Optical refrigeration

The idea of cooling a solid-state optical material by simply shining a laser beam onto it may seem counterintuitive, but this is rapidly becoming a promising technology for future cryocoolers. Here, we chart the evolution of the science of optical refrigeration in rare-earth-doped solids and semiconductors from its origins through to the present day.

MANSOOR SHEIK-BAHAE1* AND RICHARD I. EPSTEIN1,2

The term ‘laser cooling’ is most often associated with cooling dilute gases of atoms or ions to extremely low temperatures by reducing their thermal translational energy. This area of science has progressed immensely, resulting in physics Nobel prizes in 1997 and 2001 (refs 1,2). However, what many people don’t realize is that 46 years before such Doppler cooling of atoms was ever contemplated1, German physicist Peter Pringsheim suggested the possibility of cooling solids by means of fluorescence upconversion1. In the solid phase, atoms do not have relative translational motion — their thermal energy is largely contained in the vibrational modes of the lattice. Laser cooling of solids (or optical refrigeration) is similar to atom cooling: light quanta in the red tail of the absorption spectrum are absorbed from a monochromatic source followed by spontaneous emission of more energetic (blue-shifted) photons. In the case of solids, the extra energy is extracted from lattice phonons, the quanta of vibrational energy in which heat is contained. The removal of these phonons cools the solid.

Laser cooling of solids can be exploited to achieve an all-solid-state cryocooler3–6. The advantages of compactness, the absence of vibrations and moving parts or fluids, high reliability and the ability to operate without cryogens have motivated intensive research. Space-borne infrared sensors will probably be the first beneficiaries, with other applications that require compact cryocooling reaping the benefits as the technology progresses. A study by Ball Aerospace7 shows that in low-power, space-borne operations, ytterbium-based optical refrigeration could outperform conventional thermoelectric and mechanical coolers in the temperature range between 80–170 K.

The process of optical refrigeration can occur only in special high-purity materials that have appropriately spaced energy levels and emit light with high quantum efficiency. So far, optical refrigeration research has been confined to either glasses and crystals doped with rare-earth (RE) elements or direct-bandgap semiconductors such as gallium arsenide. Although laser cooling of RE-doped solids has been successfully demonstrated, observation of net cooling in semiconductors, has so far remained elusive. Figure I schematically depicts the optical refrigeration processes for a two-level system with broad ground- and excited-state manifolds. Photons from a low-entropy light source (that is, a laser) with energy, hν, excite atoms from the top of the ground state to the bottom of the excited state. The excited atoms reach quasi-equilibrium with the lattice by absorbing phonons. Spontaneous emission (fluorescence) follows with a mean photon energy, hνf, that is higher than that of the absorbed photon. This process has also been called anti-Stokes fluorescence. There were concerns that the second law of thermodynamics might be violated until Landau clarified the issue in 1946 by assigning an entropy to the radiation8.

In the above picture, the interaction rate between electrons and phonons within each manifold is assumed to be far faster than the spontaneous-emission rate, which is valid for a broad range of materials and temperatures. The cooling efficiency or fractional cooling energy for each photon absorbed is

\[ \eta = \frac{\lambda - h\nu}{h\nu} = \lambda_f \lambda_i - 1, \]  

where \( \lambda = c/\nu \) is the wavelength of the laser and \( \lambda_f \) is the mean wavelength of spontaneous emission. The invention of the laser prompted several unsuccessful attempts to observe the effect experimentally9–11. In 1995, net laser cooling in solids was first achieved by workers at Los Alamos National Laboratory12 (LANL). Two technical challenges were addressed and overcome in these experiments. The LANL researchers had to have a system in which the vast majority of optical excitations recombine radiatively and where there is a minimal amount of parasitic heating due to unwanted impurities. Both of these critical engineering issues are ignored in the idealized situation described by equation (1), but are key to experimental success.

