With all such factors optimised, it has been possible to demonstrate a new world record 20 nanometre resolution in the CL mode on a GaAs/AlAs/AlGaAs semiconductor heterostructure.  The article provides a practical guide to obtaining ultra-high cathodoluminescence (CL) and electron beam induced current (EBIC) images and linescans in the SEM. It highlights the key factors affecting the practical resolution limits of the CL and EBIC modes of the SEM, and possible pitfalls in equipment and experimental technique. With all such factors optimised, it has been possible to demonstrate a new world record 20 nanometre resolution in the CL mode on a GaAs/AlAs/AlGaAs semiconductor heterostructure.

There are relatively few techniques that are able to conveniently analyse both large and small volumes of material. The advanced modes of the SEM such as CL and EBIC have always offered great promise in this respect, being capable of quickly analysing and imaging areas ranging from several mm² down to below 1µm². It may be questioned, however, whether techniques such as these can have sufficient spatial resolution to make a significant contribution to the study of nanostructures such as individual quantum dots. Such structures have dimensions of only a few nm, and mean separations of (typically) 100nm or below.

This article describes the factors that are most influential in determining (or more importantly, in limiting) the spatial resolution in SEM-based CL and EBIC. It also describes an optimisation procedure performed on our Hitachi S4500 SEM and Gatan MonoCL system with CF302TC helium cryostat, in order to obtain the maximum resolution in the CL mode.

**FACTORS DETERMINING SPATIAL RESOLUTION**

**Beam-specimen interaction**

It is well known that in modern SEMs the spot size is much smaller than the spatial extent of the beam energy dissipation volume, even at low beam energies of 1 or 2 keV. In the absence of excessive specimen charging and the accompanying astigmatism this may introduce, the resolution is therefore determined almost entirely by the interaction of the beam with the sample. There are two principle ways in which the size of the beam energy dissipation volume can be calculated: semi-empirically using (e.g.) the Grün range [4] related method of Everhart and Hoff [5], or statistically using Monte Carlo simulations, of the type described by Joy [6] or Napchan and Holt [7]. The semi-empirical method was originally claimed to be valid only in the range 5keV < E_b < 25keV, where E_b is the beam electron energy, but in single layer systems, reasonable agreement is generally obtained between the two methods right down to beam energies as low as 1keV. It should be pointed out that when calculating electron beam penetration into multi-layer systems such as semiconductor device heterostructures, the Monte Carlo method is much more convenient.

Regarding Monte Carlo programs: as with...
any software, the results will only be as reliable as the input parameters used, so when writing or using a new program it is vital that the output is checked as far as possible against experimental data. Indeed, some of the less well known factors in the simulation such as partially screened Rutherford cross sections (which arise because the scattered electron does not experience all the nuclear charge) may be iteratively adjusted so that the calculated number and energy of electrons lost to backscatter agree most closely with experimental values.

Figure 1 shows a comparison between the electron penetration range in GaAs calculated by Monte Carlo simulations and according to Everhart and Hoff (which is in effect 1.1 times the Grün range). A reasonable agreement is found between the two methods. The region of the curves of particular interest for high resolution CL and EBIC is the low energy portion, where the penetration range falls to the level of tenths of nm only. It is clear that this region represents the operating conditions offering the highest spatial resolution.

Signal-to-noise reduction

Figure 2 however shows one of the difficulties that emerge in low beam energy work, again using GaAs as an example. As the beam energy falls below 1keV, close to 50% of the electrons and 40% of the beam energy may be lost to backscattering. This phenomenon, allied with surface recombination and near-surface band bending due to Fermi level pinning, serves to greatly reduce the CL or EBIC signal. As an example, a 100pA, 1keV beam would generate a theoretical maximum EBIC of around 20nA. Losing 40% of the input energy reduces this to 12nA, whilst surface recombination and non-100% collection efficiency (for generation outside the depletion region) can easily reduce this figure to well below 1nA. In short, the signal to noise ratio becomes very poor, and both cost and data acquisition times increase dramatically as highly specialised current amplifiers and lock-in techniques become necessary.

