STUDIES IN AERONOMY, ASTROPHYSICS AND ASTRONOMY

and

FLUCTUATIONS, NOISE AND QUANTUM ELECTRONICS

by

PAUL J EDWARDS BSc (Hons), PhD (Tas)

SELECTED PUBLICATIONS

SUBMITTED TO THE UNIVERSITY OF TASMANIA

IN FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF SCIENCE

CANBERRA

SEPTEMBER 2002
The sixty publications selected for this thesis cover a period of forty years, from my time as a PhD student in the Physics Department of the University of Tasmania up to the present in 2002 as founding Professor of Electronic Engineering and Applied Physics and Director of the Centre for Advanced Telecommunications and Quantum Electronics at the University of Canberra.

Many of the papers have been co-authored with research students and associates whose valued contributions I have acknowledged below. Unless otherwise stated I have only included accounts of work in which I have played a leading part. I have divided the papers, in some cases somewhat arbitrarily, into two groups which reflect my interests over this period:

(a) Aeronomy, Astrophysics and Astronomy; and
(b) Fluctuations, Noise and Quantum Electronics.

Within these two groups I have attempted a further subdivision in which I have selected and presented papers to illustrate the development of specific concepts and lines of approach, usually in chronological order. This has meant the inclusion of a number of abstracts, short papers and two patent descriptions.

(a) AERONOMY, ASTROPHYSICS AND ASTRONOMY

The publications in this section cover a period of twenty seven years, starting with my PhD studies at the University of Tasmania in 1960. They include work performed at the University of Adelaide, the University of Texas at Dallas, the University of Otago and finally at the University of Canberra, prior to the commencement of my work in the field of quantum electronics and quantum optics in 1989.

My interest in transient ionizing events began with upper atmosphere radiation measurements following the high altitude thermonuclear explosion “Starfish Prime” that I made as a PhD student working under the supervision of Dr Peter Fenton. I constructed the balloon borne Geiger counter instrumentation and the telemetry system used in the measurements and analysed the data that were reported in the journal Nature [1]. In the following paper [2] I collated a variety of observations made at Hobart by fellow graduate student John Reid and other observers. I attempted to link these and the radiation measurements to a common cause. I concluded that there was good evidence for a breach of the geomagnetic field leading to the widespread dispersal of radioactive fission debris in both hemispheres [3,4,5], although electron precipitation from the terrestrial radiation belts could not be excluded. This view was supported in the following two decades when observations and models of the event were finally published in the open literature.

Paper [6] is an account of measurements of bremsstrahlung radiation from auroral zone electron precipitation made at Hudson’s Bay as a member of a team from the University of Texas at Dallas working under the direction of Professor Ken McCracken, a graduate of the University of Tasmania. Paper [7] is an account of rocket measurements of solar x-ray and ultraviolet radiation, again made as a member of a team, in this instance from the University of Adelaide, working under the direction of Professor John Carver. I instrumented the x-ray payload and analysed the results. This work was a precursor to the launch of the first Australian scientific satellite, WRESAT 1, instrumented at the University of Adelaide and launched from Woomera.

After I moved to the University of Otago in 1968 this work stimulated my interest in deducing atmospheric parameters from solar radiation measurements [8-10]. I believe that PhD student Ms Mazlan Othman and I were among the first to suggest [8] that routine photometric observations at existing optical astronomical observatories could be utilised to monitor globally the atmospheric content of ozone and aerosols present in the atmosphere.
I subsequently convened an IAU working party and colloquium on the use of astronomical data to characterise variations in atmospheric turbidity due to volcanic eruptions.

While at the University of Otago I participated in a meteorological survey of surface wind characteristics as a member of a New Zealand Wind Energy Task Force established at the time of the first oil crisis. Two papers relating to wind energy generation and site characterisation are listed here [11,12]. The first of these was co-authored with my colleague and co-director Mr Keith Dawber, with whom I shared the planning and execution of a regional wind energy survey of the South Island of New Zealand. My monograph on wind power economics and technology and three co-authored formal reports published by the New Zealand Energy Research and Development Committee have not been included here. In the intervening 25 years wind power has become economically viable. Extensive wind energy farms have been selected and established worldwide on the basis of principles established in the New Zealand survey and other regional surveys. Further publications arising from this work relating to stochastic phenomena are listed below [33-36].

Before moving from the University of Adelaide to the University of Otago in 1968 I participated in auroral zone [6] x-ray measurements from balloon borne platforms in Canada, and astronomical x-ray measurements [13,14,15] from balloon borne platforms in the USA and Australia under the direction of Professor Ken McCracken.

I also performed theoretical modeling studies relevant to the interpretation of astronomical X-ray observations. In particular I devised models for transient celestial x-ray sources [16,17] that were then being observed for the first time. The first of these [16] was written in collaboration with PhD student John Harries. In two later papers written in New Zealand [18,19], I addressed the possible contribution made by a class of young active red dwarf stars known as "flare stars" towards (a) the diffuse galactic x-ray background and (b) the galactic cosmic ray background particle flux, the latter in collaboration with PhD student Malcolm McQueen. I showed that flare stars were unlikely to make significant contributions to the x-ray background. We also demonstrated that the flare star contribution to the galactic cosmic ray flux density was negligible, contrary to several long-standing suggestions in the literature. These conclusions have since been confirmed. In 1971 I was awarded the Murray Geddes Prize for astronomical research by the Royal Astronomical Society of New Zealand.

While at the University of Adelaide I had became interested in the possibility of detecting stellar x-rays from the ground based on my earlier work on ionising radiation in Tasmania and the USA [1-6]. At that time x-ray astronomical satellites were not yet in orbit and observations were only possible from rocket and balloon platforms. Following my arrival in New Zealand, I made contact with Mr Godfrey Burtt and Mr Fred Knox of the ionospheric section of the New Zealand Department of Scientific and Industrial Research with a view to searching their VLF radio propagation records for evidence of ionospheric effects caused by the x-radiation from celestial sources, similar to those due to solar flares, first documented by Professor Ronald Bracewell while studying at Cambridge.

We discovered and published the first evidence for an extra-solar x-ray influence on the terrestrial ionosphere in 1969 [20,21]. This stimulated a lively debate in the ionospheric literature and the reality of the effects was initially questioned on both theoretical and observational grounds. Following additional observation, extensive analysis and review, the ionospheric perturbations due to the celestial x-ray source Scorpius XR1 were confirmed to be significant, although somewhat smaller than we had initially reported. On this basis I was able to set conservative upper limits [22,23] to the prompt x-ray flux from the large supernova SN1987a. This provided early support for the pre-explosion blue dwarf stellar model of SN1987a which was finally adopted.

On the strength of this work I suggested that large non-solar x-ray and gamma ray bursts would be detectable by global networks of VLF radio receiving networks [24]. However, positive observations of large ionospheric effects due to transient galactic x-ray bursts did not take place until 1998. Unambiguous measurements of the effects of a large gamma ray burst on the nocturnal Pacific ionosphere were then made with networks designed to detect electron
precipitation events by Professor Richard Dowden of the University of Otago and Professor Umran Inan of Stanford University.

I was invited to present a review of these two sets of observations at the 26th General Assembly of URSI held at Toronto in 1999 [25]. A comprehensive review is in preparation but is not included in this thesis.

(b) FLUCTUATIONS, NOISE AND QUANTUM ELECTRONICS

The papers in this section have a common theme: the analysis of physical systems and processes in which random and chaotic fluctuations play an important part.

I wrote the first paper [26] while a PhD student in the Physics Department at the University of Tasmania engaged in the radio telemetry of cosmic ray data from balloons. I believe it was the first attempt to optimise the choice of bearer frequency for terrestrial VHF radio links taking galactic and other radio noise sources into account.

The second paper [27] was co-authored with graduate student Robert Hurst who performed the measurements under my supervision as part of a physics honours project at the University of Otago. It confirmed our expectation (based on the central limit theorem) that the narrow-band rectified electrical noise after low pass filtering would follow Gaussian statistics, contrary to assertions in at least one contemporary monograph on radio astronomy.

The following five papers [28-32] describe an investigation I initiated into the use of cross-correlation techniques to detect weak time-variations of physical interest in the presence of masking noise. This work was performed with the assistance of graduate students Robert Hurst, Malcolm McQueen, Grant Christie, Graham Stanley and research associates William Allen and Mervyn Thomas who were engaged in projects in the fields of radio astronomy, optical astronomy and night-sky photometry under my leadership.

The application of these techniques in these fields was considered innovative at the time of publication but has now become common practice. Paper [32] is of particular interest in that it provided an early indication of the importance of low level electron atmospheric precipitation from the radiation belts in maintaining the mid latitude ionosphere. This issue also arose in connection with my earlier discovery (with Burtt and Knox [20,21]) of weak ionospheric perturbations due to celestial x-ray sources.

Paper [33] is a short mathematical analysis of the unexpected and (until recently) unexplained power law character of horizontal wind speed gust durations discovered during the course of a survey of surface winds in New Zealand [11,12]. In this paper I also drew attention to the apparent ubiquity of the power law characteristic. At that time fractal analysis was in its infancy and level crossing statistics were usually regarded as being in the "too hard" box. In 1989 I noted [34] an interesting application of zero-crossing interval analysis in the field of spectral estimation that I later confirmed while on study leave at the CSIRO Division of Radiophysics working with Dr Warwick Wilson.

After a lapse of 20 years I revisited level-crossing phenomena in an invited presentation to a 1999 conference on "Unsolved Problems of Noise" [35], suggesting that a reanalysis of the data from a fractal viewpoint might be fruitful. I was subsequently invited to prepare a paper along these lines for publication in the journal Chaos. My former student and colleague Dr Robert Hurst had played a major role in identifying the wind characteristics in the original New Zealand wind energy survey. He kindly agreed to assist me in preparing the invited paper [36] by reanalysing and modelling some of the original data logged on magnetic tape that he was able to access in New Zealand. Largely as a result of his endeavours we were able to confirm the fractal characteristics of the horizontal wind field and to successfully simulate these to a good approximation by a Markov process. Although originally conceived in the context of aero-generator performance, this analysis also appears to have other unexpected practical applications, for example to the dynamics of bushfires.
On a more fundamental level it should help to clarify the connection between the properties of continuous-time stochastic processes and the fractal statistics of the point process generated by their level crossings.

Following my return to Australia from New Zealand in 1982 I became interested in the fields of non-classical "quantum" optics and electronics in which the quantum character of light is accessible through measurements of photon counting noise and photo-current shot noise fluctuations. I attended a seminar at the Australian National University in 1989 in which I heard that NTT researcher Dr Yoshihisa Yamamoto had recently demonstrated photonic shot noise reduction (and by implication sub-Poissonian photon counting statistics) using cooled semiconductor lasers.

I subsequently succeeded in performing the first Australian (and the second international) demonstration of photonic noise reduction using light-emitting semiconductor junction diodes [37]. I also confirmed for the first time the non-classical statistics of the emitted photons by showing that the maximum fractional shot noise reduction closely tracked the quantum counting efficiency as this latter quantity was varied [37-39]. In the latter paper, published in the Physical Review Letters, I also demonstrated for the first time the generation of correlated photon “twin beams” using semiconductor light emitters [39]. Co-author Professor Graham Pollard was responsible for an important link equation (6) in the theoretical development in this paper. He also generalised (to N beams) the equations for the conditional single-photon multiplet source derived from this concept in a later publication [60].

In 1993 I published a semi-classical interpretation of the high-impedance suppression of shot noise in light-emitting diodes [40]. More detailed models proposed by the Yamamoto group at Stanford and the Yamanishi group at Hiroshima University incorporating the dynamics of the injection process have since been shown to be superior [58]. However, this “leaky reservoir” model remains a satisfactory first-order model of shot noise reduction under conditions of high charge-carrier injection in semiconductor junctions.

My work on quantum noise suppression in light-emitting diodes has been incorporated into undergraduate physics and engineering courses at the University of Canberra and Macquarie University [41]. When working on sabbatical leave at the UK Defence Research Agency with Professor John Rarity I coined the words “quiet light” to describe the character of quantum intensity noise-suppressed light. My work on quiet light was featured in the New Scientist (June 28 1997, Forum p4).

Professor Rodney Tucker, winner of an Australia Prize for his work on laser modelling, wrote the following generous comment in assessing my 1992 research grant application in this field.

"The importance of this work was highlighted to me when I recently attended the international conference on lasers and electro-optics in California. There was a keynote invited address by Dr Yamamoto from NTT, the recognised world leader in the area of "squeezed light". In his keynote address, Dr Yamamoto referred to only one other researcher apart from those in his own group. This other researcher was Professor Edwards. This clearly indicates the high international esteem and reputation that Professor Edwards has received for his work in this area".

In 1991 after spending sabbatical leave at the NTT laboratories in Tokyo as the guest of Dr Yamamoto, I presented a paper in Vienna in which I reported for the first time the generation of positively (and negatively) correlated quantum noise fluctuations in photon “twin-beams” achieved by electrically connecting two light-emitting diodes in series (or parallel). As stated above, this was subsequently published in the Physical Review Letters [39] and was later patented [42]. Up to this time the only method used to generate correlated light beams involved parametric down conversion in non-linear optical media. It then occurred to me that if the light from a series-connected array of laser or light-emitting diodes could be efficiently collected and detected, a quantum noise-suppressed optoelectronic “photon-number amplifier” could be implemented. This concept was reported in the Electronic Letters [43] and was selected by the Optical Society of America as a principal advance in the field of quantum optics in 1993 [44].
I developed the concept of “wired” optics further at an invited plenary paper presented to an international quantum optics meeting in Rotorua [45]. In 1995, in collaboration with Dr Yong-qing Li and other members of my research group I was able to demonstrate for the first time [46] the generation of quantum-correlated twin beams using laser diodes provided by Professor Yamamoto at Stanford University.

An investigation of the second order statistics characterising the correlated fluctuations in twin-beam sources of this kind was reported in [47]. This showed that certain classical inequalities were violated in a manner that unambiguously revealed the quantum (non-classical) character of infrared and visible light fields, even in the macroscopic case. This work was carried out jointly with Dr Y-q Li and Dr Huang Xu. It showed that these non-classical characteristics were not restricted to weak fields as had often been assumed. Dr Li later co-authored a formal quantum mechanical interpretation of our joint work on correlated shot noise from semiconductor twin-beam sources [48]. At the suggestion of Dr Xiao of the University of Arkansas, Dr Li and Dr Peter Lynam demonstrated the use of amplitude-squeezed light in improving the quality of light-wave based measurements [49]. Through illness during 1996/8 I was unable to participate directly in this experimental work. However I made a major contribution to the analytic interpretation and wrote the text of the paper that subsequently appeared in the *Physical Review Letters* [49].

My earlier work on shot noise-suppressed photon number amplification [42-45] enabled me to achieve a better understanding of the origin of shot noise in electronic and photonic devices. This work formed the basis of several conference presentations, journal publications and an invited book chapter in a monograph on noise in electronic devices [55-58]. It also led to a second invention that became the subject of Australian, Japanese and US patents [42,51].

This invention was based on the application of positive feedback to the open loop amplifier of [43] to achieve increased current gain [50]. This concept was embodied in international patents [51] in which my colleague Dr William Cheung and I were cited as co-inventors. We subsequently carried out a detailed shot noise analysis of the closed loop optoelectronic photon-coupled amplifier [52] and this was incorporated into the final version of the patent specifications. Dr Y. Mizushima and colleagues at Hamamatsu Photonics facilitated this work which was featured in the journal *Laser World Focus* in 1997.

I subsequently realised that the closed loop configuration was similar to (but more general than) that of a “photon-coupled transistor” and that the phenomenology of noise suppression in photonic devices had an exact analogue in semiconductor junction transistors. This led me to develop a successful “neo-corpuscular” model of shot noise generation, propagation and suppression in electronic and photonic devices [55-58] based on the original work of Aldert van der Ziel in the 1950s. In 1999 I presented an invited paper [56] co-authored with graduate student Richard McDonald and co-supervisor Dr Cheung that applied this model to a bipolar junction transistor.

With the assistance of Professor Yamamoto and members of the semiconductor device development group at Hamamatsu Photonics we succeeded in 2000 in demonstrating for the first time the operation of a photon-coupled transistor employing a quantum noise suppressed laser, the exact photonic analogue of a bipolar junction transistor [54].

In 2001 I was invited to write a chapter on sub-Poissonian electronic and photonic noise in semiconductor junctions for a monograph on noise in electronic devices. This reviews several decades of work on the suppression of quantum noise in light-emitting junctions [58], including work on light-emitting diodes with ME student Ting-ting Zhang [55]. Paper [53] describes work on the characterisation of light-emitting diodes performed with colleagues Dr Graham French and Dr Cheung.

My most recent work in the field of quantum fluctuations has been an investigation of quantum cryptographic techniques. I initiated, supervised and helped implement the first demonstrations of quantum key distribution in Australia [59]. In 2001 I gave an invited presentation of this work and a related analysis of a conditional single-photon source [60] based on original concepts introduced in [42-46] at an international workshop held in Corsica.
Acknowledgements

I acknowledge with thanks the invaluable guidance and assistance of the academic colleagues, graduate students and research assistants, particularly those named above, with whom I shared the excitement and the frustrations of the research ventures sketched in this thesis. In the preface I have attempted to acknowledge specific contributions to the work reported here. I apologise for the inevitable inadvertent omissions.

