Vol. 2 • No. 9 • September • 2014

www.advopticalmat.de

ADVANCED OPTICAL MATERIALS

WILEY-VCH

ADVANCED OPTICAL MATERIALS
www.advopticalmat.de

Toward Smart and Ultra-efficient Solid-State Lighting

Jeffrey Y. Tsao, Mary H. Crawford, Michael E. Coltrin, Arthur J. Fischer, Daniel D. Koleske, Ganapathi S. Subramania, G. T. Wang, Jonathan J. Wierer, and Robert F. Karlicek Jr.*

Solid-state lighting has made tremendous progress this past decade, with the potential to make much more progress over the coming decade. In this article, the current status of solid-state lighting relative to its ultimate potential to be "smart" and ultra-efficient is reviewed. Smart, ultra-efficient solid-state lighting would enable both very high "effective" efficiencies and potentially large increases in human performance. To achieve ultra-efficiency, phosphors must give way to multi-color semiconductor electroluminescence: some of the technological challenges associated with such electroluminescence at the semiconductor level are reviewed. To achieve smartness, additional characteristics such as control of light flux and spectra in time and space will be important: some of the technological challenges associated with achieving these characteristics at the lamp level are also reviewed. It is important to emphasise that smart and ultra-efficient are not either/or, and few compromises need to be made between them. The ultimate route to ultra-efficiency brings with it the potential for smartness, the ultimate route to smartness brings with it the potential for ultra-efficiency, and the long-term ultimate route to both might well be color-mixed RYGB lasers.

1. A Brief History of SSL

We start, in Section 1, with a brief history of solid-state lighting (SSL):[1,2] key materials and device breakthroughs (illustrated in **Figure 1**);[3] the current state-of-the-art device and lamp architectures that those breakthroughs enabled, and; the current dominant system applications that those device and lamp architectures have enabled.

1.1. Stepping Stones: Red and Blue LEDs

Semiconductor electroluminescence was first reported by H. J. Round in 1907, and the first light-emitting diode (LED) was reported by O.V. Losev in 1927.[4,5] Not until the birth of semiconductor physics in the 1940s and 1950s, however, was scientific development of technologies for light emission possible.

For solid-state lighting (the use of semiconductor electroluminescence to produce visible light for illumination), the seminal advances were: first, the demonstration of red light emission by N. Holonyak in 1962;[6] and, second, the demonstration of a bright blue LED by S. Nakamura in 1993,[7] along with earlier materials advances by I. Akasaki and H. Amano.[8,9] Here in Section 1.1, we briefly discuss these two advances and their subsequent evolution.

1.1.1. Red LEDs: Ever-Increasing Efficiencies and Powers

As mentioned above, the first seminal advance in visible light emission was in the red, and this is the LED color that dominated the early history of LEDs. The first commercial LED lamps were introduced in 1968: indicator lamps by Monsanto and electronic displays by Hewlett–Packard. The initial performance of these products was poor, around 1 mlm at 20 mA, in part because the only color available was deep red, where the human eye is relatively insensitive. Since then, steady, even spectacular, progress has been made in efficiency, lumens per package, and cost per lumen.

As illustrated in Figure 1 (top panel), progress in efficiency was largely an outcome of the exploration and development of new semiconductor materials: first GaP and GaAsP, then AlGaAs, then, finally, AlInGaP. Luminous efficacies improved by more than three orders of magnitude: from about 0.02 lm/W in the 1970s from GaP and GaAsP LEDs; to 10 lm/W in 1990 from AlGaAs LEDs (for the first time exceeding that of equivalent red filtered incandescent lamps); to the current state-of-the-art of >150 lm/W from AlInGaP LEDs.[10]

Also, as illustrated in Figure 1 (bottom Haitz' Law panel), progress in efficiency (as well as progress in high-power packaging) then enabled tremendous progress in lumens per package and cost per lumen. In 1968, red LEDs were viewable only if competing with dim indoor lights; by 1985, they were viewable in bright ambient light, even in sunlight. Nevertheless, red LEDs at that time were still limited to small-signal indicator and display applications requiring less than 100 mlm per indicator function or display pixel. Then, around 1985, red LEDs stepped beyond those small-signal applications and entered the medium-flux power signaling market with flux requirements of 1–100 lm, beginning with the newly required center high-mount stop light (CHMSL) in automobiles. At this point

J. Y. Tsao, M. H. Crawford, M. E. Coltrin, A. J. Fischer, D. D. Koleske, G. S. Subramania, G. T. Wang, J. J. Wierer
Energy Frontier Research Center
for Solid-State Lighting Science
Sandia National Laboratories
Albuquerque, NM 87185
E-mail: jytsao@sandia.gov
Prof. R. F. Karlicek Jr.
Smart Lighting Engineering Research Center
Department of Electrical
Computer and Systems Engineering
Rensselaer Polytechnic Institute

DOI: 10.1002/adom.201400131

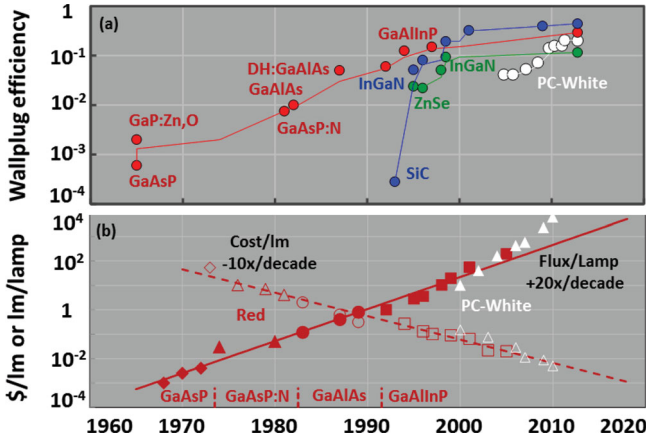

Figure 1. a) Historical evolution of the performance (lm/W) of commercial red, green, blue and phosphor-converted white LEDs. Updated from original data from M. G. Craford, Philips Lumileds. b) Historical evolution of the performance (lm/package) and cost ($/lm) for commercially available red and phosphor-converted (PC) white LEDs. Reproduced with permission.[2] Copyright 2011, Wiley.

in time, red LEDs are well into the >100 lm high-flux domain associated with lighting-class applications.

Of course, it was not just that increasingly higher efficiency enabled these increasingly higher flux applications; the needs of these higher flux applications also drove the quest for higher efficiency. In other words, there was a co-evolution of higher efficiency (technology push) and of power-signaling applications (market pull) that could make use of higher efficiency. Solutions based on large numbers of small-signal lamps were too expensive, thus demanding the development of higher-efficiency, higher-power LEDs. The development of higher-efficiency, higher-power LEDs, in turn, opened up additional stepping stone markets. The result is the Haitz' Law evolution illustrated in the bottom panel of Figure 1. In a Moore's-Law-like fashion, flux per lamp has been increasing 20x per decade while cost per lumen (the price charged by LED suppliers to original equipment manufacturers, or OEMs) has been decreasing 10× per decade.

1.1.2. Blue LEDs: Enabling White Light

As mentioned above, the second seminal advance in visible LEDs was the blue LED, and this is the color that came to dominate the subsequent history of LEDs. The initial breakthroughs came in the late 1980s and early 1990s, with the discoveries by I. Akasaki and H. Amano that a previously recalcitrant wide-bandgap semiconductor, GaN, could be p-type doped[9] and could be grown with reasonable quality on lattice-mismatched sapphire.[8] Building on these discoveries, in 1993 S. Nakamura at Nichia Chemical Corporation demonstrated a bright blue LED.[7] As illustrated in the top panel of Figure 1, efficiency improvements followed quickly, to the point where today's state-of-the-art blue LEDs, at least at low input power densities, have power-conversion efficiencies exceeding 80%.[11]

Most importantly, because blue is at the short-wavelength (high-energy) end of the visible spectrum, it proved possible to "downconvert" blue light into green, yellow and even red light using passive phosphorescent and fluorescent materials.[12] The visible spectrum could thus be filled out, white light could be produced, and general illumination applications became a possibility. Indeed, as illustrated in the bottom panel of Figure 1, Haitz' Law, developed originally for red LEDs, is continuing for white LEDs. There is now little doubt that solid-state lighting will eventually displace all conventional technologies in general illumination applications, and indeed in virtually every application in which visible light is needed.[13]

1.2. State-of-the-Art SSL Device Architecture: InGaN Blue LED + Green/Red Phosphors

At this point in time, the state-of-the-art SSL architecture is based on blue LEDs combined with green, yellow and/or red phosphors, the so-called PC-LED (phosphor-converted LED) architecture illustrated in **Figure 2**. As indicated, the sub-efficiencies of this PC-LED (at the higher current densities most desired for economical SSL, much better results have been reported at lower current densities) are approximately: blue LED (50%), phosphor+package (70%), and spectral match to the human eye response (85%). Taken together, the overall wall plug efficiency is ca. 30% (ca. $0.5 \times 0.7 \times 0.85$).

The reasons this architecture has prevailed, as opposed to a color-mixing architecture in which light from multiple LEDs with different colors is mixed (and which would potentially eliminate the inefficiencies stemming from the phosphor and spectral mismatch), are threefold.

First, improvements in the efficiency of direct electroluminescence have been uneven across the visible spectrum.[14] As illustrated in **Figure 3**,[15] the wallplug efficiencies of blue and red LEDs at the wavelengths (460 and 614 nm, respectively) desirable for general illumination are now over 50% and 30%, respectively. However, the efficiency of green and yellow LEDs at the wavelengths (535 and 573 nm, respectively) desirable for general illumination are of the order 20% or less. Thus, at least for green and yellow light, it is more efficient to produce these colors from blue-LED-pumped phosphors than directly from LEDs.

Second, color stability has proved to be important. The human visual system is extremely sensitive to the exact chromaticity of white light. Because LEDs of different colors have different operating-temperature-dependent lumen outputs, multiple independently controlled feedback loops and current drivers are necessary to achieve an operating-temperature-independent color point. In contrast, blue LEDs and conversion phosphors are far less temperature dependent, allowing open loop (no feedback) operation for many lighting applications.

Third, high-temperature operation has also proved to be quite important, enabling devices to be driven at higher input powers to achieve higher output powers. However, the efficiencies of green and yellow LEDs decrease much faster than those of blue LEDs (and of phosphor-converted green and yellow LEDs) with increasing temperature.[16]

Note, however, that there is now emerging a hybrid architecture, in which blue and red LEDs are combined with a blue-pumped green-and-yellow-light-emitting phosphor.[17] We

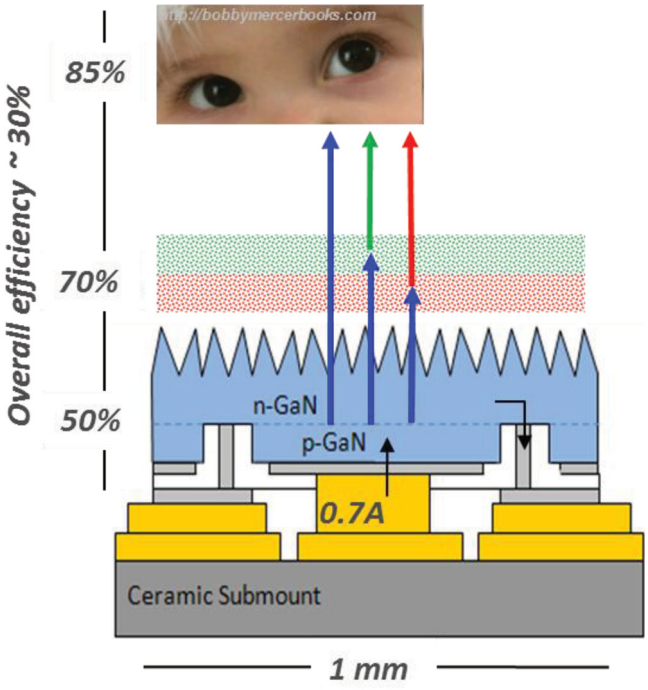

Figure 2. State-of-the-art PC-LED (phosphor-converted white LED). The blue LED is a thin-film flip-chip (TFFC) design, on top of which red and green phosphors have been coated. The TFFC schematic is reproduced with permission. Copyright 2014, Jon Wierer (Sandia National Laboratories); the photo at the top is reproduced with permission. Copyright 2009, Bobby Mercer.

discuss this architecture in Section 4.1, and consider it to be an important step beyond the current state-of-the-art and towards the smart, ultra-efficient solid-state lighting that is the central theme of this article.

1.3. State-of-the-Art SSL Lamp Architectures

Based on the PC-LED (blue LED and green+yellow+red phosphor) device architecture just discussed in Section 1.2, there are two broad classes of lamp architectures, depending on whether the white light originates from a point source or a distributed source. There are also intermediate architectures, but for our purpose here we focus on these two extremes.

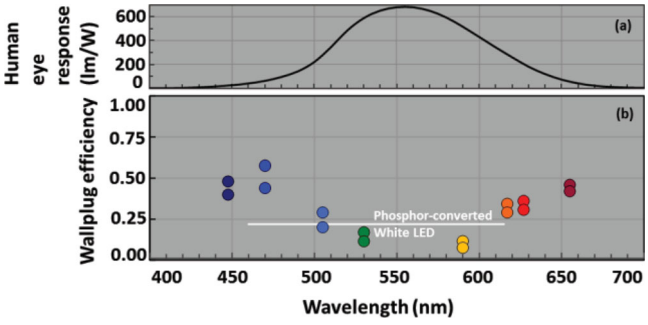

Figure 3. a) Human photopic eye response. b) Wallplug efficiency of current state-of-the-art commercial LEDs of different colors operated at room temperature and input power densities of 100–250 W/cm^2. Also shown as a horizontal white line is the wall plug efficiency of current state-of-the-art commercial warm white LEDs, similarly operated at input power densities of 100–250 W/cm^2.

1.3.1. Point-Source Lamps

In a point-source lamp, the white light originates from a very small area and hence can be relatively precisely imaged and directed with downstream optics. A simple example of such a lamp is one in which phosphors are "locally" placed directly on a blue LED with a small (say, 1 mm^2) footprint; a more complex example of such a lamp would be one in which a blue laser is focused onto a phosphor. In both examples, some of the blue light is scattered and some is phosphor-converted and re-emitted as green, yellow and/or red light, resulting in white light originating from a near-point source with relatively uniform chromaticity over its angular divergence.

Such point-source lamps are, however, difficult to achieve when "useful" amounts of light for general illumination are desired. Since all the blue light must come from a single small-area LED or laser; achieving useful amounts of light then requires high input power densities. If the LED or LED/laser-excited phosphor (particularly if a high-color-rendering-quality phosphor blend) has only moderate efficiency, heat management can be difficult and temperature rise substantial; high temperatures decrease the efficiency and reliability of both LEDs and lasers as well as of the proximal phosphors.[18]

But when such point source lamps can be achieved[19] they are the most flexible and desirable. As illustrated in the left panel of **Figure 4**, they have the flexibility, using downstream optics, to serve non-diffuse lighting applications requiring focused or directed beams. Flashlights (among the earliest of the "power" LED applications) and automotive headlights[20] (among the most recent) are good examples. As illustrated in the right panel of Figure 4, they also have the flexibility, using waveguides and light pipes, to serve diffuse lighting applications. For example, point-source lamps may be used to illuminate a waveguide with surface light-extraction features, thereby effectively creating distributed light sources. This is very similar, at least in concept, to how LEDs are used in edge-lit back lighting for liquid-crystal displays (LCDs) for televisions, computer monitors and phones/tablets. Variations on this theme can enable luminaires with the look and feel of fluorescent tube lighting. Indeed, the design flexibility associated with edge-lit waveguides can, through complex curved shapes and blends of direct and indirect illumination, enable luminaires that go far beyond the look and feel of fluorescent tube lighting.

1.3.2. Distributed-Source Lamps

In a distributed-source lamp, the white light originates from a distributed area, and hence cannot be precisely imaged or directed with downstream optics. One example of such a distributed-source lamp is an array of small point-source lamps distributed over a large area and covered with a diffuser. Another increasingly common example of such a distributed-source

lamp is one in which phosphors are "remotely" placed away from the blue LED. In Edison-socket-replacement lamps, the phosphors can coat the surface of the bulb and be excited by light from one or more blue LEDs. The blue light is scattered as well as phosphor-converted and re-emitted as green, yellow and/or red light over the distributed surface of the bulb.

Distributed-source lamps are easier to achieve than point-source lamps when "useful" amounts of light for general illumination are desired. In the remote-phosphor example, the blue light can come from multiple LEDs driven at lower input power densities with decreased heat management challenges. The phosphors are not exposed to high blue-light fluxes, and can themselves be cooled by conduction over a larger surface area.

However, distributed-source lamps do have significant disadvantages. In the remote-phosphor example, the lamp requires larger quantities of phosphor, and typically exhibit a less desirable (for some applications) non-white off-state appearance (typically yellow or orange). Perhaps most importantly, they are not compatible with non-diffuse lighting applications that require focused or directed beams. Hence, as discussed above, point-source lamps are generally more desirable: they have the flexibility to serve, using downstream optics, both non-diffuse lighting applications requiring focused or directed beams *and*, using waveguides and light pipes, diffuse lighting applications.

1.4. SSL Applications

Beginning with the humble, small-signal indicator and the display applications of early hundredth-lumen-per-lamp red LEDs, applications for the current generation of hundred-lumens-per-lamp white LEDs have grown enormously. Here, we mention two of the most important: mature applications associated with displays; and a just now emerging first wave of rapidly growing applications targeted at retrofitting general illumination lamps (both dumb and "rudimentarily smart").

1.4.1. Displays and Display Backlighting

Displays are by far the largest present-day use for LEDs: either white or red-green-blue (RGB) LEDs for backlighting liquid crystal displays (LCDs) or RGB LEDs for ultra-large video displays.

The first of these uses, backlighting, was driven initially by mobile LCDs for which small form factor was important, but now even larger LCDs for computer displays and televisions have shifted to LED backlighting. Convolved with the shift from vacuum-tube displays to LCDs, LCD backlighting has become the single largest application for LEDs. For most of this backlighting, color quality needs are high but not ultra-high, so white point-source LEDs are used. But for some backlighting, color quality needs (such as more-saturated colors and a larger color gamut) are ultra-high, and RGB point-source LEDs are used, as well as quantum-dot nanophosphors with narrow emission bands enabling wider color gamuts.[21,22]

In the long run, however, the future of LED backlighting is not clear, because the future of LCDs is not clear. Organic LED (OLED) displays have made tremendous progress in recent years, and they promise many performance advantages (higher efficiency, larger field of view, higher switching speed, compatibility with non-planar form factors) over LCDs. For small mobile displays, for which cost and reliability are less important, OLEDs are already displacing LCDs; for larger displays they may also, provided their costs continue to decrease.

The second of these uses, ultra-large displays (e.g., LED video billboards), has grown less spectacularly than has LCD backlighting, but nonetheless steadily. The reasons are two-fold. First, for ultra-large displays *every* light-emitting technology is too expensive per unit area to use directly, but must be restricted to small areas with larger pixel-sized tiles that are end-stacked into an ultra-large-scale geometric arrangement. Second, because LEDs have such high efficiencies and high power densities, they have lower cost per lumen than any other light-emitting technology, and hence are the most cost-effective choice for such light-emitting tiles. In other words, "LEDs have the lowest cost for the empty space between the pixels."[23]

In the long run, because of the above economics, the future of large LED video displays, unlike that for LED backlighting, is virtually assured. Moreover, these applications are likely to continue to grow steadily. Vastly more large outdoor displays (along streets and on building exteriors) *and* large indoor displays (inside large spaces in buildings) may be needed as humanity continues to urbanize and advance its standard of living.[24]

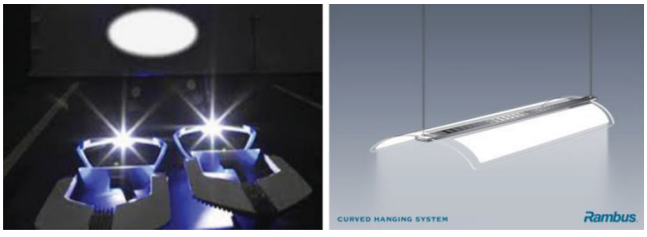

Figure 4. Point-source lamps can be used to provide non-diffuse, directed beams, such as in the laser headlights on the left (reproduced with permission from Frank Wienstroth, BMW Group, 2012), or to provide diffuse, non-directed light, such as in the LED-coupled waveguide luminaire on the right. Reproduced with permission. Copyright 2012, John K. Langevin, Rambus, Inc.

