes of the cavity, while improving the directionality of the output and narrowing the emission spectrum. U. Lancaster recently applied this architecture to non-cascaded mid-IR LEDs with InAsSb alloy active regions [193].

We finally note that amplified spontaneous emission (ASE) may provide an alternative pathway to higher radiative output power in a more directional beam with a narrower spectrum. The objective is to operate at an injection current above the gain threshold, but without front and back mirrors to provide feedback for lasing. For the typical edge-emitting geometry, one then expects the resulting output power and spectral linewidth to be intermediate between those of a laser and an LED. To our knowledge, the longest-wavelength semiconductor amplifier to employ gain from interband transitions was reported by Tampere University of Technology in Finland, who demonstrated ASE at 2.55 $\mu$m from a non-cascaded GaInAsSb QW gain region [194]. Up to 38 mW of peak power was observed at room temperature, with a spectral bandwidth of 124 nm FWHM. Limited attempts by NRL to demonstrate ASE from interband cascade devices were unsuccessful because lasing occurred rather than ASE, even when the output facet was angled to minimize feedback.

4.9. Interband Cascade Detectors

It was demonstrated in 2004 that a slight modification of the quantum cascade laser structure, when operated near zero bias, can function as a quantum cascade detector (QCD) [195]. This is because incident photons, having energy corresponding to the intersubband resonance, induce a photocurrent flowing in the opposite direction from that which produces gain under forward bias. The QCD relies on intersubband absorption like a quantum well infrared photodetector (QWIP), and a separate photon must be absorbed in each stage before a single electron is able to transit the entire cascade structure.
The following year, the first interband cascade detector (ICD) was produced by applying the same general concept to an ICL structure that did not even require modification [196]. U. Oklahoma with various collaborators led the development of ICDs [197–202], which was subsequently pursued by U. New Mexico [203,204] and several other groups [164,205–207].

Hinkey and Yang established a theoretical framework for the ICD [199]. Figure 21 schematically illustrates band profiles and wavefunctions for 1½ periods of a typical structure at reverse bias. Although an ICD can operate at zero bias, a small reverse bias can sometimes help the carriers diffuse in the desired direction. Note that the absorber in each stage of this example is a type-II InAs-GaSb superlattice rather than a single “W” QW as in the typical type-II ICL. With only a single active QW per stage, a prohibitive number of stages would be required to absorb most of the incident photons propagating vertically. Note also that while the same current must flow through all the stages, the incident beam depletes as it passes through the structure. This means that to balance the absorption per stage, the absorbers located farther from the plane of incidence should be thicker.

**Figure 21.** Schematic of the band profiles and wavefunctions for 1½ periods of an interband cascade detector structure at a small reverse bias. In this example, the absorber region of each stage is an InAs/GaSb type-II superlattice. Reproduced from Ref. [199], with the permission of AIP Publishing.

When the absorption of a photon creates an electron-hole pair, say in the stage at the right of the structure illustrated in Figure 21, the electron diffuses to the electron injector, which (close to zero bias) functions as a sink that empties into the hole injector where the electron recombines. Meanwhile, the photoexcited hole diffuses to the hole injector (needed to block electron flow in the wrong direction) of the next stage at right, and then to the interface with the electron injector of that stage where it also recombines. This leaves the stage at left with one fewer hole than had been present before the absorption event, and the stage at right with one additional hole. With further absorption, a given photoexcited charge can migrate through all the stages to the contacts, where it exits as photocurrent. Both contacts of an ICD can be n type as in an ICL.
The ICD can be advantageous over a conventional single-stage detector because its shorter absorption region (per stage) can overcome poor collection efficiency due, e.g., to a low minority hole mobility. It can have a higher resistance area product \( R_0A \) at zero bias since the separation into stages increases the voltage while decreasing the current. Although the quantum efficiency for vertical illumination is limited to the reciprocal of the number of stages because multiple absorption events are needed to transport a single charge through all of them, the dark current is similarly suppressed with further reduced noise because it depends on multiple consecutive thermal generation processes. Huang et al. pointed out that ICDs are generally expected to have higher specific detectivities than QCDs, because of the lower “saturation current density” resulting from the longer non-radiative lifetime of an interband process [202].

At 77 K, ICDs displayed specific detectivities \( (D^* ) \) of \( 6 \times 10^{12} \text{cmHz}^{1/2}/W \) at a cut-off wavelength of \( 4.0 \mu \text{m} [198] \) and \( 3.7 \times 10^{10} \text{cmHz}^{1/2}/W \) for \( \lambda_{co} = 9.2 \mu \text{m} [200] \). More recent work has emphasized operation at room temperature, where \( D^* \) up to \( 1.6 \times 10^{9} \text{cmHz}^{1/2}/W \) has been reported for \( \lambda_{co} = 3.6 \mu \text{m} [204] \) and \( 2.4 \times 10^{8} \text{cmHz}^{1/2}/W \) for \( \lambda_{co} = 9 \mu \text{m} [201] \). A particular advantage of ICDs is their potential for high speed, with response frequencies of \( 2.4 \) GHz [207] and \( >4 \) GHz [164] recently reported. Related applications of photovoltaic interband cascade devices have also been demonstrated, such as thermophotovoltaic (TPV) conversion from heat to electrical power with up to 0.65 V open-circuit voltage at room temperature [208]. The next section will discuss a further attractive aspect, namely the potential for incorporating both ICLs and ICDs into photonic integrated circuits residing on the same chip.

4.10. Photonic Integrated Circuits Incorporating ICLs and ICDs

The integration of III-V active optical components on silicon to form a photonic integrated circuit (PIC) is currently receiving a great deal of attention [209,210]. This interest is driven largely by telecommunications applications, for which relatively-mature near-IR PICs can combine lasers, detectors, resonators, modulators, couplers, multiplexers, etc. However, more recent research has begun to explore mid-IR PICs, which can potentially integrate multi-spectral lasers and detectors on a single chip to provide chemical sensing with broad spectral coverage of multiple bands, extremely compact system footprint, and low cost in volume [211]. Other applications may include remote explosives detection, thermal imaging, and free-space communications.