An additional factor in CL is that (again, using GaAs as an example and assuming isotropic photon emission) only around 2% of the photons created will succeed in leaving the sample surface: photons reaching the surface at an angle of greater than 17° to the normal will be totally internally reflected. If the signal falls below the detection limits of the system, necessitating an increase in the beam energy or beam current, then dissipation volume size or spot size will also increase resulting in a loss of resolution.

Carrier redistribution via diffusion

A popular misconception is that the spatial resolution of CL or EBIC is limited by the minority carrier diffusion length of the sample material. Similarly, a common mistake is to assume that the minimum size of the features observed (e.g. dislocation dark contrast) is a reliable measurement of the minority carrier diffusion length. Indeed, many recent publications in the gallium nitride field have repeatedly made this error. The calculations of Donolato [8] show that, even assuming a uniform generation sphere, following carrier redistribution due to diffusion the excess carrier density falls off at a rate larger than exponential outside the volume. In the real physical situation, as modelled by Monte Carlo simulations, the excess carrier density is much higher initially in the central regions of the energy dissipation volume. The subsequent carrier redistribution, even for infinite minority carrier diffusion length, will result in a volume not significantly larger than the dissipation volume being sampled. Indeed, the resulting CL or EBIC signal will be dominated by the central, higher carrier density regions.

Small improvements in spatial resolution may also be achieved by limiting the diffusion either by time-resolved measurements (limiting the measurement period to a short time period directly after excitation takes place) [9] or by Lorentz confinement [10]. In general, it may be argued that the small improvement in resolution does not justify the much greater complexity involved in such experiments.

Carriers redistribution via drift

In device structures, particularly those that rely on band structure engineering, there may be a real risk of significant carrier drift prior to recombination (CL or charge collection (EBIC)). For example, excess carriers in the border regions of vacancy intermixed GaAs/AlGaAs quantum wells were shown to drift for distances of several micrometres [11] prior to radiatively recombining. It was found, however, that fields in excess of 3 x 10^5 V/cm were required to give rise to significant drift effects.

Where drift occurs, its existence is often difficult to prove. Physical masking of the sample is a very effective way of confirming the presence of drift, though effects such as a scan rate dependence of the CL or EBIC signal may also give a clue to its presence. Carrier drift is implicit in EBIC measurements, in that carriers drift within the depletion region of the collecting barrier. Resolution of (e.g.) two closely-spaced p-n junctions would therefore only be possible if the depletion regions were very narrow.

Sample geometry

The geometrical and material properties of the specimen are vitally important in determining the resolution. Self-evidently, if a region of interest lies at a depth of, say, 5µm beneath the surface, then a beam energy sufficient to penetrate that distance into the sample would be required, and the best spatial resolution which could be expected would be of the order of 5µm. Clearly the nearer the surface a feature of interest is, the higher the spatial resolution available. The ultimate extension of this is the study of layers in cross-section. Using this method, Warwick [12] was able to demonstrate 60nm resolution using SEM-CL on a cleaved Bragg stack.

Alternatively, samples can be etched (before or after growth) to form mesas, each of which may encapsulate a feature of interest. If the mesa size is small enough, the spatial resolution may be dictated by the mesa dimensions. Using such a method, Gustafsson et al [13] were able to record spectra from, and image, individual carbon acceptors in a GaAs/AlGaAs quantum well.

In the field of EBIC, Kasai et al [14] were able to demonstrate 100nm resolution of lateral potential modulation in 2 dimensional electron gases by an array of fine (100nm width, 200nm pitch) metal electrodes atop As-doped Si interface control layers. That they were able to this with a beam energy of 10keV strongly suggests that the EBIC images they present are the result of strong signal suppression and amplification. Nevertheless, the lateral potential modulation is apparent.