1.4.2. Retrofit "First-Wave" Lighting: Dumb and Rudimentarily Smart

The most rapidly growing use for LEDs is retrofit lighting for general illumination. Although still in its infancy, past and projected future progress is such that it is now taken for granted that SSL will ultimately replace virtually all conventional (including incandescent and fluorescent) lighting technologies. Indeed, a first wave of retrofit lighting is already in motion.

Initially, this first wave will be "dumb" retrofit SSL lamps which fit into existing Edison sockets (for incandescent bulbs) or troffers (for fluorescent lamps), and which offer simple but important performance improvements in: efficiency; color rendering quality; lifetime; absence of environmental contaminants (e.g., mercury); and overall life ownership cost. Because

of the easy tailoring of these various performance attributes, solid-state lighting will eventually serve virtually any current general illumination market better than traditional technologies could.

The transition, though, will not be overnight, due to a required co-evolution of: technical and cost innovation (particularly in the lamp packaging, requiring power converters, drivers, secondary optics, sockets and heat sinks); markets and business models for lighting companies; and consumer acceptance of lighting designs that are slightly dissimilar from traditional designs. Moreover, the transition is being delayed somewhat due to the parallel transition from extremely inefficient incandescent bulbs to compact fluorescent lamps (CFLs), which until recently have beaten SSL in efficiency and are still ahead in price/lm. CFLs have overcome their initial shortcomings (too bluish and too expensive) in the 1990s and now compete well with incandescent bulbs. By 2020, however, LED consumer prices will likely be around $1–3/klm and LED efficiencies will likely be decisively higher than those of CFL lamps. CFL lamps will then suffer the same extinction as the Edison lamp, marking the beginning of the end for traditional lamps in all applications.

The second stage of this first wave will be to add rudimentary "smarts" to retrofit SSL lamps. Mainly, these capabilities will center around occupancy and daylight sensing, networked intelligence, and on/off/dimming control. These concepts in lighting are, of course, not new, but their implementation will be more pervasive than in the past. First, SSL is based on semiconductor components and hence, unlike high-intensity discharge (HID) or fluorescent lamps, is much more compatible with electronic switching circuitry. Second, the separate advance of wirelessly networked sensors, intelligence and mobile devices has in parallel created the necessary control infrastructure. Indeed, there is a tremendous amount of activity in the hardware and software infrastructure for these kinds of control: drivers based on embedded microprocessors and pulse-width modulation (PWM), wireless protocols (Zigbee integration), and distributed sensor networks. These sensors, wireless controls and protocols can easily be integrated into the LED driver electronics to enable full on/off/dimming control that responds to occupancy, daylight levels, and perhaps even user preferences. Note, though, that network latencies due to limited data handling capability of some current standards may need to be addressed.

2. Beyond the State-of-the-Art: Smart and Ultra-Efficient SSL

In Section 1, we gave a brief history of solid-state lighting, up to the current state-of-the-art device and lamp architectures, and of the system applications these device and lamp architectures have enabled. If the evolution of solid-state lighting were to end at only incremental advances beyond this state-of-the-art, and at the "first" wave of dumb and rudimentarily smart retrofit lighting, then a huge revolution in lighting would have already been accomplished.

Here in Section 2, we discuss the possibility that the evolution of solid-state lighting does not end with just incremental advances, but progresses well beyond, into a performance domain that might be called both smart and ultra-efficient (>70%). We first discuss the general characteristics of smart, ultra-efficient solid-state lighting; second discuss the potential systems applications of such lighting; and third discuss the macroeconomic and human productivity benefits of such lighting.

2.1. Characteristics: Multicolor Electroluminescence, Narrowband Spectra, High Modulation Speed

We first discuss some of the likely characteristics of smart, ultra-efficient SSL. We believe there are three such characteristics; that these three are mutually compatible and complementary; and thus that smart and ultra-efficient are themselves mutually compatible and complementary.

2.1.1. Multicolor Electroluminescence

One likely characteristic of smart, ultra efficient SSL is that, unlike in the current dominant paradigm, it does not make use of phosphor-based wavelength downconversion. This is so for two reasons.

First, because about 80% of white light power is green or red, and because the Stokes deficit on converting from blue to green and red is about 25%, SSL in which wavelength downconversion is used to produce green and red is automatically at most $0.2 + 0.8 \times 0.75 = 80\%$ efficient,[25] and likely even less due to other loss factors. So the first likely characteristic of ultra-efficient SSL is that it make use of efficient electroluminescence, rather than wavelength downconversion, across the visible spectrum. Of course, as illustrated in Figure 3, this will be challenging due to the well-known green-yellow gap in the electroluminescence efficiency of compound semiconductors.[26] However, it will be necessary to achieve ultra-efficient SSL.

Second, wavelength downconversion does not lend itself easily to simultaneous attainment of two desirable attributes of smart lighting: real-time tuning of spectra to enable tailoring of white (and perhaps non-white) light chromaticity, and real-time tuning of beam directionality and placement. In a point-source architecture, downstream optics can be used for beam directionality and placement, but the green-to-blue and red-to-blue power ratios, and hence chromaticities, in this architecture are set at the factory rather than tailorable in the field (although there is the possibility of circumventing this limitation using quantum dots as wavelength downconverters.[27] In a distributed-source architecture, multiple laterally placed and independently electrically driven monochromatic (for example, red, green, yellow, and blue) light sources can be used to tailor chromaticity but, as discussed in Section 4.1, they would not be readily compatible with downstream optics for beam directionality and placement.

2.1.2. Narrowband Spectra

Another likely characteristic of smart, ultra-efficient SSL is that, also unlike the current dominant paradigm in SSL technology,

its white light spectrum would not be broadband and continuous, but rather narrowband.

First, narrowband spectra maximize luminous efficacy of radiation (LER) for any desired color rendering quality.[28] For a lamp with a color rendering index (CRI) of 90, luminous efficacy of radiation is maximized with four narrowband spikes of light, while for a lamp with CRI of 85, luminous efficacy of radiation is maximized with three narrowband spikes of light. In both cases, so long as the narrowband spikes are optimally spaced to fill the visible spectrum, the resulting white light renders well the colors of objects typical in the environment around us (see **Figure 5**).[29a] If a different spectral quality metric is to be maximized (e.g., color quality scale [CQS][30] or gamut area index [GAI][31] or some combination of indexes,[32] then additional narrowband spikes may of course also be added.[33]

Second, narrowband spectra are compatible with chromaticity tuning. For example, by tuning the red-to-blue, green-to-blue, and yellow-to-blue power ratios, any chromaticity point within the gamut defined by the spectral spikes of the monochromatic LEDs can be created, and any chromaticity along the blackbody Planckian can of course also be created.

2.1.3. High Modulation Speed

A final likely characteristic of smart, ultra-efficient is modulation speed. Many aspects of smart lighting will require only relatively slow (seconds to minutes) time scales. Indeed, for on/off/dimming triggered by the transit of occupants through the area, fast on/off/dimming can be disconcerting and undesirable. However, some aspects of smart lighting may require relatively faster (millisecond) time-scale on/off/dimming – e.g., when used as cues for human response in rapidly changing and dangerous vehicular-traffic environments. And some aspects of smart lighting may require even faster (nanosecond) time-scale modulation – e.g., when communications and/or light-field-mapping functionality (discussed in Section 2.2) are desired.

All of these modulation speeds are compatible with the characteristics (multicolor electroluminescence and narrowband spectra) just discussed. Electroluminescent semiconductor devices (LEDs and especially lasers) have higher modulation speeds than most wavelength downconversion materials. And the use of multicolor electroluminescence can support even higher effective bandwidths using wavelength division multiplexing techniques similar to those widely used in optical communication systems.[34]

A key challenge, however, may be to maintain high efficiency while increasing modulation speed: these two characteristics are not always mutually compatible. A common way to increase modulation speed is to increase the non-radiative rate at which electrons and holes recombine; however doing so decreases simultaneously the radiative efficiency. Instead, one must find ways to increase the radiative rate, through, e.g., enhanced light-matter interactions or stimulated emission (as with lasers), or to move to wavelength-division multiplexing. In addition, as always with high-speed devices, careful attention must be paid to reducing capacitances and their associated resistance-times-capacitance (RC) time lags.

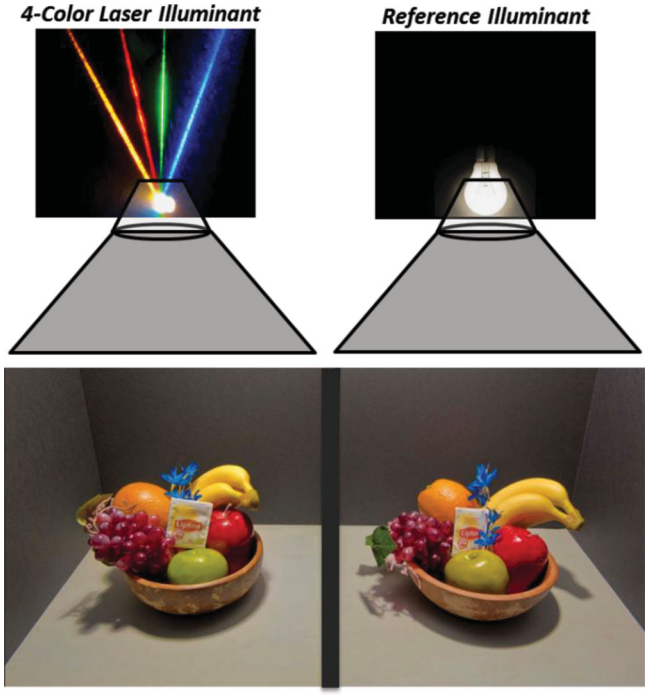

Figure 5. Schematic of a side-by-side test of the color-rendering quality of a four-color narrowband laser illuminant and a number of reference sources (including the incandescent lamp depicted). To give a sense for the color-rendering quality of the laser illuminant, the photos of the left and right bowls of fruit were illuminated, respectively, by the four-color laser illuminant and by an incandescent lamp. Although the image sensor in the camera is not the human eye, so the similarity of the photos does not automatically imply similarity in human visual experience, in fact the human visual experience was quite similar. Reproduced with permission.[29b] Copyright 2013, Springer.

2.2. Potential Future System Applications

In Section 2.1, we discussed some of the likely characteristics of smart, ultra-efficient SSL. If these characteristics can be achieved, unprecedented control of light properties would be enabled, and potential new system applications and markets would be enabled or expanded.

Here in Section 2.2, we speculate on (and illustrate in **Figure 6**) some of these potential system applications: a "second wave" of lighting which is smart[35] and feature-rich; integrated illumination and displays; human health and well-being; agriculture; communication; and light-field mapping. These potential system applications will co-evolve with other global and societal technology changes, like the increasing demand for video information and bandwidth, the critical need for energy conservation and the deployment of sustainable technologies, and the growth of highly interconnected and diversified sources of alternate energy. They will also involve a wide range of supply chain technology components, as well as new industries previously foreign to the lighting or LED marketplaces.

2.2.1. "Second Wave" Lighting: Smart and Feature-Rich

A first system application is illumination itself. As discussed in Section 1.4, the first wave of dumb and rudimentarily-smart

solid-state lighting is already beginning to penetrate illumination applications. Though important, this first wave will likely ultimately be succeeded by a second wave of solid-state lighting which makes use of features that are unique to solid-state lighting and that are not possible with traditional incandescent, HID or fluorescent lighting.

A first aspect of such smart lighting will be its networkability. With the separate advance of mobile wireless devices which sense ambient light levels and space occupancy, SSL lighting systems will enter an era in which lighting systems have the information necessary to customize light to user preferences.

A second aspect will be the unprecedented device-level control of the luminance and spectral content of light in space and time. At a simple level, it will be possible to reproduce the spectral changes associated with dimming of incandescent sources (brightness-dependent correlated color temperature [CCT]). At a more complex level, it will be possible to tailor luminance in space and time for aesthetic and safety purposes. For example, traffic safety can be influenced significantly through real-time spatial tailoring of automotive[36,37] (see, for example, the left side of Figure 4) and roadway[38,39] lighting. At an even more complex level, it will be possible to customize spectra[33,40] to improve (as discussed in more detail below) the health, well being, and productivity of individuals.

A third aspect will be the real-time tailoring of spectra to favorably impact energy efficiency. For example, in situations (e.g., parking lots viewed from a distance) for which color rendering quality is less important, energy efficiency could be improved through use of spectra that better match the human photopic (or scotopic, depending on the situation) eye response, even if they render colors somewhat less well. Alternatively, in situations (e.g., parking lots viewed up close) for which color rendering quality is more important, spectra that better fill the human photopic eye response could be used, even if energy efficiency is compromised somewhat.

A fourth aspect could be the subtle tailoring of absolute luminance levels to compensate for utility-scale fluctuations in electricity availability and consumption. Though electricity generation at the utility is difficult to change instantly to adjust to overall electricity demand by consumers, some forms of electricity demand might be tradeable for other forms, nearly instantly. In particular, because illumination is such a major percentage (ca. 20%) of the world's electricity budget, small (visually unnoticeable) changes in illumination levels could in principle be used to make such adjustments.[41] Fast-time-scale networked micro-pricing on a smart grid could even be incorporated to enable "illumination arbitrage."

2.2.2. Integrated Illumination and Displays

A second system application, displays, is more than just a potential application; it has been a key driver in the past for the development of LEDs. The ubiquity of LCD displays and their need for compact, long-lived, efficient backlights has driven tremendous progress in LEDs and phosphor-converted white

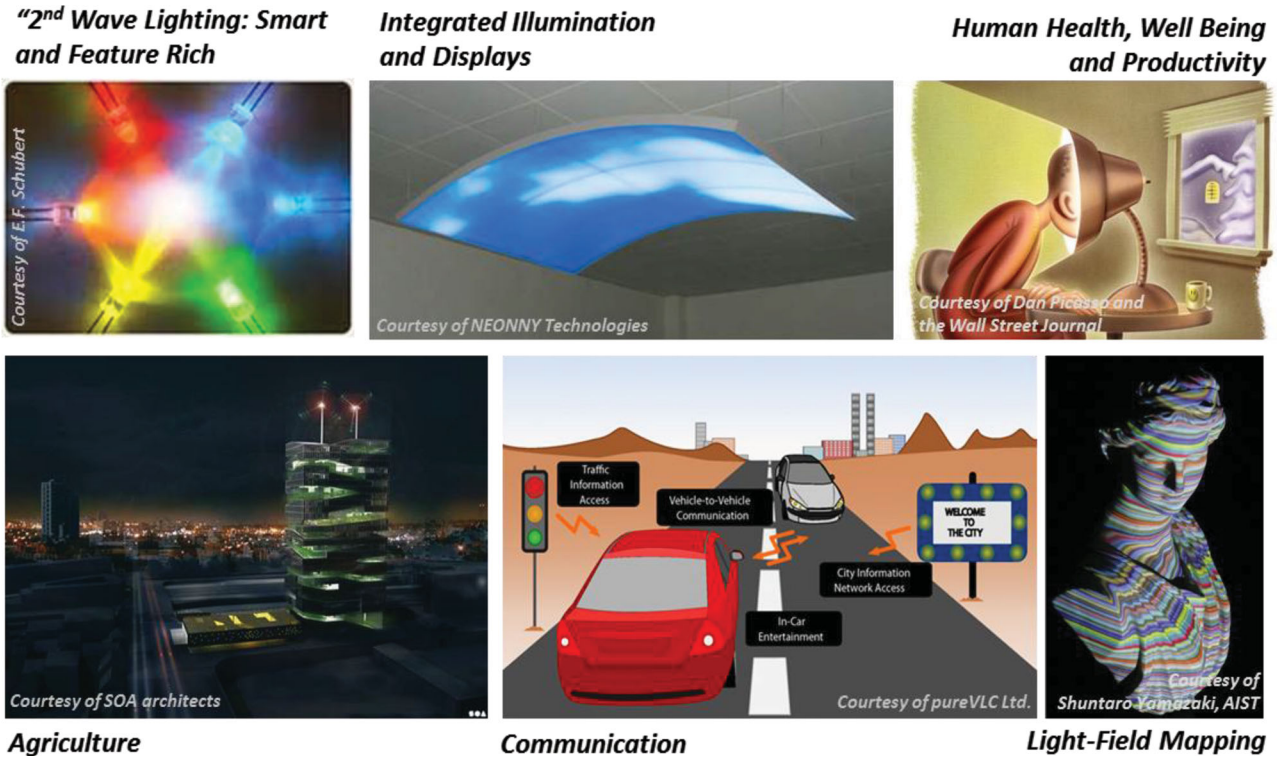

Figure 6. Potential future system applications for smart, ultra-efficient solid-state lighting. Clockwise from top left, figures are reproduced with permission from: Professor E. F. Schubert (Rensselaer Polytechnic Institute), 2013; Jason Yao (NEONNY Technologies), 2013; Dan Picasso and the Wall Street Journal, 2013; S. Yamazaki (National Institute of Advanced Industrial Science and Technology, Japan), 2013; Gordon Povey (pureVLC Ltd.), 2013; and Raffaela Schmied (SOA Architectes, France), 2013.

LEDs. Very soon, virtually every LCD display, from the smallest mobile displays to the largest television displays, is likely to be backlit by LEDs.

In the future, as displays become larger and produce more total light, their functionality might become "fused" with illumination functionality. In fact, the integration of illumination and displays has been under development for some time, and includes the use of LCD display technology for virtual windows and skylights,[42] and the use of color-tunable LED lighting panel technology to simulate illumination from an "open sky" source.[43,44] Significant technical development of both display and LED technologies will be needed to fully develop this fusion, but it is likely that future lighting systems will more closely approximate natural lighting (and the scenes that accompany it). This trend will be analogous to trends observed in telephony, where what was once known as "plain old telephone service" (or POTS) has evolved to become smart mobile communications systems with a fusion of communications, computation and display technologies.

Note that OLED technology may also be important in this vision of display and illumination functionality fusion. As mentioned in Section 1.4, OLED displays are already displacing LCDs in small mobile displays; for larger displays they may eventually also, provided their costs can continue to decrease.

2.2.3. Human Health, Well Being, and Productivity

A third system application is human health and well-being. It has been known for some time that there is a strong link between the spectral content of the light used for illumination and human health, well-being, and productivity. For example, in the late 1990s, the discovery of non-visual light receptors in the human retina[45] and the role of these receptors in setting and altering human circadian rhythm[46] led to research suggesting that future lighting systems could play an important role in human sleep management. Indeed, it is well known that lighting can mitigate some of the impact of Seasonal Affective disorder (SAD), a well-studied condition linking depression with limited exposure to daylight.[47] Since the spectral content of conventional incandescent and fluorescent lighting systems is fixed, however, detailed studies of the impact of lighting spectral content on human health have been somewhat limited. Smart solid-state lighting has the potential to enable much richer experimentation, ultimately leading to empirically sound strategies for improving human health and well-being. Or, for example, recent studies suggest new approaches for control of light spectral content, not just for human health but for enhancing human performance. Recent studies on subtracting a narrow band of blue wavelengths from conventional lighting systems show that subtractive spectral management may have a positive impact on maintaining the circadian period and on the alertness of shift workers.[48] Other studies show that the spectral qualities of light can impact cognition,[49] educational performance,[50] and the treatment of elderly patients suffering from Alzheimer's disease.[51] Finally, we note that human health and well-being are especially important in healthcare environments. Exposure to sunlight has a large impact on hospitalized patient health and well-being;[52] and exposure to color-temperature-tailored solid-state illumination may also.[53] The impact may be both intangible (patient mood and psychological state of mind) as well as tangible (patient recovery times).[54]

2.2.4. Agriculture

A fourth system application is agriculture. With the increased urbanization and localization of the world's population, there is a need for food sources to be local to urban centers (bypassing transportation costs and enabling tailoring to local preferences) and to be maximally productive all year round (extending the growing season).

Simple LED lighting is already being investigated actively for agriculture,[55] and more continues to be learned about the specific effects of wavelength on plant growth. The presence or absence of certain wavelengths can impact different phases of the plant's growth cycle, as well as the nutritional value of the produce from the plants. Production yield under optimized illumination can be higher than under natural growth conditions, though the impact of spectral distribution can be complex and can depend on plant type.[56] Moreover, the intensity and price point of LED systems are reaching a level at which they can be considered for large-scale horticultural applications.