Paul J Edwards 10-09-02
CONTENTS

(a) AERONOMY, ASTRONOMY and ASTROPHYSICS


14. X-ray Surveys of the Quiet Sun, the Magellanic Clouds and the Cygnus Region; Proc. Astr. Soc. (Aust), 1, 28 (1967).


**NOISE, FLUCTUATIONS and QUANTUM ELECTRONICS**


44. Squeezed Light-Emitting Diode Arrays, *Optics and Photonics News*, "Optics in 1993", 39, Dec 1993 (Selected Contribution to special issue on "Principal Advances in Optics in 1993").


57. Semiconductor junction noise revisited: where have all the physical noise sources gone? Fluctuation and Noise Letters 1, 3, C15-C19 (2001).


END
PART (a):

STUDIES IN AERONOMY, ASTROPHYSICS AND ASTRONOMY
Radiation Enhancement following Johnston Island Thermonuclear Explosion

We wish to report the observation, made at Hobart, Tasmania, of an enhancement in the counting-rate of a balloon-borne Geiger counter at the time of the high-altitude thermonuclear explosion above Johnston Island on July 9, 1962.

The counter, designed for operation at the temperatures encountered in night flights of balloons (down to $-60^\circ$ C), was 10 cm long and 4 cm in diameter. Thicknesses of the glass wall and internal nickel cathode were 1 mm and 0.1 mm respectively. At the time of the explosion, 0900 U.T., the balloon carrying this counter (mounted horizontally) and associated circuitry had reached an atmospheric depth of 80 g cm$^{-2}$ and was rising in altitude at the rate of 2.5 g cm$^{-2}$ min$^{-1}$. The balloon burst prematurely at 0927 U.T.

Fig. 1 summarizes the observations. The maximum error in timing is estimated at $\pm 2$ sec.

Significant features of the event are the considerable atmospheric depth at which it was observed, the delay of between 20 and 30 sec between the instant of the explosion and commencement of the enhancement, and the peak intensity (averaged over 10-sec intervals) of about 50 per cent above the normal background-level near the Pfotzer maximum of the cosmic ray transition curve. There seems little doubt that recovery to the normal background intensity was not complete until at least 0915 U.T. No effect was observed by cosmic ray neutron monitors and meson telescopes at ground-level in and near Hobart.

It is clear that if the enhancement were due to protons incident vertically at the top of the atmosphere their minimum energy was 350 MeV. However, the line of force through Hobart (geographic coordinates 43° S., 147° E., geomagnetic latitude 51° S.) crosses the equatorial plane at 2.5 Earth radii, and thus passes outside the region in which protons of energy exceeding 70 MeV were observed by Explorer VII. The possibility that other types of radiation were responsible for the enhancement must be examined, but we consider that an attempt at this
Fig. 1. Main diagram, 1 min count totals from balloon-borne Geiger counter and atmospheric pressure level versus time; Inset, variation of 10-sec count totals through time of Johnston Island explosion (0900 U.T.)

time to interpret the observations from the information available to us would be premature.

We wish to acknowledge the assistance of Mr. H. E. G. Dyer, who manned a duplicate radio telemetry station.

P. J. EDWARDS
A. G. FENTON
K. B. FENTON
J. G. GREENHILL

Physics Department,
University of Tasmania,
Hobart, Tasmania.

N. R. PARSONS

Antarctic Division,
Department of External Affairs,
Melbourne, Victoria.


Printed in Great Britain by Fisher, Knight & Co., Ltd., St. Albans.
Effects of Nuclear Explosion Starfish Prime
Observed at Hobart, Tasmania, July 9, 1962

P. J. Edwards¹ and J. S. Reid²

Department of Physics, University of Tasmania
Hobart, Australia

Abstract. Observations, at Hobart, Tasmania, of magnetic ionospheric and radiation effects of the Starfish shot are presented. The interpretation of the results is briefly discussed. Consideration of the magnitude of the ionospheric absorption that accompanied a radiation enhancement recorded by a balloon-borne Geiger counter appears to exclude atmospheric precipitation of electrons as the cause of the enhancement.

Introduction. This paper presents geophysical observations of the effects of the nuclear explosion Starfish Prime. The data were obtained at Hobart (147°E, 43°S geographic; 224.5°E, 52°S geomagnetic) on July 9, 1962.

The observations to be described include the following:

1. A radiation enhancement recorded by a balloon-borne Geiger counter at 80 g/cm² atmospheric depth.
2. The occurrence of magnetic micropulsations in the vertical (Z) component of the earth's field.
3. Fluctuations in the geomagnetic field components (X, Y, Z) recorded by a flux gate magnetograph.
4. Sudden ionospheric absorption (SCNA) recorded by wide- and narrow-beam equipments at 4.7 Mc/s.

The interpretation of the phenomena to be described is facilitated by the limited geographical distribution of the detecting instruments. It is hoped, therefore, that the present results will be of some value in elucidating the complex mechanisms that produced the varied effects observed at Hobart and elsewhere (e.g., Gregory [1962]; Durney et al. [1963]; Casaverde et al. [1963]; Basler et al. [1963]).

Radiation enhancement. A balloon-borne counter, mounted horizontally, with associated equipment was launched from Hobart before the detonation. The balloon had reached an atmospheric depth of 80 g/cm² (60,000 feet altitude) and was about 50 km southeast of Hobart at the time of the explosion, 0900.09 UT, hereafter referred to as H. The counter was 10 cm long and 4 cm in diameter. The thicknesses of the glass wall and internal nickel cathode were 1 mm and 0.1 mm, respectively. Calibration of an identical unit after the flight established the photon efficiency to be \(\eta(v) = [0.75v (\text{Mev}) - 0.25] \%, \ 0.5 < v < 3 \text{ Mev}\).

At \(H + 10\) seconds a radiation increase occurred, which reached peak intensity of more than 70% of the cosmic-ray background level at 0901 UT. This radiation enhancement, of which a preliminary account has already been given [Edwards et al., 1962], decayed rapidly to reach a roughly constant level between 0904 and 0910 and persisted until after 0915. The raw data from the balloon flight, uncorrected for the system resolving time of 2 msec, are shown in Figure 1.

Significant features of the event include the 10- to 15-second delay to the onset, the rapid rise to maximum intensity at \(H + 50\) seconds, the rather slower decline to a minimum at \(H + 200\) seconds, and the persistence of the enhancement for a period of at least 15 minutes. The magnitude of the enhancement after 0904 is somewhat uncertain because of the relatively high cosmic-ray background level. The balloon burst prematurely at 0927, having reached a depth of 40 g/cm².

Durney et al. [1962] observed a radiation burst at about the same time with the shielded...
Fig. 1. Main diagram: 1-minute count totals from balloon-borne Geiger counter and atmospheric pressure (millibars). Inset: 10-second count totals through time of Starfish explosion (0900.09 UT).

Anton 302 counter in satellite Ariel. They suggest that the satellite burst was due to hydro-magnetic redistribution of the pitch angles of naturally trapped electrons. The Hobart observations might then be interpreted in terms of the bremsstrahlung radiation from dumped electrons. We shall postpone further discussion of the balloon result and consider it in relation to the other Hobart observations, since, in the absence of local supporting data, the interpretations are necessarily open to question.

Magnetic micro pulsations. A micropulsation recorder, installed at a field station 10 miles from Hobart, was in operation at the time of the detonation. In addition to the normal chart record, the data were tape-recorded at 7¼ inches per second together with the Johnston Island countdown. The detector was a horizontal loop coupled to a galvanometer and photocell unit. The response of this unit and the associated amplifier was flat up to a frequency of 0.05 cps and dropped by 4 db per octave above this frequency.

Figure 2 shows the radiation increase and the micropulsation record plotted on the same time scale. The irregular pulsations that took place in the first 30 seconds are followed by a train of 5 quasi-sinusoids of approximately 16-second period. Particular interest attaches to these oscillations because of their narrow frequency spectrum and their time relation to the radiation burst. MacDonald's [1961] low-density model predicts a period of 16 seconds for the fundamental V mode oscillation of the field line through Hobart. Power spectrum and Chree analyses of the radiation and micropulsations data (to be presented in a later paper) confirm an association between the Geiger counting rate variations and the phase of the quasi-sinusoidal micropulsations.

In Figure 3 the onset of the artificial micropulsations is shown. The depression of the vertical (upward) component of the field within 0.1 second of the detonation is clearly apparent. This effect, presumably due to an electromagnetic pulse propagated in the earth ionosphere cavity [Wait, 1960], has been reported by other observers [Casaverde et al., 1963; Roguet et al., 1963; Odencrantz, 1963]. At \( H + 1.4 \pm 0.1 \) second a sharp increase in \( |Z| \) occurred and was the first peak in a train of three quasi-sinusoids of period 0.7 second. Saturation of the recorder \( (|Z| > \frac{3}{2} \text{ gamma/sec}) \) took place at \( H + 3.5 \pm 0.1 \) seconds, and rapid fluctuations continued until \( H + 20.9 \) seconds, when a large positive excursion lasting 3 seconds again saturated the instrument. The regular oscillations already referred to began after +30 seconds and reached peak to peak amplitudes in excess of 1 gamma. The time delay of 1.5 seconds to the first peak of the short-period oscillations is more than twice the delay \( (0.6 \pm 0.2 \text{ second}) \) observed at Wellington, New Zealand [Christoffel, 1962]. Roquet et al. [1963] have drawn attention to the worldwide observation of a magnetic impulse at \( H + 2 \) seconds. There is some suggestion of an impulse at Hobart at this time (Figure 3), but its recognition is made somewhat uncertain by the prior arrival of the short-period oscillations.

Magnetograph record. Flux gate variometers operated at Hobart by the Australian Bureau of Mineral Resources recorded variations in the \( X, Y, Z \) components of the field after the deto-
Fig. 2. Balloon Geiger counting rate, corrected for resolving time loss and the magnetic micropulsations record following the Starfish explosion.

nation (private communication, Dr. D. Parkinson). The chart speed of 1 1/2 inches per hour restricted the time resolution of the data, but it is clear that the total field strength was depressed in the first minute after the detonation and enhanced for the following period of about 15 minutes. The maximum deviations in the X (north), Y (east), Z (down) components were, respectively, −9, −13, +10 gammas during the first phase of the event and +20, +9, −9 gammas during the second phase.

This event is of more complex form than the baylike disturbance recorded by near-equatorial stations [Pisharoty, 1962; Glover, 1963; Casa-verde et al., 1963]. The mean S$_0$ variations at Hobart for July at the time of the event have

Fig. 3. Onset of the magnetic micropulsations activity, showing depression in the magnitude of the vertical field component within 0.1 second of the detonation.
been estimated by Parkinson (private communication) as approximately +6, +4, −1 gammas, respectively, in X, Y, Z. The fluctuations in the horizontal field during the second phase of the event would therefore correspond to an increase of a factor of 2 or 3 in the strength of the local $S_z$ current system.

**Ionospheric effects.** A number of cosmic noise receivers were in operation at the time of the detonation, including narrow-beam (3° by 11°) and wide-beam radio telescopes operating at a frequency of 4.7 Mc/s and a 30-Mc/s vertical wide-beam (80° by 120°) riometer.

No detectable absorption was observed with the 30-Mc/s riometer, but a doubly peaked absorption event of amplitude 6 db was recorded with the narrow-beam antennas (declinations of −52°, −47°, −32°) that had been receiving appreciable cosmic noise before the detonation. Figure 4 shows a typical chart record. The recovery after the initial (4-db) absorption peak was probably greater and more rapid than that observed because of the long return time constant (60 seconds) of the minimum reading circuitry. Owing to the slow chart speed it is difficult to assign an accurate onset time to these effects, but the results are consistent with a time coincidence between the first absorption peak and the peak of the radiation burst. There is little evidence for a second radiation increase at the time ($H + 2.5$ minutes) of the second absorption. Within the limits of accuracy ($±0.5$ db, $±30$ seconds), all narrow-beam equipments recorded equal absorptions at the same times. The wide-beam absorption ($5 ± 0.5$ db) was certainly not greater than the narrow-beam absorption and although of the same duration (15 minutes) did not show a double peak.

As the Hobart ionosonde was not operating at the time of the explosion, no critical frequencies are available for this time. Comparison with the maximum galactic noise level at 4.7 Mc/s, however, shows that the noise level was depressed by several decibels throughout the night of July 8 and until about 1300 UT (2300 local time) on the night of July 9. The normal galactic intensity was recorded after $H + 4$ hours and also during the following night of July 10.

**Conclusion.** The broad features of the time relation between the Geiger counter, riometer, and magnetograph events are shown in Figure 5. Although a number of different mechanisms were probably involved it is clear that the observations impose quite severe restrictions on the interpretation of the radiation increase.

The Hobart observations of the ionospheric effects following the Starfish explosion suggest that the enhanced Geiger counting rate observed at balloon altitudes was due to radiation of a considerably more penetrating nature than that normally observed during auroral dumping or low-energy solar-flare events. Observations [e.g., Pfotzer et al., 1963; Webber, 1962] of such phenomena show that radiation increases of comparable magnitude at the same atmospheric depth would normally be accompanied by greater $D$-layer absorption than was in fact recorded during this event.

A detailed analysis [Edwards, 1964] shows that interpretation of the radiation increase at 80 g/cm² in terms of bremsstrahlung X rays from dumped electrons with energies less than...
count for the radiation increase would, however, not be expected to cause D-layer absorption in excess of that observed. Protons with energies between 0.4 and 2 gev, the upper energy limit being imposed by the lack of a detectable increase in the Mount Wellington neutron monitor rate, are therefore not excluded by the riometer observations as a possible cause of the Geiger counting rate enhancement. Gamma rays, for example, the delayed gamma radiation from fission products, are also a possible cause of the enhancement.

Acknowledgments. We wish to thank Professor G. R. A. Ellis for providing the radio telescope data, and Dr. D. Parkinson and Mr. C. Bisdee for making the magnetograph records available.

REFERENCES


(Manuscript received December 30, 1963; revised May 29, 1964.)
Discussion of Paper by A. G. Petschek,  
‘Interpretation of Satellite Detector Counter Rates’

P. J. EDWARDS¹

Department of Physics, University of Tasmania
Hobart, Tasmania, Australia

Petschek [1963] has computed a value of 2·10⁻⁴ for the fission β counting efficiency of the lead-shielded Anton Geiger counter in the Injun satellite. Since the efficiency of heavily shielded counters to nonpenetrating electrons is of importance in the interpretation of satellite radiation measurements [Van Allen et al., 1963; Hess, 1963], I wish to draw attention to several errors in Petschek’s paper. I wish also to comment briefly on the thick target electron bremsstrahlung process, as it is relevant to the problem of space radiation hazards and of the interpretation of auroral X-ray measurements as well as to the efficiency of shielded Geiger counters.

Petschek takes the differential energy spectrum of photons radiated by an electron of initial kinetic energy \( E \) Mev stopping in a thick target to be

\[
n(\nu, E) \cdot d\nu = C \frac{d\nu}{\nu} \left( 1 - \frac{\nu}{E} \right)
\]

and then proceeds to calculate the photon yield for an electron energy spectrum \( N(E) \) as

\[
n(\nu) \cdot d\nu = d\nu \cdot \left[ \int_{\nu}^{\infty} n(\nu, E) \cdot N(E) \, dE \right]
\]

where \( N(E) \) is the number of electrons of energy greater than \( E \) that fall on the lead.' This statement is clearly incorrect, and \( N(E) \, dE \) must be taken to be the differential energy spectrum, that is, the number of electrons with energies in the interval \((E, E + dE)\).

Another point requiring comment is the calculation of the bremsstrahlung radiation energy yield from the Bethe-Heitler [1934] relation

\[
\frac{(dE/dx)_{\text{RAD}}}{(dE/dx)_{\text{ION}}} \approx \frac{ZE_0}{800} = \frac{Z(E + \mu)}{800}
\]

¹ Now at Department of Physics, University of Adelaide, Adelaide, South Australia.
The experimental results of Buechner et al. [1948] and Petrauskas et al. [1943] for relativistic energies are in good agreement with equation 4.

Petschek's equation for the counting efficiency \( \epsilon \) for nonpenetrating electrons with normalized differential energy spectrum \( N(E) \) may then be rewritten as

\[
\epsilon \approx \frac{Z}{800} \int_{v_0}^{E} \frac{E}{\nu} \left( 1 - \frac{\nu}{E} \right) \cdot N(E) \cdot \eta(\nu) \cdot d\nu
\]

where \( \eta(\nu) \) is the efficiency of the counter for photons of energy \( \nu \) Mev. Neglecting photoproduction of electrons in the steel counter wall, Petschek sets the photon efficiency to zero for energies below \( \nu_0 = 0.5 \) Mev. In the present case this does not lead to a large underestimate of the efficiency because of the absorption of lower-energy photons in the lead shield. In general, however, photoproduction may make a significant contribution to the efficiency, particularly for a high-Z, thin-walled counter shielded by a light material. Evaluation of equation 5 using a figure for the photon efficiency derived from experimentally determined values for counters with wall materials of similar atomic number [Curran and Craggs, 1949, p. 89], and putting \( \nu_0 = 0.5 \) Mev, gives a value for the fission \( \beta \) efficiency of about \( 6 \cdot 10^{-4} \) for the Anton counter shielded by 3 g/cm\(^2\) of lead. This figure agrees well with the experimental value of \( 10^{-4} \) obtained by Motz and Carter [1963].