PRACTICAL SOURCES OF RESOLUTION LOSS

There exist several reasons why theoretical optimum resolution may not be achieved in practice. These are, in general, related more closely to the standard of the equipment available for the experiment and the degree of care taken in setting up the experiment.
Contamination
The most damaging form of contamination is that by carbon or hydrocarbons, most often arising from rotary pump oil back-streaming. Regardless of the front-end pump type and quality, tests using residual gas analysers invariably reveal traces of pump oil. Such contamination is highly inconvenient as it is, in effect, permanent, and not easily removed from the sample. The best solution is to use an oil-free pumping system. Many types of high specification dry pumps are available within the semiconductor industry, but they tend to be costly. A moderate cost option is the use of a two-stage turbo/molecular drag pump backed by a diaphragm pump. Such a system is virtually oil-free and is perfectly capable of obtaining chamber pressures (i.e. at the sample) in the mid $10^{-7}$ Torr range. The only drawback which may be encountered in the use of diaphragm pumps is an increased propensity for vibration (see section below on mechanical vibration), especially as the pump ages. This can usually be cancelled out by extra vibration damping of the vacuum lines to the SEM. Contamination of the chamber can also occur when hydrocarbon-containing substances are inserted. If, for instance, the SEM is also used for e-beam lithography then increased contamination can result from the beam resist layers. This may persist for weeks unless the chamber is baked.

Contamination results in resolution loss in two ways: by increased astigmatism as the contamination charges, and by rendering low beam energy measurements impossible because the beam cannot deliver sufficient energy to the sample through the contaminating layer. This problem becomes exacerbated when performing measurements at liquid helium temperature as atmospheric gases, unavoidably present in the chamber unless metal seals are used, will rapidly condense onto the sample. The use of a liquid nitrogen cold finger will greatly reduce the water and carbon dioxide levels in the vacuum, but will only serve to slow down (not stop) the rate of contamination by other species.

Mechanical vibration
This is sometimes a problem where SEMs are used with many after-market add-ons, such as cryo-stages or cooled detector housings. Cryo-stages, particularly liquid helium stages, have a requirement for good thermal insulation from the main body of the SEM, and so their top-end design is often flimsier than the standard stages normally supplied with an SEM. Simple design flaws such as not anchoring the base of the stage at both the front (usually the door) and back of the SEM chamber can lead to specimen vibrations of amplitudes as large as 20nm. In addition, such stages necessarily have cryogen inlets and outlets (typically steel capillary tubes) and electrical connections (for heaters and sample contacts), each of which will have its own natural frequency and may transmit vibration to the sample holder. This can usually be minimised by tethering any such tubes and wires together with fine wire or PTFE tape, thereby damping the vibration.

The initial source of the vibrations can be something as innocuous as a cooling fan in the housing of a detector, lock-in amplifier, voltage source, or similar. Even when such electronics are not in direct contact with the SEM chamber, vibration can still be transmitted up connectors such as BNC cables. Since fans tend to rotate at speeds which are some multiple of the mains frequency (50 or 60Hz), small vibrations caused by them can easily be mistaken for mains pick-up. Vibration input from the local environment may also increase with time as vacuum pumps age and their bearings or other moving parts wear, so regular monitoring is advisable.

Electrical noise
Since very few SEMs are operated in strict isolation from other equipment, a certain amount of electrical interference is to be expected. Different SEM columns are susceptible to such interference to different degrees. It is well known that electrical interference becomes most evident for beams of low energy at large work distances. These, unfortunately, represent the beam conditions most often used for high resolution CL, namely, low energy for high resolution, but larger than ideal working distance in order to accommodate the collecting mirror. The only realistic solution is to keep electrical equipment (computers, monitors, power supplies, etc) as far as possible from the SEM column. Fluorescent lighting should not be used, and any electrical power cables running beneath the floor or above the ceiling should be re-routed. If noise pick-up remains a problem, a field cancellation unit should be installed. Where a magnetically-levitated turbo is used to pump the chamber, there is a danger of pick-up within the chamber itself, in (e.g.) wires carrying EBIC connections. Such turbos typically rotate at speeds around 48000 rpm, which can lead to a characteristic 800Hz interference.