A related emerging use of LED lighting is to grow genetically engineered plants for production of pharmaceutical products (sometimes referred to as Pharming.[57] Such transgenic plants could be engineered to require particular illumination wavelengths at particular phases of their life cycle. Those illumination wavelengths could then be preferentially supplied by tailored LED illuminants, thereby optimizing pharmacological yield. At the same time, the relative absence of those illumination wavelengths in sunlight would help ensure the lack of viability of such transgenic plants in natural environments.

2.2.5. Communication

A fifth system application is communication. Communicating with light is of course already commonplace in the infrared, both in free-space (low-bandwidth remote control of televisions) and in fibers (high-bandwidth fiber optic communications). But with the advent of LEDs that can be modulated at high speed even while providing illumination functionality, a new area called visible Light communication (VLC) is emerging.[58] Multicolor electroluminescent devices would be ideal for such dual illumination+communication functionality, since LEDs can be modulated at relatively high speed (faster than the human eye can resolve) and VLC applications might be partitionable by color. But even the current phosphor-converted white LED systems can be used: the blue LED (some of whose blue light leaks through the phosphors) can be modulated at high speed even if the phosphors themselves cannot due to their long radiative lifetimes. Indeed, high speed data links using phosphor-converted white LEDs have now been demonstrated.[59]

Sometimes called LiFi (like the RF based WiFi), this new technology promises to circumvent limitations of radio frequency (RF) based wireless communications and to satisfy

the rapidly increasing demand for mobile bandwidth.[60,61] Lighting systems designed to supply high speed wireless data and illumination simultaneously could relax demand for more sophisticated high frequency WiFi systems, could provide wireless connectivity in RF-restricted locations, and could provide a somewhat higher level of security (due to the need for line-of-sight) than is possible with RF wireless services.

Other advantages of VLC over RF-based wireless systems include: unrestricted allocation of bandwidth (though many challenges need to be worked out with respect to hand-offs between zones); easy integration with fixtures already in place for illumination purposes; and high bandwidth on the downlink where it is needed (though this requires integration with RF up-links). Much of this work leverages IR free-space optical communications work,[62] and preliminary standards for VLC applications are under development.[63] Ultimately, integrated combinations of VLC and RF wireless communications may provide the best solution to the seemingly endless growth in demand for wireless data and mobile connectivity.[64]

Note that one of the obvious limitions of LiFi, that it is line-of-sight, is similarly a limitation of future generations of WiFi. As RF based WiFi moves to higher bandwidth 60 GHz operating frequencies, WiFi will also to some extent become line-of-sight, due to strong absorption by a wide range of common materials. These limitations will require special designs and implementations,[65] many of which will be similar to those required by VLC systems.

2.2.6. Light-Field Mapping

A sixth system application is light-field mapping. This application builds experimentally on the high-bandwidth communications functionality of illumination: the ability to transmit and receive data. It will also build conceptually on the knowledge domains of time-of-flight sensing, structured light for coarse 3D mapping of surfaces,[66] and light transport analysis developed in graphic arts research.[67]

The proximate goal is illumination systems equipped with simple light detector networks that perform light-based sensing of occupancy[68] and of the overall architecture of the illuminated space.[69] If each light source in an illumination system had an identification code and setting information (dimming, color point, etc.), receivers embedded in these same light sources would enable the lighting system to characterize the rough geometry of the illuminated space, including both inanimate and animate objects. Lighting systems would, for the first time, "see" where emitted light goes, and image the surfaces that scatter or reflect that emitted light.

The ultimate goal of such systems is real-time knowledge of the illuminated space, how it is being illuminated, and who is observing the illumination, so that lighting controls can adaptively tailor the timing, placement, and type of illumination required. The system would know how much of the light is artificial (with an identification code) versus how much of the light is natural (without an identification code). It would autonomously interact with other light-emitting systems with their own digital light signature (e.g., video systems), tailoring the illumination as required. Such intuitive and even autonomous illumination control systems would go well beyond earlier generations based on simple camera networks.[70] We note, though, that they may be just as subject to privacy concerns as those previous generations, though perhaps not at the same level if lower pixel densities (1 to 5 per m^2) make it difficult to reconstruct high-resolution images.

Finally, we mention here for completeness that virtually all light-field mapping applications rely on, and are unified conceptually by, the so-called plenoptic function (PF). The 7-dimensional PF represents the time-dependent intensity, directionality and chromaticity of the light observed at any position in a 3- dimensional space. It is the complete time-dependent volumetric light field and contains all information needed for lighting design, and visible-light communications. Through an understanding of how a distribution of light sources (accent lighting, task lighting, diffuse-conference-room lighting, etc.) determine the plenoptic function, accompanied by distributed and networked light sensors that characterize the plenoptic function, the plenoptic function can be adjusted dynamically to provide the highest quality task-dependent lighting and data at the lowest possible energy cost.

This then leads to the concept of a universal lighting fixture. A networked array of such instrumented fixtures, emitting light as well as monitoring the local light field, would construct the PF for the illuminated space and then alter the emitted light to minimize the error between the desired and actual PF. The sensor network contained in the illumination system would likely be augmented by other sensors (including mobile devices of occupants) not mounted in the fixtures to fully characterize the ambient light field.

2.3. Benefits: "Effective" Efficiency, Consumption of Light, and GDP

Assuming that smart, ultra-efficient solid-state lighting could be achieved, let us now discuss what the energy and human productivity benefits might be.

2.3.1. "Effective" Efficiency of Lighting

Of the two characteristics, smartness and ultra-efficiency, the benefits to ultra-efficiency are easiest to quantify. Ultra-efficiency translates directly to energy consumption given a certain lighting consumption, or directly to lighting consumption given a certain energy consumption.[71]

The benefits to smartness are more difficult to quantify. How does one quantify, for example, the various aspects of human productivity, just discussed in Section 2.2, that are likely to be impacted by the ability manipulate the spectral content of illumination in time and space: human health and well-being, communications and sensing, and even agriculture?

One benefit to smartness, however, might perhaps be quantified: how light is provisioned within a given space. In typical use, the fraction of photons leaving a lamp that finally strikes the retina of a human eye is probably less than a millionth, even in an enclosed space such as an office.[72] The loss factors include: the less-than-100% reflectance of non-white objects

being illuminated; the small entrance aperture and field of view of the human eye relative to the area and solid scattering angle of the illuminated space; and the less-than-100% portion of time an object is actually in someone's field of view while it is being illuminated. The first two of these loss factors are difficult to decrease. However, the third factor can clearly be decreased.

To estimate the magnitude of the potential reduction in this third factor, note that per capita light consumption in the US in 2001 was 136 Mlmh/year,[73] the equivalent of the average person in the US being surrounded during his or her waking hours by seventeen 100 W incandescent light bulbs, or by about 24 klm. In contrast, a comfortable 500 lm/m^2 illuminance in a typical office-sized 3-m-radius half circle in front of a person corresponds only to about 8 klm. Although in many situations a person views a much larger area, in these same situations a person often shares light with many others. Assuming these effects cancel approximately, one can estimate that there is room for potential reduction by a factor 3 = 24 klm/8 klm.

Such reductions could in principle be brought about through aggressive and sophisticated application of sensor-based real-time controls. SSL is characterized by fast switching speeds and in principle can be turned off not only when there are no occupants in a room (if indoors) or area (if outdoors), but can be turned off in zones within a room or area that no occupant is looking at.

In other words, ultra-efficiency improves the efficiency of the production of light while smartness, in addition to its difficult-to-quantify (but important, as discussed in Section 2.2) benefits, would improve the efficiency of the utilization of light. Hence, by defining an "effective" efficiency, one can to zeroth-order account for the benefits of both. Note that in terms of headroom for future improvement, both are of the same order of magnitude. Solid-state lighting is currently roughly 25–40% efficient (when operated at "economical" input power densities, as discussed in section 3) and thus has a factor ca. 3× before it reaches its 100% limiting production efficiency. Smart solid-state lighting could be yet another factor ca. 3× more "efficient."

2.3.2. Consumption of Light and Increase in GDP

Including both ultra-efficiency and smartness, then, it may be possible to improve "effective" efficiency by ca. 9×. What would be the economic consequences of such an increase?

One possible answer to this question is sketched on the left side of **Figure 7**. This is a scenario in which light is considered to be a factor of production in the global economy. Light is thus considered to be both desirable (more is better) and useful (productivity enhancing). Thus, if the effective efficiency with which light is produced or utilized were to increase, the effective cost of light would decrease, more light would be consumed, and world GDP would increase (above where it otherwise would be, all other things in the world held constant).

This scenario has been analyzed in some detail using a simple Cobb–Douglas model which treats light as a factor of production in the global economy.[74] The result of that analysis is that there is no direct energy (or environmental) benefit to increased efficiency: since the consumption of light increases as efficiency increases, energy consumption doesn't decrease. Indeed, because GDP increases as light consumption increases, there is even a slight increase in energy consumption. There is, however, a huge benefit of increased efficiency in terms of an increase in global GDP. Although there is a diminishing return to the increase in GDP, if the effective efficiency of lighting were to increase from 0.5 to 1.5, global GDP would increase by a staggering $680B/yr.

Although we do not know with certainty that this scenario will come to describe the future, if the past is any indication, it might. The reason is illustrated in the plot[75] on the right side of Figure 7. The vertical axis of this plot is consumption of light, in Plmh/year. The horizontal axis of the plot is a fixed constant, β, times the ratio between gross domestic product (GDP) and cost of light (CoL). If we use as our units for GDP billions of dollars per year, and for cost of light $/Mlmh, we see that this ratio has the same units, Plmh/year, as the units of the vertical axis. And if we choose the fixed constant, β, to be 0.0072, we see that the empirical data fall very closely along a line of slope unity and zero offset. These results are exactly what one would predict from the scenario illustrated on the left side of Figure 7, in which the more efficiently light is produced, the more we consume and the more productive we are.

Again, we do not know whether the future will be like the past but, if there is even some small probability that it is, then this analysis shows that it is worth the effort to continue to improve SSL "effective" efficiency through both ultra-efficiency *and* smartness.[76]

3. Ultra-efficient SSL Lighting: Towards Multi-color Semiconductor Electroluminescence

In Section 2, we discussed the likely characteristics and potential benefits of smart, ultra-efficient solid-state lighting. In Section 3, we discuss the prospects for achieving ultra-efficient solid-state lighting; in the following Section 4, we discuss the prospects for developing smart solid-state lighting.

Because multicolor electroluminescence is at the heart of realizing ultra-efficient solid-state lighting, we discuss, in turn, the state-of-the-art and challenges associated with semiconductor electroluminescence of the three primary colors (blue, green and red) necessary for white light production.[77] As mentioned in Section 2.1, it is possible that more than three colors will ultimately be needed, but it is these core colors that we focus discussion on here.

3.1. Blue Materials and Devices

As discussed in Section 1.1, since the demonstration of reasonably efficient blue InGaN LEDs in 1993,[7] spectacular progress has been made: state-of-the-art blue LEDs now have power-conversion efficiencies exceeding 80%.[78] To put this in perspective, the power-conversion efficiency of an LED, ε_{tot}, can be written as the product of four sub-efficiencies:[79]

$$\varepsilon_{tot} = \varepsilon_{inj}\varepsilon_{rad}\varepsilon_{ext}\left(\frac{h\upsilon}{eV}\right) \quad (1)$$

where ε_{inj} is the injection efficiency, ε_{rad} is the radiative efficiency, ε_{ext} is the extraction efficiency, and $[(h\upsilon)/(eV)]$ is the

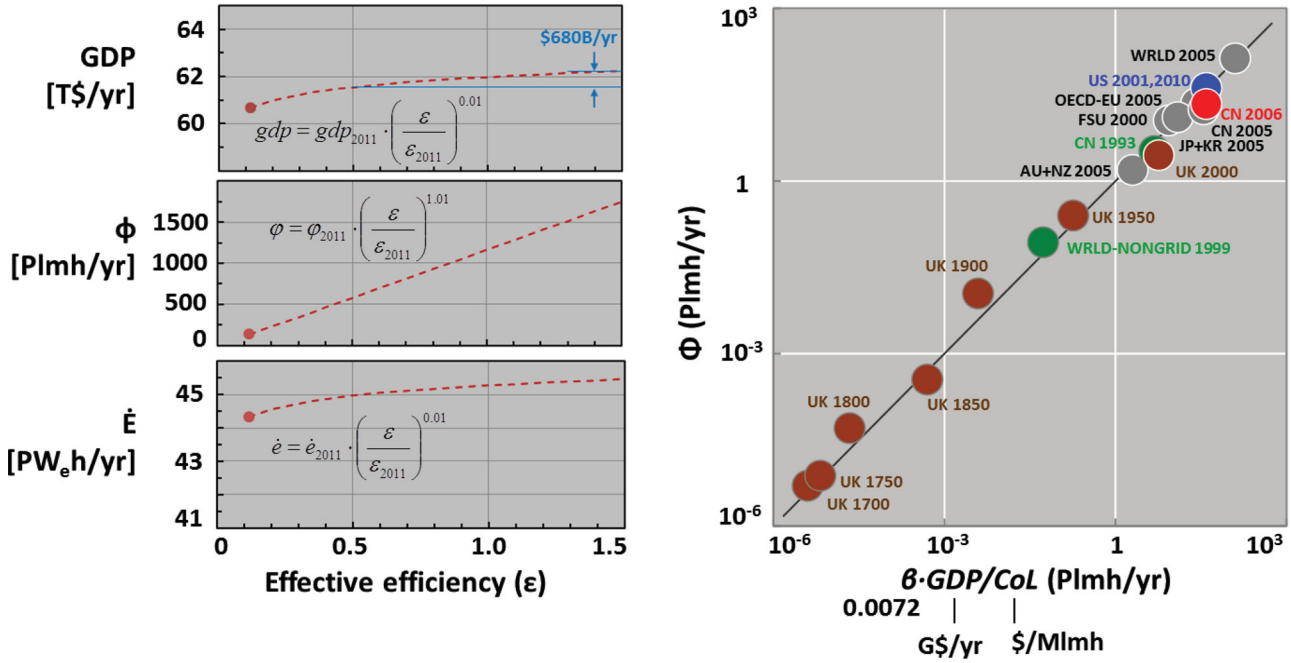

Figure 7. Left: Economic scenario for world gross domestic product (*GDP*), world consumption of light (Φ) and world consumption of energy ($\dot{E}$) assuming that light is a factor of production in the economy. Right: Consumption of light versus the ratio between gross domestic product (*GDP*) and cost of light (*CoL*) for various countries at various times in history.

Joule efficiency. In the last term, h is Plank's constant, υ is the photon frequency, e is electronic charge and V is the operating voltage. An overall efficiency that exceeds 80% means that each of the four sub-efficiencies is at least 80% and, because they are in a multiplicative product, must be very nearly 100%.

The injection efficiency, ε_{inj}, is the fraction of carriers injected from the contacts into the semiconductor heterostructure that eventually recombine in the active region. In principle, it is a complicated function of the epitaxial layer structure, the resulting energy band structure, carrier lifetimes, and internal and externally applied fields.[80,81] In practice, from comparisons between photoluminescence and electroluminescence yields, and from measurements of injection efficiencies in blue lasers with similar structures, it has been seen that injection efficiencies are very high, on the order of 95%,[82,83] and also relatively independent of input power density.

The radiative efficiency, ε_{rad}, is the fraction of electron-hole pairs in the active region that recombine to produce blue light (as opposed to recombining and not producing blue light). It is the ratio between the radiative and all the recombination rates, often characterized by the heuristic equation:

$$\varepsilon_{rad} = \frac{Bn^2}{An + Bn^2 + Cn^3} \tag{2}$$

where An is the Shockley–Read–Hall (SRH) defect-mediated recombination rate, Bn^2 is the spontaneous (bimolecular) radiative recombination rate, Cn^3 is the Auger recombination rate,[84,85] and n is the carrier density in the light-emitting active layer(s). For state-of-the-art blue LEDs, ε_{rad} rises at small n, peaks at near-100% at an intermediate n, and then decreases at large n. This is so even when the non-constancy of the recombination rate coefficients A, B, and C with carrier density n is taken into account (at high carrier densities, degeneracy effects enter in, as well as screening of the polarization fields that, in c-plane InGaN heterostructures, decrease electron and hole wave function overlap and recombination rates).

The extraction efficiency, ε_{ext}, is the fraction of light produced in the active region that escapes into free space. Because of the high index of refraction of semiconductors, the emitted light can easily be trapped withing the device through total internal reflection and eventually absorbed by paracitic losses. Therefore, all advanced LED chip designs rely on either textured surfaces or ergodic shapes to randomize internal photon trajectories and/or remove the substrate to reduce confined modes. In a thin-film device where the LED material is placed on a reflective surface and has a rough exit surface, the internal light behaves as a photon gas with an extraction efficiency given by the ratio between the fraction of light that escapes on each attempt ($T/[4n^2_{opt}]$) and the fraction of light that either escapes ($T/[4n^2_{opt}]$) or is irreversibly absorbed in the active region ($[1 - \varepsilon_{rad}]\alpha d$) or at the back mirror ($[1 - R]/4$) between attempts:[86]

$$\varepsilon_{ext} = \frac{\frac{T}{4n^2_{opt}}}{\frac{T}{4n^2_{opt}} + (1-\varepsilon_{rad})\alpha d + \frac{1-R}{4}} \tag{3}$$

Here, T is the angle-averaged transmission coefficient though the front exit surface, n_{opt} is the optical index of the semiconductor, α and d are the optical absorption coefficient and thickness of the active region, and R is the angle-averaged reflectivity of the back mirror. In advanced structures, the fraction that

escapes on each attempt is much larger than the fraction that is absorbed between attempts, so the extraction efficiency is very high, on the order of 85%.

The Joule efficiency, $(h\upsilon)/(eV)$, is the ratio between the energy of the photon emitted and the energy of the electron injected into the device at voltage V. That voltage is the sum of two voltages:

$$V = V_D + IR \tag{4}$$

The first voltage, the diode voltage, is commonly characterized as $V_D = (n_f kT/e)\ln(I/I_o)$, where n_f is the diode ideality factor, k is Boltzmann's constant, T is temperature, and I_o is the reverse-bias saturation current. The second voltage, IR, is the voltage drop caused by the series resistance, R, experienced by carriers enroute to the active region of the device. Typically, R is dominated by the spreading resistance in the n-type GaN layers, because the p-type layers, despite their much lower conductivity, are very thin, with relatively low voltage drop when supporting vertical hole transport directly to the metal contact.[87] At low input currents, the voltage is very close to the turn-on voltage of the diode (i.e., the bandgap of the active region) and the energy of the emitted photon; in this case the Joule efficiency approaches unity. At high input currents, the voltage increases due to the series resistance and Joule efficiency, and therefore total efficiency, decreases.

The dependences of all of these sub-efficiencies on input power density for a state-of-the-art blue LED are plotted in **Figure 8**. While there is room for improvement in all of the sub-efficiencies, the most critical is the radiative efficiency. This efficiency rises at low input power density, peaks, and then decreases at modest input power densities. This decrease, called "efficiency droop," is perhaps the single biggest challenge in blue materials and devices.

The reason why it is the biggest challenge is: if blue LEDs can only be operated at the low input power densities at which efficiency is maximum, then blue LEDs will be limited to low output light fluxes per unit chip area. Running at such low input power densities is likely to be cost-prohibitive, even with the most aggressive projections of cost per unit chip area.

Going forward, then, there are a number of approaches being considered for economically viable ultra-high-efficiency lighting; we discuss these approaches next. One approach is to "live with" efficiency droop, by moving to nanostructures with very high active surface areas: in this way, input power densities can be low per unit active nanostructure surface area, but high per unit planar surface area on which the nanostructures are fabricated. Another approach is to delay the onset of efficiency droop. Some progress in this direction has been made through non-square quantum well designs[88] and alternative materials[89,90] on the usual polar c-plane crystal orientation, but perhaps the most striking progress has been through the use of semi- and non-polar crystal orientations. Yet another approach is to mitigate efficiency droop by moving to lasers, with their stimulated emission radiative recombination process that can outcompete Auger recombination at high-enough current density. Because lasers suffer from their own inefficiencies, mostly due to resistive losses, a final approach is to explore low-threshold-current-density lasers, for which such resistive losses would be significantly reduced. In the remainder of this Section 3.1, we discuss each of these approaches in turn.