In conclusion we note that the efficiency of a shielded counter to nonpenetrating electrons of energy \( E \) Mev may be written in more general form as

\[
\mathcal{E} (E) = \int_{0}^{E} n(\nu, E) \cdot T(\nu) \cdot \eta(\nu) \cdot d\nu
\]

where \( n(\nu, E) \), the photon spectrum, has already been defined and \( T(\nu) \) is the transmission of the shield to photons of energy \( \nu \). If the shield thickness is such that the transmission is either 1 or 0 according as \( \nu \) is either greater or less than some photon energy \( \nu_0 \), and if Compton production provides the main contribution to the photon efficiency, we have \( \eta(\nu) = K(Z) \cdot \nu \) and equation 6 becomes

\[
\epsilon(E) = \int_{\nu_0}^{E} \frac{Z \cdot K(Z)}{1600} \cdot \frac{E - \nu_0}{1660} \approx \frac{K^2E^2}{1600} \text{ for } E \gg \nu_0
\]

Acknowledgment. I wish to acknowledge useful discussions on the content of this letter with Dr. K. B. Fenton.

REFERENCES


Petrauskas, A. A., L. C. Van Atta, and F. E. Myers, Measurements on X-ray production and absorption in the range 0.7 to 2.5 Mv, Phys. Rev., 63, 389-399, 1943.


(Received November 29, 1963.)
Discussion of Letter by Oscar P. Manley and Jack W. Carpenter, 'Starfish Debris Measurements from Cosmos 5'

P. J. Edwards

Graduate Research Center of the Southwest, Dallas, Texas

In a recent letter Manley and Carpenter [1966] use Cosmos 5 observations [Galperin and Bol'zunova, 1964] of a γ radiation burst on July 9, 1962, to deduce the quantity of fission debris above the satellite horizon over Johnston Island ($L > 1.25$) immediately following the Starfish nuclear detonation. They calculate at least 0.1 MT of fission debris at altitudes in excess of 1200 km above Johnston Island 1 second after detonation.

I wish to point out that the same deduction from the Cosmos 5 data has already been made [Edwards, 1965] and that, in the same paper, γ-ray observations from Ariel 1 [Durney et al., 1964] and from a stratospheric balloon [Edwards et al., 1962] were used to shed further light on the initial motion of the Starfish debris.

At the instant of detonation the Ariel, Cosmos, and balloon-borne Geiger counters were situated to the west of the Johnston Island magnetic meridian. All three detectors were approximately 8000-km great circle distance from the explosion. Ariel 1 was located at an altitude of 820 km at $L = 4.8$, several degrees west of the meridian. The balloon, at an altitude of 60,000 feet at $L = 2.5$ was 32°W of the meridian.

By assuming the radiation bursts to be due to the delayed γ activity of fission debris, one may reach the following conclusions regarding the dispersion of the debris:

1. The two satellite observations are consistent with the transport of $\sim 10\%$ of the fission products above 1200 km and $\sim 1\%$ above 2000 km within 1 second of detonation. Containment and deposition of a large fraction of the debris on the magnetic shell ($L = 1.12$) through the detonation site as evidenced by conjugate point observations [D'Arcy and Colgate, 1965] is not necessarily in disagreement with this model. During transit to the southern deposition area, magnetically guided debris reached altitudes of at least $2.10^6$ km ($L > 1.3$), rose above the Cosmos and Ariel horizons, and may have caused the fast rise and fall in the γ-ray intensity observed with Cosmos and also saturated the Ariel geiger counter.

2. To account for the continued saturation of the Ariel counter till $H + 200$ seconds, significant injection of debris at higher $L$ values must occur. Precipitation of the order of 1% of the fission products at $L \sim 2.5$, would, for example, be consistent with the Ariel result.

3. The balloon observations (at $L = 2.5$) of a γ-ray burst beginning 15 seconds after detonation imply westward drift of debris toward the balloon meridian. The time delay is consistent with the longitudinal drift of magnetically trapped positively charged ions with energies per unit charge between 5.5 Mev/Z (corresponding to the onset time) and 2 Mev/Z (corresponding to the peak of the radiation increase) followed by the arrival of debris with much lower energy for the following 15 minutes.

At the beginning of their range, fission products have kinetic energies per unit charge peaking at 5 Mev/Ze and 3 Mev/Ze for the light and heavy groups, respectively. The balloon results, supported by riometer and magnetometer measurements [Edwards and Reid, 1964], are then most easily interpreted [Edwards, 1964] in terms of the γ activity of a small high-energy fraction of debris, injected at $L \sim 2.5$ and deposited above the balloon at an altitude of 100 km. A minimum of $10^5$ fission product nuclei, $3 \times 10^{-4}\%$ of the Starfish yield (assumed to be a fission device) would be required to account for the magnitude of the

---

*On leave from the University of Adelaide, Adelaide, Australia.*
radiation increase. If no westward drift is assumed, the amount of debris above the balloon horizon (10,000 km above Johnston Island) is unrealistically large and is inconsistent with both satellite results.

4. The intensity of hard radiation reaching Cosmos 5 began to increase at $H + 2$ minutes. This increase is also most simply explained by a westward drift of debris toward the satellite. As pointed out by Manley and Carpenter the interpretation of the increase in counting rate of the Cosmos soft particle detector at $H + 8$ minutes, $L = 1.4$, $35^\circ \text{W}$ of the Johnston Island meridian is not clear. However, if this increase were due to the detection of fission $\beta$ particles from trapped debris, it is natural to associate the apparent absence of debris between $L = 1.4$ and $L = 1.8$ with (a) the latitude gap in the occurrence of white ray structure, radio-wave, and magnetic disturbance between the conjugate area and New Zealand [Gabites and Rowles, 1962] and (b) the corresponding minimum in energetic electron intensities at $L \sim 1.5$ [Durney et al., 1964] measured with Ariel within a day of the Stafish detonation.

In summary, Cosmos 5 measurements considered jointly with Ariel 1 and with balloon measurements of $\gamma$ radiation strongly suggest prompt injection, temporary trapping, and deposition of about 10% of Starfish debris near $L = 1.25$ and perhaps an order of magnitude less near $L = 2.5$. This interpretation may help resolve some of the outstanding problems [Kenney and Willard, 1964] associated with the high-latitude effects of Starfish.

References


(Received March 4, 1966; revised April 15, 1966.)
Asymmetric Plasma Expansion from the 'Starfish' Explosion

P. J. Edwards and R. F. Entwistle

Department of Physics, University of Otago, New Zealand

Recent astronomical measurements of solar and celestial X-radiation have led to renewed interest in strong shock-wave phenomena arising from explosions which occur on an astrophysical scale as in solar flares, novae and in supernovae. Although on a much smaller scale, the 'Starfish' nuclear explosion at an altitude of 400 km over Johnson Island in 1962, provides a useful test of current understanding in this field.

We have interpreted observations of delayed gamma radiation from 'Starfish' carried out from a balloon by the Hobart Cosmic-Ray Group and have concluded that a large fraction of the hot plasma produced in the explosion was not magnetically confined but instead expanded anisotropically in the form of a 1000 km/s jet at least 1.5 earth-radii long into the southern hemisphere. This anisotropy in the motion of the detonation products is quite contrary to theoretical expectations for an expanding 'snow plough' shock of the type investigated by Colgate in connection with 'Starfish'. Similar asymmetries in expanding shells of novae and supernovae may well be important.

We have determined the motion of the Starfish debris on the basis of a suggestion by Manley et al. that the fission decay chain $^{87}$Se (16s) $\rightarrow$ $^{87}$Br (55s) provides a relatively long-lived source of high energy (5 MeV) gamma rays which allow the motion of fission debris to be traced by detectors situated at considerable atmospheric depth. In the case of the Hobart balloon-borne detector, the time profile of the gamma-ray intensity indicates that $^{87}$Br gamma rays predominated over softer gamma rays. This immediately
sets an upper limit of $30^\circ$ to the elevation angle of the gamma-active debris. We have therefore re-examined the response of the Cosmos 5, Ariel 1 and Hobart balloon detectors to debris travelling outward in the plane of the magnetic meridian of the explosion.

We conclude that the luminous jet observed to leave the site of the detonation in a direction parallel to that of the local magnetic field transported at least 10\% of the plasma mass across the geomagnetic field shell ($L = 2$) and gave rise to the unexpectedly high latitude effects observed after the explosion.

This work was supported by the New Zealand University Grants Committee under UGC 68/153.

Microburst Phenomena

1. Auroral-Zone X Rays

D. Venkatesan,$^1$ M. N. Oliven,$^2$ P. J. Edwards,$^3$ K. G. McCracken,$^3$,$^4$ and M. Steinbock$^1$,$^3$

High-time resolution X-ray equipment flown from Fort Churchill, Manitoba, Canada, on August 11, 1965, provides evidence for species of auroral-zone X-ray microbursts with rise times of 20-30 msec. These microbursts are characterized by a rise of the form $1 - e^{-tr_F}$, where $r_F$ is about 30 msec, and a decay of the form of $e^{-tr_D}$, where $r_D$ is about 200 msec, and a typical peak flux for the largest events of $J_\infty(E_{X,\gamma} > 60$ kev) $\sim 1 \times 10^5$ photons cm$^{-2}$ sec$^{-1}$ at 10 g/cm$^2$.

An episode of these rapidly rising bursts was observed in the early morning hours (after 4:30 local time), and an episode of the more common slower rising (temporally more symmetric) microbursts began after 9:30 local time. The fast rise times and the lack of dispersion $\geq 10$ msec in the X-ray bursts observed at different energies imply restrictions on the nature and propagation of the parent electron bursts.

INTRODUCTION

Numerous rocket and balloon studies in the auroral zone have investigated the bremsstrahlung X radiation produced by electrons impinging upon the upper atmosphere. With the availability of data from charged-particle detectors on polar satellites, measuring precipitated electrons, it is now possible to investigate the behavior of both the parent electrons and the daughter bremsstrahlung X rays.

This study reports on a portion of the data obtained during a series of seven balloon flights made from Fort Churchill, Canada, of balloon-borne scintillator-photomultiplier X-ray detectors. The geographic coordinates of the balloon launch site are 58.75°N and 94.09°W ($L = 8.66$). The data of interest were obtained during a balloon flight that commenced at 0225 UT on August 11, 1965, during a period of relatively low geomagnetic activity. The international planetary index for geomagnetic activity, $K_p$, was 10 or less during the entire flight period.

The balloon reached a ceiling altitude of 110,000 feet at 2315 local time (0515 UT) and thereafter drifted in a west-southwesterly direction. It remained above 106,000 feet during the period of 15 hours, during which useful data were acquired at the telemetry receiving station. Table 1 gives the positions of the balloon during the period of interest, namely 0700-1700 UT (0100-1100 LT) and the $L$ coordinate corresponding to the positions. Also given, for comparison, are the $L$ values for places from which other X-ray experiments have been flown before.

INSTRUMENTATION

The detection system, illustrated in Figure 1, consisted of a 5-inch diameter, half-inch thick NaI (Tl) scintillation crystal, viewed by a Dumont 6364 photomultiplier of the same diameter. The geometric factor of the detector for X rays isotropic over the whole of the upper hemisphere, was 476 cm$^2$ ster. A lucite light pipe, of the same dimensions as the crystal, separated the crystal and the phototube by 2 cm, to minimize the effects of photocathode nonuniformities upon the energy resolution of the system.

The photomultiplier EHT was supplied from a regulated dc converter, which, in turn, obtained its power from a series of mercury batteries. The whole high voltage system was vacuum potted to eliminate the possibility of corona. A $\mu$-metal shield enclosed the photomultiplier dynode chain and the photocathode.
and this minimized the perturbation of the electron orbits and hence the photomultiplier gain, by the earth's magnetic field. The complete detector system was observed to yield a resolution of 50% FWHM for the 32-kev X-ray from Cs133.

After linear pulse amplification, the photomultiplier pulses were applied to a three-window pulse height analyzer, with contiguous windows set to 20-40 kev, 40-60 kev, and >60 kev. The energy calibration of each window was adjusted before flight, using known monochromatic X-ray sources, and a 400-channel pulse height analyzer operating in the coincidence mode. Each of the outputs from the three window pulse height analyzer was fed to identical logarithmic count-rate meters, and also to a three position subcommutator, which sampled each of the three pulse rates for a minute at a time. The output from the subcommutator was fed to a binary scaler with outputs at scaling factors of 64 and 1024. The counting rate meter outputs, and the two scaler outputs were fed to separate FM subcarrier oscillators, which were used to modulate the telemetry transmitter, delivering a power of 0.25 watt into a vertically polarized coaxial sleeve antenna at 72 Mc/sec.

The response time of the counting rate meter circuits was determined to be 3 msec for an instantaneous increase in pulse input rate. Each count rate meter was individually calibrated before flight, and the two binary scaler outputs

<table>
<thead>
<tr>
<th>Location</th>
<th>Latitude</th>
<th>Longitude</th>
<th>L Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>700 UT</td>
<td>58.50</td>
<td>266.20</td>
<td>8.457</td>
</tr>
<tr>
<td>800 UT</td>
<td>58.50</td>
<td>265.90</td>
<td>8.426</td>
</tr>
<tr>
<td>900 UT</td>
<td>58.50</td>
<td>265.30</td>
<td>8.362</td>
</tr>
<tr>
<td>1000 UT</td>
<td>58.50</td>
<td>265.10</td>
<td>8.192</td>
</tr>
<tr>
<td>1100 UT</td>
<td>58.30</td>
<td>264.50</td>
<td>8.128</td>
</tr>
<tr>
<td>1200 UT</td>
<td>58.30</td>
<td>263.90</td>
<td>8.064</td>
</tr>
<tr>
<td>1300 UT</td>
<td>58.40</td>
<td>263.40</td>
<td>8.081</td>
</tr>
<tr>
<td>1400 UT</td>
<td>58.40</td>
<td>262.90</td>
<td>8.025</td>
</tr>
<tr>
<td>1500 UT</td>
<td>58.40</td>
<td>262.30</td>
<td>7.957</td>
</tr>
<tr>
<td>1600 UT</td>
<td>58.40</td>
<td>261.80</td>
<td>7.969</td>
</tr>
<tr>
<td>1700 UT</td>
<td>58.50</td>
<td>261.20</td>
<td>7.897</td>
</tr>
</tbody>
</table>

(1) Indicates low time resolution equipment flown from these points.

(2) Indicates fast time resolution equipment flown from these points.

![Block diagram of the entire system associated with the X-ray detector.](image-url)
were used to provide an inflight check of the calibrations. An inflight calibration of the subcarrier oscillator stability was also provided by sampling a stabilized square wave input periodically throughout the flight.

Observations

The first six hours of X-ray data obtained after the balloon attained ceiling altitude were rather featureless. Beginning at 1028 UT (0428 LT) and over a period of 6½ hours, however, the data revealed a great variety and number of fast time variations of auroral-zone X-ray activity. The magnetometer records at Fort Churchill showed no appreciable magnetic activity during this period, nor did the riometer register any event. Since these X-ray events occurred during daylight hours, no visual confirmation of the presence of aurorae was possible.

Figures 2 and 3 present a number of typical fast rise time X-ray bursts. They appear in large numbers, both as isolated events emerging from the background and in trains occurring in close succession with an apparent periodicity of approximately 0.6 second. When two follow each other very closely, the result is a compound structure in which the two leading edges can usually be discerned. Furthermore, there are also periods of complex activity during

\[ E > 60 \text{ KeV} \]

\[ \text{TIME} \]

\[ 20 \quad 15 \quad 10 \quad 5 \quad 0 \text{ SECONDS} \]

\[ 20 \quad 15 \quad 10 \quad 5 \quad 0 \text{ SECONDS} \]

\[ \text{TIME} \]

Fig. 2. Typical appearance of fast rise time X-ray microbursts. In the top data segment, compound events are seen, whereas in the bottom two examples, combs or trains and single bursts can be identified. Vertical scales of bursts are analog.
which the absolute X-ray flux is so high that any attempt at isolation of such bursts is futile, although one does see indications of the superposition of individual bursts. The over-all features of these bursts are, in general, similar to those that have been reported and discussed previously by Anderson and Milton [1964], Anderson [1965], and Anderson et al. [1966], who have named them microbursts, Parks [1967], and several other authors. In view of the fact that the bursts observed during the flight reported herein are also devoid of substructure and are of the same time scale, we adopt the same terminology of microbursts, since there is enough general evidence to show that the bursts herein described belong to the same general category as those observed by Anderson and Milton [1964].