Mid-IR PICs can be constructed on platforms employing silicon waveguides (i.e., silicon photonics) or by processing multiple active devices coupled by passive waveguides on a native III-V chip, often referred to as monolithic integration. The primary attraction of silicon photonic integration is the potential to dramatically reduce costs with large scale fabrication on wafers up to 300 mm in diameter. The relative maturity of silicon fabrication techniques can be leveraged to construct low loss waveguides and high-performance phase-sensitive optical elements, like arrayed waveguide gratings (AWGs). Despite these advantages, the integration of active III-V elements on silicon is complex compared to the alternative monolithic III-V integration approach. Still, the majority of passive mid-IR integrated circuit elements have so far been demonstrated with silicon or Ge rather than III-V waveguides [212], while the integration of active mid-IR devices on silicon remains a relatively new area of research.

Two primary methods for integrating active III-V devices on silicon platforms have been demonstrated: (1) Grow the III-V epitaxial layer by MBE or MOCVD on its native substrate, and then heterogeneously bond the III-V material to a pre-patterned silicon chip; or (2) Grow the III-V epitaxial layer directly on the silicon substrate. In 2016, UCSB, NRL, and U. Wisconsin demonstrated the first integration of quantum cascade lasers on silicon, using the heterogeneous bonding approach [213,214]. The III-V structure with conventional lower InP cladding layer and QCL active stages, but only a thin upper cladding layer, was heterogeneously bonded epitaxial-side-down to a silicon-on-nitride-on-insulator (SONOI) substrate that was already patterned with passive waveguides and DFB gratings etched into the silicon. All of the III-V device processing was then performed after
the bonding step, following etch removal of the InP substrate. The active section of each laser cavity comprised a hybrid waveguide in which the III-V gain material and underlying silicon waveguide shared the optical mode. At each end of the gain section, the hybrid III-V/Si waveguide was tapered for coupling into a Si-based passive waveguide, from which light was emitted at a cleaved facet. Unfortunately, the tapers were found to couple only a small fraction of the propagating light to the Si-based waveguides. However, after one taper was polished away to allow output from the hybrid waveguide, these QCLs on Si ($\lambda \approx 4.8$ $\mu$m) displayed attractive thresholds of $\leq 1$ kA/cm$^2$, with peak power $> 200$ mW for pulsed operation at 20 $^\circ$C and lasing up to 100 $^\circ$C, although the slope efficiencies were lower than for state-of-the-art QCLs processed on the native III-V substrates. Multispectral mid-infrared light was also generated on a single chip by combining three heterogeneously integrated QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using growth on silicon [216], heterogeneous bonding [217], and epoxy-assisted transfer printing [218]. U. Montpellier reported InAs-based QCLs grown on silicon that performed comparably to devices grown on the native substrate, with pulsed thresholds $\approx 1$ kA/cm$^2$ and $> 50$ mW peak power at 300 K, and operation to 380 K [219].

In 2018, UCSB and NRL demonstrated the first ICLs integrated on silicon, which were heterogeneously bonded to a silicon on insulator (SOI) platform using methods analogous to those applied earlier to the integrated QCLs [220]. Figure 22a shows a top view of the cavity, in which the central ICL mesa provides gain within a hybrid waveguide, and the tapers at each end are designed to provide coupling to passive silicon waveguides. Feedback is then provided by reflection from the polished outer facets. Figure 22b shows a cross-sectional view of the gain region, with the ICL gain stages indicated in red and the underlying silicon waveguide in gray. Figure 22c illustrates that the $TE_{00}$ optical mode profile is to some extent shared between the III-V and silicon portions of the hybrid waveguide, whereas the $TE_{10}$ mode resides almost entirely within the III-V material.

Figure 23a shows the $L$-$I$ characteristics of a Fabry-Perot ICL integrated on silicon, with ridge width 11 $\mu$m and underlying silicon ridge width 1 $\mu$m, at a series of temperatures between $-100$ $^\circ$C and 50 $^\circ$C. Over this temperature range, the centroid emission wavelength increased from 3.34 $\mu$m to 3.74 $\mu$m. As in the case of QCLs integrated on silicon, coupling from the hybrid to the silicon waveguide at the tapers was quite weak. Therefore, one taper was polished off to allow output from the hybrid waveguide. Unsaturated peak powers of 12 and 7 mW were measured at $-20$ $^\circ$C and 20 $^\circ$C, respectively. The plot of threshold current density vs. temperature in Figure 23b can be fit with a characteristic temperature $T_0 = 43$ K at higher temperatures. The value 1.1 kA/cm$^2$ at 20 $^\circ$C is 5–10 times higher than is usual for a high-quality ICL processed on the native GaSb substrate (see Figure 4), and the slope efficiency is much lower. It was concluded from a comparison of $L$-$I$ characteristics for three different ridges with widths of 6, 8, and 11 $\mu$m that this most likely resulted from severe current leakage at the ridge sidewalls. The threshold currents of 310–360 mA were nearly identical for all three devices, and the $I$-$V$ turn-on was also much softer than is normally observed. The poor sidewall quality probably resulted from device processing that employed a BCl$_3$ ICP etch to define the ridges, rather than the CH$_4$-RIE, which usually gives the most consistent yield for ICL ridges processed on the native III-V substrate. Far-field characterization confirmed that lasing was primarily in the $TE_{10}$ mode, possibly because of the high current that flows near the ridge sidewalls. Much better performance may be expected once a fully-optimized processing protocol has been developed and coupling to the silicon waveguide is improved. With flip-chip bonding to a heat sink, cw operation will be much less challenging for integrated ICLs than for QCLs, since the threshold power density is 1–2 orders of magnitude lower.
photons respectively. The plot of threshold current density vs. temperature in Figure 23b can be fit with a $3.74 \times 10^7$ width 11 μm and underlying silicon ridge width 1 μm.

in the TE10 mode, possibly because of the high current that flows near the ridge sidewalls. Much better performance may be expected once a fully-optimized processing protocol has been developed and characterized.