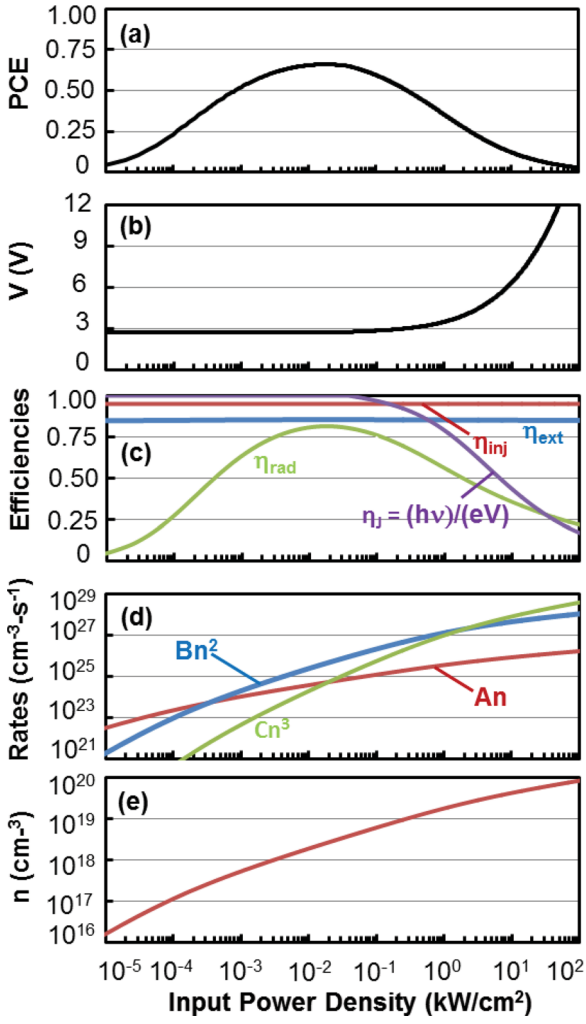

Figure 8. Dependence on input power density of a typical blue LED (a) power-conversion efficiency (PCE), (b) operating voltage (V), (c) sub-efficiencies, (d) carrier recombination (PCD) rates, and (e) carrier density (n). Adapted with permission.[79a] Copyright 2013, Wiley.

3.1.1. "Living with" Efficiency Droop: High-Surface-Area Nanostructures

A first approach to affordable ultra-high-efficiency is to "live with" efficiency droop, by increasing the volume of light-emitting material, thereby reducing carrier densities at a given injection current. To date, some reduction in efficiency droop has been achieved using double heterostructure devices with thicker bulk InGaN active layers,[91] wider quantum wells in non-polar materials[92] or improving the carrier distribution among several quantum wells by engineering of the quantum barrier bandstructure for more facile carrier transport.[93] A more effective, "next-generation" solution might be a move to nanostructures with very high active surface areas. In this way, input power densities can be low per unit nanostructure active

surface area, but high per unit planar surface area on which the nanostructures are fabricated.

For example, for an array of c-axis-oriented core-shell nanowires with hexagonal cross-sections arranged in a hexagonal lattice, the ratio between the nanowire array sidewall area and the planar substrate area on which the nanowire array is fabricated is:

$$\frac{A_{sidewall}}{A_{substrate}} = \frac{6rh}{\sim 2.6R^2} \sim 4.6F \cdot \chi \quad (5)$$

where r is the nanowire radius (distance from its center to a sidewall vertex at the active region), h is the nanowire height, R is the center-to-center distance of the nanowires in the hexagonal layout on the substrate, $F = r^2/R^2$ is the fill factor (the fraction of the substrate area the nanowires occupy), and χ is the aspect ratio $h/2r$ of each nanowire.

Thus, high aspect ratios can lead to a higher ratio of light-emitting to substrate area compared to planar architectures. For a fill factor of $F = ½$ and an aspect ratio of $\chi = 4$, both well within practical achievability, the sidewall-to-substrate-area ratio would be about ten. With all other costs and input power densities (per unit heterojunction area) being equal, a nanowire LED array architecture could support either a 10x higher substrate areal cost or a 10x lower input power density (per unit heterojunction area) than a planar LED architecture could. In addition, nanowire facets might also provide access to nonpolar and semi-polar orientations for reduction of the polarization effects discussed in the next subsubsection below, without the need for semi- or non-polar GaN substrates.

There are, however, several practical challenges to realizing an efficient radial heterostructure nanowire LED with a high active area. In an ideal geometry, the active region would be confined to the (possibly non-polar or semipolar) sidewalls, with uniform current injection directed radially from coaxial electrical contacts. However, in practice this architecture would be difficult to achieve. Growth of the active region can also occur at the nanowire tops and bases, as well as in the planar region between nanowires. And, even if confined to the nanowire sidewalls, the growth can be non-uniform along the length of the sidewalls. Moreover, current injection along the inner and outer coaxial conductors may not be uniform. These present non-trivial fabrication challenges.

3.1.2. Delaying the Onset of Efficiency Droop: Semi- and Non-polar Orientations

A second approach to affordable ultra-high-efficiency may be to push efficiency droop out to higher input power densities by moving to semi- or non-polar orientations for the quantum wells. This approach is based on the idea that, on the usual polar (c-plane) orientation on which InGaN blue LEDs are fabricated, polarization fields are large and cause the electron and hole wave functions to shift in opposite directions within the quantum wells in the active regions. All the recombination rates—Shockley–Read–Hall (SRH), radiative and Auger—depend on electron-hole wave-function overlap,[94] hence all are lower in c-plane-oriented than in semi-polar-oriented or non-polar-oriented quantum wells.

If, instead, non-polar (e.g., *m*-plane (1–100) or *a*-plane (11–20)) or semi-polar (e.g., (11–22)) crystal orientations were used, recombination rates would be higher, carrier densities (n) would be lower for a given input power density, and at these lower carrier densities Auger recombination, with its n^3 dependence, would be slower relative to radiative recombination, with its n^2 dependence. The result is the possibility of pushing the efficiency peak out to higher input power densities.

Note, though, that this route to addressing the detrimental effects of efficiency droop hinges on further developments.

First, whether or not the efficiency peak is indeed "pushed out" to higher input power densities depends on the details of how the three recombination mechanisms depend on polarization field, and these dependencies are not yet fully understood.

Second, the ability to fabricate blue LEDs on semi- and non-polar orientations is itself challenging. Unlike c-plane orientated GaN, which can be grown heteroepitaxially on readily obtained sapphire or even Si (111) substrates, semi- and non-polar orientated GaN cannot be grown heteroepitaxially on any known substrate while still yielding high optical efficiency. Indeed, all attempts thus far (e.g., growth of *a*-plane GaN on *r*-plane sapphire) have resulted in unacceptably high densities of extended defects, including threading dislocations (>10^{10} cm^{-2}) and stacking faults (>10^5 cm^{-1}).[95]

Thus, the most promising current route to semi- or non-polar orientations appears to be the development of bulk crystal growth of large GaN boules from which arbitrarily oriented wafers can be sawed and on which LEDs can be epitaxially grown. Although much remains to be done, the development of native GaN substrates for LED growth has made great strides. GaN substrates are already being used for blue and green GaN-based laser diode production, and have been used to demonstrate nonpolar InGaN LEDs with better wavelength stability and claims of higher efficiency at high current densities compared to those of *c*-plane LEDs.[96] The key present-day challenge is to develop technologies for substrates that are large enough and low enough in cost to be economical for solid-state lighting.

3.1.3a. Mitigating Efficiency Droop: Lasers

A third approach to affordable ultra-high-efficiency lighting is based on the idea that radiative recombination rates can be much higher if based on stimulated rather than spontaneous emission.[79] A higher radiative recombination rate results in a lower carrier density for a given input power density, and thus also results in slower Auger recombination. A consequence is the possibility of much higher radiative efficiencies at high input power densities. Indeed, high input power densities and thus high output light fluxes per unit chip area imply a supportable cost per unit chip area comparable to and perhaps even higher than that already achieved by high-power high-efficiency lasers in the infrared. Moreover, though the achievable single-chip heat-sink-limited output light fluxes are lower than those for LEDs (because of the smaller volume within which heat must be dissipated), they are nevertheless in a practical range (1 klm) for solid-state lighting.

However, this route suffers from one key drawback: at the very high input power densities normally required to achieve

lasing threshold, resistive *IR* losses become very large, so the peak power-conversion efficiency of blue lasers (about 30–40%) is significantly lower than the peak power-conversion efficiency of blue LEDs (about 80%). Thus, this route to ultra-high-efficiency also hinges on a number of developments.

First, because resistive losses are the dominant contributor to reduced power-conversion efficiency, methods for reducing resistance would be extremely beneficial. Possible methods include: increasing p-type doping concentrations (e.g., by operating at higher temperatures [while maintaining reliability] to overcome the high energies required to activate the p-type dopants), tunnel junctions to eliminate p-type doping entirely, or reducing the laser threshold current density (discussed in more detail next).

Second, because photons in a laser cavity have a much longer lifetime than in an LED structure, there are many more opportunities for optical absorption losses to be incurred. Methods for decreasing these losses would also be extremely beneficial. Possible methods include: increased confinement of the optical mode to decrease the optical absorption associated with the p-layer, or spatially confining the p-dopants to the nodes of the optical standing wave.

3.1.3b. Reducing Losses in Lasers: Low Current Density Thresholds

As mentioned above, for high efficiency lasers we are particularly interested in reducing resistive losses which means lowering operating (and threshold) current *density*, not just absolute current. To see this, note that the *IR* drop can be rewritten as:

$$IR = \frac{I}{A} \cdot RA = \frac{I}{A} \cdot \frac{\rho h}{A} A = j \cdot \rho h \tag{6}$$

Here, we have written $j = I/A$ as the current density, or the current normalized to the area A of the layer through which current passes enroute to the active region. We have also written the resistance as the product of the electrical resistivity ρ and thickness h of the layer through which current passes divided by the area A. For a given material electrical resistivity and thickness, the *IR* drop is thus lower the lower the operating current density,[97] which in turn is lower the lower the current density at lasing threshold. Equivalently, forcing more current through a smaller area will always require a higher voltage. Thus, reducing the threshold current density will improve efficiency by reducing resistive loss.

To see how the threshold input current density of a laser might be decreased, let us write it as:

$$j_{th} = \frac{ed}{\varepsilon_{inj}} \left[Bn_{tr}^2 e^{2\frac{\alpha_m + \alpha_i}{\Gamma g_o}} + Cn_{tr}^3 e^{3\frac{\alpha_m + \alpha_i}{\Gamma g_o}} \right] \tag{7}$$

where e is electron charge, d is the active layer thickness, ε_{inj} is the injection efficiency, n_{tr} is the transparency carrier density, α_m and α_i are the mirror and optical absorption losses, Γg_0 is the maximum modal gain, and B and C are the radiative and Auger recombination coefficients.[98] In general, we would like to reduce B, C, and α_i, while maximizing ε_{inj} and Γg_0. The modal gain will be strongly dependent on geometry but might be increased by placing gain at cavity anti-nodes (particularly for vertical cavity architectures). It is also worth noting that, for a given set of parameters, there will be an optimal value of α_m that maximizes efficiency.

Of the terms in Equation 7 that might be altered to reduce threshold current density, reducing optical absorption loss α_i is one of the most important. Optical absorption loss is caused by a number of different mechanisms, including: residual below band gap semiconductor absorption, scattering from waveguide imperfections, absorption from dopant centers, and free-carrier absorption. Improving materials quality and interface sharpness, and tailoring dopant distributions, could thus yield useful improvements. Note, though, that for certain mechanisms such as free-carrier absorption there may be a trade-off between low optical absorption and low electrical resistance. For example, VCSELs with doped distributed Bragg reflectors (DBRs) have such a trade-off.[99] In an analogous manner, high-efficiency edge emitters must be designed to carefully balance optical and resistive losses.

Another of the terms in Equation 7 that might be altered to reduce threshold current density is the modal gain, Γg_0. Quantum dot (QD) lasers are particularly interesting in this regard. The spectral collapse of density of states in quantum dots enables higher modal gain and potentially lower threshold current density. Indeed, in the infrared, where quantum dot technology is more advanced, threshold current densities for continuous wave (CW) room temperature operation can be as low as 10 A/cm^2.[100] A key challenge to QD lasers in the visible, however, will be to develop an InGaN quantum dot fabrication technology, particularly one with a relatively narrow quantum-dot size distribution and hence a relatively narrow inhomogeneous gain linewidth.

Finally, we mention here that other, more exotic, lasing mechanisms such as polariton lasing[101] with even lower thresholds may be beneficial. This equates to reducing n_{tr} in Equation 7: polariton lasing is known to occur at carrier densities 3 to 4 orders of magnitude lower than photon lasing occurs at.[102] However, polariton lasing (coherent emission of light from a polariton condensate) occurs in the strong-coupling regime, and requires careful attention to cavity architecture, including extremely high Q. Nevertheless, this exotic form of lasing might also have potential for ultra-high efficiency.

3.2. Green Materials and Devices

As just discussed in Section 3.1, much progress has been made in blue InGaN LED performance. Progress is also being made in green InGaN LED performance, and InGaN is considered the most promising semiconductor system for green light-emitting devices as well.

Note that two other materials have been considered in the past for green light emitting devices, but InGaN has fundamental advantages over both. First is AlInGaP materials. As the energy bandgap of InGaN decreases from the blue to the green with increasing In content, the InGaN bandgap remains direct; while as the energy bandgap of AlInGaP increases from the red to the green with increasing Al content, the AlInGaP bandgap

becomes indirect. Second is II–VI alloys. The medium-In-content InGaN materials which luminesce in the green are just as resistant to photoinduced defect production as the low-In-content InGaN materials which luminesce in the blue; while II–VI materials which luminesce in the green and have weaker bond strengths are not resistant to photoinduced defect production.

At this point in time, state-of-the-art commercial InGaN LEDs at 530 nm have demonstrated 68 lm/W performance (at 70 A/cm^2).[16] This level of performance, however, is far from that required for ultra-efficient solid-state lighting. To achieve power-conversion efficiencies of the order 70% or higher, internal quantum efficiencies (IQEs) ≥ 90% will be necessary. Perhaps more importantly, even longer wavelengths into the deep green will be necessary for high-color-rendering-quality white light: 548 nm is optimal for RGB white light, while 535 nm and 573 nm (straddling 548 nm) are optimal for RYGB white light.[103]

Though increasing IQE from 30% to 90% and shifting wavelengths from 530 nm to wavelengths near 548 nm should in principle be possible, thus far this has not been achieved. Indeed, the efficiency drop-off with increasing wavelength seen in Figure 3 is a major focus of SSL research efforts, and has come to be known as the "green gap." This efficiency loss has been attributed to a number of issues, two of which are illustrated in **Figure 9**.

The first issue is lattice mismatch and strain. As illustrated in the bottom panel of Figure 9, the medium-In-content InGaN alloys[104] necessary for deep-green light emission are highly lattice mismatched to the GaN buffer layers typically grown on sapphire or other substrates that are currently in use. This lattice mismatch leads to high strain, high internal piezoelectric fields, quantum-confined Stark effects, potentially high defect densities, and low radiative recombination rates.[105]

The second issue is the narrow window in epitaxial growth temperature associated with the medium-In-content InGaN alloys necessary for deep-green light emission.

On the one hand, as illustrated in part by the upper left panel of Figure 9, thermodynamics would suggest higher growth temperatures have advantages. Higher growth temperatures would be desirable both to circumvent microscopic-strain-induced InGaN immiscibility[106] (the concave curve labeled $T_{spinodal}$) as well as to maintain adatom surface mobility typically necessary for high crystalline quality (the diagonal line labeled T_{epi}). Higher growth temperatures are also desirable to adequately dissociate the ammonia (NH_3) that is the most common source of nitrogen in the metalorganic chemical vapor deposition of InGaN.

On the other hand, as illustrated in part by the rightmost two panels of Figure 9, lower growth temperatures also have advantages. The pyrolysis temperature of trimethylindium, the most commonly used source for indium in the metalorganic chemical vapor deposition of InGaN, is relatively low,[107] and exceeding it significantly can cause parasitic pyrolysis and particle nucleation in the boundary layer above the growing surface.[108] In addition, the evaporation rate of In increases rapidly with temperature,[109] so to grow InGaN with significant In content also requires lower growth temperatures.[110] Finally, overall thermal budget favors lower growth temperatures: active region quality, e.g., can degrade during subsequent growth of p-side contact layers.[111]

Options for circumventing these two issues (first, lattice mismatch and strain, and, second, the small window in epitaxial growth temperature) fall into four categories: i) "living with" the lattice mismatch by moving to crystal orientations with lower or vanishing piezoelectric fields; ii) reducing the lattice mismatch by moving to substrate/epitaxy combinations which are (more closely) lattice matched; iii) moving to lower dimensional substrate/epitaxy combinations that facilitate coherent strain relaxation; iv) novel epitaxial growth techniques that might open the window in epitaxial growth temperature. In the remainder of Section 3.2, we discuss each of these approaches in turn.

3.2.1. "Living with" Substrate-Epitaxy Lattice Mismatch: Non- and Semi-polar Orientations

As just discussed above, InGaN materials, typically grown lattice-mismatched and coherently to GaN buffers layers, are highly strained and, on the also typical c-plane orientation, have large piezoelectric (>1 MV/cm) fields. These fields cause a spatial separation of the electron and hole wavefunctions in quantum well (QW) active regions, thereby reducing recombination rates and pushing peak efficiency to lower current densities. Because the piezoelectric fields depend on strain, all recombination rates are lower the higher the In-content and the longer (moving towards the green and yellow) the emission wavelength.

As discussed in Section 3.1 in the context of blue InGaN light emitters, one route to overcoming these piezoelectric fields is similarly to "live with" the substrate/epitaxy lattice mismatch and hence "live with" the high strain associated with medium-In-content green-emitting materials, but to move to non-polar or semi-polar crystal orientations with lower piezoelectric coefficients. Then, all recombination rates would be higher.

As one potential benefit, efficiency droop, which is exacerbated at these green wavelengths relative to blue, might be pushed out to higher input power densities. Again, as discussed in Section 3.1, carrier densities would be lower for a given input power density, and at these lower carrier densities Auger recombination, with its n^3 dependence, would be slower relative to radiative recombination, with its n^2 dependence. As another potential benefit, the quantum-confined Stark shifts in emission wavelength due to these carrier-density-dependent and hence input-power-density-dependent piezoelectric fields would also be reduced.

Early efforts to realize these potential benefits focused on non-polar (m-plane) LEDs, where the piezoelectric fields are entirely eliminated. However, indium incorporation was found to be substantially lower than for c-plane,[112] making the demonstration of high performance longer wavelength LEDs extremely challenging. In contrast, semipolar planes, particularly (11–22) and (20–2–1), were found to have higher indium incorporation efficiencies. Higher indium incorporation efficiencies enable higher growth temperatures to be employed to achieve the same indium composition, with the accompanying potential for higher quality material.