The microburst activity appears in two time intervals, separated by 134 hours of little activity, the two periods containing bursts having significantly different characteristics. During the earlier episode of activity, 1027-1355 UT (0427-0755 LT), the bursts have hard spectra, very fast rise times, and slower decay times and, as such, are somewhat different from the slower

---

Fig. 3. Two examples of fast rise time asymmetric microbursts as they appear, singly, in all three energy channels. All vertical scales of bursts are analog.
rising, more symmetric microbursts discussed by earlier authors. Although others have observed microbursts with a variety of temporal asymmetries [e.g., Anderson and Milton, 1964, and Anderson et al., 1966], the rapid rise times of the order of 20–30 msec observed by us [Edwards et al., 1966] have not previously been reported. Because of the very rapid rise time of these bursts and the correspondingly much longer decay time, the over-all appearance of these bursts is one of marked asymmetry. Later, during the period 1540–1707 UT (0940–1107 LT), the temporal character of the observed microbursts is essentially symmetric, that is, correspondingly longer rise times (of the order 100 to 200 milliseconds) than those of the rapid rise time bursts and decay times comparable to the rise times. The designation of symmetry does not imply that the rise and decay are exactly equal, but rather that they have similar forms and orders of magnitude. These events exhibit characteristics similar to the microbursts reported by Anderson and Milton [1964]. An example of each of the two types, together with one of less severe asymmetric features (an intermediate type) observed in the late morning, is presented in Figure 4.

Preceding the initial onset of the period of

Fig. 4. Comparison of the rapidly rising asymmetric microburst seen in the early morning hours, a more symmetric type of burst seen later in the morning, and the more slowly rising symmetric type of microburst seen late in the morning and similar to the type commonly observed by other experimenters [e.g., Anderson and Milton, 1964]. The lower two bursts have been normalized to the top burst.
microburst activity, a gradual increase in the background counting rate of the 20-40, 40-60, and > 60-keV energy channels was observed. Such increases before microburst epochs have also been reported by Anderson and Milton [1964]. It should be pointed out that the peak intensities of the rapidly rising bursts observed by us are similar to those of the type seen by Anderson and his co-workers. In common with them, more than half the peak fluxes were less than 20 cm$^{-2}$ sec$^{-1}$ above background and only about 1% of the rapidly rising asymmetric bursts had peak fluxes in excess of 60 cm$^{-2}$ sec$^{-1}$ at an atmospheric depth of 10 g/cm$^2$.

Certain gross features of the X-ray events emerge from a general study of the entire data. The occurrence of fast rise time microbursts is seldom isolated but continues over an extended period of time. There seems to be no evident relationship between the onset of the rapidly rising microbursts and the nature of the preceding X-ray activity. In short, the onset of these microbursts appears to be sudden and impulsive. It is observed that the fast rise bursts are more predominant in the $E > 60$-keV channel than are the slower rising bursts, indicating a harder photon spectrum. A peak flux above previous background level of $J_0(E > 60$ keV) $\approx 10^6$ photons cm$^{-2}$ sec$^{-1}$ can be observed in the $E > 60$-keV channel during a large rapid rise time event. The individual microbursts maintain their identity (i.e., the rise and decay times are not changed) despite superposition upon one another, or upon a smooth background of X-ray activity. The shortness of the characteristic times of the microbursts, especially the rapidly rising variety, would appear to be of importance in the understanding of the dynamic processes responsible for the parent precipitation phenomenon. A logical step forward would be to investigate the parent particle precipitation, using detectors on high-latitude satellites, and this has been done and is discussed in a companion paper [Oliven et al., 1968].

The present study reveals a wide variability in the rise time characteristics of X-ray microbursts. The fast rise time asymmetric microbursts observed during early morning hours are characterized by rise times, $\tau_R$, of about 20–30 msec and by decay times, $\tau_D$, of about 200 msec, where the rising and decay phases are represented by $(1 - e^{-t/\tau_R})$ and $e^{-t/\tau_D}$, respectively. Throughout the period of activity, one common feature, the time of apparent duration of a rapid rise burst (the period during which it is visible above background from the rise to the apparent end of the decay), was almost always between 300 and 500 msec. This duration is in agreement with a similar characteristic of the slower rising bursts observed by Anderson and Milton [1964] whose flights were from Flin Flon at $L = 6.1$.

One hundred of the large rapidly rising bursts were studied in detail; the frequency distribution of rise times for these events appears in Figure 5. About 60% of the cases have rise times $\leq 50$ msec. A composite of 75 of these microbursts is shown in Figure 6, and a decay time constant of 200 msec is apparent. Figure 7, an example of an individual burst, shows a rise time of 30 msec and a decay time of 200 msec. The $E > 60$-keV energy window is displayed in all these diagrams.

Figure 8 gives samples of all three energy channels during the late morning hours when the slower rising symmetric bursts were seen. Anderson et al. [1966] have also made flights recently at $L \approx 8.0$, but they have not observed the very rapidly rising type of bursts during their flights, but have seen events similar to those predominant in the later part of our flight. Although one does see the rapidly rising

![Fig. 5. Distribution of rise times for 100 large fast rise time microbursts.](image-url)
bursts occasionally in the lower energy channels, it is possible to generalize that, in most of the examples of fast rise time asymmetric microbursts observed during our flight, the highest energy channel contained the highest counting rate of all the channels (see Figure 3). The longer rise, more symmetric bursts, on the other hand, were observed to exhibit comparable responses in all three channels.

To study the photon energy spectrum of the rapid rise microburst, we have computed the ratios of counting rates, above background of:

\[ \frac{\text{X rays, energy} > 40 \text{ kev}}{\text{X rays, energy} > 60 \text{ kev}} \]

\[ \frac{\text{X rays, energy} > 20 \text{ kev}}{\text{X rays, energy} > 60 \text{ kev}} \]

\[ \frac{\text{X rays, energy} > 20 \text{ kev}}{\text{X rays, energy} > 40 \text{ kev}} \]

at 10-msec intervals, throughout the entire duration of individual bursts. Figure 9 shows the largest observed burst, with the analog output at the top, and the ratios at the bottom. The ratios have the lowest values within 10 msec of the occurrence of peak counting rates in all three channels, and this indicates relatively small dispersions (\(\leq 10 \text{ msec}\)) in the arrival of the parent precipitating particles of different energies at the production layer.

Although our emphasis in this paper has been on the newly observed fast rising microbursts, it is worthwhile to make some comparison with the better known types of slower rising bursts, observed during the latter half of the flight. A frequency distribution of the rise times for both types of events is presented in Figure 10, and it can be seen that there is no prominent peaking in the lower histogram; that is, the rise times of the slower rising symmetric microbursts display a wide variability. Specifically, only 22\% of the symmetric events have rise times \(\leq 50 \text{ msec}\) in contrast with 71\% in the case of rapidly rising events.

Figures 11 and 12 display composite rapid rise asymmetric and longer rise symmetric microbursts for the energy ranges 40–60 kev, and \(>60 \text{ kev}\), and it can be seen that the more symmetric microburst clearly possesses the softer spectrum of the two species of events. The e-folding energies for both types of events are significantly higher than those obtained in the vicinity of \(L \approx 5–6\) by other investigators. The composite fast rise time burst spectrum has an e-folding energy, \(E_e \sim 360 \text{ kev at peak}\).
intensity. The spectrum slowly softens with $E_0$ falling to ~200 keV near the end of the events. The spectrum of the more slowly rising symmetric bursts hardens during the rise to peak intensity, at which time $E_0 \sim 160$ keV, and this softens with $E_0$ eventually falling below 60 keV. In both cases, the event spectra are substantially harder than those of the background radiation. It must be remembered here that with such large values for $E_0$ and in consideration of the restrictions, which a limited two- or three-point determination of $E_0$ impose, the true significance of these e-folding energies is lost. The values stated are just presented to give a numerical answer and do not necessarily imply a spectrum of this form, characterized by these values of $E_0$.

**Summary and Conclusions**

The primary object of the study has been the investigation of auroral-zone microburst X-ray fluxes, with high-time resolution balloon-borne equipment at high latitudes (magnetic shell parameter $L \sim 8$). These balloon results are presented herein, and those correlating the microbursts with satellite observations are presented in the following companion paper (Oliven et al., 1968). A third paper (Oliven and Gurnett, 1968), establishing a connection between microbursts and VLF phenomena, also follows.

A new type of microburst has been observed and reported herein with the characteristics of a hard energy spectrum, fast rise time, of the order of 30 msec, and a decay of time constant ~200 msec. Microbursts, having longer rise times, and similar time structure to those reported by others were also observed during the latter half of the flight. The latter occurred at a slightly lower $L$ value and at a later time (local time 9:30 in contrast with 4:30) than the rapidly rising asymmetric microbursts.

The fast rise time microbursts have a harder spectrum than the more symmetric events (of slower rise times), as is evidenced by the large response in the high-energy channel ($E > 60$ keV) as compared with that in the lower channels. In contrast, more slowly rising bursts are characterized by comparable responses in the lower energy channels (20–40 and 40–60 keV). A typical peak flux for the fast rise time events in the >60-keV channel is given by $J_0(E_{x, \text{typ}} > 60 \text{ keV}) \approx 20$ photons cm$^{-2}$ sec$^{-1}$.

The extremely hard photon spectra found for the larger events (e-folding energies of the order of hundreds of keV) place rather stringent restrictions on the interpretation of the events.
If the precipitating particles responsible for these bursts are assumed to be electrons, then the X-ray spectra must be scrutinized in terms of the differential energy spectra of the parent electrons, within the restrictions imposed by the range of the three channels and the two- or three-point determination of the X-ray spectra. It appears necessary to postulate relativistic electrons with an insignificant flux at lower energies. Differential electron spectra with a high-energy peak $>400$ keV have been deduced by Mozer and Bruston [1966], and rapid fluctuations in the integral flux above 400 keV have been seen by Blake et al. [1966], and the impulsive bursts reported herein may be an extension of these relativistic electron populations.

The spectra of the harder rapidly rising bursts is not as easily interpreted in terms of electron bremsstrahlung radiation. One of the possible explanations could be that of the proton-excited gamma radiation. If the events were due to proton-excited gamma radiation [Hoffmann and Winckler, 1963], then no special assumptions about the proton energy spectrum are necessary. As an example, a flux of $\sim 10^5$ protons cm$^{-2}$ sec$^{-1}$ above 0.5 Mev (the Injun 3 Geiger counter threshold) having a differential power law spectral index of $-5$ could account for the larger fast rise time asymmetric bursts. These protons might augment the precipitating electron population. At the present time, we have no direct evidence for the contribution protons.

Fig. 9. Largest observed rapidly rising microburst for three energy windows and the ratios between the various energy channels. Vertical scale of burst is analog.
Fig. 10. Comparison between the distribution of rise times for fast rise time asymmetric and slower rising symmetric microbursts during the two epochs of microburst activity.

Fig. 11. Comparison curves of the energy channels $E > 60$ keV and $40$ keV $\leq E \leq 60$ keV for 25 fast rise time microbursts.

Fig. 12. Composite curves of the energy channel $E > 60$ keV and $40$ keV $\leq E \leq 60$ keV for 25 slowly rising symmetric microbursts.

To summarize, the detailed study of the newly observed rapidly rising asymmetric X-ray microbursts, in addition to the better known more slowly rising symmetric microbursts, has provided a supplement to the knowledge of fast temporal variations in auroral-zone X rays. The fast rise times and hard spectra impose rather stringent conditions on the mechanisms involved in this impulsive particle precipitation, and in turn upon the causative plasma instabilities within the magnetosphere. In particular, we note that the fast rise time microburst indicates dispersion times of $\leq 10$ msec for the particles responsible for the microburst, implying either a limit on the distance from the source particle to the earth, or that there is a significant particle-particle or particle-field interaction, so that the particles remain bunched while propagating through the geomagnetic field.

Acknowledgments. The authors wish to express their thanks for the advice and support-offered in this project by Professor J. A. Van Allen of the University of Iowa.

We wish to acknowledge substantial design contributions made by Messrs. C. Adair and L. Brooks, and the assistance of the various personnel associated with the ONR Skyhook expedition of 1965.
This work was supported at The University of Calgary by National Research Council grant A-3865, at the University of Iowa by Office of Naval Research contract 1509(06), and at the Southwest Center for Advanced Studies by NASA contract NAS-r-198.

REFERENCES


(Received September 15, 1967; revised December 12, 1967.)
Solar absorption photometry and the determination of atmospheric composition

Department of Physics, University of Adelaide, Australia

Bryan Rofe and S. G. Johnson
Weapons Research Establishment, Salisbury, Australia

(Received 25 September 1968)

Abstract—Molecular oxygen and total number densities in the altitude range 70–120 km have been determined at Woomera by rocket measurements of the atmospheric absorption of solar ultra-violet and X-radiation using non-dispersive detectors carried in a Long Tom rocket fired to coincide with a pass over Australia of the N.R.L. solar radiation satellite Explorer 30. Comparison of the measured total and molecular oxygen number densities indicates a significant change in atmospheric composition above 100 km. Absolute measurements of the solar flux for several wavelength bands within the range 1050–1680 Å have been obtained and are compared with Explorer 30 values.

INTRODUCTION

ABSORPTION spectroscopy using rocket-borne detectors provides a powerful method for studying the composition of the neutral atmosphere and the method is particularly suitable for the determination of the density profiles of molecular oxygen and ozone. In previous rocket firings from Woomera (Carver et al., 1965; 1966), molecular oxygen densities have been determined to 90 km by measuring the absorption of solar Lyman-α radiation, and high altitude ozone distributions at night have been determined by measuring the absorption of moonlight in the 2500–2900 Å band. The present report describes the results of a Long Tom rocket flight from Woomera on 11th November 1966 fired to coincide with a pass over Australia of the N.R.L. solar radiation satellite Explorer 30. A number of non-dispersive u.v. and X-ray detectors were used in the Long Tom flight to determine molecular oxygen and total atmospheric densities between 70 and 120 km. The flight also provided absolute determinations of the solar u.v. flux for several wavelength bands within the range 1050–1670 Å.

INSTRUMENTATION

The spectral characteristics of the various detectors carried in the rocket are listed in Table 1. The two X-ray detectors were mica-windowed Ne A (Halogen) filled Geiger counters (Philips Type 18504). Their scaled counting rates were separately telemetered. The quantum efficiency of the detectors was measured in the laboratory at a number of discrete wavelengths and the spectral response was calculated in terms of the mass absorption coefficients of the gas and window materials. The effective spectral response for the present application was 8 Å. At this wavelength all constituents of the neutral atmosphere absorb the radiation so that the X-ray

* Present address: Department of Physics, University of Dunedin, New Zealand.
† South Australian Institute of Technology, Adelaide, S.A.
detector may be used to measure total atmospheric density. The effective absorption cross section (Byram et al., 1956) is given in Table 1.

The construction and properties of the u.v. ion chambers used in the rocket flight have been described previously (Carver and Mitchell, 1964; 1967). Before the flight, the quantum efficiency and spectral response of each ion chamber was determined in the laboratory using a $\frac{1}{3}$ metre vacuum monochromator. These ion chambers are sensitive to bands which are strongly absorbed by molecular oxygen and may therefore be used to determine molecular oxygen profiles in the atmosphere over height ranges which depend on the effective molecular oxygen absorption cross sections for the radiation to which each detector is sensitive. The effective absorption cross sections were determined by combining the measured spectral response of each detector with the molecular oxygen absorption cross sections (Watanabe, 1958; Blake et al., 1966) together with Tousey's (1963) solar spectrum data. This is a simple and unambiguous procedure when the absorption cross section varies only slowly over the response range of the detector as is so for the BaF$_2$-toluene, sapphire-xylene and quartz-triethylamine ion chambers. In the case of the LiF-nitric oxide chamber, the Lyman-$\alpha$ line so dominates the solar spectrum in the wavelength region to which the detector responds that, at least for altitudes below 90 km, the effective cross section is that at the Lyman-$\alpha$ wavelength. The effective molecular oxygen absorption cross sections listed in Table 1 for other ion chambers correspond approximately to unit optical depth in the atmosphere, i.e. to a height range over which the ion chamber current is expected to be changing most rapidly with height. For the LiF-ethyl chloride and LiF-ethyl bromide detectors, which are sensitive in the molecular oxygen "window" region, the absorption cross section varies so strongly over the wavelength response of the chambers that it is not possible to define an effective absorption cross section even for the limited height range near unit optical depth. The data from these two ion chambers were therefore used for solar flux measurements only and not for the determination of atmospheric densities.