It is also important that spontaneously emitted photons escape the material without being trapped and re-absorbed, which would effectively inhibit spontaneous emission13,14. This is a critical issue for high-refractive-index semiconductors, where total internal reflection can cause strong radiation trapping. In the absence of radiation trapping, the fraction of atoms that decay to the ground state by the desired radiative process is known as the quantum efficiency, \( \eta_q = W_{\text{rad}}/(W_{\text{rad}} + W_{\text{nr}}) \), where \( W_{\text{rad}} \) and \( W_{\text{nr}} \) are radiative and non-radiative decay rates, respectively. An external quantum efficiency, which includes a fluorescence escape efficiency, \( \eta_x \), is defined by \( \eta_x = \eta_q W_{\text{rad}}/(\eta_q W_{\text{rad}} + W_{\text{nr}}) \). This describes the efficiency by which a photo-excited atom decays into an escaped fluorescence photon. In a similar fashion, an absorption efficiency, \( \eta_{\text{abs}} = \alpha/(\alpha + \alpha_s) \), is defined to account for the fraction of pump-laser photons that are engaged in cooling15. Here \( \alpha \) is the
resonant absorption coefficient and $\alpha_s$ is the unwanted parasitic (background) absorption coefficient. The combination of all these effects defines the cooling efficiency, $\eta_c = \eta_{\text{ref}} \eta_{abs} (\lambda/\lambda_p) - 1$, where the product $\eta_{\text{ref}} \eta_{abs}$ indicates the efficiency of converting an absorbed laser photon to an escaped fluorescence photon. Note that $\eta_{abs}$ is frequency-dependent and falls off rapidly below photon energy $h\nu - k_B T$ where $k_B$ is Boltzmann’s constant and $T$ is the lattice temperature. At very low pump-photon energy, $\eta_{abs}$ will decrease enough to make $\eta_c < 0$ and laser cooling unattainable. The above analysis defines the condition needed for laser cooling\textsuperscript{15}:

$$\eta_{\text{ref}} \eta_{abs} > 1 - \frac{k_B T}{h\nu}.$$  

This relation quantifies the needed efficiencies: cooling a material from room temperature with a nominal energy gap (pump photon) of 1 eV from room temperature demands that $\eta_{\text{ref}} \eta_{abs} > 97\%$. Although suitable lasers were available in the early 1960s, more than three decades of progress in material growth were needed to satisfy this condition.

COOLING RARE-EARTH-DOPED SOLIDS

The advantages of RE-doped solids for laser cooling have been foreseen for decades. Kastler and Yatsiv suggested these materials could be used for optical cooling as long ago as 1950 (ref. 9) and 1961 (ref. 11), respectively. The key optical transitions in RE-doped ions involve 4f electrons that are shielded by the filled 5s and 5p outer shells, which limit interactions with the surrounding lattice. Non-radiative decay due to multi-phonon emission is thus suppressed. Hosts with low phonon energy (such as fluoride crystals and glasses) further diminish non-radiative decay and hence boost quantum efficiency. In 1968, Kushida and Geusic\textsuperscript{10} attempted to cool a Nd\textsuperscript{3+}:YAG crystal with 1.064-nm laser radiation. They reported a reduction of heating, but no cooling; it is unclear whether they observed any anti-Stokes cooling effects. Laser cooling of a solid was first demonstrated experimentally in 1995 with the ytterbium-doped fluorozirconate glass ZBLAN:Yb\textsuperscript{3+} (ref. 12). Laser-induced cooling has since been observed in a range of glasses and crystals doped with Yb\textsuperscript{3+} (ZBLANP (refs 16–19), ZBLAN (refs 20,21), CNBZn (refs 22,23), BIG (refs 23,24), KGd(WO\textsubscript{4})\textsubscript{2} (ref. 22), KY(WO\textsubscript{4})\textsubscript{2} (ref. 22), YAG (ref. 25), Y\textsubscript{2}SiO\textsubscript{5} (ref. 25), KPB\textsubscript{2}Cl\textsubscript{3} (refs 23,26) and BaY\textsubscript{2}F\textsubscript{8} (refs 27,28)).