At this point in time, semipolar LEDs have demonstrated impressive performance over a wide wavelength range. Blue (458 nm) LEDs employing a [10–1–1] orientation and with

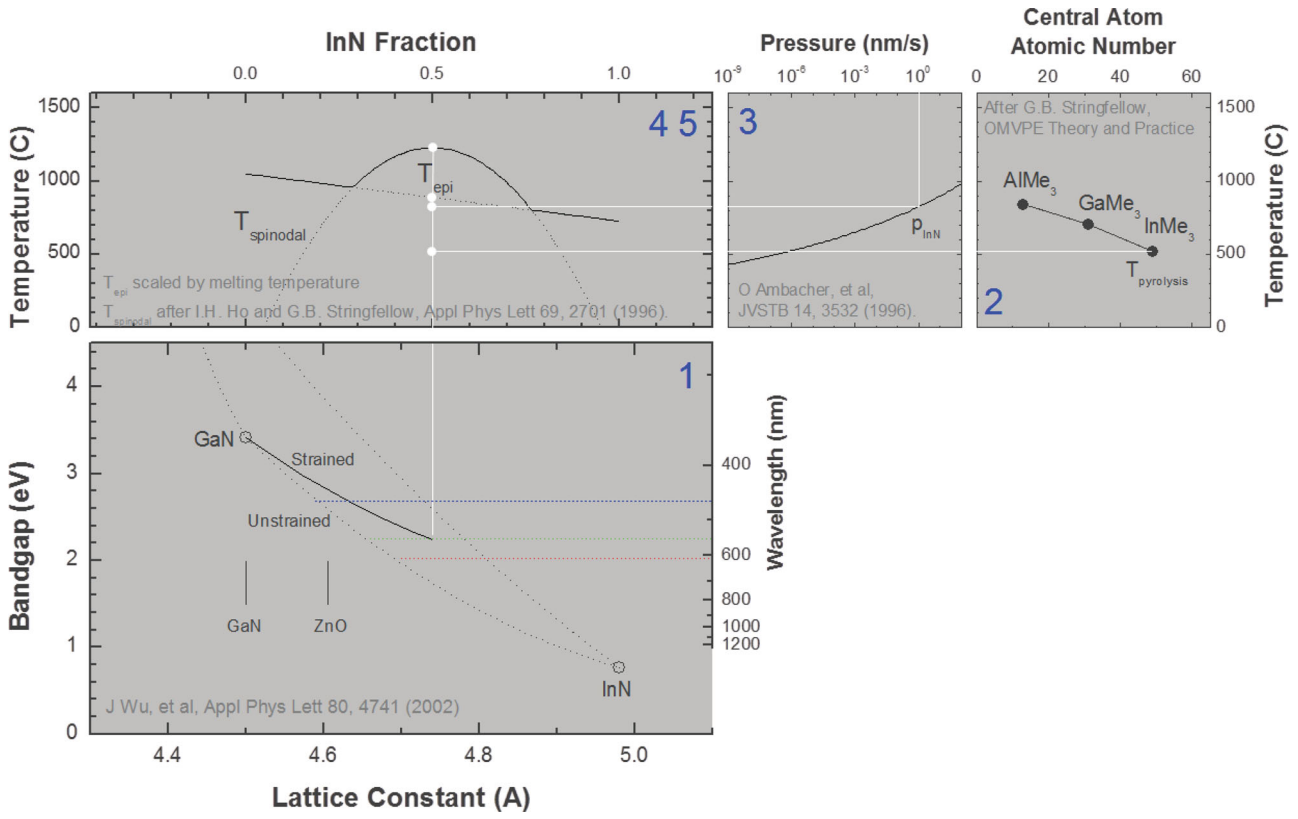

Figure 9. Some of the possible origins of the "green gap." (1) InGaN bandgap decreases as In content and lattice constant increase; at the bandgap required for deep-green emission, the InGaN lattice mismatch relative to GaN is very large. (2) Pyrolysis temperatures of trimethylaluminum, trimethylgallium and trimethylindium. (3) InN vapor pressure over InGaN, in units of thickness (nm) of material evaporated per second. (4–5) The temperature $T_{spinodal}$ above which InGaN will not phase separate, and the epitaxial growth temperature T_{epi} desirable for high crystal quality. The vertical white line in panels 1 and 4 is the lattice constant of unstrained InGaN at the composition necessary for green light emission at 548 nm. It would be desirable to minimize spinodal decomposition and maximize epitaxial material quality, and thus to grow at temperatures above the intersection of this vertical line and $T_{spinodal}$ and T_{epi}. The vertical white line in panel 3 is the temperature at which InN evaporation is comparable to typical nm/s growth rates. It would be desirable to grow at temperatures below the intersection of this rate and p_{InN}.

optimized light extraction designs have yielded peak EQE > 50%, comparable to state-of-the-art c-plane LEDs.[113] The use of [20,21] templates resulted in EQEs of 20% at green (516 nm) and 12% for green-yellow (552 nm) emission.[114] Indeed, further studies revealed the (20–2–1) semipolar orientation as having beneficial properties beyond the high In incorporation efficiency mentioned above. In particular, LEDs employing this orientation were found to have reduced linewidths, reduced wavelength shift with injection current, and reduced efficiency droop compared to LEDs in other semipolar orientations.[115,116] It is hypothesized that at least two factors contribute to these superior properties.

First, improved material homogeneity, from reduced indium composition fluctuations and/or well width fluctuations, may substantially reduce carrier localization. Because localization of carriers reduces the effective active region volume and leads to higher carrier densities at a given current,[117] avoiding this phenomenon may delay efficiency droop. Notable performance includes 424 nm, blue (20–2–1) LEDs with 52% peak EQE, and EQE as high as 45% at a current density of 200 A/cm^2, representing only 14% efficiency droop.[118]

Second, unlike for c-plane, (11–22), and (20–21) orientations, piezoelectric fields in (20–2–1) QWs are in a direction that opposes the p-n junction field, resulting in near-flat-band conditions at low bias conditions. This fortuitous field alignment largely avoids QCSE-related wavelength shifts and wavelength shifts as small as ca. 2 nm up to 200 A/cm^2 have been reported.[115]

3.2.2. Reducing Substrate/Epitaxy Lattice Mismatch: Alternative Substrates

Instead of "living with" the high strain associated with the substrate/epitaxy lattice mismatch, another route toward higher efficiency is to reduce the lattice mismatch by moving to substrate/epitaxy combinations that are lattice matched.

Homoepitaxial Substrates (InGaN Buffer Layers and Epitaxial Lateral Overgrowth [ELO]): For example, one might imagine replacing the GaN buffer-layer template normally grown on sapphire or Si (111) substrates with InGaN buffer-layer templates more closely lattice-matched to the medium-In-content InGaN QW structures necessary for green light emission. While this approach sounds straightforward, achieving thick, strain-relaxed InGaN templates with both low defect densities and smooth surface morphologies presents a tremendous challenge. Unlike GaN, direct growth of InGaN on sapphire substrates using a

thin, intermediary buffer layer typically yields films with high dislocation densities (>10^{10} cm^{-2}), rendering them unsuitable as templates for high performance LEDs.[119]

An alternative approach is "standard" growth of InGaN on lattice mismatched GaN-on-sapphire templates combined with strategies to relax InGaN strain while mitigating the accompanying defect formation. One of the recent advances in this area is epitaxial lateral overgrowth (ELO) of InGaN on periodically grooved GaN templates.[120] ELO involves epitaxial growth on templates (e.g., GaN) that are patterned, typically with dielectric stripes or by etching a stripe pattern directly into the template. Vertical growth combined with lateral growth between the stripes results in a coalesced epitaxial film, and provides a mechanism for turning or redirecting threading dislocations away from the overlying active region. Application of these approaches to GaN has been highly successful, yielding thick, coalesced GaN films with dislocation densities <5×10^7 cm^{-2}.[121] It might also be possible to do additional dislocation filtering afterwards through subsequent nanopatterned overgrowth.[122,123]

Application of these approaches to InGaN has been more difficult. InGaN ELO on *c*-plane and *a*-plane GaN orientations, e.g., leads to faceting of the InGaN surface.[124] Nevertheless, some progress has been made with ELO on *m*-plane GaN templates, which yield significantly smoother surface morphologies. Thick (7 μm) coalesced $In_{0.07}Ga_{0.93}N$ films with almost complete strain relaxation and low threading dislocation densities (<10^8 cm^{-2} in select regions) have now been realized on such *m*-plane GaN templates.[125] Furthermore, the photoluminescence intensity of InGaN multiple quantum wells on these templates is reportedly higher than that on GaN ELO templates (which require a relaxed InGaN overlayer). This relative photoluminescence intensity improvement increases with increasing indium composition (and related strain) in the quantum well layers, reaching a factor of 2.1x improvement for ca. 535 nm green-emitting quantum wells.[126] Beyond these experimental results, modeling of spontaneous emission rates of green and red emitting InGaN quantum wells on $In_{0.15}Ga_{0.85}N$ templates predict va. 2.5–3.2× enhanced performance over conventional GaN templates. These enhancements are largely due to reduced piezoelectric fields.[127]

Heteroepitaxial Substrates (e.g., ZnO): Instead of InGaN buffer-layer templates grown on the usual sapphire or Si (111) substrates, one might imagine replacing the sapphire or Si (111) substrates with alternative substrates better lattice-matched to the medium-In-content InGaN compositions necessary for green light emission. ZnO is perhaps the most interesting candidate, having a similar wurtzite crystal structure to InGaN, and it is lattice-matched to $In_{0.18}Ga_{0.82}N$ alloys grown in the [0001] crystal orientation.[128] However, a major roadblock to achieving high-quality InGaN on ZnO is the thermal incompatibility of the two materials, as metal-organic chemical vapor deposition (MOCVD) of InGaN is ideally performed at temperatures >700 °C to achieve higher material quality, whereas ZnO decomposition and formation of interfacial alloys occurs at temperatures <600 °C. Although this incompatibility seems formidable, three recent areas of exploration have been fruitful.

First, non-polar ZnO substrates have been shown to resist decomposition to higher growth temperatures than *c*-plane ZnO. Accordingly, MOCVD-grown $In_{0.25}Ga_{0.75}N$ films on both *a*-plane and *m*-plane ZnO have shown no degradation in structural quality up to 650 °C.[129] Growth on non-polar substrates provides an additional benefit in eliminating piezoelectric fields.

Second, the use of lower-growth-temperature "intermediary" layers between the ZnO substrate and InGaN film has provided significant benefit.[130] In particular, previous work employing pulsed laser deposition has shown that an intermediary layer of GaN grown at room temperature can reduce the interdiffusion of Zn, O and other elements.[131] Employment of this intermediary layer has led to 500-nm-thick m-plane $In_{0.33}Ga_{0.67}N$ films, grown at 660 °C, that maintain a significant degree of strain and show no signs of phase separation. Further studies using a GaN layer grown at 200 °C explored indium compositions from 0.24–0.43 and reported a critical thickness for strain relaxation some 10x higher than InGaN films grown by other techniques.[132] Additional studies have shown that an Al_2O_3 interlayer effectively reduced diffusion of Zn into the InGaN epilayer.[133]

Third, pulsed laser deposition (PLD) has demonstrated some success with InGaN/ZnO growth. Here, an In_xGa_{1-x} eutectic target was laser ablated in the presence of atomic nitrogen. As one benefit, the ablation process is thought to impart enhanced kinetic energy to group-III atoms on the growing surface, enabling improved materials quality at lower growth temperatures and even atomically smooth films at room temperature.[134] PLD growth of GaN on *m*-plane ZnO further suggests that GaN or InGaN epilayers deposited at room temperature are viable buffer layers, suppressing ZnO decomposition during subsequent growth at temperatures up to 700 °C.[135] Overall, these reports present clear progress toward viable InGaN templates, but further studies of the structural quality of these films are still needed.

3.2.3. Facilitating Substrate/Epitaxy Strain Relaxation: Nanowires

Instead of either "living with" the high strain associated with the substrate/epitaxy lattice mismatch or reducing the lattice mismatch by moving to substrate/epitaxy combinations that are lattice matched, another route is to move to lower dimensional substrate/epitaxy combinations that facilitate strain relaxation.

In particular, one-dimensional nanowires (or nanorods) with small radial dimensions (ca. 10–200 nm) and typical axial dimensions >1 μm, are one such candidate currently under investigation.

During axial growth, highly anisotropic nanowires can relax elastically in the lateral direction, thereby eliminating strain nearly entirely and accommodating much higher initial substrate/epitaxy lattice mismatches than could be accommodated in 2D planar heterostructures before misfit dislocations are formed.[136] Indeed, single-crystal nanowires entirely free of threading dislocations have been demonstrated on a wide range of highly lattice-mismatched substrates, including silicon.[137] Single-crystal $In_xGa_{1-x}N$ nanorods have been demonstrated over the entire compositional range[138] with photoluminescence intensities decreased by only a factor of 1.7× in going from In fractions of x = 0 to x = 0.6 (spanning near-UV to red wavelengths).

During radial growth of a lattice-mismatched nanowire shell on a nanowire core, if the core diameter is extremely small (10 nm or so), the lattice mismatch can be accommodated at least in part by the core rather than by the shell. Such "compliant substrate" phenomena can in principle increase critical shell diameters before the onset of misfit dislocations, even to infinity for small enough core diameters.[139] The experimental challenge here is to create nanowire cores with such small diameters.[140]

Note that, beyond strain relaxation, nanowires have other potential advantages.

One such potential advantage of nanowires is higher optical efficiency. Using RF plasma molecular beam epitaxy, Nguyen et al. have demonstrated "dot-in-a-wire" heterostructures where the InGaN emitting layers are embedded in a GaN matrix, providing an energy barrier between these layers and the nanorod sidewall. This superior carrier confinement greatly reduces carrier interaction with any surface defects that might be present, and has led to LEDs with emission into the deep red region (ca. 650 nm), with reported internal quantum efficiencies of >30%.[141]

Another such potential advantage of nanowires is the ability to tailor emission wavelengths laterally across arrays of such nanowires – a potential path toward monolithic integration of multi-color LEDs on a single chip.

In axial-quantum-well geometries, Sekiguchi et al. have achieved this control by selective area growth of InGaN/GaN nanorods through Ti-masks of controlled aperture sizes, thereby varying the In and Ga precursor fluxes to the quantum well.[142] Photoluminescence wavelengths across the visible spectrum, spanning from blue (479 nm) to red (632 nm), were demonstrated for a 137–270 nm range of nanorod diameters. Thus, in a single growth run, spatially separated nanorod LED arrays with different emission wavelengths can be demonstrated with appropriate design of Ti-mask apertures, presenting a pathway for monolithic RGB emitter arrays.[143] In radial-quantum-well geometries, InGaN-QW shells on GaN cores have been investigated due to the potentially high active region area on the m-plane sidewalls when compared to axial-QW nanorod or planar architectures. Hong et al. reported core-shell nanorod LEDs with multi-color tunable emission in the range 500–690 nm as the forward bias was decreased from 10 to 3.0 V.[144] At lower forward biases, the long-wavelength emission was determined to originate from the c-plane QWs (In ≈ 0.6) at the nanorod tips, whereas at larger voltages the shorter wavelengths were observed from the m-plane sidewall QWs (In ≈ 0.15). High-In-content-shell QWs on the m-plane sidewalls have not been demonstrated to date, although strong emission into the infrared has been observed for thick InGaN shell layers grown on a-axis triangular nanowires.[145]

The ability to tailor emission properties of nanowire structures has further led to notable advances in "phosphor-free" white LEDs with high color rendering properties. Connie et al. have applied dot-in-a-wire heterostructure designs to achieve CRI > 90% in both warm and cool white regions.[146] Similar designs have been applied to achieve white LEDs with internal quantum efficiencies >56%.[147]

Given these impressive advances, two primary roadblocks to more manufacturable nanorod-based white LEDs remain. First, while significant progress has been made in controlling the emission wavelength of nanorod LED arrays, e.g., through selective area growth using controlled aperture sizes, the nanowire emission spectra remain relatively broad for a given "color" nanorod array. Second, low-cost processing of these nanostructured materials into a viable LED device is a challenge. To date, numerous approaches have been developed to planarize nanorod LED arrays for more facile processsing. One strategy involves growth of InGaN "quantum disks" along the growth direction of a GaN nanowire followed by lateral growth and coalescence of the overlying p-GaN layers to present a planar top layer for electrical contacting.[148] Similar axially grown LEDs use a spin-on-glass layer to encase the nanowire array, thereby providing a support structure for top metal contacting.[149] Finally, LEDs with a radial core-shell geometry have employed lateral growth of p-type overlayers to enable a fully coalesced LED heterostructure.[150]

Despite still being largely in the research realm, these demonstrations show the long-term potential for nanostructured InGaN LED solutions in the green-yellow-red spectral region. Even in the near term, lower-dimensional structures present opportunities for studying fundamental properties of InGaN heterostructures with light emission across the entire visible region, with distinctly different defect, strain, and polarization properties compared to 2D planar structures.

3.2.4. Novel Epitaxial Growth Techniques

A final difficulty is the small window in epitaxial growth temperature associated with the medium-In-content InGaN alloys necessary for deep-green light emission. On the one hand, as discussed above, lower growth temperatures are preferred to avoid parasitic pyrolysis and particle nucleation during metalorganic chemical vapor deposition, and to avoid evaporation of InN. On the other hand, higher growth temperatures are preferred to circumvent InN-GaN immiscibility, to maintain adatom surface mobility typically necessary for high crystalline quality, and to dissociate the ammonia (NH_3) that is the most common source of nitrogen in the metalorganic chemical vapor deposition of InGaN.

Opening up the window on the higher growth temperature end may require new In-containing precursors with higher pyrolysis temperatures as well as higher pressures to suppress InN evaporation. Opening up the window on the lower growth temperature end may require new N-containing precursors with lower pyrolysis temperatures, alternative (e.g., catalyst) methods for cracking ammonia, slower growth rates to accommodate lower adatom surface mobility, and device architectures that are less sensitive to phase separation.

3.3. Red Materials and Devices

As illustrated in Figure 3, presently, the highest-efficiency semiconductor materials system for red emission is quaternary $(Al_xGa_{1-x})_{1-y}In_yP$. In the deep red (650 nm or so), these semiconductors have near-100% internal quantum efficiency. For ultra-efficient SSL, however, the red wavelength must be much

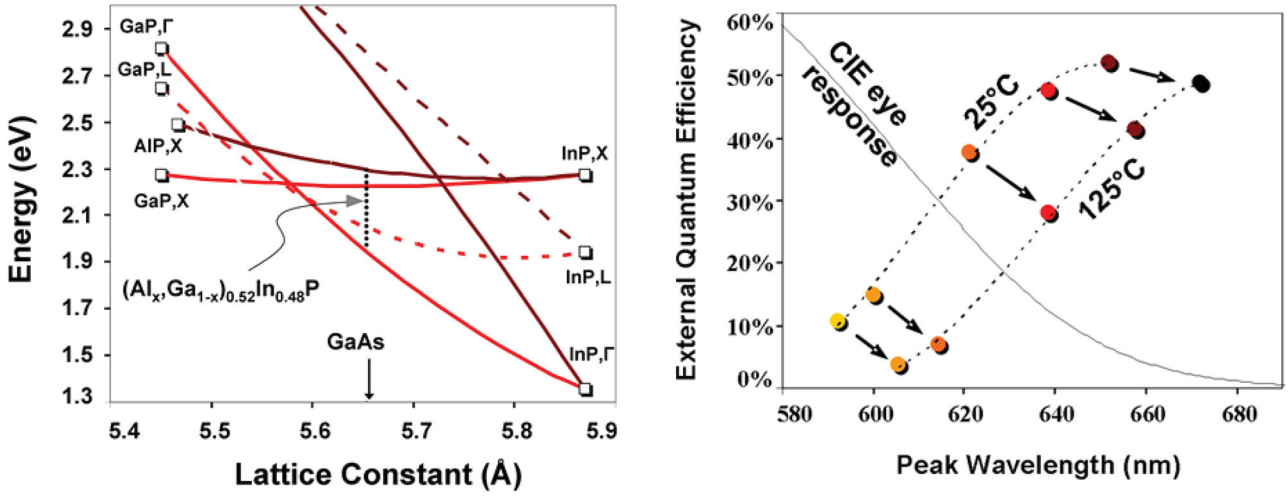

Figure 10. Left: Bandstructure of the $(Al_xGa_{1-x})_{1-y}In_yP$ material system. The system is lattice-matched to GaAs at y = 0.48. Right: External quantum efficiency of AlGaInP truncated-inverted-pyramid (TIP) LEDs (350 mA) as a function of peak emission wavelength, both for room temperature and 125 °C junction temperature. Also shown is the relative change in eye sensitivity with wavelength. Adapted with permission.[72] Copyright 2007, Wiley.

shorter, on the order of 614 nm. Though LEDs in this material system span the ca. 650 (deep red) to ca. 580 (amber-yellow) nm range, their efficiencies drop off rapidly at wavelengths shorter than 630–640 nm. Three issues are responsible for this drop-off.

The first issue is related to bandstructure, and is illustrated in the bandgap vs. lattice-constant diagram in the left panel of **Figure 10**.[103] The composition of the quaternary $(Al_xGa_{1-x})_{1-y}In_yP$ is characterized by two parameters, y and x. The first parameter, *y*, is normally fixed near $y = 0.48$, as this is the composition which is latticed matched to the GaAs substrates normally used for growth. The second parameter, *x*, determines the bandgap energy and type: as x increases the bandgap becomes wider but, importantly, also undergoes a crossover from direct (Γ valley) to indirect (X valley). The crossover occurs at $x \approx 0.55$, corresponding to an energy ca. 2.24 eV and an emission wavelength ca. 553 nm.[151] For compositions beyond $x \approx 0.55$, carriers preferentially occupy the (indirect) X valley, hence radiative recombination rates are slow, and alternate non-radiative recombination pathways become competitive. Even worse, because of higher densities of states in the indirect valleys than in the Γ valley, occupation of the indirect valleys (particularly at room or higher temperatures) is significant, even at alloy compositions much less than $x \approx 0.55$.