Far-field characterization confirmed that lasing was primarily to the silicon waveguide is improved. With flip-chip bonding to a heat sink, cw operation is automatically matched at 15 °C [223]. Hitaki et al. avoided the wavelength mismatch issue by growing a separate layer on a silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhere, using QCLs with a silicon arrayed waveguide grating (AWG) acting as a beam combiner, in a PIC on a single silicon chip [215]. InP-based QCLs were subsequently integrated on silicon elsewhi...
Besides lasers, a PIC may also incorporate one or more detectors. A distinct advantage for PICs involving ICLs is their compatibility with ICDs processed from the same epitaxial structure. This is not the case for a QCD processed in parallel with a QCL, due to the “extraction” subband lying about one optical phonon energy ($\hbar \omega_{0}$) below the lower lasing level to provide rapid depopulation following a stimulated emission event. At zero bias, most electrons populate the extraction subband, and the detector’s absorption peak occurs closer to $\hbar \omega + \hbar \omega_{0}$ than to the lasing energy $\hbar \omega$ when one takes into account a small Stark shift of all subbands.

The 7-stage ICL narrow ridge ($\lambda_{0} = 3.1 \, \mu m$ [221], later used to detect ethanol occupying a gap between the laser and detector [222], and subsequently optimized to produce up to 1 W of cw emission from the laser at 15 °C [223]. Hitaki et al. avoided the wavelength mismatch issue by growing a separate QCD on top of the QCL, so that the design of each could be optimized independently [224]. The detector region on top was etched away in the laser section of the waveguide.

Fortunately, wavelength mismatch does not present a significant issue for ICL structures, where any shift is small and due only to the Stark effect. Consequently, the ICD cut-off wavelength automatically matches the emission wavelength of an ICL processed from the same wafer material. U. Oklahoma demonstrated an ICL and ICD integrated on the same native GaSb chip [225], although separated by an air gap rather than a passive waveguide, as will be required for versatile integration on a PIC. For top illumination, the detector with $\lambda_{co} = 3.1 \, \mu m$ displayed $D^{*} = 1.05 \times 10^{9} \, cmHz^{1/2}/W$ at 20 °C.

Ref. [166] describes how an ICL structure processed on the native GaSb substrate can be patterned to form ICLs, ICDs, and other active and passive optical components connected by passive waveguides. It also proposes on-chip sources that do not experience feedback from external optics, for substantial enhancement of the temporal and spectral stability.

4.11. Linewidth and Stability

The sensitivity and resolution of a chemical sensing system based on mid-IR spectroscopy is limited by the photon intensity and frequency stability of the laser source.

The power emitted at a given time: $P(t) = P_{avg} + \delta P(t)$ varies due to random fluctuations of the spontaneous emission, carrier generation, and carrier recombination (as well as imperfect temperature and current stability). We define the frequency-dependent power spectral density (PSD):

$$S_{f}(f) = \int_{0}^{\infty} < \delta P(t) \delta P(t + \tau) > e^{2\pi f \tau} \, d\tau$$

from which we can quantify the relative intensity noise: $RIN = 10 \log(S_{f}/P_{avg}^{2})$. In 2008, Patrick Folkes at ARL measured the first RIN for an ICL, which was operated in cw mode at 30 K and 100 K [226]. Based on comparison with a model developed for QCLs [227], he concluded that thermal fluctuations of the nonradiative recombination dominated the noise above the threshold. The RIN at 100 K decreased non-monotonically with current, from −112 dB/Hz at threshold ($l/th = 1$) to −152 dB/Hz at $l/th = 1.6$ and −134 dB/Hz at $l/th = 2.9$, but was relatively independent of frequency up to the detection limit of 100 kHz.

More recently, the Shanghai Technical University (STU) measured the RIN up to much higher frequencies [228]. The 7-stage ICL narrow ridge ($\lambda = 3.39 \, \mu m$ when operated cw at RT. Following the subtraction of non-laser noise sources, the RIN gradually decreased with increasing frequency above 100 MHz. However, the intrinsic RIN level of the ICL was not reached at the detector bandwidth limit of 450 MHz. At that frequency, the RIN decreased from −115 dB/Hz at $l = 80 \, mA$ (just above the threshold) to −130 dB/Hz at $l = 150 \, mA$. Values for the gain coefficient and transparency carrier density ($N_{tr}$) were obtained by fitting the RIN data to a simple rate equation theory. Because $N_{tr}$ was overestimated by more than an...
order of magnitude compared to the accepted value, it was concluded that the intrinsic RIN was no higher than $-140 \text{ dB/Hz}$.

STU also reported characterization of the linewidth broadening factor ($\alpha$) [229]. This factor quantifies the spectral linewidth broadening (by a factor of $1 + \alpha^2$) compared to the Schawlow-Townes limit, and governs other laser characteristics such as the stability against optical feedback, chirp. Measurements were performed on a 7-stage narrow-ridge ICL provided by U. Oklahoma, which was nominally identical to that whose RIN characteristics were described above. Using the Hakki-Paoli method, a value $1.1-1.4$ was extracted from the ASE spectrum just below the lasing threshold. This is somewhat smaller than the typical values for shorter-wavelength quantum-well diode lasers, but similar to the result $\alpha \approx 1.3$ obtained previously by NRL from Hakki-Paoli characterization of a non-cascade “W” diode laser emitting at $\lambda \approx 3.2 \mu\text{m}$ when operated at 78 K [230]. STU then used the self-mixing interferometry method to determine $\alpha$ above the threshold. The result $\alpha = 2.2$ was nominally independent of the injection current, which was attributed to pinning of the carrier concentration. This contrasts other reports that the carrier concentration in an ICL does not pin, as will be discussed in Section 5.2 below. The higher $\alpha$ value above the threshold was attributed to carrier heating and spectral hole burning, by analogy with strained quantum-well lasers at shorter wavelengths.