In 2000, laser cooling in Tm\textsuperscript{3+}-doped ZBLANP was reported at $\lambda = 1.9$ $\mu$m (ref. 29). The significance of this result was twofold. First, it verified the scaling law of equation (1) by demonstrating an enhancement of nearly a factor of two in the cooling efficiency (compared with Yb-doped systems) to a value in agreement with the ratio of the corresponding cooling transition wavelengths. Second, it was the first demonstration of laser cooling in the presence of excited-state absorption. A record cooling power of approximately
73 mW was later reported in this material by using a multipass geometry\(^\text{36}\). More recently, cooling of Er\(^{3+}\)-doped glass (CNBZn) and crystal (KPbCl\(_3\)) has been reported at \(\lambda = 0.870 \, \mu\text{m}\) (ref. 31). It is interesting to note that the cooling transition used in these experiments is between the ground state and the fourth excited state \(\left(4S_{3/2}\right)\) of Er\(^{3+}\), and not the first excited state as illustrated in Fig. 1. Although such high-energy transitions have a lower cooling efficiency (equation (1)), they offer an effectively higher quantum efficiency owing to their low non-radiative decay rates to the ground state. Furthermore, as in the case of Tm\(^{3+}\) (ref. 29), the presence of higher excited states in Er\(^{3+}\) may prove advantageous because the energy transfer up-conversion transitions are endothermic (at the cooling wavelengths of the main transition) and have a high quantum efficiency\(^\text{31,32}\).

The initial proof-of-principle experiments in ZBLANP:Yb\(^{3+}\) achieved only 0.3 K below ambient temperature\(^\text{12}\). More recently, the LANL group has cooled ZBLANP:Yb\(^{3+}\) from room temperature to 208 K (ref. 19). Although progress is being made, optical refrigerators need to be more efficient and operate at lower temperatures to be competitive with other solid-state coolers, such as thermoelectric (Peltier) devices (see Fig. 2). Several studies\(^\text{25,33,34}\) have shown that ytterbium- or thulium-doped solids can potentially provide efficient cooling at temperatures well below 100 K.

There are several factors that limit the cooling of RE-doped solids both in theory and in practice owing to the nature of realizable materials. The most significant factor is the choice of laser-cooling medium. The ideal cooling efficiency (equation (1)) shows that there is an advantage if an optical refrigerator is pumped with lower-energy photons. This fact was part of the motivation for investigating thulium-doped cooling materials, because their ground- and excited-state manifolds are separated by about 0.6 eV, compared with 1.2 eV in ytterbium-doped solids. Weighing against this advantage are two other considerations. First is the choice of the pump laser, with fewer sources readily available near 0.6 eV. Although not a fundamental consideration, it needs to be kept in mind for commercialization plans that can be realized in the near future. A second and more general reason lies in the ratio of radiative to non-radiative relaxation decays. The rate of non-radiative, heat-producing, multi-phonon decay decreases exponentially with the separation between the two manifolds; this is the well-known energy-gap law. In practical terms, this means that because of the relatively large energy of the excited level in ytterbium-doped materials, non-radiative decays do not significantly decrease the quantum efficiency. However, for a pure thulium-doped material, non-radiative decays can overwhelm the anti-Stokes cooling, depending on the properties of the host material. For materials with low maximum phonon energies, such as ZBLANP (and other fluoride hosts), the non-radiative decays are relatively slow, whereas for many thulium-doped oxide crystals and glasses rapid non-radiative decays prevent laser cooling.

Another consideration in the choice of cooling medium is the width of the ground-state manifold. According to the Boltzmann distribution, the lower energy levels in the manifold are more populated than the higher ones. As the temperature falls and \(k_B T\) becomes small compared with the energy width of the ground-state manifold, the upper levels become depopulated, leading to a decrease in the material opacity at lower frequencies. The net effect is that at low temperatures the cooling material...
cannot be pumped at frequencies below $\nu_t$ and cooling becomes impossible, see equation (1). The width of the ground-state manifold is a result of crystal field splitting and depends on both the dopant ion and the host material. By choosing ions and a host that give narrow ground-state manifolds, the material can cool to lower temperatures before the low-frequency transparency condition sets in.