Table 1. Ultra-violet and X-ray detectors

<table>
<thead>
<tr>
<th>Detector (ion chamber)</th>
<th>Spectral response</th>
<th>Effective absorption cross sections (molecular oxygen)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LiF-nitric oxide</td>
<td>1050–1340 Å</td>
<td>$1.0 \times 10^{-20}$ cm$^2$ (Lyman-$\alpha$)</td>
</tr>
<tr>
<td>LiF-ethyl bromide</td>
<td>1050–1200</td>
<td>highly variable</td>
</tr>
<tr>
<td>LiF-ethyl chloride</td>
<td>1050–1130</td>
<td></td>
</tr>
<tr>
<td>CaF$_2$-ethyl iodide</td>
<td>1220–1320</td>
<td>$1.0 \times 10^{-18}$ cm$^2$</td>
</tr>
<tr>
<td>CaF$_2$-benzene</td>
<td>1220–1340</td>
<td>$1.0 \times 10^{-18}$ cm$^2$</td>
</tr>
<tr>
<td>BaF$_2$-toluene</td>
<td>1350–1410</td>
<td>$1.4 \times 10^{-17}$ cm$^2$</td>
</tr>
<tr>
<td>Sapphire-xylene</td>
<td>1420–1470</td>
<td>$1.5 \times 10^{-17}$ cm$^2$</td>
</tr>
<tr>
<td>Quartz-triethylamine</td>
<td>1570–1680</td>
<td>$3.0 \times 10^{-18}$ cm$^2$</td>
</tr>
<tr>
<td>Geiger counter</td>
<td>8 Å</td>
<td>All constituents</td>
</tr>
<tr>
<td>Philips Type 18504</td>
<td></td>
<td>$5 \times 10^{-20}$ cm$^2$</td>
</tr>
</tbody>
</table>

* The figure in brackets indicates the number of detectors of each type carried in the rocket.
Among the Explorer 30 photometers were LiF-nitric oxide and CaF$_2$-nitric oxide ultra-violet ion chambers, which provided information for comparison with the rocket determination of absolute fluxes over the range 1050–1350 Å and 1225–1350 Å. The spectral response and quantum efficiency of the Explorer 30 chambers were supplied by N.R.L.

A Long Tom rocket carrying the sensors listed in Table 1 was fired from Woomera (Lat: 30°35'S Long: 136°31'E) on 11th November 1966 at 08.35 hr local time when the solar zenith angle was 48°. This time corresponded to an Explorer 30 pass over the satellite receiving station at Carnarvon, W.A. (Lat: 24°S Long: 107°E). The rocket carried sunslits and other visible light photocells of known angular response to sunlight, so that the aspect angle between the detectors and the sun could be determined. The rocket was tracked by FPS16 radar and reached a maximum altitude of 121.2 km. Above 70 km the solar ultra-violet radiation was observed each time the rocket, which was rolling with a period of 4 sec, pointed the detectors in the direction of the sun. Information from the solar aspect measuring devices was used to correct these observations to zero aspect angle. The rocket motion was very satisfactory for the purposes of this experiment and the aspect corrections which had to be applied to the ion chamber observations were mostly less than 15% and in no case, greater than 30 per cent.
Density profile

Figure 1, which includes data from both the upward and downward portions of the flight, shows, as a function of altitude, the measured response of the different u.v. and X-ray detectors after correction to zero aspect angle.

Molecular oxygen and total atmospheric densities were determined from the u.v. and X-ray absorption curves of Fig. 1 using the cross sections listed in Table 1. For the purposes of this analysis it was assumed that the atmosphere could be divided into layers of height, \( h \), within which the number density, \( n \), was constant and that the attenuation of the solar flux over a height interval \( h \) was given by

\[
\exp \left( -hn \sigma \sec \theta \right)
\]

where \( \sigma \) is the absorption cross section and \( \theta \) the solar zenith angle (45° in the present case). The densities derived in this way for a layer height \( h = 2.5 \) km are shown in Fig. 2. The total atmospheric density derived from the X-ray absorption measurements (upper histogram) is compared in Fig. 2 with the 1962 U.S. standard atmosphere values (full curve) and it may be seen that there is fair agreement in the altitude range 90–110 km. Below an altitude of 90 km the derived molecular oxygen densities depend on the measurements made with the LiF-nitric oxide (Lyman-\( \alpha \)) ion chamber.
At higher altitudes there is very good agreement between the oxygen density determinations obtained independently from the BaF$_2$-toluene, CaF$_2$-benzene, CaF$_2$-ethyl iodide, sapphire-xylene and quartz-triethylamine chambers; the lower histogram shows the average of the molecular oxygen densities determined by all ion chambers. The observed molecular oxygen density profile below 90 km runs very nearly parallel to the standard atmosphere total density curve but at higher altitudes the effects of molecular dissociation are clearly evident. The dotted curve in Fig. 2 shows the atomic oxygen density profile implied by the measured molecular oxygen and total density profiles (the last is not significantly different from the 1962 U.S. standard). The present results suggest that the atomic oxygen concentration exceeds the molecular oxygen concentration above 100 km.

**Solar flux**

The data shown in Fig. 1 have also been used to determine absolute solar ultraviolet fluxes in terms of the laboratory calibrations of each of the ion chambers. Some of the ion chambers (e.g. BaF$_2$-toluene, sapphire-xylene) are sensitive to radiation which is very strongly absorbed in the atmosphere and did not reach saturation currents at the maximum rocket altitude of 121 km. It was necessary therefore to extrapolate the measurements to zero optical depth and this was done by assuming that the observed molecular oxygen density curve of Fig. 2 could be extended to higher altitudes with constant scale height.

The solar u.v. fluxes determined in this way are shown in Fig. 3 where they are compared with the fluxes corresponding to various solar black body temperatures. Figures 4–6 indicate the variation from the mean flux value for the Explorer 30.
Solar absorption photometry and the determination of atmospheric composition

both to solar u.v. emission lines and to continuum radiation. In the case of the LiF nitric oxide chamber the solar Lyman-\(\alpha\) line is by far the strongest radiation source within the band pass of the detector. This line coincides in wavelength with a "window" in the oxygen absorption cross section thereby allowing the Lyman-\(\alpha\) radiation to penetrate more deeply into the atmosphere than other radiation within the band pass of the detector. At higher altitudes there is a small but significant ion chamber response to other radiations as shown by the rise in ion chamber current above 95 km. The Lyman-\(\alpha\) contribution to the ion chamber current can be clearly identified from the atmospheric absorption curve and amounts to approximately 85 per cent of the total current which would be observed above the atmosphere. For

![Graph](image)

**Fig. 6. Explorer 30 determination of variation from mean of 8-16 Å solar X-ray flux.**

the Explorer 30, LiF-nitric oxide chamber the Lyman-\(\alpha\) contribution was calculated to be 81 per cent. The measured Lyman-\(\alpha\) flux is found to be 4.2 erg/cm\(^2\)/sec, which is in excellent agreement with Explorer 30 flux value of 4.25 erg/cm\(^2\)/sec.

Emission lines make a less dominating contribution to the response of the other ion chambers. Tousey's (1963) measurements of the solar spectrum suggest that for wavelengths longer than 1340 Å the solar flux is predominantly in the continuum. If we ignore the (small) contribution from emission lines the present results indicate that from 1400 Å to 1600 Å the continuum flux corresponds to that of a black body at a temperature of 4500°K; between 1000 Å and 1100 Å the equivalent black body temperature has increased to 5300°K.

**Acknowledgements**—This work was supported in part by grants to the University of Adelaide from the Australian Research Grants Committee and the Department of Supply. We are indebted to Mr. R. W. Kreplin of the U.S. Naval Research Laboratory who supplied information about the instrumentation of the Explorer 30 satellite.
<table>
<thead>
<tr>
<th>Editors/Authors</th>
<th>Year</th>
<th>Journal/Magazine</th>
<th>Volume</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Byram E. T., Chubb T. A. and Friedman H.</td>
<td>1956</td>
<td>J. geophys. Res.</td>
<td>61</td>
<td>251</td>
</tr>
<tr>
<td>Carver J. H., Mitchell P., Murray E. L. and Roffe B.</td>
<td>1965</td>
<td>Space Research VI</td>
<td></td>
<td>373</td>
</tr>
<tr>
<td>Carver J. H., Horton B. H. and Burger F. G.</td>
<td>1966</td>
<td>Space Research VII</td>
<td></td>
<td>1020</td>
</tr>
<tr>
<td>Tousey R.</td>
<td>1967</td>
<td>J. opt. Soc. Am.</td>
<td>57</td>
<td>738</td>
</tr>
<tr>
<td>Watanabe K.</td>
<td>1963</td>
<td>Space Sci. Rev.</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>J. H. Carver et al.</td>
<td>1958</td>
<td>Advances Geophys.</td>
<td>5</td>
<td>157</td>
</tr>
</tbody>
</table>
PROPOSAL FOR MONITORING ATMOSPHERIC TURBIDITY IN NEW ZEALAND AT SELECTED ASTRONOMICAL OBSERVATORIES*

P. J. EDWARDS and M. OTHMAN

Physics Department, University of Otago

INTRODUCTION

The World Meteorological Organisation has established a network of background air pollution stations. The stated purpose of this network is to: (a) Measure current levels of pollution, and (b) Detect and measure long-term trends in the levels of significant atmospheric constituents which may result in climatic change (WMO, 1974).

Two categories of stations are called for:

(a) **Baseline air-pollution stations**: These are regarded as research stations having as their first priority measurements directed towards documenting long-term changes in atmospheric composition of particular significance to weather and climate.

(b) **Regional air-pollution stations**: These have as their primary purpose the documentation of long-term atmospheric composition changes which may be related to changes in regional land use or other regional activity.

The siting criteria established for the two categories (WMO, 1974) are more stringent for the Baseline stations than for the Regional stations but in both areas isolated rural environments are required, "located away from major population centers, major highways and air routes, preferably oil small isolated islands or on mountains above the tree line" (Baseline) or on "the summits or higher slopes of elevated areas" (Regional). The general similarity between these sites and those of astronomical observatories is obvious. Freedom from significant changes in anticipated land use for 50 years within 100 km is also required for Baseline station sites.

The New Zealand Meteorological Service has indicated its particular interest in a proposal to use astronomical observatories to monitor atmospheric turbidity and hence aerosol content. High-altitude rural observatories at Mt John and Black Birch, for example, could function much like WMO Baseline stations and could provide valuable information about the background level of particulate pollution.

*Otago Physics Department Research Report Astrophysics 2 / 76.

*Southern Stars* 26: 189-95 (1976)
Turbidity measurements at lower altitude observatories such as the Auckland, Beverly-Begg, and Signal Hill Observatories would, on the other hand, provide data on local urban aerosol levels. Also of considerable interest to the Meteorological Service (Othman, 1974).

One of the writers (P.J.E.) has proposed that the Physics Department of the University of Otago set up a turbidity monitoring network using photoelectric stellar photometers at selected observatories. This paper aims to inform the astronomical community of these proposals with a view to defining more closely the operation of a long term programme of the type envisaged. We also solicit suggestions and amendments to the proposals in order to maximize their mutual benefit to astronomy and meteorology.

ATMOSPHERIC EXTINCTION

The basis of the present proposal is that routine photometric measurements of stellar extinction at selected observatories be collated and analysed to provide a continuing measure of the atmospheric turbidity, which is that component of extinction due to particulates ("aerosols").

The extinction of a pencil of monochromatic light of wavelength \((\lambda)\) from a source at zenith angle \((z)\) can be conveniently written in either of the forms

\[
\Delta (\lambda, z) = A (\lambda) M (z) \text{ or } \tau (\lambda, z) = k (\lambda) M (z)
\]

where \(\Delta m (\lambda, z)\), the extinction in stellar magnitudes, is equal to 1.086 times the optical depth \(\tau (\lambda, z)\). Here \(A(\lambda)\) is the astronomical and \(k(\lambda)\) is the exponential extinction coefficient and \(A(\lambda) / k(\lambda) = 1.086\). The atmospheric air mass, \(M(z)\), at the zenith is \(M(0) = 1\).

The spectral radiance \(I (\lambda, \tau)\) of the pencil is given by

\[
I (\lambda, \tau) = I (\lambda, 0) \exp[-\tau (\lambda, z)]
\]

The spectral irradiance \(I (\lambda, \tau)\) in a solid angle small compared with the characteristic scattering solid angle is then

\[
F(\lambda, \tau) = F(\lambda, 0) \exp[-\tau (\lambda, z)]
\]

A measurement of \(F(\lambda, \tau)\) serves to determine \(\tau(\lambda)\) provided the flux at zero depth is known or can be inferred. Since \(\tau (\lambda, z) = k(\lambda) M(z) = k (\lambda) \sec z\) for \(z \approx 0\) and \(\tau(\lambda, 0) = k(\lambda)\), the optical depth in the zenith \((z = 0)\) can be deduced and can characterize the extinction at a particular site and wavelength.

Alternatively, the logarithmic slope of a spectral irradiance-air-mass plot, derived from

\[
\frac{\partial m (\lambda, M(z))}{\partial M(z)} \bigg|_{\tau(\lambda, 0)} = A (\lambda)
\]

or

\[
\frac{\partial \ln F \big(\tau (\lambda, 0), M (z)\big)}{\partial M(z)} = - \tau(\lambda, 0) = - k(\lambda)
\]
provides the extinction coefficients directly without requiring a priori knowledge of the zero depth flux.

A logarithmic plot of irradiance against air-mass, which can be subjected to linear regression analysis to yield best fit slope \( \tau(\lambda, 0) \) and zero mass intercept \( m(\lambda, 0) \), is the basis of the proposed method. When applied to sun photometer measurements such a plot is called a Langley Plot. In stellar astronomy it is usually called a Bouguer Plot. An example is given in Fig. 1 from which \( \tau(\lambda, 0) \) and \( F(\lambda, 0) \) at five wavelengths may be deduced.

**Fig. 1**

**LANGLEY PLOT**

O. U. Physics Dept.
Dunedin
16.9.75
Depending on the degree of stability of the photometer and associated filters, the Langley method provides either a periodic updating of photometer calibration or, at the other extreme, becomes mandatory for every extinction determination. For photoelectric stellar photometers using photomultipliers, it is desirable to use a modified Bouguer method for all extinction measurements. Figure 2 shows typical $\tau(\lambda)$ values determined from solar and stellar "Langley" plots.
ATMOSPHERIC TURBIDITY

The total optical depth may be written (Shaw et al., 1972)

\[ \tau(\lambda) = \tau_R(\lambda) \frac{p}{p_0} + \tau_a(\lambda) + \tau_p(\lambda) \]

where:

\[ \tau_R(\lambda) = \text{Rayleigh scattering optical depth for standard atmosphere with sea level pressure } p_0 \text{ and atmospheric pressure at site } p. \]

\[ \tau_a(\lambda) = \text{optical depth due to absorption by gases (ozone, water vapour, etc.).} \]

\[ \tau_p(\lambda) = \text{optical depth due to scattering and absorption by airborne atmospheric particulates.} \]

The Rayleigh scattering term may be accurately subtracted from \( \tau(\lambda) \) if the surface pressure is known. Standard atmosphere values of ozone and precipitable water content are usually adequate to enable the removal of the gas absorption term.

The residual optical depth \( \tau_p(\lambda) \) is a measure of turbidity. Turbidity coefficients (WMO, 1974), \( B = 0.434 \tau_p(0.5 \mu m, 0) \) and \( \beta = \tau_p(1.0 \mu m, 0) \) are usually deduced from solar measurements, subject to certain assumptions. Schott glass filters having bandwidths similar to those of the UBV system are conventionally used. Simple Bouguer and Langley plots show similar scatter, both giving typical rms extinction errors of 5% leading to typical turbidity coefficient errors of 25% at high altitude. See Fig. 3 for a typical example of a Bouguer plot. Most of this scatter is due to time varying extinction and can be largely eliminated by using a modified Langley/Bouguer method.

PROPOSED IMPLEMENTATION

(1) In the first instance the Otago Physics Department will act as a central data collection centre for UBV and other primary extinction data.

(2) UBV (and other) extinction data will be compared with Otago Physics Department and N.Z. Meteorological Service sun photometer data obtained at the same site on days bracketing the stellar observation times.
(3) Transformation equations relating stellar and solar extinction will be developed.

(4) Routine turbidity reduction procedures will be established.

(5) Standardization of observing techniques; extinction intercomparisons with portable narrow band photoelectric photometer. Purchase of new observatory equipment where necessary.

(6) Periodic calibration checks to verify stability of the network.

\[ \tau_p (\lambda, 0) = \beta \lambda^\alpha \]

\[ \beta = 0.0008 \pm 0.0002 \]

\[ \alpha = 4.4 \pm 0.4 \]
Items (1), (2), and (3) of the proposed implementation scheme are already being actioned through the courtesy of B. Marino, W. S. G. Walker and G. Christie of Auckland Observatory, R. Austin of Mt John University Observatory, and J. Trodahl and D. Sullivan of Victoria University. We thank these observers for data already received and for assistance in interpretation.

We also take this opportunity to invite enquiries from interested observatories and observers.

REFERENCES


WMO, 1974. WMO Operations Manual /or Sampling and Analysis Techniques /or Chemical Constituents in Air and Precipitation, WMO No. 299.
UTILIZATION OF NEW ZEALAND ASTRONOMICAL
OBSERVATORIES FOR MONITORING ATMOSPHERIC TURBIDITY

P.J. EDWARDS

Introduction

The World Meteorological Organisation has established a network of background air pollution stations. The stated [1] purpose of this network is to (a) measure current levels of pollution, and (b) detect and measure long term trends in the levels of significant atmospheric constituents which may result in climatic change.

Two categories of stations are specified:

(a) Baseline air-pollution stations

These are regarded as research stations having as their first priority measurements directed towards documenting long term changes in atmospheric composition of particular significance to weather and climate.

(b) Regional air-pollution stations

These have as their primary purpose the documentation of long term atmospheric composition changes which may be related to changes in regional land use or other regional activity.