In fact, frequency noise has more effect than intensity noise on the sensitivity of chemical sensing based on laser spectroscopy. Quite recently, an Italian team used an ultra-low-noise current driver to characterize the frequency noise power spectral density (FNPSD) of a DFB ICL from Nanoplus ($\lambda = 4.6 \mu\text{m}$) [231]. The results spanning 6 orders of magnitude in frequency (10 Hz–10 MHz) were compared to analogous data for a DFB QCL emitting at $\lambda = 4.3 \mu\text{m}$. While $1/f$ noise dominated the FNPSD for both lasers at frequencies up to $\approx 10 \text{ kHz}$, the ICL dependence began to deviate at 10–100 kHz. They attributed the deviation to thermal effects that also occur in other semiconductor lasers with relatively slow gain dynamics, but not in QCLs. The FNPSD for the ICL flattened above 1 MHz due to white noise, which begins to dominate at a somewhat higher frequency in the QCL, in part owing to its lower $\alpha$ factor. The ICL’s $1/f$ component at low frequencies was found to increase monotonically with both current and temperature, which is expected in an interband laser. On the other hand, the higher-frequency white noise component increased with temperature but decreased with current. The white noise level is proportional to the laser’s intrinsic linewidth, which according to Schawlow-Townes theory is inversely proportional to the optical power. For this ICL the dependence indicated an intrinsic linewidth of $\approx 10 \text{ kHz}$ at the maximum output power of a few mW. This value is higher than for a QCL, but generally lower than for interband lasers emitting at shorter wavelengths. While an earlier frequency-modulation study of DFB ICLs [232] found a much larger intrinsic linewidth ($> 300 \text{ kHz}$), that measurement may have been influenced by noise from the driver current [233].

STU also quite recently reported measurements of the FNPSD for two DFB ICLs, in this case Nanoplus devices emitting at $\lambda \approx 3.39 \mu\text{m}$ [234]. The lasers powered by batteries showed a slight improvement of the FNPSD at frequencies $> 1 \text{ MHz}$. The measured spectral linewidth at room temperature was 284 kHz when observed over 1 ms, which is narrower than the typical results for QCLs, even though this relationship is reversed for intrinsic linewidths. The extrapolated intrinsic linewidth of 12 kHz was similar to the Italian result, and removing the effect of the linewidth enhancement factor, it was as narrow as 1.6 kHz.

Because the noise decreases with increasing frequency, techniques such as wavelength modulation spectroscopy (WMS) and frequency modulation spectroscopy (FMS) can provide much more sensitive chemical detection [232]. When a small sinusoidal modulation is added to the slow current sweep of a tunable laser, the magnitude and shape of the multiple harmonic orders that are produced can provide information about the concentration and temperature of the absorbing species. However, the sensitivity of these techniques is limited by any residual noise in the laser output. In addition to the importance of using an ultra-low-noise current driver, Ref. [231] suggests that the high-frequency
noise can be minimized by locking, with feedback to the drive current. Those authors also recommend a systematic study of carrier density fluctuations and other internal dynamics of the ICL.

5. Mechanisms Limiting Device Performance

5.1. Mechanisms Limiting Threshold Current Density

The carrier lifetime in a semiconductor laser is usually limited by some combination of radiative, Auger, and Shockley-Read recombination processes. In diodes emitting near $\lambda \sim 1 \mu m$, for example, the fundamental radiative recombination mechanism often determines the threshold current density. As the emission moves toward the mid-IR, however, the energy gap narrows, and Auger processes become energetically favored. Calculations indicate that the typical radiative recombination rate in an ICL at threshold is only a few percent of the Auger rate. While the Shockley-Read non-radiative lifetime $\tau_{SR}$ has not been determined for ICL active QWs at RT, the data for related type-II InAs/GaInSb superlattices used in mid-wave and long-wave IR photodetectors [235–237] suggest that $\tau_{SR}$ is most likely about two orders of magnitude longer than the typical Auger lifetime of $\approx 0.5$ ns. Thus, for most practical purposes, Auger recombination is the only relevant mechanism determining the ICL threshold at practical operating temperatures.

Since the Auger rate for non-degenerate statistics scales as the carrier density cubed, it was long considered unlikely that semiconductor lasers emitting at $\lambda > 3 \mu m$ would ever realize RT operation. Fortunately, the Auger coefficient in a type-II QW is lower than the original estimates based on data for bulk materials with similar bandgaps [238]. Furthermore, a population inversion can be reached at threshold carrier densities lower than those in GaAs- and InP-based devices, due to the smaller band-edge densities of states in strained narrow-gap QWs. This subsection will examine what is known about the Auger coefficients in ICL active regions, and how they compare with those in bulk-like materials.

While numerous approaches to characterizing the Auger coefficients $\gamma$ in type-II SLs and QWs have been discussed in the literature [238], the data have shown significant spread, and no systematic agreement with theoretical calculations has been realized [71,72]. Since the RT ICL threshold is strongly dominated by Auger recombination, $\gamma$ can be extracted from the $J_{th}$ data provided the internal loss and internal efficiency for a given device are sufficiently well known, and an accurate model for the modal optical gain is available to allow extraction of the threshold electron and hole densities. An added complication is that, whereas the electron and hole densities are roughly equal in most other interband semiconductor lasers, this assumption does not necessarily apply to the ICL [70]. Nevertheless, the optical gain provided by given electron and hole densities may be estimated using standard techniques starting from the band structure solver [239]. The ICL band structure can be calculated self-consistently by isolating 1.5 periods of the active region [64,70] and assuming that the active hole well and the hole injector of one stage share a common Fermi level with the electron injector of the next stage. This computational approach was used to obtain the results discussed below.