The second issue is related to leakage of injected electrons, particularly from the active layers into p-type confining layers (typically Mg-doped layers with alloy compositions of x ~ 0.7 to 1). The cause is the fairly light effective mass of electrons in combination with increasingly smaller conduction band offsets, which range from ~240 meV (the offset between the conduction bands of InGaP and AlInP) to ca. 90 meV for 580 nm (the offset between the conduction bands of $(Al_{0.3}Ga_{0.7})_{0.52}In_{0.48}P$ and AlInP). This can lead to a high level of diffusive and/or drift-based current leakage, depending on the operating current density of the device.[152]

The third issue is related to strong and deleterious temperature effects. The right panel of Figure 10 shows the external quantum efficiency of truncated inverted-pyramid (TIP) LEDs as a function of peak emission wavelength.[103] The left-most curve is the trend at room temperature, while the right-most curve is the trend at 125 °C, a typical operating temperature for high-power LEDs used for solid-state lighting. The drop in external quantum efficiency with increasing temperature is due to a combination of indirect-minima population and electron leakage, as mentioned above. Moreover, as also seen from Figure 10, there is a significant wavelength shift with 100 °C change in temperature, due to crystal lattice dilation and temperature-dependent electron-phonon interactions. The shift is significant, leading, for example, to an 18-nm shift from 630 to 648 nm as temperature increases from 25 °C to 125 °C, and a significant decrease in red lumens due to reduction in eye sensitivity. In order to "recover" the desired wavelength at operating temperatures (and to maintain equivalent lumens), a higher AlInP mole fraction active layer is required. For example, for 620 nm emission at operation, the room temperature target wavelength might be 605 nm. However, the higher AlInP mole fraction results in a significant drop in efficiency as just discussed.

In the remainder of Section 3.3, we discuss two routes, both challenging, for circumventing these three issues.

3.3.1. "Living with" AlInGaP: Enhanced Spontaneous Emission

A first route might be to "live with" AlInGaP materials as they currently exist; after all, there have been striking performance improvements in recent years, including research results from Osram of >50% wallplug efficiency (168 lm/W at 35 A/cm^2 and a dominant wavelength of 609 nm).[153] Nevertheless, continued performance improvements may require new methods for altering the competition between spontaneous emission in the direct valley and either non-radiative recombination in the indirect valley or carrier leakage. Here we discuss three possible methods.

Strain Engineering: A first method is strain engineering, which might enable band offsets to be shifted so as to reduce

carrier leakage, or might enable the relative positions of the direct and indirect valleys to be shifted so as to enhance direct-band radiative recombination.

One possibility might be to use substrates with a lattice constant different from that of GaAs, and hence to enable growth of alloys with compositions different from $y = 0.48$. There has been much recent success with metamorphic lattice-constant-shifted substrates for solar cells; perhaps similar lattice-constant-shifted substrates could be used for AlInGaP light emitters.

Another possibility might be to use compliant nanostructures, like nanowires, which can accommodation large strains and hence also compositions different from $y = 0.48$. We discussed this possibility in Section 3.2 for green-emitting InGaN nanowires; there may be similar opportunities for shallow-red-emitting AlInGaP nanowires.

Plasmonics: A second method is plasmonics, which is being explored to enable enhanced light extraction and improved IQE at visible wavelengths.[154,155] These devices circumvent radiative recombination in the QW region and instead invoke ultra-fast energy transfer from the QW to surface-plasmon (SP) modes in a neighboring metal surface coating.[156] With appropriate nanoscale structure in the metal, the SP excitations can then scatter, lose momentum, and couple to radiation modes. Through this SP-mediated process, LED efficiency can potentially be enhanced in multiple ways. Coupling to SP modes aids in extracting photons that would have otherwise suffered total internal reflection.[157] In addition, the fast energy transfer to the SP, promoted by the high SP density of states at resonance, may increase IQE by allowing more effective competition with non-radiative processes in the QW.

Whether SPs ultimately enhance or reduce the overall emission efficiency is critically dependent upon the losses of the SP mode and the efficiency with which radiation can be extracted from SP modes compared to the intrinsic radiative efficiency of the semiconductor material. In the blue and green, Joule losses, even in the best metal, Ag, are significant. In the red, however, Joule losses may be manageable. Much more work has been reported thus far on blue and green LEDs than red LEDs, however, so this would appear to be an open and relatively unexplored area.

Note that, for the purposes discussed in this article, it is not sufficient to increase efficiencies from the few percent range to the tens of percent range, something that many groups have demonstrated at various wavelengths. One must increase the efficiency of red LEDs from the tens of percent range to the ultra-efficient 70%-or-higher range, and this is a much more daunting, though perhaps surmountable, challenge. Moreover, even if in idealized geometries (very small semiconductor volumes optimally placed with respect to a nearby Ag antenna) ultra-efficient red light emission could be demonstrated, transitioning to practical structures (with larger semiconductor volumes) adds formidable challenges.

First, all LEDs employ thick (>100 nm) p-type layers between the QWs and the surface metal. Substantially thinner p-type layers would likely be needed to place the QW within the fringing field of the SP mode; however, thin p-type layers can lead to device shorting and electrical performance degradation.

Second, the resonance-like nature of the enhancements imply that, for large active semiconductor volumes, some of the volume might be enhanced and some might not be. Small and tailored active semiconductor volumes might thus be necessary. However, light emitters in the regime of solid-state lighting are typically heat-sink limited, not active-volume limited,[158] so there is likely room for decreasing active volumes and increasing local current densities while maintaining large-scale current densities in a heat-sink-manageable regime.

Photonic Crystals: A third method is the use of photonic crystals to enhance spontaneous emission.[159] Photonic crystal LEDs (PXLEDs) are fabricated by periodic patterning of the semiconductor material, or an overlying dielectric layer, with feature sizes on the order of the wavelength of emitted light.[160] Such structures have the potential to modify the photonic density of states and to thereby enhance spontaneous recombination rates (and the related IQE). We note that such structures also have the potential for improving light extraction,[161] but here we focus on their more important potential for improving radiative efficiency.[162] For this purpose, either 2D "bulk" PhC or cavity PhC based LED designs[163] are possible.

In the case of PhC cavities, the radiative-rate enhancement is achieved using the Purcell effect.[164] The Purcell enhancement factor (F_p) is given by:

$$F_p = \frac{3Q\lambda^3}{4\pi^2 V_m n_{eff}^3} \tag{8}$$

where Q is the quality factor, V_m is the mode volume, and n_{eff} is the effective index of the cavity. With a realistic Q of ca. 500 and mode volume of $0.5\lambda^3/n^3$, a Purcell factor ~75 can be achieved. However, due to a confined mode distribution, only the emitters at the cavity antinode participate in emission, while the rest do not. As discussed above in the context of plasmonic enhancements, however, small cavity volumes might not be as disadvantageous as one might think, as in most solid-state lighting architectures light emission is typically heat-sink-limited, not active-volume-limited.

In the case of bulk PhCs, more emitters can be involved in the emission process. The radiative rate of an emitter given by Fermi's golden rule is proportional to the local photonic density of states (LPDOS):

$$\Gamma_r = \frac{\pi\omega\mu^2}{3\varepsilon h}\rho_L(\omega, r) \tag{9}$$

where μ is the dipole moment, ε is the dielectric constant, r is the propagation direction, ω is the angular frequency, and $\rho_L(\omega,r)$ is the local photonic density of states. By increasing the LPDOS, one can boost the radiative rate relative to the non-radiative rate, increasing the radiative efficiency.

We note, though, that large enhancements may require difficult-to-fabricate 3-dimensional logpile PhCs.[165] An attractive, albeit lower LPDOS alternative, might be a 2 dimensional (2D) structure. Also note that, regardless of dimensionality, for all PhC LEDs, cavity or bulk, a major difficulty will be achieving electrical injection. Since photonic crystals are typically etched into the semiconductor, there is potential for degradation of

LED performance through current leakage, surface recombination, or damage to the QW layers.

3.3.2. Beyond AlInGaP: Alternative Red-Emitting Semiconductors

A second route is to go beyond AlInGaP materials, to alternative semiconductor materials with potential for greatly improved performance. Such materials must have a few key properties: high radiative efficiency; stable emission wavelength and efficiency at elevated operating temperatures and high current densities; and peak wavelengths of ~614 nm and narrow linewidths. In the following, we examine three promising (though extremely challenging) red-emitting semiconductor materials.

InGaN: Just as it is for the deep green, InGaN is a potential candidate for the shallow-red: unlike AlInGaP, InGaN is a direct bandgap material throughout the visible region. Indeed, early work by Nichia demonstrated InGaN LEDs at amber (594 nm) and deep red (675 nm) wavelengths,[166] and amber InGaN LEDs do not lose efficiency with increasing temperature as much as do amber AlInGaP LEDs (due to increased carrier confinement and less carrier loss to indirect energy bands).

However, achieving the high indium content necessary for such long wavelength emission is challenging. Long-wavelength planar InGaN LEDs are highly strained, defective, and radiatively inefficient. They also suffer from broad (≥50 nm) linewidths and significant blue-shifting of emission wavelength with injection current (e.g., 675–580 nm over the 20–100 mA range). Strong compositional inhomogeneities contributed to the broad linewidths, and both bandfilling of low-energy states (In-rich regions) and screening of large internal polarization fields contribute to the large wavelength shifts with current.

Nonetheless, methods similar to those being considered for green InGaN LEDs may enable reduction in strain and improvement in performance. Among the most promising methods is InGaN nanowires. As highlighted previously, axial geometry NWs have reached red emission wavelengths by tuning of NW diameter,[109] and "dot-in-a-wire" structures at 650 nm have internal quantum efficiencies >30%.[141] More recently, "disks-in-nanowire" LEDs, grown on Si, have shown emission to 650 nm and achieved radiative efficiencies up to 52% with optimized sidewall passivation.[167] Such progress is encouraging, though a remaining challenge is the broad linewidths (ca. 50–70 nm) that markedly degrade luminous efficacy and that are largely due to NW inhomogenieties.

GaPN: GaP is a well-known indirect bandgap semiconductor. Undoped, its optical emission efficiency is low. Doped with isoelectronic nitrogen (N) impurities, however, highly localized states emerge slightly below the conduction band minimum in GaP, leading to "quasi-direct" bandgap behavior and improved radiative efficiency.[168] Moreover, as N concentrations increase from relatively low doping levels ($N < 10^{17}$ cm^{-3}) to the several percent range, large bandgap bowing enables emission at yellow-amber-red wavelengths,[169] with red LEDs at 670 nm demonstrated for N ~ 1.1%. Among the benefits to GaPN (over AlInGaP) alloys are an acceptable lattice mismatch with transparent GaP substrates, and their reduced bandgap shift with temperature.[170]

Despite these advances and potential benefits of GaNP LEDs, significant challenges remain to meet the requirements the "ideal" red-emitter for SSL. Radiative efficiency remains low, and spectral linewidths are extremely broad (>50 nm vs. ca. 15 nm for AlInGaP).[171] As N concentration increases, there is a complex evolution of N-related optical transitions through a heirarchy of impurity complexes including NN pairs and N-clusters.[172] There is also a predicted transition from an indirect to direct bandgap at a concentration of ca. 0.03,[173] with evidence for direct-bandgap-like characteristics and strong room temperature photoluminescence at concentrations as low as 0.0043.[174] Understanding and controlling these various N-concentration-dependent phenomena will not be trivial.

II–VI Semiconductors: Among the II–VI semiconductor alloys, the CdMgZnSe alloy is particularly promising. It can be lattice matched to InP, has direct energy bandgaps, and emits at wavelengths that span the entire visible spectrum. Most important for the ideal red emitter for SSL, its emission linewidth is narrow: ca. 15 nm.

Indeed, heterostructures grown from these alloys have been demonstrated by Haase et al. to be excellent color converters: they can be optically pumped by a blue LED to yield green (545 nm), yellow (581 nm) and red emission (626 nm).[175] Due to the high internal quantum efficiency of these materials, color-conversion efficiencies in the 60–70% range have been demonstrated, limited mostly by the Stokes loss associated with all color converters as well as some losses related to parasitic absorption of the blue LED pump light.

For ultra-efficient SSL requiring high-brightness electroluminescence, however, the main challenges are two. First, p-type doping[176] must be developed; and second, precise control of strain during epilayer growth must be developed, to avoid stacking faults and related dark line defects which compromise reliability.[177]

Trade-Offs Amongst Alternative Materials: Overall, each of these three alternative red materials (InGaN, GaPN, CdMgZnSe) has potential, but has significant challenges if it were to become the "ideal" red emitter. Both InGaN nanowires and GaPN planar films have demonstrated electroluminescence in the amber-red range and reduced bandgap sensitivity to temperature compared to AlInGaP alloys; however, their inhomogeneous broadening is much too large. CdMgZnSe alloys have demonstrated high radiative efficiency, bandgap tunabilty throughout the visible region and narrow emission linewidths; however, they cannot yet be p-type doped.

4. Smart Solid-State Lighting: Towards Control of Flux and Spectra in Time and Space

In Section 2, we discussed the characteristics, possible system-level applications, and economic benefits of smart and ultra-efficient solid-state lighting. Then, in Section 3, we focused on ultra-efficiency: its prospects and challenges at the materials, physics and device level. In Section 4, we turn our attention to smartness: its prospects and challenges at the device and lamp architecture level. Throughout, we assume that the ancillary technologies necessary to enable smart lighting will be making parallel and perhaps even faster progress. These technologies

include the distributed wireless sensors needed to monitor light fluxes, and the networked communications and intelligence needed to act on the sensory information to control the quantity and quality of light in space and time.

We note that the device and lamp architectures that would best enable realization of smart solid-state lighting will necessarily be different from those currently being developed for the presently dominant phosphor-converted white LED architecture. However, we by no means mean to disparage such architectures. There is currently rapid and disruptive technical innovation throughout the solid-state lighting supply chain as LED light bulbs (incandescent replacements) and LED panel lighting (fluorescent fixture replacements) begin to enter the market. We envision in Section 4, however, a world in which the architectures intended to displace existing lighting applications will themselves be displaced with more powerful architectures intended to provide new smart-lighting functionality: architectures that enable real-time digital electronic control of the spectral, spatial, and temporal properties of light.

To achieve these functionalities in a cost-effective manner, we envision device and lamp architectures in which the most important performance aspects are integrated into small packages. This does not necessarily mean that the final light source or luminaire will be small: as discussed in Section 2.2, small devices and lamps can feed light into waveguides incorporated into much larger fixtures for diffuse lighting. But if the light engine is small, the challenges associated with smartness—controlling light flux and spectra in time and space—become easier.

In the following, we discuss in turn three levels of integration that would enable smartness in small packages: optical integration, optoelectronic integration, and optomechanical integration.

4.1. Optical Integration: Mixing Colors while Maintaining Low Etendue

The first kind of integration, optical, has to do with lamp architectures that are consistent with control of spectra in space.

In fact, there are already now emerging lamp architectures which give control of spectra (but not of spectra in space). Among the most promising are those which employ both blue and red (AlInGaP) LEDs combined with a green phosphor, with the red LED added to tune the color temperature of the illumination source.[178] This architecture does already require some level of feedback (especially temperature feedback, but also chromaticity feedback) due to differential temperature and aging performance of the blue and red LEDs. Currently, this architecture presents a price/performance tradeoff, but as the cost of solid-state lighting systems continues to drop and new applications for solid-state lighting develop, incorporation of active color-control systems in multi-wavelength LED lighting modules will likely become routine.

Control of spectra in space using multiple LEDs of different colors, however, requires both some form of color mixing as well as some way of maintaining a small effective source size whose light can be subsequently collimated or imaged using secondary optics. This is a significant challenge, not least because the different colors must be blended efficiently without color breakup (color fringing around shadows, for example). Among the possible approaches, here we mention three.

First, a variant of the current architecture would be to arrange three or more discrete LED chips of the primary colors red (R), green (G), blue (B) side-by-side, but extremely closely, so that a very small diffusing element could be placed around them. For LEDs of approximately 1 mm^2 size scale, the diffusing element (reflective mixing elements are typically most efficient) would need to be several mm^2 to perhaps cm^2 scale to avoid color inhomogeneity and colored edge effects in the beam.[179] This may be sufficient for some applications, but is less than ideal. Indeed, optical mixing losses associated with this solution are likely to be high, perhaps as high as 20% (though it may be possible to moderate some of these losses with graded-refractive-index optical systems).

Second, one might consider a vertically-stacked-LED strategy with red at the bottom and blue at the top. Because the multi-quantum wells (MQWs) in the upper LED chips are transparent to the lower LED chips due to their bandgaps, absorption losses might be reasonably small while maintaining good color mixing. Such an architecture was recently studied using LED chips with truncated inverted pyramid geometries fabricated by laser micromachining.[180] To maintain good efficiency, sidewall-emitted light was "recycled" into top-emitting light by coating the sidewalls of the red, green and blue LED chips with a thin layer of highly reflective silver film.

Third, if highly collimated (e.g., laser) sources are used for the individual colors, one might consider dichroic or multichroic beam combiners, followed by imaging optics.[181] Or, one could imagine a Cassegrainian architecture, in which a small diffusing element is placed at the focus of a conical parabolic reflector, then struck by three or four laser beams of different colors which enter through small holes in the reflector.

Other approaches are surely possible. It is not unlikely that approaches that are compatible with backlighting for displays will advance most quickly (e.g., edge-lit waveguides work very well as color mixers for diffuse lighting solutions). The color gamut of displays is improved significantly when backlit by narrowband rather than broad band spectra, so there will likely be an evolution of illumination packages that can provide such narrowband spectra from multicolor LEDs. However, integration with planar displays requires solving different optics problems than integration with imaging and collimation optics for illumination.

4.2. Optoelectronic Integration: Reliability, Functionality, and Cost

A second kind of integration, optoelectronic, has to do with lamp architectures that are consistent with independent sensing, control and driving of multiple LEDs. Sensor, control and driver electronics separate from the LEDs will certainly be the short-term approach. However, there may be potential to reduce system cost through integration of electronics and optoelectronics on the same die.

One approach is the consideration of LED die structures integrated with silicon wafers which would have the potential

for creating a light emitting integrated circuit (an LEIC). Of course, Silicon device processing is not compatible with GaN device processing, so the integration might not be monolithic, but rather at the wafer-scale packaging level (where the processed Si wafer perhaps would have LEDs flipped into on-wafer sockets to achieve integration).

Another approach would be to explore monolithic integration of GaN electronics and LEDs, since GaN is now being developed independently for a host of power electronics and high speed switching applications.[182] It may be possible to integrate simple electronics functionality and LED operation into the same wafer structure.[183] In the limit that GaN epitaxial materials continue to get less expensive, monolithic integration of light emission and simple control electronics in the same structure would lead to system simplification, improved reliability and reduced costs in much the same way as has been true for the integration of different functionality (e.g., combined analog and digital functions) in increasingly complex silicon integrated circuits. Because the level of complexity of a monolithically integrated LEIC fabricated completely from AlInGaN materials would be many orders of magnitude less than that for current silicon ICs, the development of such integrated structures seems quite feasible.

4.3. Optomechanical Integration: Control of Flux in Space

A third kind of integration, optomechanical, has to do with lamp architectures that directly control the directionality and focus of the lamp emission, bypassing the need for large and expensive mechanical manipulation at the luminaire level. Indeed, lamp-level distribution, or at least pre-distribution management, of light, has been shown to be very effective in directing light to where it is needed rather than in trapping light within the luminaire or sending part of it into undesirable directions ("light pollution"). In one example studied, the difference was 2x over a conventional light source.