Astigmatism
The presence of insulating materials on an EBIC specimen holder may cause the beam astigmatism to change with time. Similarly, the presence of low levels of contamination around the hole of a CL mirror may also be subject to charging resulting in a change of the beam shape over time. In severe cases of contamination, this astigmatism may become uncorrectable, requiring the mirror to be
DEMONSTRATION OF RECORD RESOLUTION

Having outlined the factors affecting resolution, can they really be optimised to produce a new standard of SEM-CL resolution? With this specific purpose in mind, the equipment in our EM lab was optimised (and up graded where necessary) over a period of time in the following ways:

1. Rotary pumps in Hitachi S-4500 vacuum system replaced by turbo/molecular drag pumps backed by diaphragm pumps. Oil-contaminated vacuum lines replaced. Chamber pumped by magnetically-levitated turbo pump. Chamber vacuum level and composition monitored using a residual gas analyser.
2. Stage locking plate added to cryostat to defeat fine-scale vibration.
3. Damping of all helium inlet lines and electrical connections to SEM chamber.
4. Field cancellation unit used to avoid EM field interference.
5. Iterative optimisation of CL collection optics, cleaning of mirrors, etc. to maximise CL collection efficiency.

A GaAs/AlGaAs sample was designed and purpose-grown by molecular beam epitaxy (MBE), containing seven sets of three quantum well/barrier combinations, as shown schematically in Fig 3. The thickness of the wells and barriers in each set were 30nm, 20nm, 10nm, 5nm, 60nm, and 80nm. The MBE growth rate was carefully calibrated and the thickness of the well/barrier combinations was checked by microscopy. The layer thickness was confirmed to be accurate to within ±1nm in all cases.

Figure 4a shows the 1.5keV secondary electron image of the cleaved cross-section, taken with the Hitachi ‘E X B’ detector. The contrast: therefore probably arises both from chemical and doping contrast. The sets of wells are visible, right down to the 10nm well/barrier combination. Figure 4b shows the 3keV CL pseudocolour image of a small region of the cross-section containing the 40nm, 30nm, 20nm and 10nm wells/barriers. At this beam energy the 30nm wells are clearly resolved, but the 20nm wells appear only partly so. The successful spatial resolution of the 20nm wells becomes apparent only at a beam energy of 1.5keV, as shown in the linescan of Fig 5. Lower beam energies were attempted, but a trade-off exists between the beam current necessary to generate sufficient CL signal to measure a linescan with a reasonable signal-to-noise ratio, and the ability of the SEM to focus a beam of such (relatively) high current (a few tens of pA) into a small spot size. At 2keV, the dissipation volume size limits the spatial resolution, whereas at 3keV, it is effectively the spot size, given the requirement for higher than optimum beam current. In the case of the sample described above, 1.5keV was found to give the optimum spatial resolution.

It is one thing to demonstrate an ultra-high resolution, but can useful information really be gained from the sample on these length scales? In the example quoted above, it was noted that the first well grown in each set gave significantly lower CL emission (by approximately half — see Fig 5) than the other two, despite high resolution dynamic SIMS profiling showing equal doping concentrations in each well [15]. This effect was visible in wells down to 30nm in width. This was interpreted as evidence of unexpectedly high interfacial roughness of the lowestmost hetero-interface, arising during the growth of the intervening high-alumina containing layers. Thus, useful characterisation has been demonstrated on a length scale of 30nm.

CONCLUSIONS

20nm resolution probably represents the maximum that can be achieved in the present generation of SEMs with the collection and detection systems currently available. A simple extrapolation of the experiment suggests that to do significantly better would require a dedicated UHV instrument, optimised for ultra-low energy beam generation, and very probably in-situ sample preparation, at the very least, in-situ cleaving. In addition, such a system would require the simplest (and therefore most efficient) possible collection optics and more efficient detectors. Under these conditions, however, the technique would effectively become a surface technique, with a very low sample throughput. It is therefore arguable that the improvement in the available spatial resolution would be rendered pointless by the unrepresentative electrical properties of surfaces as compared to bulk or buried layers.

The present work has outlined the many factors affecting the spatial resolution available in SEM-based CL and EBIC. It is important to note that lack of attention to any single one of these factors may completely preclude obtaining the maximum theoretical resolution. Following a strict process of equipment optimisation and sample design, a new world record SEM-CL resolution of just 20nm has been demonstrated.

REFERENCES