For the material systems studied so far, the limitation on achieved cooling is not due to the reasons outlined above, but is more likely to be the result of parasitic heating from unwanted impurities in the bulk of the cooling material or on its surface. As can be seen in Fig. 2b, the cooling efficiencies of ZBLANP:Yb$^{3+}$ available at present are far below that for an ideal material with no parasitic heating. One important source of heating in this material is quenching of excited ytterbium ions by impurities, such as iron and copper. The radiative decay time of an excited Yb$^{3+}$ ion is about 1 ms. During this time, the excitation migrates through the glass by transferring energy to neighbouring ions. If the excitation encounters an impurity atom, the energy can be transferred to this atom and be rapidly converted into heat. A detailed study by Hehlen and co-workers$^{14}$ found that close to ideal cooling efficiency could be achieved when the concentration of impurities is 0.01 p.p.m. for Cu$^{2+}$ and 0.1 p.p.m. for Fe$^{2+}$.

An additional source of parasitic heating is absorption in the mirrors that trap the pump radiation in the cooling element. In the LANL experiments the cooling glass has a pair of high-reflectivity mirrors deposited on two surfaces, as indicated in Fig. 3. Pump light is reflected many times by each mirror, so that even relatively low absorption of 0.0001 per reflection produces significant heating. Depositing a higher-quality dielectric mirror may obviate this problem. An alternative approach is to use an optical-cavity configuration to enhance the pump absorption. Both intra-laser-cavity$^{21}$ and external-resonant-cavity$^{35}$ geometries have been demonstrated. The external-resonant-cavity approach used at the University of New Mexico has been shown to be capable of achieving pump absorption exceeding 90% (ref. 35). It has also been proposed that photon localization in nanocrystalline powders can be exploited to enhance laser-pump absorption in the cooling of RE-doped systems$^{36}$.

**WILL SEMICONDUCTORS COOL?**

In addition to the studies of optical refrigeration in RE-doped materials, researchers have examined other condensed-matter systems including semiconductors$^{37-40}$. Semiconductor coolers provide more efficient pump-light absorption, the potential for

---

**Figure 5** The external quantum efficiency. a. The GaAs/GaInP double heterostructure is bonded to a nearly index-matched dome (ZnS or ZnSe) to enhance its luminescence extraction. b. External quantum efficiency measured at room temperature in a bonded sample of the heterostructure in a for various GaAs thicknesses ($d$), with the GaInP layer thickness fixed at 750 nm. c. Measured $\eta_{\text{ext}}$ versus lattice temperature for a GaAs thickness of 1 µm. d. Fractional heating of 1-µm-thick GaAs sample as a function of excitation wavelength for a fixed (optimum) electron–hole density at a starting temperature of 100 K. The extrapolation of short-wavelength data determines the zero-crossing wavelength from which a record $\eta_{\text{ext}}$ of about 99% is deduced. Excitation at longer wavelengths causes heating due to background parasitic absorption$^{15,51}$ as indicated by the solid line calculated assuming $\alpha_e = 10$ cm$^{-1}$. b, c, d Reprinted with permission from ref. 15, copyright (2006) SPIE, and ref. 51.
Semiconductors should achieve a higher cooling power density (rate of heat removal) compared with RE-materials. The maximum cooling power density is about \( N \frac{k_B T}{r_e} \), where \( N \) is the photo-excited electron (hole) density and \( r_e \) is the radiative recombination time. The excitation density does not provide an advantage; in fact, optimal \( N \) is limited owing to many-body processes, and does not exceed that of moderately doped RE systems. However, we can gain five to six orders of magnitude in cooling power density as the radiative recombination rates in semiconductors are much faster than in RE ions.

There have been a number of theoretical investigations as well as a few experimental attempts to achieve laser cooling with semiconductors. A feasibility study by the authors of this review outlined the conditions for net cooling based on fundamental material properties and light management. Researchers at the University of Arizona studied luminescence up-conversion in the presence of partially ionized excitons, and developed a microscopic theory for laser cooling that provides valuable predictive capabilities for low-temperature operations. The possible enhancements of laser cooling by including the effects of the photon density of states as well as luminescence coupling schemes based on surface plasmon polaritons were recently introduced by J. B. Khurgin at Johns Hopkins University. The first thorough experimental effort was reported by the University of Colorado. No net cooling was achieved, despite realization of an impressive external quantum efficiency of 96% in GaAs. These experiments used a high-quality heterostructure in optical contact with a ZnSe dome structure for enhanced luminescence extraction. A report of local cooling in AlGaAs quantum wells by a European consortium was later attributed to misinterpretation of spectra caused by Coulomb screening of the excitons.