Since the ICL gain is provided by isolated QWs, the Auger decay is most appropriately specified in terms of a 2D rather than 3D Auger coefficient. However, it is convenient to somehow relate the derived 2D Auger coefficients to bulk 3D values so a comparison may be made to bulk-like mid-IR materials [238]. This requires a normalization length, which may be derived using the wavefunctions for the lowest electron and highest heavy-hole subbands [73]. The resulting relationship of 3D and 2D coefficients is:

$$\gamma_{3D} = \gamma_{2D} \frac{\int |\psi_e|^2 dz \int |\psi_{hh}|^2 dz}{|\psi_{el_{max}}|^2 |\psi_{hh_{max}}|^2}$$

If the emission wavelength is varied by adjusting the thicknesses of the active InAs QWs, the normalization length (given by the square root of the 3D/2D coefficient ratio) is nearly fixed at a value close to the width of the GaInSb hole well.
Figure 24 updates the data reported previously [73] by plotting the Auger coefficients derived from pulsed threshold current densities at 300 K for broad area devices processed from 70 different ICL wafers (see Figure 4), along with bulk 3D values for bulk III-V and II-VI materials with energy gaps in the same range. For simplicity, a fixed internal efficiency of 80% was used in analyzing all the data, although it has not been measured for ICLs operating at $\lambda > 4 \, \mu m$. In the 3–4 $\mu m$ spectral window, the $\gamma_{2D}$ values fall in the narrow range $2.5$–$3.5 \times 10^{-15} \, \text{cm}^3/\text{s}$, which nominally corresponds to $\gamma_{3D} = 2$–$3 \times 10^{-28} \, \text{cm}^6/\text{s}$. The Auger decay in type-II ICLs is seen to be suppressed significantly relative to bulk materials with similar bandgaps, and the increase of $\gamma$ with wavelength is also modest. These data show $\gamma_{2D}$ rising very gradually to $\approx 10^{-27} \, \text{cm}^6/\text{s}$ at $\lambda \approx 6 \, \mu m$. Evidence from the carrier-rebalancing studies discussed above suggests that the Auger rates for multi-electron and multi-hole processes are comparable in the 3–4 $\mu m$ range [70].

The Auger rate is usually assumed to scale with the cube of the carrier density (for equal electron and hole densities). Although this is rigorously correct only in the low-density limit, some theoretical work indicates that it may hold reasonably well at the typical threshold carrier densities in an ICL [72]. Proceeding for now with the cubic form and assuming constant internal loss and internal efficiency, the ICL gain calculation may be used to estimate the temperature dependence of $J_{th}$. The results can be fit with reasonable accuracy to the logarithmic-gain model:

$$g_{th}(J_{th}, T) \approx \Gamma_a g_0 \ln \left( \frac{J}{J_{300}} \right) = \Gamma_a g_0 \left[ \ln \left( \frac{J}{J_{300}} \right) - \frac{T - 300}{T_{th}} \right]$$

where $T_{th} = 78 \, \text{K}$, $g_0 = 140 \, \text{cm}^{-1}$ is independent of $T$, $J_{300} = 32/\eta_i$ ($\gamma_{2D}/10^{-15} \, \text{cm}^3/\text{s} \, \text{A/cm}^2$ is the ICL’s “transparency” current density at $T = 300$ K, and $\eta_i$ is the internal efficiency. While the temperature variation of the transparency current density is not necessarily exponential, this approximation is nonetheless useful over a limited temperature range (here, 300–350 K).
The expression indicates that the magnitude of the current density is set by the Auger coefficient, while its temperature dependence is that of the transparency carrier density raised to the third power. Note that in this model, the gain available at a given current density $J$ decreases linearly with operating temperature. Since the internal loss tends to increase and the internal efficiency decreases with increasing temperature, generally $T_{\text{op}} > T_0$. Indeed, we have seen above that the typical $T_0$ values for state-of-the-art ICLs do not exceed $\approx 55$ K. We next consider the mechanisms that govern the internal loss and internal efficiency in an ICL.

Ever since Auger suppression in type-II QWs was predicted [71] and confirmed experimentally [238] in the 1990s, it has been taken for granted that type-II structures generally have smaller Auger coefficients than type-I QWs with the same mid-IR bandgaps. However, when the threshold analysis described above was applied recently to type-I InGaAsSb QW diode lasers grown on GaSb, the typical 2D Auger coefficients near $\lambda \approx 3.0 \, \mu m$ were actually found to be somewhat smaller than in the type-II ICLs [11]. However, because this is mitigated by the population of multiple electron subbands in the wider type-I QW, near $3.0 \, \mu m$ the lasing thresholds for type-I diodes and type-II ICLs are roughly the same. Although Ref. [11] did not explicitly consider type-I ICLs, we expect similar considerations to apply. It was also found that near $\lambda \approx 2.0 \, \mu m$, $\gamma_{2D}$ (and $J_{\text{th}}$) in the type-I QWs increased rapidly with decreasing wavelength due to the onset of resonance between the energy gap and split-off gap, which induced rapid Auger recombination involving two holes and a single electron.