The siting criteria established [1] for the two categories are more stringent for the Baseline stations than for the Regional stations but in both areas isolated rural environments are required, "located away from major population centres, major highways and air routes, preferably on small isolated islands or on mountains above the tree line" (Baseline) or on "the summits or higher slopes of elevated areas" (Regional). The general similarity between these sites and those of astronomical observatories is obvious. Freedom from significant changes in anticipated land use for 50 years within 100 km is also required for Baseline station sites.

High altitude astronomical observatories could function much like WMO Baseline stations and provide information on total ozone and the background level of particulates. Turbidity measurements at lower altitude observatories on the other hand could provide data on local urban aerosol levels.

The Physics Department of the University of Otago, under contract to the New Zealand Meteorological Service, is setting up a turbidity monitoring network using photoelectric stellar photometers at selected observatories [2]. This paper outlines the methods to be used in data collection and analysis.
Atmospheric Extinction

Routine photometric measurements of stellar extinction at selected observatories will be collated and analysed to provide a continuing measure of the atmospheric turbidity.

The extinction of a pencil of monochromatic light of wavelength ($\lambda$) from a source at zenith angle ($z$) is specified by the optical depth,

$$\tau(\lambda, z) = k(\lambda) \cdot M(z)$$

where

- $k(\lambda)$ = exponential extinction coefficient,
- $M(z)$ = atmospheric airmass ($M(0) = 1$).

The spectral irradiance in a solid angle which is small compared with the characteristic scattering solid angle is then

$$J(\lambda, T) = J(\lambda, 0) \exp[-\tau(\lambda, z)].$$

Since $\tau(\lambda, z) = k(\lambda) \cdot M(z)$ and $\tau(\lambda, 0) = k(\lambda)$ the optical depth in the zenith ($z=0$) serves to characterise the extinction at a particular wavelength.

The logarithmic slope of a monochromatic spectral irradiance–airmass plot,

$$\frac{\Delta \log J(\lambda, T)}{\Delta M(z)} = -\tau(\lambda, 0) = -k(\lambda)$$

provides the extinction coefficients directly without requiring a prior knowledge of the zero depth flux.

A logarithmic plot of stellar irradiance against airmass can be subjected to linear regression analysis to yield best fit slope ($\tau(\lambda, 0)$) and zero mass intercept ($J(\lambda, 0)$). Applied to sun photometer measurements such a plot is called a Langley Plot. In stellar astronomy it is usually called a Bouguer Plot. The optical depth at the zenith ($z=0$) due to aerosols $\tau_p(\lambda, 0)$ may then be found in the usual manner:

$$\tau_p(\lambda, 0) = \tau(\lambda, 0) - \tau_R(\lambda, 0) \frac{p}{p_0} - \tau_a(\lambda, 0)$$

where

- $\tau_R(\lambda, 0)$ = Rayleigh scattering optical depth for standard atmosphere with surface pressure $p_0$;
- $p$ = atmospheric pressure at site;
- $\tau_a(\lambda, 0)$ = optical depth due to absorption by gases (ozone, water vapour, etc.);
- $\tau_p(\lambda, 0)$ = optical depth due to scattering and absorption by airborne atmospheric particulates.
The Rayleigh scattering term may be accurately subtracted from $\tau(\lambda,0)$ if the surface pressure is known. Standard atmosphere values of precipitable water content are usually adequate but total ozone should preferably be known or else be deduced from the extinction data in order to adequately evaluate the gas absorption term. The residual optical depth $\tau_p(\lambda,0)$ is the turbidity measure.

For photo-electric stellar photometry using photomultipliers it is customary to use the Bouguer method to calibrate the photometer in terms of an international system such as the Johnson-Morgan UBV system [3] by observing a sequence of standard stars covering a wide range of irradiance and spectra. Study of variable stars for example is then carried out by including several nearby comparison standard stars in the observations. Astronomical photometry can be performed with a precision approaching 0.001 stellar magnitudes (0.1%). The existence of such sets of internationally accepted standard candles and photometric systems makes astronomical observations ideal for on-going international surveys of extinction from which turbidity and ozone content may be deduced. Extinction measures from a number of astronomical observations have already been published and analysed for turbidity [4].

In New Zealand we propose to use data from the network of five observatories listed in Table 1. The data will in the main be obtained as part of the observatories' normal photometric programs. The reductions will be carried out on comparison star observations typically made at airmasses, $M(z) < 3$.

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Altitude</th>
<th>Photometric System</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auckland Observatory</td>
<td>174°47'E, 36°54'S</td>
<td>80 m</td>
<td>UBV</td>
<td>Urban (Auckland)</td>
</tr>
<tr>
<td>Beverly-Begg Observatory</td>
<td>170°30'E, 45°52'S</td>
<td>120 m</td>
<td>UBV</td>
<td>Urban (Dunedin)</td>
</tr>
<tr>
<td>Carter Observatory at Black Birch</td>
<td>173°48'E, 41°45'S</td>
<td>1400 m</td>
<td>UBV narrow-band</td>
<td>Mountain ridge</td>
</tr>
<tr>
<td>Mt. John University Observatory</td>
<td>170°28'E, 43°59'S</td>
<td>1030 m</td>
<td>UBV narrow-band</td>
<td>Mountain top</td>
</tr>
<tr>
<td>Signal Hill University Observatory</td>
<td>170°33'E, 45°51'S</td>
<td>400 m</td>
<td>narrow-band</td>
<td>Hilltop near Dunedin city</td>
</tr>
</tbody>
</table>
Data Reduction [5]

Each observatory will provide raw photometric readings, usually in the form of stellar, sky background and dark current readings at known universal times. These observations, together with star name, star colour and star coordinates, additional meteorological information (pressure, temperature, humidity, cloud) and the photometric parameters form the primary data set from which the extinction and turbidity coefficients are computed.

The stellar data will usually consist of observations of stars of astronomical interest together with observations of nearby comparison stars. The comparison stars will normally be chosen from an International list of standard stars to be similar in spectrum (colour) and irradiance to that of the stars under astronomical investigation. The turbidity coefficients are computed from the standard star observations.

The data are punched on IBM cards and the following operations carried out:

1. Subtraction of sky background and photometer dark current from the reading.
2. Conversion to a logarithmic scale.
3. Calculation of the hour angle of the star at the time of the reading.
4. Calculation of the zenith angle.
5. Calculation of the airmass.

After a set of data has been read and processed as above, any or all of the following operations may be required to be carried out on the accumulated data:

6. Preliminary X-Y plot of log (photometer reading) against airmass for editing purposes.
7. Weighted linear regression analysis of log (photometer reading) against airmass. Readings are weighted by (M(z))^{-1}.
8. Subtraction of calculated Rayleigh extinction.
10. Calculation of turbidity parameters.

Photometric Systems

Some observations will be made with narrow-band (<5 nm) and intermediate band (uvby) interference filters. Such data requires no correction for the dependence of effective wavelength upon star colour and airmass. However the internationally accepted UBV (wide-band) system, like the Volz system, does require such a correction.

For the systems considered here, the airmass and spectrum corrections can be adequately accounted for by rewriting Equation (1) in the form [7]

$$\frac{\partial \lambda \lambda_0 M}{\partial M} = -k(\lambda_o)[1-n(\lambda_o)\left(\frac{\phi(\lambda_o T)}{\lambda_o} - 5 - \frac{n+1}{2} + nM k(\lambda_o)\right)]$$
\[ J_{\lambda_0} (M) = \text{the observed irradiance at airmass, } M; \quad \lambda_0 = \int \lambda S(\lambda) d\lambda / \int S(\lambda) d\lambda, \text{ the mean photometer wavelength}; \quad \sigma^2 = \int (\lambda - \lambda_0)^2 S(\lambda) d\lambda / \int S(\lambda) d\lambda, \text{ the variance of the photometer response, } S(\lambda); \quad n = \text{the exponent in the power law representation of the wavelength dependence of } k(\lambda), \text{ the monochromatic extinction coefficient at } \lambda = \lambda_0; \quad \text{and } \phi(\lambda_0, T) = 5\lambda_0 - \frac{3\ln J(A_0)}{81/\lambda^3}, \text{ the absolute gradient of the stellar spectrum above the atmosphere.} \]

The size of the bracketed correction terms is proportional to both the extinction exponent, \( n \), and the square of the bandwidth, \( \sigma^2 \), of the photometer. Table 2 compares the bandwidth parameter \((\sigma^2/\lambda_0^2)\) for the U, B, V and Volz Photometric Systems and presents the correction factors for a typical sea level observation of a solar type star. It may be seen that the corrections required are of the same order for both systems. In the case of the Volz Photometer effective wavelengths are quoted [6] so that the corrections referred to here have already been made. The airmass dependence is of the order of several per cent per unit airmass and may need to be taken into account for the more precise astronomical measurements. The colour correction is typically several per cent for a sunlike star but can exceed 10% for blue and red stars.

<table>
<thead>
<tr>
<th>( \lambda_0/\mu \text{m} )</th>
<th>( \sigma^2/\lambda_0^2 )</th>
<th>( \lambda_{\text{eff}}/\mu \text{m} )</th>
<th>( \text{Volz Sun Photometer} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( U )</td>
<td>( B )</td>
<td>( V )</td>
<td>( B )</td>
</tr>
<tr>
<td>0.350</td>
<td>0.447</td>
<td>0.556</td>
<td>0.44</td>
</tr>
<tr>
<td>3.8x10^{-3}</td>
<td>6.5x10^{-3}</td>
<td>5.2x10^{-3}</td>
<td>3.4x10^{-3}</td>
</tr>
<tr>
<td>Colour Correction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solar Type Star</td>
<td>+1%</td>
<td>+3%</td>
<td>+2%</td>
</tr>
<tr>
<td>Airmass Correction (per unit airmass)</td>
<td>-4.5%</td>
<td>-2%</td>
<td>-0.3%</td>
</tr>
</tbody>
</table>

The variation of the correction terms with star colour and airmass can be regarded as due to a change in effective wavelength. For analytical purposes however it seems desirable to quote final extinction and turbidity parameters at fixed wavelengths \( \lambda_0 \), rather than at effective wavelengths which differ slightly from one to the next. For the UBV system the effective wavelength varies by less than 0.005 \( \mu \text{m} \) over a 2:1 range in stellar colour temperature.

**Summary**

Astronomical observations of turbidity offer a number of advantages. Among these are:
(1) Absence of local solar heating effects.

(2) Extension of data acquisition to night-time, allowing diurnal variations to be determined.

(3) Precise photoelectric measurements with small acceptance angles to agreed international standards.

(4) Multiple airmass observations which minimise errors due to time dependent extinction [8].

(5) Utilization of existing world-wide facilities at well selected sites.

The N.Z. network, when fully operational will provide data compatible with that from international observatories, both solar and stellar.

References


The New Zealand atmospheric extinction monitoring programme

P.J. Edwards

Physics Department, University of Otago, Dunedin, New Zealand

The use of existing optical observatories in New Zealand to monitor atmospheric extinction is described. The instrumental and data reduction procedures required to provide extinction data from which the atmospheric "turbidity" can be deduced are outlined. Preliminary results are compared with solar derived data.

Synthetic extinction coefficients in the UBV and $uvby$ systems have been computed for a wide range of stellar spectra for a Rayleigh atmosphere containing ozone, water vapour and dust, with the total content of each of these three constituents as parameters.

Stellar spectra for over 50 spectral type/luminosity classes as tabulated by the Vilnius group have been used in the analysis.

INTRODUCTION

The Physics Department of the University of Otago, under contract to the New Zealand Meteorological Service, is setting up an extinction monitoring network using photoelectric stellar photometers at selected observatories. This paper outlines the methods to be used in data collection and analysis.

In a previous paper (Edwards & Othman 1976) it was suggested that high altitude astronomical observatories could function much like World Meteorological Organization Baseline airpollution stations and provide information on total atmospheric ozone content and the background level of particulates. Measurements at lower altitude observatories, on the other hand, could provide data on local urban aerosol levels. Astronomical observatories are well suited to the former task for a number of reasons:

(1) The siting criteria for WMO Baseline stations and optical observatories are similar.

(2) Photometric measurements of extinction using standard international photometric systems (like the UBV system) are routine at many observatories.

(3) These observations have been made for several decades on a global basis and therefore constitute a valuable data base for long term climate studies.

The use of astronomical observatories for studies of atmospheric particulates is not new. Extinction measures from a number of observatories have already been published and analysed (Dachs et al. 1965, de Vaucouleurs & Angione 1973) for 'turbidity' — that component of extinction due to particulates. The present New Zealand study differs from these in that we are attempting to set up a network of stations from which data is analysed on a routine basis with particular attention to calibration of the 'natural' photometric systems in use at each observatory.

We hope to use data from the network of five observatories listed in Table 1. The data will in the main be obtained as part of the observatories' normal photometric programmes. The reduction will be carried out on comparison star observations typically made at airmasses $M(z)<4$ during variable star observation programmes.

ATMOSPHERIC EXTINCTION

Routine photometric measurements of stellar extinction at selected observatories will be collated and analysed to provide a continuing measure of the atmospheric turbidity.

The extinction of a pencil of monochromatic light of wavelength $\lambda$ from a source at zenith angle $z$ is specified by the optical depth,

$$\tau(\lambda,z) = k(\lambda) M(z)$$

where

$$k(\lambda) = \text{exponential extinction coefficient, and}$$

$$M(z) = \text{relative atmospheric airmass (} M(0) = 1) \text{.}$$

The spectral irradiance in a solid angle which is small compared with the characteristic scattering solid angle is then

$$J(\lambda, \tau) = J(\lambda,0) \exp(-\tau(\lambda,z))$$

Since $\tau(\lambda,z) = k(\lambda) M(z)$ and $\tau(\lambda,0) = k(\lambda)$ the optical depth in the zenith ($z = 0$) serves to characterise the extinction at a particular wavelength.

The logarithmic slope of a monochromatic spectral irradiance-airmass plot,

$$\frac{\partial \ln J(\lambda, \tau)}{\partial M(z)} = -\tau(\lambda,0) = -k(\lambda)$$

(1)
Table 1  Proposed turbidity stations.

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Altitude (m)</th>
<th>Photometric system</th>
<th>Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auckland Observatory</td>
<td>174°47'E 36°54'S</td>
<td>80</td>
<td>UBV</td>
<td>Urban (Auckland)</td>
</tr>
<tr>
<td>Beverly-Begg Observatory</td>
<td>170°30'E 45°52'S</td>
<td>141</td>
<td>UVBY</td>
<td>Urban (Dunedin)</td>
</tr>
<tr>
<td>Carter Observatory at Black Birch</td>
<td>173°48'E 41°45'S</td>
<td>1400</td>
<td>UBV, VUW scanner</td>
<td>Mountain ridge</td>
</tr>
<tr>
<td>Mt John University Observatory</td>
<td>170°28'E 43°59'S</td>
<td>1030</td>
<td>UBV, UVB, VUW scanner</td>
<td>Mountain top</td>
</tr>
<tr>
<td>Signal Hill University Observatory</td>
<td>170°33'E 45°51'S</td>
<td>400</td>
<td>Narrow-band</td>
<td>Hilltop near Dunedin city</td>
</tr>
</tbody>
</table>

provides the extinction coefficients directly without requiring a priori knowledge of the zero depth flux.

A logarithmic plot of stellar irradiance against air mass can be subjected to linear regression analysis to yield best fit slope \( r_\infty(X,0) \) and zero mass intercept \( J(X,0) \).

\[
\tau_\lambda(X,0) = k_R(X)R_\infty + k_o(X) \frac{X}{X_0} + \tau_w \left( \frac{W}{W_0} \right)^{1/2} + \tau_p(X,0)
\]

\( k_R(X) \) = Rayleigh scattering extinction coefficient for standard atmosphere with standard pressure \( po \);
\( p \) = atmospheric pressure at site;
\( k_o(X) \) = ozone extinction coefficient for total content \( x_o \);
\( x \) = total ozone content;
\( \tau_p(X,0) \) = optical depth due to scattering and absorption by airborne atmospheric particulates;
\( \tau_w(X) \) = optical depth due to water vapour of total content \( w_o \); and
\( w \) = total water vapour content.

The size of the Rayleigh scattering term \( k_R(X) \) has for many years been a matter of controversy. It is now generally accepted, however, that earlier laboratory values are too high (Hoyt 1976), by as much as 4.5%. Since molecular scattering is large compared with that due to aerosols, this error has a proportionally large influence on the values of aerosol extinction deduced from solar and stellar observations. Adoption of the most recent laboratory data leads to an upward revision in aerosol content and removes many anomalous results, possibly including some associated with the calculated response of the Johnson U filter system (Dachs et al. 1965).

In our reduction programme we use the expression

\[
\tau_\infty(X,0) = k_R(X) = aX^{-4} + bX^{-s}
\]

with \( a = 8.337 \times 10^{-3} \), and \( b = 1.044 \times 10^{-4} \) to calculate the Rayleigh extinction coefficient with a precision of \( \pm 0.0005 \) over the range \( \lambda = 0.35-0.8 \mu m \) for a standard atmosphere with \( p_0 = 1013.25 \text{ mb} \).