Going beyond simple lamp-level distribution to lamp-level distribution control would of course be challenging. This control must be inexpensive and hence the beneficiary of chip-level integration. It must be multi-functional: capable of real-time control of all the features described above, and at the same time wired to receive power, and wirelessly connected to receive control signals from the smart electricity grid. As illustrated in **Figure 11**, it would essentially enable a new generation of smart lamps with the functionality of high-end theater luminaires, but with volumes and costs amenable to unobtrusive arrayed placement in a wide range of indoor (offices, factories, stores, homes) and outdoor (streets, parking lots, security zones) spaces. To put this in perspective, a typical theater luminaire weighs ca. 25 kg, with dimensions ca. 40 × 40 × 60 cm, and costs ca. $1–3k. Microsystem luminaires could potentially weigh <10 gm, have volumes of a few cubic centimeters, and cost <$10.

The ultimate microsystem luminaire would provide the two key functions of a high-end theatre luminaire: it would shape and steer the light, and it would change chromaticity. One might imagine two approaches: a "hybrid" approach which is on a centimeter scale; and an "intimately integrated microsystem" approach that is on a millimeter scale. The first approach might involve hybrid integration of separately optimized Si (microelectronics, sensors, and MEMS) and GaN (LEDs) functional modules; while the second approach might involve intimate integration of Si with GaN (e.g., wafer bonding and thin-film transfer) combined with increased functionality of GaN (e.g., GaN MEMS).

Clearly, there is significant future opportunity in this area (e.g., sensors, micro-opto-electro-mechanical systems

Figure 11. Left: Theater luminaire that rotates, tilts, focuses, contains gel filters so that it can change the chromaticity of the light, and is fully digitally controlled. Adapted with permission. Copyright 2013, Neo-Neon LED Lighting International Ltd. Right: a "dream" solid-state-lighting microsystem luminaire with the same functionality as the theater luminaire but at a dramatically scaled-down volume and cost.

(MOEMS), human visual perception, lighting architectures and controls), and one can anticipate more attention will be paid to it in the coming decades.

5. Summary and Conclusions

Solid-state lighting has made tremendous progress over the past decade, with the potential to make much more progress over the coming decades. In particular, we can imagine a future in which solid-state lighting is both "smart" and ultra-efficient. Such smart, ultra-efficient solid-state lighting would enable: very high (>150%) "effective" light production and utilization efficiencies; a host of new system applications (including feature-rich illumination, integrated illumination/displays, human health/well-being/productivity, agriculture, communications, and light-field mapping); and huge potential increases in human economic productivity.

To achieve ultra-efficiency, phosphors must eventually give way to multi-color semiconductor electroluminescence in the blue, green and red. In the blue, the major challenge is overcoming InGaN light-emitter efficiency droop to achieve ultra-high-efficiency at the input power densities necessary for economical production of light. In the green, the major challenges are: overcoming the severe lattice mismatch between known substrates and the medium-In-content InGaN necessary for green light emission, and the low growth temperatures which appear necessary for medium-In-content InGaN but which also appear to compromise material quality. In the red, the major challenge is overcoming the reduced radiative efficiency caused by the direct/indirect bandgap crossover in the medium-Al-content AlInGaP currently used for red light emission.

To achieve smartness, additional characteristics such as the ability to control light flux and spectra in time and space will be important. Devices and lamp architectures that achieve this in a cost-effective manner will likely mean integration into small packages. Three levels of integration can be envisioned: optical integration for mixing colors while maintaining the low etendue necessary for beam directing; optoelectronic integration for achieving reliability and low cost while simultaneously achieving multiple functions (sensing, current driving and control, and light emission); and optomechanical integration for placement of light in space.

Importantly, smart and ultra-efficient are not either-or characteristics of future solid-state lighting but are mutually compatible and even synergistic. The ultimate route to ultra-efficiency (multi-color electroluminescence from, e.g., RYGB lasers) brings with it the potential for smart (a high level of optical, optoelectronic and optomechanical integration). The ultimate route to smartness (a high level of optical, optoelectronic and optomechanical integration) also brings with it the potential for ultra-efficiency (multi-color electroluminescence from, e.g., RYGB lasers).

Acknowledgements

The authors are grateful to Wiley for permission to publish this review article as an adapted version of a book chapter in the upcoming work: D. L. Andrews, Ed., "Photonics Volume 3: Photonics Technology and Instrumentation" (Wiley, 2014). The authors also acknowledge the larger solid-state lighting and wide-bandgap semiconductor groups at Sandia and RPI for day-to-day intellectual stimulation, and especially Jerry Simmons, Charles Barbour and Julie Phillips for their ongoing encouragement.

Work at Sandia National Laboratories was supported by Sandia's Solid-State-Lighting Science Energy Frontier Research Center, funded by the U.S. Department of Energy, Office of Basic Energy Sciences. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04–94AL85000. Work at Rensselaer Polytechnic Institute was performed at the Smart Lighting Engineering Research Center and was supported by the National Science Foundation under cooperative agreement EEC-0812056 and by New York State under NYSTAR contract C090145.

Received: March 24, 2014
Revised: May 22, 2014
Published online: June 27, 2014

[1] Solid-state lighting can in principal be based on inorganic or organic semiconductors. Although we will mention organic semiconductors in the context of display technology, in this article we emphasize the inorganic semiconductors which, for solid-state lighting, are currently the most advanced.
[2] R. Haitz, J. Y. Tsao, *Phys. Status Solidi A* **2011**, *208*, 17–29.
[3] a) Data for the top panel (a) of Figure 1 were compiled from: A. Bergh, M. G. Craford, A. Duggal, R. Haitz, *Phys. Today* **2001**, *54*, 42–47; b) J. Edmond, H. Kong, C. Carter, *Phys. B: 7th Trieste ICTP-IUPAP Semiconductor Symposium* **1993**, *185*, 453–460; c) N. Nakayama, S. Kijima, S. Itoh, T. Ohata, A. Ishibashi, Y. Mori, *Opt. Rev.* **1995**, *2*, 167–170; and Philips Lumileds datasheets, http://www.philipslumileds.com/support/documentation/datasheets, accessed: December 2012.
[4] E. F. Schubert, *Light-Emitting Diodes*, Cambridge University Press, Cambridge, UK **2006**.
[5] N. Zheludev, *Nat. Photonics* **2007**, *1*, 189–192.
[6] N. Holonyak, S. F. Bevacqua, *Appl. Phys. Lett.* **1962**, *1*, 82–83.
[7] S. Nakamura, T. Mukai, M. Senoh, *Appl. Phys. Lett.* **1994**, *64*, 1687–1689.
[8] H. Amano, N. Sawaki, I. Akasaki, Y. Toyoda, *Appl. Phys. Lett.* **1986**, *48*, 353–355.
[9] H. Amano, M. Kito, K. Hiramatsu, I. Akasaki, *Jpn. J. Appl. Phys. 2* **1989**, *28*, L2112–L2114.
[10] Osram Opto Semiconductors GmbH of Regensburg, Germany, recently announced research results with a record efficiency of 61% for a red high-power LED. The 1 mm^2 chip, housed in a laboratory package, emits at a dominant wavelength of 609 nm with a luminous efficiency of 201l m/W at an operating current of 40 mA. At a typical operating current of 350 mA its luminous efficacy is still 168 lm/W, so even at this high wattage more than half of the electrical energy is converted into light.
[11] Y. Narukawa, M. Ichikawa, D. Sanga, M. Sano, T. Mukai, *J. Phys. D* **2010**, *43*, 354002.
[12] Y. Shiimizu, K. Sakano, Y. Noguchi, T. Moriguchi, US Patent 6614179, **2003**.
[13] R. Haitz, J. Y. Tsao, *Opt. Photonik* **2011**, 26–30.
[14] M. R. Krames, O. B. Shchekin, R. Mueller-Mach, G. O. Mueller, L. Zhou, G. Harbers, M. G. Craford, *J. Disp. Technol.* **2007**, *3*, 160–175.
[15] Data for Figure 3 were taken from "Luxeon Rebel and Luxeon Rebel ES Color Portfolio," Technical Datasheet DS68 (Philips Lumileds, Dec 19, 2012), accessed December, 2012.

[16] "Luxeon Rebel and Luxeon Rebel ES Color Portfolio," Technical Datasheet DS68 (Philips Lumileds, May, 2014), accessed May, 2014.
[17] "Osram Opto unveils Brilliant-Mix LED mixing concept," LEDs Magazine (May, 2011), accessed May, 2014.
[18] Note that phosphor cooling can be especially problematic, though progress is being made with higher thermal conductivity conformal phosphor coatings, so that the phosphor can be cooled reasonably effectively through the LED die.
[19] "Luxeon T High-Efficacy, Illumination Grade LED light source," Technical Datasheet DS106 (Philips Lumileds, March 22, 2013); and "Cree XLamp XM-L2 LEDs," Product Family Data Sheet CLD-DS61 Rev 1A (Cree, 2012–2013), accessed March, 2013.
[20] M. Conner, *Electron. Des. News* **2011**, *56*, 14. See also Proceedings of the 10th International Symposium on Automotive Lighting, Darmstadt, September 2013.
[21] E. Jang, S. Jun, H. Jang, J. Llim, B. Kim, Y. Kim, *Adv. Mater.* **2010**, *22*, 3076–3080.
[22] S. Coe-Sullivan, W. H. Liu, P. Allen, J. S. Steckel, *ECS J. Solid State Sci. Technol.* **2013**, *2*, R3026–R3030.
[23] R. Haitz, personal communication, unpublished.
[24] Within limits, of course: too much visual information might distract and even impair cognitive function.
[25] The factor of 0.2 is the fraction of white light that is blue, with the blue assumed to be produced with 100% efficiency; the factor of 0.8×0.75 is the fraction of white light that is red or green, with the red and green light produced with 80% efficiency due to the Stokes deficit. See, e.g., J. Y. Tsao, M. E. Coltrin, M. H. Crawford, J. A. Simmons, Proc. IEEE **2010**, *98*, 1162–1179.
[26] M. H. Crawford, *IEEE J. Sel. Top. Quantum Electron.* **2009**, *15*, 1028–1040.
[27] J. Y. Tsao, I. Brener, D. F. Kelley, S. K. Lyo, *IEEE J. Disp. Technol.* **2013**, *9*, 419–426.
[28] P. C. Hung, J. Y. Tsao, *IEEE J. Disp. Technol.* **2013**, *9*, 405–412.
[29] a) A. Neumann, J. J. Wierer, Jr., W. Davis, Y. Ohno, S. R. J. Brueck, J. Y. Tsao, *Opt. Express* **2011**, *19*, A982–990; b) J. Y. Tsao, J. J. Wierer Jr., L. E. S. Rohwer, M. E. Coltrin, M. H. Crawford, J. A. Simmons, P. C. Hung, H. Saunders, D. S. Sizov, R. Bhat, C. E. Zah, in III-Nitride Based Light Emitting Diodes and Applications, Topics in Applied Physics, Vol. 126, (Eds: T. Y. Seong, J. Han, H. Amano, H. Morkoc), Springer, 2013.
[30] W. W. Davis, Y. Ohno, *Proc. SPIE* **2005**, *5941*, 594111–594117.
[31] M. S. Rea, J. P. Freyssinier-Nova, *Color Res. Appl.* **2008**, *33*, 192–202.
[32] A. Zukauskas, R. Vaicekauskas, P. Vitta, A. Tuzikas, A. Petrulis, M. Shur, *Opt. Express* **2012**, *20*, 5356–5366.
[33] S. Paolini, presented at Taiwan Solid-State Lighting, June 2012.
[34] H. Ishio, J. Minowa, K. Nosu, *J. Lightwave Technol.* **1984**, *2*, 448–463.
[35] E. F. Schubert, J. K. Kim, *Science* **2005**, *308*, 1274–1278.
[36] L. Ulrich, *IEEE Spectrum* **2013**, *50*, 36–56.
[37] C. Basu, M. Meinhardt-Wollweber, B. Roth, *Adv. Opt. Technol.* **2013**, *2*, 313–321.
[38] J. D. Bullough, E. T. Donnell, M. S. Rea, *Accid. Anal. Prevention* **2013**, *53*, 65–77.
[39] M. S. Rea, J. D. Bullough, Y. Zhou, *Lighting Res. Technol.* **2010**, *42*, 215–241.
[40] Philips Hue, http://meethue.com/, accessed June 2014.
[41] F. Rubinstein, S. Kiliccote, "Demand Responsive Lighting: A Scoping Study," Lawrence Berkeley National Laboratory Technical Note LBNL-62226 (January, 2007), http://drrc.lbl.gov/publications/demand-responsive-lighting-scoping, accessed March 2013.
[42] B. Witherspoon, M. Petrick, "Scientific Research and Sky Image Ceilings," Sky Factory White Paper. http://www.skyfactory.com/files/SkyFactory_White_Paper_032408.pdf, accessed March 2013.
[43] "Sky Light Sky Bright – In the Office," Fraunhofer Press Release, http://www.fraunhofer.de/en/press/research-news/2012/january/sky-light-sky-bright.html, accessed March 2013.
[44] Such panels are now commercially available, see: http://www.neonny.com/news/html/?464.html, accessed March 2013
[45] D. M. Berson, F. A. Dunn, M. Takao M, *Science* **2002**, *295*, 1070.
[46] R. G. Stevens, D. E. Blask, G. C. Brainard, J. Hansen, S. W. Lockley, I. Provencio, M. S. Rea, L. Reinlib, *Environ. Health Perspect.* **2007**, *115*, 1357.
[47] N. E. Rosenthal, D. A. Sack, R. G. Skwerer, F. M. Jacobsen, T. A. Wehr, in *Seasonal Affective Disorders and Phototherapy*, (Ed: N. E. Rosenthal, M. S. Blehar), Guilford Press, New York **1989**, pp. 273–294.
[48] S. A. Rahman, S. Marcu, C. M. Shapiro, T. J. Brown, R. F. Casper, *J. Physiol. Endocrinol. Metab.* **2011**, *300*, E518.
[49] I. Knez, *J. Environ. Psychol.* **2001**, 201–208.
[50] T. Goven, T. Laike, P. Raynham, E. Sansal, "Influence of Ambient Light on the Performance, Mood, Endocrine Systems and Other Factors of School Children," presented at CIE 27th Session, Sun City, South Africa, 112 (2011).
[51] A. Satlin, L. Volicer, V. Ross, L. R. Herz, *Am. J. Psychiatry* **1992**, *149*1028–1032.
[52] J. M. Walch, B. S. Rabin, R. Day, J. N. Williams, K. Choi, J. D. Kang, *Psychosom. Med.* **2005**, *67*, 156–163.
[53] "New Philips lighting solution increases duration of sleep for hospital patients," Philips Study at: http://www.newscenter.philips.com/main/standard/news/press/2011/20111122-healwell.wpd, accessed March 2013
[54] L. Edwards, P. Torcelli, "A Literature Review of the Effects of Natural Light on Building Occupants," *National Renewable Energy Laboratory Technical Report (NREL TP-550–30769)* July, **2002**.
[55] R. C. Morrow, *Horticulture Sci.* **2008**, *43*, 1947.
[56] S. W. Hogewoning, P. Douwstra, G. Touwborst, W. V. Leperen, J. Harbinson, *J. Experimental Botany* **2010**, DOI:10.1093/jxb/erq005.
[57] M. A. Abdullah, A. U. Rahmah, A. J. Sinskey, C. K. Rha, *Open Med. Chem. J.* **2008**, *2*, 49–61.
[58] The Visible Light Communications Consortium was established in Japan in 2003. See: http://www.vlcc.net/modules/xpage3/, accessed June 2014.
[59] J. Vucic, C. Kottke, S. Nerreter, K. D. Langer, J. W. Walewski, *J. Lightwave Technol.* **2010**, *28*, 3512.
[60] M. Kavehrad, *IEEE Commun. Mag.* **2010**, *48*, 66.
[61] P. Daukantas, *Opt. Photonics News* **2014**, 34–41.
[62] S. Williams, in *IEEE Personal Communications*, **2000**, *7*, 11–19.
[63] See the IEEE802.15.7 draft standard for visible light communications.
[64] M. B. Rahaim, A. M. Vegni, T. D. C. Little, in *Proc. 2nd IEEE Globecom 2011 Workshop on Optical Wireless Communications (OWC 2011)*, **2011**, 818–822.
[65] B. V. Quang, V. Prasad, I. Niemegeers, *IEEE Commun. Surveys Tutorials* **2012**, *14*, 64–85.
[66] J. Geng, *Adv. Opt. Photonics* **2011**, *3*128.
[67] S. N. Pattanaik, S. P. Mudur, *J. ACM Trans. Graphics* **1995**, *14*, 77.
[68] F. Chiabrando, R. Chiabrando, D. Piatt, F. Rinaudo, *Sensors* **2009**, *9*, 10080.
[69] F. Brida, J. Duha, M. Krasnovsky, in *IFIP International Federation for information Processing: Personal Wireless Communications*, (Eds. B. Simak, R. Bestak, E. Kozowska), Springer, Boston, MA, USA **2007**, pp. 423–432.
[70] G. R. Newsham, C. D. Arsenault, *Lighting Res. Technol.* **2009**, *41*, 143.
[71] Note that ultra-efficiency also translates to a lower capital cost of light (the cost of the lamp) by simplifying heat management.