5.2. Mechanisms Limiting Internal Loss and Internal Efficiency

An ICL’s external differential quantum efficiency per stage does not reach 100%, because some injected electrons do not produce additional electron-hole pairs in the active region (i.e., the internal efficiency $\eta_i$ is less than unity), and also because some generated photons are lost to parasitic absorption in the active core or another part of the ICL waveguide before they are emitted from a facet (internal loss $\alpha_i > 0$). The EDQE is proportional to $\eta_i$, while $J_{\text{th}}$ is inversely proportional. The more complicated dependences of these quantities on $\alpha_i$ are generally sublinear, except very close to the maximum operating temperature. A standard approach to characterizing $\eta_i$ and $\alpha_i$ is to measure the external quantum efficiency as a function of cavity length (or facet reflectivity), which varies the mirror loss while assuming that internal loss and internal efficiency do not change. Of course, care must be taken because neither $\alpha_i$ nor $\eta_i$ remains perfectly constant as the cavity length changes, e.g., owing to the dependence of free-carrier absorption on the threshold carrier density. Nevertheless, cavity-length measurements are the most straightforward and widely-used method for determining $\alpha_i$.

NRL used this approach to characterize ICLs from two 7-stage ICL wafers with RT centroid wavelengths of $\lambda \approx 3.11 \, \mu m$ (Wafer A) and $\lambda \approx 3.45 \, \mu m$ (Wafer B) in pulsed mode to minimize lattice heating [87]. Figure 25a plots the pulsed inverse external differential quantum efficiency per stage as a function of cavity length for a broad-area ICL processed from Wafer B operating at $J = 1.2 \, \text{kA/cm}^2$ and three different temperatures (300, 320, and 345 K). In order to analyze the efficiency droop that will be discussed below, the EDQE is extrapolated to $J = 0$. A fit to the linear dependence allows $\eta_i$ to be determined from the intercept and $\alpha_i$ from the slope. The internal efficiency results of $\eta_i = 80, 76, \text{and 68} %$ at $T = 300, 320, \text{and 345 K}$, respectively, were nearly independent of injection current density. The corresponding internal losses, extrapolated to zero current density, were $\alpha_i = 2.9, 3.2, \text{and 3.5 cm}^{-1}$ if $R = 41%$ is assumed for the cleaved uncoated facet, and $\alpha_i = 3.8, 4.2, \text{and 4.6 cm}^{-1}$ for $R = 31%$. While the precise facet reflectivity is unknown, these values represent limits that may depend on imperfections in the cleave quality.
A number of cavity-length studies of ICLs emitting in the $\lambda = 3–4 \mu m$ spectral window have found internal efficiencies of $\approx 80–90\%$ near the lasing threshold, with modest decreases at higher currents and temperatures. Actually, it is unclear why the internal efficiency should depart noticeably from 100%, since carrier escape from the active QWs seems unlikely (due to the large conduction and valence band offsets), while injection directly into the electron injector is blocked by a very thick barrier. Furthermore, for reasonable energy relaxation times, the effect of carrier heating (that reduces $\eta_i$ by requiring a higher carrier density, and hence higher current, to maintain the same optical gain) should also be quite small near the lasing threshold.

It is not understood why the estimated internal loss increases from $\approx 4 \text{ cm}^{-1}$ at $J = 0$ (extrapolated) to $\approx 12 \text{ cm}^{-1}$ at $J = 2.5 \text{ kA/cm}^2$ [87]. This manifests itself as the “efficiency droop” shown in Figure 25b for the devices of Figure 25a with $L_{cav} = 2 \text{ mm}$. Note that between the threshold and $J = 2.8 \text{ kA/cm}^2$, the EDQE decreases by nearly a factor of 3 at all temperatures. However, owing to the large electron density already present in the ICL active core, due to carrier rebalancing, it may result from an increase of the hole density above the threshold (i.e., the density does not pin). This conjecture is also consistent with the observation that the spontaneous emission intensity fails to saturate above the threshold [88], although the linewidth broadening factor data discussed in Section 4.11 above provided evidence for pinning [229]. Non-pinning of the carrier concentration and emission intensity above the threshold is also characteristic of type-I mid-IR QW lasers [11,89]. Further studies are needed to understand the origin of the non-pinning phenomenon, and to develop approaches for combating its deleterious consequences.

While it would be impractical to repeat cavity-length measurements on all ICL designs emitting at various mid-IR wavelengths, the dependence of $\alpha_i$ on $\lambda$ estimated from the pulsed slope efficiency data for each device, assuming fixed $\eta_i = 80\%$ and $R = 41\%$, was shown above in Figure 7. Because the loss displays a minimum around $\lambda = 3.3–3.8 \mu m$, it is not surprising that the highest cw output powers are realized in this region. Since initial cavity-length measurements indicate that the internal efficiency remains high as $\lambda$ approaches 3 $\mu m$, the lower EDQE cannot be attributed to a decrease of $\eta_i$. The pronounced increase of $\alpha_i$ at wavelengths beyond 4 $\mu m$ may indicate a substantial increase of the free-carrier absorption cross-section, possibly due to intervalence transitions.
The efficiency droop can be quantified further by plotting the ratio of pulsed EDQE at \( J = 2 \text{ kA/cm}^2 \) to the extrapolated value at \( J = 0 \). Figure 26 plots this ratio as a function of emission wavelength for a number of ICLs with both 5 and 7 stages. The ratio remains nearly constant at 65–70% for \( \lambda > 3.5 \mu\text{m} \), but exhibits a significant decline at shorter wavelengths, reaching \( \approx 30\% \) at \( \lambda = 3.0 \mu\text{m} \). This is quite surprising, since the ICL design is altered only incrementally as \( \lambda \) decreases. Furthermore, we do not expect such mechanisms as carrier heating [89], which are cited as possible causes of carrier non-pinning (and, potentially, efficiency droop), to change discontinuously with wavelength.