Total ozone should preferably be known or deduced from a narrow band extinction measurement near \( \lambda = 0.600 \mu m \), the peak of the Chappuis band, at which \( \tau_o(X,0) \) is typically 0.04. Ozone absorption contributes significantly to Stromgren u and Johnson U system extinction via the Huggins bands and to Stromgren y and Johnson V via the Chappuis band.

The residual optical depth \( \tau_p(X,0) \), found by subtracting the gas absorption and scattering terms from \( \tau(X,0) \), is assumed to be due to aerosols. Fig.1 shows an example of the residual extinction deduced from the spectrum scanner observations of the Victoria University Group (Trodahl et al. 1973). These observations require little correction for the dependence of 'effective', that is, equivalent monochromatic wavelength, upon star colour and airmass. However, Stromgren u and Johnson UBV data do need such a correction.

For the broad band systems considered here, the airmass and spectrum corrections can be accounted for by rewriting Equation 1 in the form (Golay 1974)

\[
\frac{\ln J_{\omega_\infty}(X)}{\Delta M} = -k(\omega)\left[|\omega - n\left(\frac{\omega}{\omega_0}\right)^{\gamma} \varphi(\omega, T) - \frac{n+1}{2} + nMk(\omega)\right]
\]

where

\( J_{\omega_\infty}(X) \) = the observed irradiance at airmass \( M \);
\( \omega_0 \) = \( \int S(\lambda)\lambda d\lambda / \int S(\lambda) d\lambda \), the mean photometer wavelength;
\( \sigma \) = \( \int (\lambda - \lambda_o)^2S(\lambda)\lambda d\lambda / \int S(\lambda) d\lambda \), the variance of the photometer response, \( S(\lambda) \);
\( n \) = the exponent in the power law representation of the wavelength dependence of \( k(\lambda) \), the monochromatic extinction coefficient; and
Edwards: New Zealand atmospheric extinction

Fig. 1 Optical depth in the zenith due to ozone and aerosols deduced from Victoria University of Wellington scanner observations (Trodahl et al. 1973). The turbidity parameters are defined by \( \tau_{p}(\lambda,0) = \tau_{p}(1,0)\lambda^{-\alpha} \) and \( B = 0.434\tau_{p}(0.5,0) \).

\[
\phi(\lambda_0, T) = 5\lambda_0 - \left( \frac{\partial \ln J(\lambda,0)}{\partial \lambda} \right)_{\lambda = \lambda_0}, \text{ the 'effective' absolute gradient of the stellar spectrum above the atmosphere.}
\]

In cases where line blanketing is significant in the neighbourhood of discontinuities such as the Balmer Jump, the effective gradient must be found from published spectra. We have numerically integrated over the UBV and uvby passbands in order to calculate Rayleigh and ozone extinction for more than 50 spectral type/luminosity classes. The 50 A scanner data published by Straizys & Sviderskeine (1972) and the UBV spectral responses adopted by Buser (1978) were used.

Following from the operational definition of effective optical depth \( \tau_{\lambda_0} \):

\[
\tau_{\lambda_0}(X, \phi(\lambda_0, T)) = \ln \left( \frac{J_{\lambda_0}(0)}{J_{\lambda_0}(X)} \right)
\]

where \( X \) defines the absolute values of the air mass parameters: pressure, ozone content, water content, etc., and \( \phi(\lambda_0, T) \) is a single parameter representation of the stellar spectrum, we have calculated effective Rayleigh, ozone, and water vapour extinction coefficients from equations of the form

\[
k_{i,j,\lambda_0} = \frac{\tau_{j,\lambda_0}}{X_j}
\]

with

\[
X_i = M_i(z) \frac{P_0}{P}; \quad X_2 = M_2(z) \frac{x}{X_0}; \quad \ldots;
\]

where

\[
\sum_{j=1}^{n} \tau_{j,\lambda_0} = \ln \left[ \frac{E(\lambda,0)}{S(\lambda)} \right] \frac{d\lambda}{\sum_{j=1}^{n} \Pi(T_j(\lambda))} d\lambda
\]

\( E(\lambda,0) = \) stellar spectrum above atmosphere, and \( T_j(\lambda) = \) atmospheric transmission for the \( j \)th constituent.

Fig. 2 illustrates the spectral dependence of the extinction.
After subtraction of the calculated extinction, Equation 1 is used iteratively to evaluate \( \tau_p(\lambda,0) \). The variation of the correction terms with star colour and airmass can be regarded as due to a change in effective wavelength. For analytical purposes it is desirable to quote final extinction parameters at fixed wavelengths rather than at effective wavelengths which differ between observatories and from one observation to the next. We have chosen these fixed wavelengths to be the mean wavelengths (\( \lambda_0 \)) of the system. These can be deduced from the colour transformation coefficients (Hardie 1962) for each observatory.

It is hoped that implementation of this technique will be of astronomical as well as meteorological value in the areas of photometric system calibration and astronomical site quality evaluation.

ACKNOWLEDGMENT
I thank M. Othman for calculating the data for Fig.2

REFERENCES
AN INVESTIGATION OF WIND-ENERGY PROSPECTS IN THE OTAGO REGION OF NEW ZEALAND

P.J. EDWARDS and K.R. DAWBER
Physics Department, University of Otago, Dunedin (New Zealand)

Summary

The investigation reported here covers the Otago region of the south island of New Zealand east of 169° east longitude, an area of $2.5 \times 10^4$ hectares centred near 45° latitude. Three years of data from an anemometer network installed as part of a national wind-energy resource survey [1] have been analyzed to provide (a) an inventory of the wind energy available annually in the region; (b) a description of the characteristics of prospective aerogeneration sites. The total resource is estimated to exceed 80 TW h per annum (9 GW average power) of which 10% might be utilised without either significant land use conflict or severe environmental impact. Characterisation of prospective sites in terms of wind-speed frequency, gust and lull distributions, wind-speed height profile and longitudinal turbulence using data with high time resolution yields parameters similar to those found in other temperate maritime and continental climatic regimes.

Notation

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZMS</td>
<td>New Zealand Meteorological Service</td>
</tr>
<tr>
<td>$V_a(H)$</td>
<td>annual mean wind speed at height $H$ metres above ground (m/s)</td>
</tr>
<tr>
<td>$H_s$</td>
<td>elevation (altitude above sea level) of terrain (km)</td>
</tr>
<tr>
<td>$V_R$</td>
<td>rated wind speed at generator hub height (m/s)</td>
</tr>
<tr>
<td>$\bar{V}_a$</td>
<td>annual mean wind speed at generator hub height (m/s)</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>overall power coefficient at rated speed</td>
</tr>
<tr>
<td>$P_A$</td>
<td>average power output per unit swept area (W/m$^2$)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>air density</td>
</tr>
<tr>
<td>$&lt;V^3&gt;$</td>
<td>mean cube wind speed</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>power-law exponent</td>
</tr>
<tr>
<td>$&lt;V&gt;$</td>
<td>mean wind speed (m/s)</td>
</tr>
<tr>
<td>WEFP</td>
<td>wind energy pattern factor $&lt;V^3&gt;/&lt;V&gt;^3$</td>
</tr>
<tr>
<td>$E$</td>
<td>wind energy flux (W/m$^2$)</td>
</tr>
<tr>
<td>$p(V)$</td>
<td>probability density function (s/m)</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>Weibull shape factor</td>
</tr>
<tr>
<td>$c$</td>
<td>Weibull scale velocity (m/s)</td>
</tr>
<tr>
<td>$t$</td>
<td>time interval</td>
</tr>
<tr>
<td>$V_T$</td>
<td>total horizontal velocity component (m/s)</td>
</tr>
</tbody>
</table>
1. Introduction

The dominating orographic feature of the South Island of New Zealand is the series of ranges running parallel to the western coastline. The eastern ranges of Otago merge into rolling foothills comprising slightly tilted fault blocks with large areas of flat terrain. The average elevation of the $2.5 \times 10^6$ hectares of land in the Otago region east of $169^\circ$ east longitude is 500 metres. 60% of this land lies between 300 and 1200 metres elevation.

The region is subjected to a synoptic scale weather pattern of westerly origin. The extensive high country inland is well exposed to the prevailing westerly winds. In the lee of the ranges, the associated east coast landforms experience sea breezes from the easterly quarter and south-west air flows associated with cyclonic motion to the west of the island.

Much of the inland high country is accessible tussock grassland with shallow relief used only for grazing purposes and, as such, appeared to be potentially well suited to wind energy farming. Existing wind data for the region were limited to those from a small number of NZMS stations generally located at low-altitude sheltered sites, completely unrepresentative of the region.

2. Methodology

Selection of anemometer stations

In order to enlarge the data base for the purposes of a realistic resource inventory, a network of 30 anemometer stations was established and maintained for a period of 3 years, from 1975 through 1978. The stations were chosen to provide, together with those NZMS stations already in existence, a representative sample of the wind field in the region. Figure 1 shows their location. Particular attention was paid to installing stations in the less-accessible inland high country and at exposed coastal sites.

The stations used in the investigation fall into four categories. The first, type (a), are exposed coastal sites and as such are representative of areas of eastern Otago commonly subjected to sea breezes. The second, type (b), are elevated inland sites and are representative of much of the inland area of Otago. Type (c) are locally augmented sites and type (d) are baseline sites.

(a) Exposed coastal sites

Type (a) sites are situated mainly in rugged regions of the east coast or on coastal hilltops. The intention with these sites was to obtain measurements of
Fig. 1. South Island of New Zealand showing the location of anemometer stations installed during the investigations.

wind at various elevations in the coastal area before substantial modification occurs by relief.

(b) Elevated inland sites

These sites lie sufficiently far inland for winds to be substantially modified by relief and elevation. As with type (a), the type (b) sites were selected for good exposure to winds with no obstruction by trees, buildings, etc., but as far as possible without the topographic features likely to cause the local augmentation discussed in the next section.
(c) Locally augmented sites

The flow over slight or moderate relief is discussed by Davidson et al. [2], and in particular the "speed-up factors" which may operate to make some locally elevated sites windier than the statistically "normal" increase of wind with land elevation might suggest. We have chosen several sites with geomorphology characteristic of the simple landforms proposed by Meroney [3] which are likely to cause exceptional augmentation above that already apparent for high-country type (b) sites. In particular, we selected examples of rounded hills with unobscured exposure in all directions, deep escarpments, and narrow gaps between high hills.

(d) Baseline sites

The NZMS stations in Otago for which records are available for at least eight years were used as baseline stations.

A detailed description of all sites is given in [4].

As the main interest of the survey was in viable wind energy, some areas were not instrumented. These included sheltered horticultural areas, low-lying valleys away from the coast and other sites with obvious environmental or land-use conflict with regard to aerogeneration.

Instrumentation and data processing

All stations were equipped with conventional cup-type anemometers located in the main at a height of 10 m above ground. A number of different recording instruments were used. Mechanical and electrical digital wind speed integrators ("cup counters") provided wind run records from which monthly and annual mean wind speeds were obtained. More-detailed records were obtained from paper chart records with hourly resolution and from magnetic tape records (Edwards [5]) with a resolution of a few seconds limited by anemometer inertia. Details of the instrumentation, calibration, data processing and analysis are described by Edwards et al. [6].

Figure 2 summarizes the data flow from the stations. Referring to this figure, cup counter readings were used to prepare annual and monthly means. At stations where both wind run and either tape or chart records were kept, cross checks were made in order to minimise errors caused by occasional winter freezing at the higher altitudes or due to other malfunction. Hourly wind speeds were read manually from charts as 10-minute means. True hourly and shorter-term means were computer analysed from taped data as were most wind-speed frequencies, and other statistical results.

3. Wind-energy resource inventory

Annual mean wind speeds $\bar{V}_a(H)$ at height $H$ above ground were used as the basis of the inventory. After examination of eight years of Meteorological Service Station data, a base year, 1976, was chosen as being representative of the last decade. The 1976 mean wind speeds averaged over these stations is
within 3% of the 8-year average value. Because of the high correlation between
the monthly and annual means at all stations in the region, it was possible to
use data gathered in the 3-year survey to fill gaps in the data coverage of the
base year.

Altitude dependence

The mean speeds at exposed inland sites were found to be well correlated
with station altitude. For example, the 1976 annual means at exposed non-
coastal stations is given by

$$\bar{V}_a(10) = 2.9 + 3.0 H_s$$

with a 0.9 correlation coefficient. Stated as a rule of thumb, the mean speed
thus doubles in the first 1000 metres rise in elevation. This is in general agreement with northern hemisphere experience for exposed areas (Davidson [2]). This relation is plotted as a broken line in Fig. 3. The relationship becomes

$$\bar{V}_a(10) = 2.4 + 3.4 H_s$$

when low-altitude, partially shielded stations (type (d) in Fig. 3) are included. This latter relation, plotted as a solid line in Fig. 3, is a better representation of the altitude dependence in the region. The two lines do not differ significantly in the potentially important range of elevations between 300 and 1200 metres. Coastal (type (a)) and locally augmented (type (c)) stations were excluded in this initial analysis.

Fig. 3. Annual (1976) wind speeds at Otago stations showing the altitude dependence for inland sites.

The annual wind energy available in the region was estimated in the following way:

(i) The annual mean wind speeds measured at 10 m height were first extrapolated to generator hub height by multiplying them by the factor 1.25. This is a conservative assumption. It is discussed further in Section 4.

(ii) A plant utilisation factor (average power/rated power) of 0.2 was assumed for the notional generators installed in the region. This assumption is consistent with the expected performance of optimally rated \( (V_R = 2 \bar{V}_a) \) generators in a Rayleigh-distributed wind-speed regime (Allen and Bird [7]; Brown and Higgin [8]). It was further assumed that the maximum overall power coefficient at rated speed, \( \epsilon = 0.25 \), so that the average power output per unit swept area, \( P_A = 0.2\epsilon(1/2 \rho V_R^3) \), becomes \( 0.2(\rho \bar{V}_a^3) \). Since the kinetic
energy flux $= \frac{1}{2} \rho <V^3> = \rho V_a^3$ for the wind regime typical of the region (see Section 4), these assumptions together imply that about 20% of the potential energy flux is extracted. The available power, taking into account the decrease in air density with altitude $H_s$ is then (in W/m$^2$)

$$P_A = 0.24 \, V_a^3 (1 - H_s/9)$$

(iii) In order to estimate the resource, it was necessary to determine the energy available per unit land area. A conservative assumption is a minimum spacing between generators of 10 rotor diameters (Ljungstrom [9]). We therefore assume a total swept area equal to 1% of the land area, that is 100 m$^2$/hectare. The average available specific power in kW/ha is then 0.024 $V_a^3 (1 - H_s/9)$.

(iv) Finally, the magnitude of the resource was calculated using this last equation with $V_a(10)$ given by the solid regression line of Fig. 3 to obtain the specific power yield as a function of elevation. The power and energy available in each elevation range were then calculated from a knowledge of the land area in each range.

**Magnitude of the resource**

Table 1 shows the results of applying this equation to the exposed inland region. It can be seen that the average power density increases rapidly with elevation, reaching $\sim 9$ kW/ha (0.9 W/m$^2$) at 1000 m altitude above sea level. The total power available is 8.6 GW.

Finally, as a possible strategy for utilization, the available average power and annual energy yield are evaluated assuming arbitrarily that 10% of the terrain between 300 m and 1200 m is used. This inventory excludes coastal and augmented sites. The resulting total available power is 660 MW. The installed (rated) power under the foregoing assumptions would be five times this, 3.3 GW.

Although the coastal winds are high, typically 7-8 m/s (see Fig. 3, type (a)), the land area of the coastal strip is only about $10^5$ ha. With the same assumption of 10% occupancy the average power is 100 MW. The potential contribution of the augmented sites is more difficult to estimate but seems likely to be of a similar magnitude. The grand total, close to 900 MW average power for the whole region, equivalent to 8 TW h (30 PJ) of annual energy, is approximately 40% of current New Zealand power consumption. Removing the 10% land use restriction, the total potential resource is 9 GW average power.

**Utilization of the resource**

The resource estimate described above is based on general principles and does not assume specific hardware for its utilization. Nevertheless, in the light of the present state of the art, it is possible to outline possible implementation.