Figure 4b shows anti-Stokes luminescence in a GaAs heterostructure, where excitation at \( \lambda = 890 \text{ nm} \) produces broadband luminescence with a mean wavelength of \( \lambda = 860 \text{ nm} \). Each luminescent photon carries away about 40 meV more energy than an absorbed photon adds, so cooling might be expected. Why then have we not been able to observe laser cooling in this material or any semiconductor? To answer this question we have to go back to the cooling condition of equation (2), where \( \eta_{\text{ext}} \) is modified to include the presence of many-body interactions, yielding an expression dependent on \( N \):

\[
\eta_{\text{ext}} = \frac{\eta_B N^2}{A N + \eta_B N^2 + C N^3} \tag{3}
\]

where \( A, B \) and \( C \) are decay coefficients. The radiative recombination rate \( BN^2 \) indicates the bimolecular nature of electron–hole recombination (again, a consequence of ‘indistinguishability’). The non-radiative recombination rate is determined by \( A \) and is dominated by the surface recombination in a high-purity semiconductor. At high \( N \), luminescence efficiency degrades owing to the three-body Auger recombination, determined by the coefficient \( C \). It follows from equation (3) that there is an optimal carrier density that maximizes the external quantum efficiency, a condition not present in RE-materials. A further complication is the poor luminescence escape efficiency (\( \eta_e \)) in high-index semiconductors, such as GaAs. Total internal reflection causes the majority of the spontaneous emission to get trapped and re-absorbed. A similar problem limits the efficiency of LEDs. For a bulk pure semiconductor, the coefficients \( B \) and \( C \) are fixed, fundamental properties. Researchers can manipulate \( \eta_e \) and \( A \) by fabrication and growth procedures to satisfy the cooling condition. Highly controlled epitaxial growth techniques, such as metal–organic chemical vapour deposition (MOCVD) can produce very low surface recombination rates \( (A < 1.5 \times 10^{1} \text{ s}^{-1}) \). This involves a double heterostructure of GaAs/GaInP as shown in Fig. 5a, where the lattice-matched cladding layers provide surface passivation as well as carrier confinement. To deal with extraction efficiency, geometric coupling schemes (such as nearly index-matched dome lenses, as shown also in Fig. 5a) have been used to enhance \( \eta_e \) to 15–20%. This has resulted in a measured \( \eta_{\text{ext}} \) of 97% at room temperature, which is not quite sufficient for net cooling. The optimum GaAs thickness is found to be much lower than the \( 2 \mu \text{m} \) common in planar diode lasers.
about 1 μm, determined by excessive luminescence re-absorption for thicker layers and surface recombination for thinner layers89, as shown in Fig. 5b. Improvements in luminescence extraction as well as surface passivation are needed to cool semiconductors from room temperature.

The temperature dependence of \( \eta_{\text{ext}} \) is an important factor. It is fortuitous that \( B \) increases while \( A \) and \( C \) decrease as the temperature drops. Low temperatures thus shift the external quantum efficiency closer to unity, according to equation (3). A record efficiency of 99% has been observed in a GaAs/GaInP double heterostructure at \( T = 100 \, \text{K} \) (Fig. 5c,d and ref. 51). The experiment involved measuring the temperature of the heterostructure that was van der Waals bonded to a ZnS dome as the Ti:sapphire pump laser was tuned around the mean luminescence wavelength. The excitation density was kept constant by monitoring the total luminescence density.

InGaAs/GaAs heterostructures are being explored to investigate cooling, based on the potential of the material for cooling RE-based materials is fast approaching cryogenic operation. In semiconductors, much promising technology. Cooling of RE-based materials is fast integrated with a high-quality GaAs heterostructure to investigate cooling. Recently, we have fabricated GaAs nanogap structures that will be integrated with a high-quality GaAs heterostructure to investigate cooling.

Optical refrigeration clearly have the potential to make a significant impact on electronics and photonics in the future.

**Acknowledgements**

This work has been supported by the Air Force Office of Scientific Research (MURI program), the National Aeronautics and Space Administration (NASA) and the US Department of Energy. The authors thank M. P. Hasselbeck and D. Seletskiy for reading the manuscript.