![Figure 26](image-url)  
*Figure 26. Ratio of the external differential quantum efficiency at \( J = 2 \text{ kA/cm}^2 \) to the value extrapolated to \( J = 0 \) for 5-stage and 7-stage ICLs as a function of emission wavelength at room temperature.*

### 5.3. Optimal Number of Stages and Comparison to the Quantum Cascade Laser

The optimal number of stages in an ICL depends on the device operating conditions, as discussed above in Section 3.2. For example, a single-mode DFB laser to be employed for chemical sensing may only need to emit a low cw power on the order of 1 mW (unless the sensing is remote). Then the primary parameter to be optimized is most likely the power dissipated at the operating point, which is typically very close to the lasing threshold. On the other hand, if maximum power or brightness is needed, the laser operates at a current density well above the threshold. The external efficiency may be nearly irrelevant in the former case, while in the latter the threshold power density is of little concern.

The ICL optimization for minimum threshold power density was summarized in Figure 3 of Ref. [73] and the accompanying discussion. The key parameter is the ratio of transparency current density to threshold current density, \( r \). If this ratio is close to unity, it is preferable to employ only a few stages to minimize the bias voltage, since the extra gain needed to reach threshold will not require much additional current density. However, the optimum number of stages increases dramatically when \( r \) is small (as when the active-core optical confinement factor is low or the internal losses are high). This is seen in Figure 27, which plots the optimal number of stages for both ICLs and QCLs as a function of \( r \). Unpublished NRL cavity-length studies indicate that for state-of-the-art ICLs emitting in the \( \lambda = 3–4 \mu\text{m} \) band, \( r \approx \frac{1}{2} \). Referring to the solid blue curve in the figure, this implies that only 2–3 stages may minimize the threshold power density. Since the actual operating power is somewhat higher, and a single-mode DFB structure will not perform quite as well, the actual optimum probably falls between 2–3 and the 5–7 stages employed in most NRL ICLs to date. The threshold power of
an edge-emitting ICL can be reduced further by employing a short optical cavity and depositing a partially-transmitting coating with $R >> 30\%$ on the output facet in conjunction with an HR coating on the back facet, by analogy with a strategy demonstrated for QCLs [240]. This approach may yield ICLs with threshold powers as low as 10 mW.

![Figure 27](image_url)

**Figure 27.** Minimum pulsed room-temperature threshold power densities and the corresponding optimum number of stages for ICLs (solid) and QCLs (dashed) emitting at $\lambda = 4.0 \, \mu m$, plotted vs. the ratio of the transparency to threshold current densities. Experimental threshold power densities for state-of-the-art ICLs and QCLs, extrapolated to the 4.0-µm wavelength, are indicated. Reproduced from Ref. [73], with the permission of IEEE.

More stages are needed when the objective is to maximize the cw output power, brightness, or wall-plug efficiency. Then one must trade off the higher slope efficiency obtained with more stages against the additional lattice heating due to a higher voltage drop. When more stages are added, the internal loss can be minimized by thickening the low-doped SCLs and reducing the active-core confinement factor $\delta_a$. Nevertheless, weaker gain and a higher threshold power density eventually limit this strategy. Preliminary modeling indicates that ICLs optimized for high power and WPE in the 3–4 µm wavelength range should have ≈10 stages, although further experiments are needed to confirm and refine this projection.

While the differential gain per unit carrier density is only somewhat higher in QCLs than in ICLs (by a factor of ≈4), the modal gain per unit current density per stage is nearly two orders of magnitude lower due to the very short (ps-scale) lifetime of the QCL’s upper lasing subband [73]. As a result, the optimal stage multiplicity for QCLs (dashed blue curve in Figure 27) is nearly an order of magnitude higher to ensure that most of the lasing mode resides in the active core. QCLs also require a larger extra voltage drop per stage ($>>k_B T$), in order to minimize backfilling of the lower lasing subband. Typically, the extra margin needed to overcome the internal loss in an ICL does not exceed $k_B T$.

The solid and dashed red curves in Figure 27 represent theoretical minima for the threshold power densities ($P_{th}$) in ICLs and QCLs, respectively. For similar transparency/threshold current density ratios, the values $P_{th} \approx 0.2 \, kW/cm^2$ for ICLs and $\approx 7 \, kW/cm^2$ for QCLs should be achievable. These estimates are consistent with experimental results (green horizontal lines), which are only $\approx 50\%$ higher [14,64]. For similar cavity dimensions and characteristics, both the theoretical and experimental threshold
powers of ICLs are more than an order of magnitude lower than those of QCLs [64]. On the other hand, because of its larger number of stages and far weaker temperature sensitivity for Pth and EDQE, not to mention the much lower thermal resistance of its cladding materials (InP vs. InAs/AlSb SLs), the QCL can generate much higher cw output power from a single narrow ridge [14,241]. Nonetheless, in view of the QCL’s relatively high threshold current density and larger parasitic voltage drop per stage, the maximum cw WPEs appear comparable for the two mid-IR laser classes. The experimental values reported to date for cw operation at \( T = 25 \degree C \) are only slightly higher for the QCL (22%) [15] than the ICL (18%) [109].

5.4. Prospects for Future Improvements

The preceding sections have reviewed the design, operating characteristics, and factors limiting the performance of a variety of mid-IR ICL configurations. In order to summarize our discussion and give a better idea of the current possibilities for the ICL performance, we have constructed a model that accounts for the measured efficiency droop and optical gain as a function of emission wavelength, assuming that the scattering loss added when narrow ridges are etched is limited to 0.5 cm\(^{-1}\). The results of the model are instructive, since narrow-ridge ICLs with good performance have not yet been fabricated from all the wafers grown in this wavelength range. In Figure 28, the projected maximum cw output power (blue) and wall-plug efficiency (red) are presented for state-of-the-art 7-stage ICLs as a function of emission wavelength. For wavelengths between 3.25 and 4.5 \( \mu \text{m} \) (stars), it is also required that the output power exceed 200 mW and the WPE exceed 10%. Outside this range (circles), the absolute maximum values are shown. It may be observed that the best performance is expected in the 3.5–4 \( \mu \text{m} \) window, as should be expected from the preceding discussions. The drop-off at shorter wavelengths results from the efficiency droop phenomenon discussed in the previous section. The degradation at longer wavelengths arises primarily from the higher internal loss characteristic of state-of-the-art devices at those wavelengths. Figure 28 does not account for potential future improvements in the internal loss and efficiency droop, or the possible benefits of employing a different number of stages.