Lower cost, in turn, translates to decreased energy consumption through increased market penetration.
[72] J. M. Phillips, M. E. Coltrin, M. H. Crawford, A. J. Fischer, M. R. Krames, R. Mueller-Mach, G. O. Mueller, Y. Ohno, L. E. S. Rohwer, J. A. Simmons, J. Y. Tsao, *Laser Photonics Rev.* **2007**, *1*, 307–333.
[73] Navigant Consulting, "US Lighting Market Characterization Volume I: National Lighting Inventory and Energy Consumption Estimate" (US Department of Energy, Office of Energy Efficiency and Renewable Energy, Building Technologies Program, September, 2002).
[74] J. Y. Tsao, H. D. Saunders, J. R. Creighton, M. E. Coltrin, J. A. Simmons, *J. Phys. D* **2010**, *43*, 354001.
[75] This plot has been adapted from: J. Y. Tsao, P. Waide, "The World's Appetite for Light: A Simple Empirical Expression Spanning Three Centuries and Six Continents, LEUKOS 6, 259–281 (2010). The only changes are: it plots per capita light consumption and gross domestic product rather than total light consumption and gross domestic product; and a data point for the US in 2010 has been added from Navigant, "2010 US Lighting Market Characterization" (US Department of Energy, Office of Energy Efficiency and Renewable Energy, Building Technologies Program, 2012).
[76] H. D. Saunders, J. Y. Tsao, *Energy Policy* **2012**, *49*, 477–478.
[77] *Basic Research Needs for Solid-State Lighting: Report of the Basic Energy Sciences Workshop on Solid-State Lighting*, (Eds: J. M. Phillips, P. E. Burrows), Department of Energy Office of Basic Energy Sciences, **2006**.
[78] Y. Narukawa, M. Ichikawa, D. Sanga, M. Sano, T. Mukai, *J. Phys. D* **2010**, *43*, 354002.
[79] a) We follow the simplified but internally self-consistent treatment described in: J. J. Wierer, J. Y. Tsao, D. S. Sizov, *Laser Photonics Rev.* **2013**, *7*, 963; b) However, we note that there are subtleties to a quantitative analysis of the various sub-efficiencies, particularly the injection efficiency, that we do not treat here. See, e.g., J. Cho, E. F. Schubert, J. K. Kim, *Laser Photonics Rev.* **2013**, *7*, 408.
[80] H. Zhao, G. Liu, J. Zhang, R. A. Arif, N. Tansu, *J. Display Technol.* **2013**, *9*, 212–225.
[81] I. E. Titkov, D. A. Sannikov, Y. M. Park, J. K. Son, *AIP Adv.* **2012**, *2*, 032117.
[82] G. Chen, M. Craven, A. Kim, A. Munkholm, S. Watanabe, M. Camras, W. Gotz, F. Steranks, *Phys. Status Solidi A* **2008**, *205*, 1086–1092.
[83] A. David, M. J. Grundmann, J. F. Kaeding, N. F. Gardner, T. G. Mihopoulos, M. R. Krames, *Appl. Phys. Lett.* **2008**, *92*, 053502.
[84] K. T. Delaney, P. Rinke, C. G. Van de Walle, *Appl. Phys. Lett.* **2009**, *94*, 191109.
[85] J. Iveland, L. Martinelli, J. Peretti, J. S. Speck, C. Weisbuch, *Phys. Rev. Lett.* **2013**, *110*, 177406.
[86] I. Schnitzer, E. Yablonovitch, C. Caneau, T. J. Gmitter, A. Scherer, *Appl. Phys. Lett.* **1993**, *63*, 2174–2176.
[87] M. R. Krames, G. Christenson, D. Collins, L. W. Cook, M. G. Craford, A. Edwards, R. M. Fletcher, N. F. Gardner, W. K. Gotz, W. R. Imler, E. Johnson, R. S. Kern, R. Khare, F. A. Kish, C. Lowery, M. J. Ludowise, R. Mann, M. Maranowski, S. A. Maranowski, P. S. Martin, J. O'Shea, S. L. Rudaz, D. A. Steigerwald, J. Thompson, J. J. Wierer, J. G. Yu, *Light-Emitting Diodes: Research, Manufacturing, and Applications IV*, (Eds: H. W. Yao, I. T. Ferguson, E. F. Schubert), SPIE Proceedings Vol *3938* (2000), pp. 2–12.
[88] R. A. Arif, Y-K. Ee, N. Tansu, *Appl. Phys. Lett.* **2007**, *91*, 091110.
[89] C. K. Tan, J. Zhang, X-H. Li, G. Liu, B. O. Tayo, N. Tansu, *J. Display Technol.* **2013**, *9*, 272–279.
[90] A. Kimura, C. A. Paulson, H. F. Tang, T. F. Kuech, *Appl. Phys. Lett.* **2004**, *84*, 1489.
[91] a) N. F. Gardner, G. O. Muller, Y. C. Shen, G. Chen, S. Watanabe, W. Gotz, M. R. Krames, *Appl. Phys. Lett.* **2007**, *91*, 243506-1; b) A. Laubsch, W. Bergbauer, M. Sabathil, M. Strassburg, H. Lugauer, M. Peter, T. Meyer, G. Bruderl, J. Wagner, N. Linder, K. Streubel, B. Hahn, *Phys. Status Solidi C* **2009**, *6*, S885–S888.
[92] N. F. Gardner, G. O. Muller, Y. C. Shen, G. Chen, S. Watanabe, W. Gotz, M. R. Krames, *Appl. Phys. Lett.* **2007**, *91*, 243506-1.
[93] D. Zhu, A. N. Noemaun, M. F. Schubert, J. Cho, E. F. Schubert, M. H. Crawford, D. D. Koleske, *Appl. Phys. Lett.* **2010**, *96*, 121110 .
[94] E. Kioupakis, Q. Yan, C. G. Van de Walle, *Appl. Phys. Lett.* **2013**, *101*, 231107 .
[95] M. D. Craven, S. H. Lim, F. Wu, J. S. Speck, S. P. DenBaars, *Appl. Phys. Lett.* **2002**, *81*, 469–471.
[96] D. F. Feezell, J. S. Speck, S. P. DenBaars, S. Nakamura, *J. Display Technol.* **2013**, *9*, 190.
[97] R. Michalzik, K. J. Ebeling, *Vertical-Cavity Surface-Emitting Laser Devices*, (Eds: H. Li, K. Iga), Springer, Berlin, Germany **2002**, Ch. 3.
[98] For simplicity, this equation omits the *A* contribution from SRH defect recombination, which would normally be overwhelmed by the *B* and *C* terms. However, for low threshold current density devices, the A contribution might be significant and would need to be considered.
[99] F. Z. Jasim, K. Omar, Z. Hassan, *Optoelectron. Adv. Mater., Rapid Commun.* **2009**, *3*, 1136.
[100] D. G. Deppe, K. Shavritranuruk, G. Ozgur, H. Chen, S. Freisem, *Electron. Lett.* **2009**, *45*, 54.
[101] H. Deng, G. Weihs, D. Snoke, J. Bloch, Y. Yamamoto, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15318.
[102] H. Deng, H. Haug, Y. Yamamoto, *Rev. Mod. Phys.* **2010**, *82*, 1489
[103] J. M. Phillips, M. E. Coltrin, M. H. Crawford, A. J. Fischer, M. R. Krames, R. Mueller-Mach, G. O. Mueller, Y. Ohno, L. E. S. Rohwer, J. A. Simmons, J. Y. Tsao, *Laser Photonics Rev.* **2007**, *1*, 307–333.
[104] J. Wu, W. Walukiewicz, K. M. Yu, J. W. Ager, E. E. Haller, H. Lu, W. J. Schaff, *Appl. Phys. Lett.* **2002**, *80*, 4741–4743.
[105] However, note that the In content can be reduced somewhat (to 20%) if InGaN alloys are grown strained on c-plane GaN: the quantum-confined Stark effect red-shifts the emission wavelength from ~470 nm (for unstrained InGaN) to 530 nm (for fully strained InGaN).
[106] I. H. Ho, G. B. Stringfellow, *Appl. Phys. Lett.* **1996**, *69*, 2701–2703. Note, however, that this is countered by the relative lack of mobility of surface adatoms at typical growth temperatures, and thus a tendency to freeze in random rather than phase-separated InGaN alloys. See, e.g., M. J. Galtrey, R. A. Oliver, M. J. Kappers, C. J. Humphries, D. J. Stokes, P. H. Clifton, A. Cerezo, *Appl. Phys. Lett.* **2007**, *90*, 061903.
[107] G. B. Stringfellow, *Organometallic Vapor-Phase Epitaxy*, 2nd Ed., Academic Press, Boston, MA, USA **1998**.
[108] J. R. Creighton, G. T. Wang, W. G. Breiland, M. E. Coltrin, *J. Cryst. Growth* **2004**, *261*, 204–213.
[109] O. Ambacher, M. S. Brandt, R. Dimitrov, T. Metzger, M. Stutzmann, R. A. Fischer, A. Miehr, A. Bergmaier, G. Dollinger, *J. Vacuum Sci. Technol.* **1996**, *B14*, 3532–3542.
[110] This imposes constraints not just on growth, but annealing temperatures as well. As annealing temperature is increased for InGaN films with an 18% indium concentration, InGaN film decomposes into metallic indium and gallium prior to reaching the binodal temperature ~1100 °C for this indium composition. See, e.g., G. T. Thaler, D. D. Koleske, S. R. Lee, K. H. A. Bogart, M. H. Crawford, *J. Cryst. Growth* **2010**, *312*, 1817–1822.
[111] W. Lee, J. Limb, J. H. Ryou, D. Yoo, T. Chung, R. D. Dupuis, *J. Electron. Mater.* **2006**, *35*, 587–591.
[112] Y. J. Zhao, Q. M. Yan, C. Y. Huang, S. C. Huang, P. S. Hsu, S. Tanaka, C. C. Pan, Y. Kawaguchi, K. Fujito, C. G. Van de Walle,

J. S. Speck, S. P. DenBaars, S. Nakamura, D. Feezell, *Appl. Phys. Lett.* **2012**, *100*, 201108.
[113] Y. J. Zhao, J. Sonoda, C. C. Pan, S. Brinkley, I. Koslow, K. Fujito, H. Ohta, S. P. DenBaars, S. Nakamura, *Appl. Phys. Express* **2010**, *3*, 102101.
[114] S. Yamamoto, Y. J. Zhao, C. C. Pan, R. B. Chung, K. Fujito, J. Sonoda, S. P. Denbaars, S. Nakamura, *Appl. Phys. Express* **2010**, *3*, 122102.
[115] C. C. Pan, S. Tanaka, F. Wu, Y. J. Zhao, J. S. Speck, S. Nakamura, S. P. DenBaars, D. Feezell, *Appl. Phys. Express* **2012**, *5*, 062103.
[116] D. F. Feezell, J. S. Spect, S. P. Denbaars, S. Nakamura, *J. Disp. Technol.* **2013**, *9*, 190.
[117] H. Y. Ryu, D. S. Shin, J. I. Shim, *Appl. Phys. Lett.* **2012**, *100*, 131109.
[118] Y. J. Zhao, S. Tanaka, C. C. Pan, K. Fujito, D. Feezell, J. S. Speck, S. P. Denbaars, S. Nakamura, *Appl. Phys. Express* **2011**, *4*, 082104.
[119] a) Y. Nanishi, Y. Saito, T. Yamaguchi, T. Araki, T. Miyajima, H. Naoi, *Phys. Status Solidi* **2004**, *C1*, 1487–1495; b) H. K. Cho, J. Y. Lee, K. S. Kim, G. M. Yang, *Appl. Phys. Lett.* **2000**, *77*, 247–249.
[120] M. Iwaya, A. Miura, R. Senda, T. Nagai, T. Kawashima, D. Iida, S. Kamiyama, H. Amano, I. Akasaki, *J. Crystal Growth* **2008**, *310*, 4920–4922.
[121] O. H. Nam, M. D. Bremser, T. S. Zheleva, R. F. Davis, *Appl. Phys. Lett.* **1997**, *71*, 2638–2640.
[122] Q. Li, J. J. Figiel, G. T. Wang, *Appl. Phys. Lett.* **2009**, *94*, 231105.
[123] Y.-K. Ee, J. M. Biser, W. Cao, H. M. Chan, R. P. Vinci, N. Tansu, *IEEE J. Sel. Top. Quantum Electron.* **2009**, *15*, 1066–1072.
[124] R. Senda, A. Miura, T. Hayakawa, T. Kawashima, D. Iida, T. Nagai, M. Iwaya, S. Kamiyama, H. Amano, I. Akasaki, *Jpn. J. Appl. Phys.* **2007**, *46*, L948–L950.
[125] M. Iwaya, A. Miura, R. Senda, T. Nagai, T. Kawashima, D. Iida, S. Kamiyama, H. Amano, I. Akasaki, *J. Crystal Growth* **2008**, *310*, 4920–4922.
[126] R. Senda, T. Matsubara, D. Lida, M. Iwaya, S. Kamiyama, H. Amano, I. Akasaki, *Appl. Phys. Express* **2009**, *2*, 061004.
[127] J. Zhang, N. Tansu, *J. Appl. Phys.* **2011**, *110*, 113110.
[128] A. Kobayashi, J. Ohta, H. Fujioka, *J. Appl. Phys.* **2006**, *99*, 123513.
[129] Y. Kawai, S. Ohsuka, M. Iwaya, S. Kamiyama, H. Amano, I. Akasaki, *Proc. SPIE* **2008**, *6889*, 88904–88904.
[130] K. Shimomoto, A. Kobayashi, K. Ueno, J. Ohta, M. Oshima, H. Fujioka, *Appl. Phys. Express* **2010**, *3*, 061001.
[131] A. Kobayashi, S. Kawano, Y. Kawaguchi, J. Ohta, H. Fujioka, *Appl. Phys. Lett.* **2007**, *90*, 041908.
[132] H. Tamaki, A. Kobayashi, J. Ohta, M. Oshima, H. Fujioka, *Appl. Phys. Lett.* **2011**, *99*, 061912.
[133] N. Li, S.-J. Wang, C.-L. Huang, Z. C. Feng, A. Valencia, J. Nause, C. Summers, I. Ferguson, *J. Cryst. Growth* **2008**, *310*, 4908.
[134] A. Kobayashi, J. Ohta, H. Fujioka, *J. Appl. Phys.* **2006**, *99*, 123513.
[135] A. Kobayashi, S. Kawano, Y. Kawaguchi, J. Ohta, H. Fujioka, *Appl. Phys. Lett.* **2007**, *90*, 041908.
[136] E. Ertekin, P. A. Greaney, D. C. Chrzan, T. D. Sands, *J. Appl. Phys.* **2005**, *97*, 114325 .
[137] K. Kishino, A. Kikuchi, H. Sekiguchi, S. Ishizawa, *Proc. SPIE* **2007**, *6473*, 64730T.
[138] T. Kuykendall, P. Ulrich, S. Aloni, P. Yang, *Nat. Mater.* **2007**, *6*, 951–956.
[139] a)Y. Liang, W. D. Nix, P. B. Griffin, J. D. Plummer, *J. Appl. Phys.* **2005**, *97*, 043519; b)S. Raychaudhuri, E. T. Yu, *J. Vac. Sci. Technol. B* **2006**, *24*, 2053–2059.
[140] S. A. Dayeh, W. Tang, F. Boioli, K. L. Kavanagh, H. Zheng, J. Wang, N. H. Mack, G. Swadener, J. Y. Huang, L. Miglio, K. N. Tu, S. T. Picraux, *Nano Lett.* **2013**, *13*, 1869–1876.
[141] H. P. T. Nguyen, S. Zhang, K. Cui, A. Korinek, G. A. Botton, Z. Mi, *IEEE Photonics Technol. Lett.* **2012**, *24*, 321
[142] H. Sekiguchi, K. Kishino, A. Kikuchi, *Appl. Phys. Lett.* **2010**, *96*, 231104.
[143] K. Kishino, K. Nagashima, K. Yamano, *Appl. Phys. Express* **2013**, *6*, 012101.
[144] Y. J. Hong, C. H. Lee, A. Yoon, M. Kim, H. K. Seong, H. J. Chung, C. Sone, Y. J. Park, G. C. Yi, *Adv. Mater.* **2011**, *23*, 3284.
[145] QM Li, GT Wang, *Appl. Phys. Lett.* **2010**, *97*, 181107.
[146] A. T. Connie, H. P. T. Nguyen, S. M. Sadaf, I. Shih, Z. Mi, *J. Vac. Sci. Technol. B* **2014**, *32*, 02C113.
[147] H. P. T. Nguyen, S. Zhang, K. Cui, X. Han, S. Fathololoumi, M. Couillard, G. A. Botton, Z. Mi, *Nano Lett.* **2011**, *11*, 1919–1924.
[148] K. Kishino, A. Kikuchi, H. Sekiguchi, S. Ishizawa, *Proc. SPIE* **2007**, *6473*, 64730T.
[149] H.-M. Kim, Y-H. Cho, H. Lee, S. I. Kim, S. R. Ryu, D. Y. Kim, T. W. Kang, KS. Chung, *Nano Lett.* **2004**, *4*, 1059–1062.
[150] G. T. Wang, Q. M. Li, J. Wierer, J. Figiel, J. B. Wright, T. S. Luk, I. Brener, *Proc. SPIE* **2012**, *8278*, 827816.
[151] J. Nelson, E. Jones, S. Myers, D. Follstaedt, H. Hjalmarson, J. Schirber, R. Schneider, J. Fouquet, V. Robbins, K. Carey, *Phys. Rev. B* **1996**, *53*, 15893–15901.
[152] D. Bour, D. Treat, R. Thornton, R. Geels, D. Welch, *IEEE J. Quantum Electron.* **1993**, *29*, 1337–1343.
[153] "Osram red LED prototype breaks 200 lm/W efficiency barrier", Semiconductor Today, Oct. 11, 2011, http://www.semiconductor-today.com/news_items/2011/OCT/OSRAM_111011.html, accessed: May 2014.
[154] I. Gontijo, M. Boroditsky, E. Yablonovitch, S. Keller, U. K. Mishra, S. P. DenBaars, *Phys. Rev.* **1999**, *B60*, 11564.
[155] K. Okamoto, I. Niki, A. Shvartser, Y. Narukawa, T. Mukai, A. Scherer, *Nat. Mater.* **2004**, *3*, 601.
[156] J. Vuckovic, M. Loncar, A. Scherer, *IEEE J. Quantum Electron.* **2000**, *36*, 1131–1144.
[157] A. Kock, E. Gornik, M. Hauser, W. Beinstingl, *Appl. Phys. Lett.* **1990**, *57*, 2327–2329.
[158] J. J. Wierer, J. Y. Tsao, D. S. Sizov, *Laser Photonics Rev.* **2013**, *7*, 963–993.
[159] A. David, H. Benisty, C. Weisbuch, *Rep. Prog. Phys.* **2012**, *75*, 126501.
[160] E. Yablonovitch, *Phys. Rev. Lett.* **1987**, *58*, 2059–2062.
[161] a) J. J. Wierer, A. David, M. M. Megens, *Nat. Photonics* **2009**, *3*, 163–169; b) A. David, T. Fujii, E. Mattioli, R. Sharma, S. Nakamura, S. P. DenBaars, C. Weisbuch, H. Benisty, *Appl. Phys. Lett.* **2006**, *88*, 073510-1; c) C. F. Lai, H. C. Kuo, P. C. Yu, T. C. Lu, C. H. Chao, H. H. Yen, W. Y. Yeh, *Appl. Phys. Lett.* **2010**, *97*, 013108 .
[162] a) J. D. Joannopoulos, *Photonic Crystals: Molding the Flow of Light*, 2nd Ed., Princeton University Press, Princeton, NJ, USA **2008**; b) K. Busch, N. Vats, S. John, B. C. Sanders, *Phys. Rev. E* **2000**, 62, 4251–4260.
[163] Y. S. Choi, K. Hennessy, R. Sharma, E. Haberer, Y. Gao, S. P. DenBaars, S. Nakamura, E. L. Hu, C. Meier, *Appl. Phys. Lett.* **2005**, *87*, 243101.
[164] E. M. Purcell, *Phys. Rev.* **1946**, *69*, 681–681.
[165] G. Subramania, Q. M. Li, Y. J. Lee, J. J. Figiel, G. T. Wang, A. J. Fischer, *Nano Lett.* **2011**, *11*, 4591–4596.
[166] T. Mukai, M. Yamada, S. Nakamura, "InGaN-based uv/blue/green/amber/red LEDs" Proc. SPIE 3621, Light-Emitting Diodes: Research, Manufacturing, and Applications III, 2-13 (April 14, 1999); DOI:10.1117/12.3444641999.
[167] S. Jahangir, M. Mandl, M. Strassburg, P. Bhattacharya, *Appl. Phys. Lett.* **2013**, *102*, 071101.
[168] D. G. Thomas, J. J. Hopfield, *Phys. Rev.* **1966**, *150*, 680–689.
[169] V. A. Odnoblyudov, C. W. Tu, *Appl. Phys. Lett.* **2006**, *89*, 191107.
[170] V. A. Odnoblyudov, C. W. Tu, *J. Vac. Sci. Technol. B* **2006**, *24*, 2202–2204.

[171] H. P. Xin, R. J. Welty, C. W. Tu, *Appl. Phys. Lett.* **2000**, *77*, 1946–1948.
[172] Y. Zhang, B. Fluegel, A. Mascarenhas, H. P. Xin, C. W. Tu, *Phys. Rev. B* **2000**, *62*, 4493–4500.
[173] L. Bellaiche, S.-H. Wei, A. Zunger, *Phys. Rev. B* **1997**, *56*, 10233–10240.
[174] H. P. Xin, C. W. Tu, Y. Zhang, A. Mascarenhas, *App. Phys. Lett.* **2000**, *76*, 1267–1269.
[175] M. A. Haase, J. Xie, T. A. Ballen, J. Zhang, B. Hao, Z. H. Yang, T. J. Miller, X. Sun, T. L. Smith, C. A. Leatherdale, *Appl. Phys. Lett.* **2010**, *96*, 231116.
[176] W. Lin, S. P. Guo, M. C. Tamargo, *J. Vac. Sci. Technol.* **2000**, *18*, 1534–1537.
[177] Y. Zhu, S. McKernan, J. Xie, T. J. Miller, M. A. Haase, X. Sun, T. L. Smith, C. A. Leatherdale, *J. Appl. Phys.* **2010**, *108*, 123104.
[178] A. P. Van de Ven, G. H. Negley, US Patent 8,029,155, **October 4, 2011**.
[179] A. Cvetkovic, R. Mohedano, O. Dross, M. Hernandez, P. Benitez, J. C. Miñano, J. Vilaplana, J. Chavesa, in *Nonimaging Optics: Efficient Design for Illumination and Solar Concentration IX*, Proc. SPIE Vol. 8485 84850Q-1--84850Q-11 (Eds: R. Winston, J. M. Gordon), Proc. SPIE, DOI: 10.1117/12.931508.
[180] K. N. Hui, K. S. Hui, *Curr. Appl. Phys.* **2011**, *11*, 662–666.
[181] R. P. van Gorkom, M. A. van As, G. M. Verbeeka, C. G. A. Hoelen, R. G. Alferink, C. A. Mutsaers, H. Cooijmans, *Proc. SPIE* **2007**, *6670*, 66700E-10.
[182] T. P. Chow, Z. Li, in *GaN and ZnO based Materials and Devices*, (Ed: S. J. Pearton), Springer, Berlin, Germany **2012**.
[183] Z. Li, J. Waldron, T. Detchprohm, C. Wetzel, R. F. Karlicek Jr., T. P. Chow, *Appl. Phys. Lett.* **2013**, *102*, 192107.