![Figure 28](image-url)

**Figure 28.** Projected maximum room-temperature continuous-wave output powers (blue) and wall-plug efficiencies (red) for 7-stage ICLs with the best state-of-the-art characteristics measured for broad-area pulsed devices. For wavelengths between 3.25 and 4.5 \( \mu \text{m} \) (stars), the maximum output power is presented for WPE > 10%, while the maximum WPE is shown for an output power exceeding 200 mW. Outside this wavelength range (circles), the absolute maximum values are displayed. The ridge width and cavity length were varied to obtain these results, with the maximum cavity length fixed at 4 mm. An HR coating was assumed for the back facet, and a 3% AR coating for the front facet.
Although this review has provided only a snapshot of the ICL development, the basic broad area and narrow ridge devices appear to have reached at least an intermediate level of maturity for the most investigated spectral range of $\lambda = 3.2$–$4.0 \mu m$. However, we do anticipate further opportunities for improvements.

For example, if the goal is either to: (a) minimize threshold drive power or (b) maximize output power and brightness, the optimal number of stages is probably either fewer (a) or greater (b) than what has been employed in most of the ICL structures grown and tested to date. The corresponding designs should feature modified SCL thicknesses that tailor the mode overlap with the active core. While the threshold current density can be reduced nearly to its “transparency” value by employing more stages, or the same number of stages with a higher optical confinement factor, the more important threshold power density for state-of-the-art ICLs is already no more than 50% higher than its nominal theoretical limit of 200 W/cm$^2$. The internal loss is also unlikely to see additional dramatic reduction, since in some devices it has already reached a minimum of several cm$^{-1}$ at RT when extrapolated to zero current density.

However, selected variations on the basic ICL design may still have an appreciable effect on the device performance, once they are attempted and optimized experimentally. For example, simulations suggest that a thinner electron injector, with only moderate $n$-doping, should rebalance the electron/hole density ratio as effectively as the current designs with relatively thick injectors that must be doped quite heavily. In particular, if the origin of the substantial efficiency droop at high injection current densities can be understood and eliminated, the cw output powers for narrow-ridge ICLs may immediately improve by as much as a factor of 2. Unfortunately, our investigations to date have failed to identify any clear correlations of this phenomenon with incremental, or in some cases major, changes to a broad range of active and waveguide design parameters. Similarly, the upturn of the threshold current density in Figure 4 at wavelengths shorter than $\approx 3.2 \mu m$ remains unexplained, since basic theoretical considerations predict that both the Auger coefficient and the internal loss should continue to decrease with decreasing wavelength. Were this trend to be reversed, type-II ICLs may remain advantageous over type-I diode lasers down to wavelengths considerably shorter than 3 $\mu m$.

From a practical standpoint, ICLs are known to be quite robust so long as they are not subjected to excessive voltages or currents. In one accelerated aging study, ICLs that were operated in cw mode at $T = 90 \degree C$ for 4000–10,000 h each did not show any appreciable performance degradation [242]. In another investigation, four DFB ICLs driven with $\geq 200$ mA of cw current at 40 $\degree C$ for $>8000$ total hours showed no measurable degradation [92]. Anecdotally, NRL has found that ICLs stored for several years nearly always operate as they had previously when taken off the shelf and tested. Furthermore, PNNL and NRL found that when ICLs were exposed to a cobalt-60 gamma ray dose of 500 krad, which exceeds that normally encountered in a shielded spacecraft, their performance did not change [243].

Naturally, the more specialized architectures now under development still have considerable headroom for further improvement. We expect promising new applications to result from much more efficient and resonant cavity ICLEDs, ultra-fast ICDs for rf beatnote detection and free-space communications, DFB ICLs with broad tunability and emitting higher power, single-mode ICVCSELs requiring ultra-low drive power (<5 mW), broadband ICL frequency combs for multi-species chemical sensing, mid-IR laser spectroscopy with resolution approaching the intrinsic linewidth, and ICL-based (or multi-spectral) PICs for ultra-compact on-chip sensing and other applications that require low cost and small footprint. In general, the richness of ICL device physics will offer attractive additional possibilities for future breakthroughs.

**Author Contributions:** J.R.M. coordinated the research, performed analysis and most of the literature review; W.W.B. and C.D.M. performed experiments and data analysis; C.L.C. grew ICL materials by MBE; C.S.K. and M.K. processed ICL devices; I.V. performed most of the simulations, analyzed/interpreted data, coordinated the research, and performed portions of the literature review. All authors have read and agreed to the published version of the manuscript.
Funding: This work was supported by the Office of Naval Research.

Acknowledgments: The authors thank Rui Yang, Alexander Spott, Lukasz Sterczewski, and Mahmood Bagheri for critically reviewing the manuscript and offering constructive comments.

Conflicts of Interest: The authors declare no conflict of interest.

References


47. Hill, C.J.; Wong, C.M.; Yang, B.; Yang, R.Q. Type-II interband cascade lasers emitting at wavelength beyond 5.1/spl mu/m. *Electron. Lett.* 2004, 40, 878–879. [CrossRef]


145. Veerabathan, G.K.; Sprengel, S.; Andrejew, A.; Amann, M.-C. Room-temperature vertical-cavity-surface-emitting lasers at 4 \mu m with GaSb-based type-II quantum wells. *Appl. Phys. Lett.* 2017, 110, 071104. [CrossRef]


188. Wassweiler, E.; Toor, F. Gallium antimonide texturing for enhanced light extraction from infrared optoelectronics devices. *AIP Adv.* **2016**, *6*, 065018. [CrossRef]


© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).