First, as regards competition for land use, there appears no major problem. Most of the land is presently used for high-country, low-density sheep farming. This use is unlikely to be unduly perturbed by the installation of some 2000 2-MW machines grouped with a minimum separation of 10 rotor diameters in
# TABLE 1

The power and energy from elevated inland sites

<table>
<thead>
<tr>
<th>Height range (m)</th>
<th>Height mid-point (m)</th>
<th>Land area (ha)</th>
<th>( \bar{V}(10) ) (m/s)</th>
<th>( \bar{V}_a ) (m/s) (est. at hub-height)</th>
<th>( E ) (W/m²)</th>
<th>( P_A ) (W/m²)</th>
<th>Average available power per hectare (kW/ha)</th>
<th>Average avail. power overall (MW)</th>
<th>Average avail. power (using 10% of land)* (MW)</th>
<th>Annual avail. energy (using 10% of land)* (TW h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–300</td>
<td>150</td>
<td>810000</td>
<td>2.9</td>
<td>3.6</td>
<td>56</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300–600</td>
<td>450</td>
<td>780000</td>
<td>3.9</td>
<td>4.9</td>
<td>133</td>
<td>27</td>
<td>2.7</td>
<td>2100</td>
<td>210</td>
<td>1.8</td>
</tr>
<tr>
<td>600–900</td>
<td>750</td>
<td>420000</td>
<td>5.0</td>
<td>6.2</td>
<td>256</td>
<td>52</td>
<td>5.2</td>
<td>2200</td>
<td>220</td>
<td>1.9</td>
</tr>
<tr>
<td>900–1200</td>
<td>1050</td>
<td>260000</td>
<td>6.0</td>
<td>7.5</td>
<td>432</td>
<td>89</td>
<td>8.9</td>
<td>2300</td>
<td>230</td>
<td>2.0</td>
</tr>
<tr>
<td>1200–1500</td>
<td>1350</td>
<td>140000</td>
<td>7.0</td>
<td>8.7</td>
<td>668</td>
<td>136</td>
<td>14.0</td>
<td>2000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1500–1800</td>
<td>1650</td>
<td>50000</td>
<td>8.0</td>
<td>10.0</td>
<td>965</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total available power and energy from elevated inland sites

<table>
<thead>
<tr>
<th>Total available power</th>
<th>8600 MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total available energy</td>
<td>660 MW</td>
</tr>
<tr>
<td></td>
<td>5.7 TW h per annum</td>
</tr>
</tbody>
</table>

*See Section 3.*
a number of central energy "farms". For example, in the elevation range 600–1200 metres in which the annual mean speed at 10 metres height exceeds 5 m/s and the wind energy flux exceeds 250 W/m² at hub height (Fig. 3 and Table 1) machines rated above 12.4 m/s (44 km/h) at hub height with average power outputs $P_A > 50$ W/m² (corresponding to a conservative rated power coefficient of 0.25) could be used. Machines similar to the NASA/DOE MOD 1/MOD 2 designs (Robbins and Thomas [10] spaced ~1 km apart in the lower elevation range and ~0.5 km apart at higher elevations (and in the coastal strip) appear well suited to the wind regime and to the terrain. With a specific average power yield between 5 and 10 kW/hectare (Table 1), a typical energy farm rated at 100 MW yielding 20 MW average power would occupy an area of 2000–4000 hectare (20–40 km²). It is interesting to note that the specific energy yield of the wind farms envisaged here is of the same magnitude as that of hydroelectric lake storage areas presently operating in New Zealand and the Australian state of Tasmania.

4. Site characterization

The foregoing discussion makes use of three meteorological parameters accessible to measurement. These are

(i) the annual mean wind speed at anemometer height,

(ii) the wind-speed frequency distribution,

(iii) the exponent $\alpha$ in the power-law wind speed—height relation.

In what follows we briefly discuss the latter two of these together with other parameters relevant to the siting and performance of wind-energy conversion systems in the region under study.

Wind-speed frequency distribution

Wind-speed frequency distributions are a first step in the processing of wind data to give the mean $<V>$, the mean energy flux $E = \frac{1}{2} \rho <V^2>$, the wind energy pattern factor $<V^3>/<V>^3$ and for the modelling of aerogenerator performance (Lindley and Chin [11]; Edwards and Hurst [12]). Data logged in this survey have been routinely analysed to obtain these distributions. Figure 4 shows a typical result for one month of data from a station located at an altitude of 950 m and having a 1976 mean $\bar{V}_a(10) = 5.8$ m/s. This distribution yields parameters: $<V> = 5.1$ m/s; $E = 150$ W/m²; WEPF = 1.9.

The method of using fitted theoretical distributions to characterize experimental wind-speed distributions is well established. In accordance with widespread practice, we have used a two-parameter Weibull distribution for this purpose:

$$p(V) = (k/c) (V/c)^{k-1} \exp -(V/c)^k$$

where $k$ is a dimensionless shape parameter, and $c$ is a scale velocity.

A maximum likelihood criterion was used to fit values of the parameters $k$ and $c$. The maximum likelihood Weibull distribution fitted to the data of
Fig. 4. Typical wind-speed frequency distribution for one month of data (time resolution = 112 s) from an exposed inland station (altitude 950 m). \( \langle V \rangle = 5.1 \, \text{m/s} \); best fit \( k = 2.2 \).

Fig. 4, with \( k = 2.23 \) and \( c = 6.01 \, \text{m s}^{-1} \), is plotted in the figure. Other methods of fitting Weibull parameters, for instance matching the moments of the experimental and Weibull distributions, or using the method described by Justus et al. [13], may give slightly different results.

For monthly periods of wind data from the Otago stations, fitted values of \( k \) generally lie in the range 1.5–2.5, the higher values implying a smaller spread of wind speeds about the mean and vice versa. Successive months of data from a single site typically show an r.m.s. spread of 0.2 in fitted \( k \)-values.

Analysis of annual periods with a resolution of 112 seconds showed that the most representative value of \( k \) for the region was close to 2 with an energy pattern factor of 2.0, slightly higher than that for the Rayleigh distribution \( (k = 2) \) commonly found in temperate maritime and continental climates [13] from analysis of hourly data. The justification for adopting a pattern factor of 2 in making the inventory from annual means can be seen in Fig. 5, compiled from two years of data from a typical inland station. It illustrates the influence of wind-speed averaging time upon the energy pattern factor. It is evidence that little additional energy flux is available if the wind speeds are averaged over periods shorter than one hour.

The time resolutions used in collecting wind data have ranged from less
than 1 sec (Lindley and Chin [11]) to days or months for wind-run data. We have investigated the changes in the wind-speed statistics which take place as the time resolution is shortened from one day, when successive samples are largely uncorrelated, to 10 seconds, when much of the structure is resolved. Table 2 summarizes the statistical parameters of the distributions obtained for a typical month over the time-resolution range considered. As the time resolution is shortened an increased proportion of samples are found at very high and very low speeds. This is reflected in the increase in WEPF in Fig. 5 and Table 2 and the decrease in Weibull $k$ parameter in Table 2.

TABLE 2

Statistical parameters as a function of time resolution

<table>
<thead>
<tr>
<th>Time resolution</th>
<th>Mean m s(^{-1})</th>
<th>Wind energy pattern factor</th>
<th>Fitted Weibull parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$k$</td>
</tr>
<tr>
<td>1 day</td>
<td>5.90</td>
<td>1.37</td>
<td>3.49</td>
</tr>
<tr>
<td>6 hours</td>
<td>5.90</td>
<td>1.76</td>
<td>2.27</td>
</tr>
<tr>
<td>1 hour</td>
<td>5.90</td>
<td>1.94</td>
<td>2.14</td>
</tr>
<tr>
<td>15 minutes</td>
<td>5.90</td>
<td>1.98</td>
<td>2.10</td>
</tr>
<tr>
<td>5 minutes</td>
<td>5.90</td>
<td>2.01</td>
<td>2.07</td>
</tr>
<tr>
<td>100 seconds</td>
<td>5.90</td>
<td>2.05</td>
<td>2.04</td>
</tr>
<tr>
<td>10 seconds</td>
<td>5.90</td>
<td>2.15</td>
<td>1.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$c$ (m s(^{-1}))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6.46</td>
</tr>
</tbody>
</table>
The wind speed—height dependence

For the purposes of the inventory, 10-m anemometer means were extrapolated to hub height by raising them by a factor of 1.25, thus doubling the energy flux estimate. For hub heights between 30 m and 60 m this is equivalent to assuming a power-law exponent $\alpha$ between 0.2 and 0.12, respectively, in the empirical relation $V(H_1)/V(H) = (H_1/H)^\alpha$ used by De Lisle [14] for NZMS stations. It is a conservative assumption in the light of surface boundary-layer theory and tower measurements made at a typical inland site, particularly for hub heights appropriate to the MOD-1 (43 m) and MOD-2 (61 m) machines.

Measurements made from a 30-m tower located at an inland hill site 600 m above sea level at which the surface roughness parameter was estimated to be $\sim 10^{-2}$ m (next Section), indicate a mean $\alpha = 0.18$ for $V(30) < 5$ m/s and $\alpha = 0.15$ for $V(30) \geq 5$ m/s. These values are similar to those tabulated by Davenport [15] for open level terrain and are in general agreement with ESDU predictions [16]. Wind shear at smooth sites of this type is therefore less likely to present problems for generator operation than at coastal sites in the region which generally have rougher fetch terrain.

Turbulence intensity

The magnitude of the comparatively rapidly fluctuating component ($\tau < 100$ s) of the wind speed due to atmospheric boundary-layer turbulence is important in considering the interfacing of aerogenerators with an electrical utility network (Johnson and Smith [17], and for estimation of extreme winds at a site. The turbulence intensity, conventionally defined as $\sigma/<V>$ for r.m.s. fluctuation $\sigma$ about a mean wind speed $<V>$, is therefore a useful site parameter.

A considerable body of empirical and theoretical work on atmospheric turbulence has been summarized in the ESDU publications [18, 19]. These results are for neutral atmospheric stability and homogeneous terrain. They have been used to “fill out” direct measurements of turbulence obtained from the data from the inland sites. For some other Otago sites, particularly coastal ones, the terrain is clearly not homogeneous, and the ESDU data may not be valid. Neutral stability is ensured for strong winds because of thorough mixing, but in light winds the validity of the ESDU data will also depend on the temperature profile.

Two aspects of the turbulence measurements must be considered. Firstly, the 3-cup anemometers used in the survey respond to the total horizontal wind speed, $V_T$. This contains contributions from the mean wind speed $<V>$, and from both the longitudinal and transverse turbulence velocities $u$ and $v$, i.e.

$$V_T = \sqrt{(<V> + u)^2 + v^2}$$

$$= (<V> + u) \left[ 1 + \left( \frac{v}{<V> + u} \right)^2 \right]^{1/2}$$
Since $\sigma_u$ is generally less than $0.15 < V >$ [19], the transverse turbulence affects $V_T$ by typically less than 1%, which is negligible when $\sigma_u$ is greater than $0.1 < V >$, as is usually the case. Therefore, for practical purposes, the fluctuations in the rate of a 3-cup anemometer are a measure of only the longitudinal turbulence component.

Secondly, the survey anemometers are a heavy-duty type with a distance constant of about 10 metres. They therefore do not resolve the most rapid turbulence structure. However, the factor by which the variance of the longitudinal component, $\sigma_u^2$, is reduced by the anemometer smoothing and the finite time resolution (usually 28 seconds) of the data readout, may be obtained from data in ESDU 74031 [19].

The analysis of the Otago data to obtain turbulence intensities has consisted of:

(1) For successive short blocks of data (usually one hour) calculation of the mean wind speed $< V >$.

(2) Subtraction of a 5-minute sliding mean from the data.

(3) Calculation of the standard deviation of the residues.

Figure 6 shows an example of the results obtained in this way for 10 days of data from a typical smooth hill site. The vertical bars indicate the range of

![Fig. 6. Standard deviation of longitudinal turbulence component as a function of mean wind speed $< V(10) >$. Solid curves are ESDU predictions [19] for (A) $Z_o = 0.1$; (B) $Z_o = 0.01$; (C) $Z_o = 0.001$ corrected for averaging time.](image)
values of $\sigma_u$ found for the corresponding values of $\langle V \rangle$. (The full length of each bar is twice the r.m.s. spread in $\sigma_u$ values.) Also plotted are curves derived from ESDU data, for values of $Z_0$, the surface roughness length, of $10^{-1}$ m (curve A), $10^{-2}$ m (B) and $10^{-3}$ m (C). These theoretical curves have been corrected for the same smoothing and high-pass filtering to which the experimental data were subjected. A value of $Z_0 = 10^{-2}$ m fits the observations for $\langle V \rangle$ above 5 m s$^{-1}$. At lower wind speeds, departure from neutral stability may account for the discrepancy between the theoretical and observed results.

A value of $Z_0 = 10^{-2}$ m for this site implies the following turbulence characteristics (ESDU 75001) [18] in strong winds: at height 10 m, $x_{Lu}$ (scale size of $u$-component in $x$ (longitudinal) direction) = 75 m, $y_{Lu} = 35$ m, $z_{Lu} = 25$ m; at height 30 m, $x_{Lu} = 95$ m, $y_{Lu} = 40$ m, $\sigma_u/\langle V \rangle = 0.14$.

**Episode distributions**

Aerogenerators generally have a cut-in wind speed, typically a few m s$^{-1}$, below which they cannot be operated, and a furling speed, typically a few tens of m s$^{-1}$, above which they must be shut down. In order to formulate operating strategies for aerogenerators, it is desirable to know the statistical distributions of the lengths of episodes during which the wind speed exceeds or falls below these thresholds. This is because there is likely to be some minimum time for which it is economic to start up the aerogenerator, in the event of the wind briefly rising above cut-in speed or falling below furling speed.

We have used data from the Otago survey to determine the form which these distributions take. Since the cut-in and furling speeds vary among different types of existing machines, and are not known for future machines, a fairly general investigation was made. The distributions were determined for wind-speed excursions both above and below a wide range of thresholds (Table 3).

It was found (Hurst et al. [20]) that the results could be easily generalized. The experimental points (from all stations) are well described by a power law,

$$N(T) = N_0 (T/100)^\alpha$$

where $N$ is the number per month (of 30 days) of episodes lasting longer than

**TABLE 3**

Parameters of a typical episode distribution

<table>
<thead>
<tr>
<th>Threshold (m s$^{-1}$)</th>
<th>$N_0$</th>
<th>Power-law exponent, $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. per month with $T &gt; 112$ s</td>
<td>Excursions above threshold</td>
</tr>
<tr>
<td>4</td>
<td>575</td>
<td>$-0.56$</td>
</tr>
<tr>
<td>8</td>
<td>389</td>
<td>$-0.66$</td>
</tr>
<tr>
<td>12</td>
<td>200</td>
<td>$-0.70$</td>
</tr>
<tr>
<td>16</td>
<td>73</td>
<td>$-0.72$</td>
</tr>
<tr>
<td>20</td>
<td>17</td>
<td>$-0.84$</td>
</tr>
</tbody>
</table>
With appropriate values of $N_0$ and $\alpha$, this power law is valid for $T$ in the range 100 s to 1 day, for episodes both above and below the thresholds. A similar power-law behaviour was found for thresholds 8, 12 and 16 m s$^{-1}$. Table 3 lists the parameters $N_0$ and $\alpha$ for these thresholds calculated from 6 months data from an inland site. Similar results have been obtained by Corotis et al. [21] in the U.S.A. for episodes longer than one hour. The value of the exponent $\alpha$ for the median speed is the same for episodes above threshold as for episodes below threshold. It is typically between $-0.5$ and $-0.7$.

The power-law form implies the following. If the wind speed has been above (or below) some threshold of interest for time $t_1$, then the conditional probability of it remaining above (or below) the threshold for a further time $t_2$ is given by $(1 + t_2/t_1)^{\alpha}$ for the appropriate value of $\alpha$ (assuming $t_1$ and $t_2$ lie within the range of validity of the power law). This estimate can be made in the absence of any other information with which to forecast the future behaviour of the wind.

For values of $T$ longer than a few days the probability deviates below the power-law relation leading to a finite expectation length of the episodes which would otherwise be infinite for $\alpha$ less negative than $-1$.

5. Conclusion

The magnitude of the resource and the characteristics of potential aero-generator sites in the Otago region of New Zealand have been investigated. The prospects for viable wind energy farming in the more elevated areas and on the coast appear to be favourable.

The installation of experimental wind generating plant would seem to be the next step.

Acknowledgement

The work reported here was supported under New Zealand Energy Research and Development Committee Contract 3096 as part of a national wind-energy resource survey.

References

9 O. Ljungstrom, Large scale WECS design and installation as affected by site wind energy characteristics, grouping arrangement and social acceptance, Wind Enging., 1 (1) (1977) 36–56.
14 J.F. de Lisle, Structure of the wind over New Zealand, NZMS Technical Information Circular No. 147, May 1975.
16 ESDU 72026, Characteristics of wind speed in the lower layers of the atmosphere near the ground: strong winds (neutral atmosphere), Engineering Sciences Data Unit, London, 1972.
18 ESDU 75001, Characteristics of atmospheric turbulence near the ground, III — Variation in space and time for strong winds (neutral atmosphere), 1975.
19 ESDU 74031, Characteristics of atmospheric turbulence near the ground, II — Single point data for strong winds (neutral atmosphere).
WIND ENERGY RESOURCE STUDIES IN NEW ZEALAND

PAUL J. EDWARDS
Department of Electronics Engineering and Applied Physics,
School of Information Sciences and Engineering, Canberra College of Advanced Education,
PO Box 1, Belconnen ACT 2616, Australia

(Received for publication 1 February 1989)

Abstract—The New Zealand Energy Research and Development Committee was established in 1974 in response to the oil crisis. The Wind Energy Resource Survey of New Zealand started in 1974. A Wind Energy Task Force coordinated a cross-disciplinary team of engineers and physicists drawn from universities and government departments. A number of aerogenerator demonstrations were carried out during the 10 year funding of wind power research and a Small-Scale Wind Power Report was published in 1986. The survey confirmed the conventional wisdom that New Zealand, by virtue of its topography and location, is an island well endowed with wind energy resources. However, small scale homestead use of wind energy is limited by the extensive penetration of the NZ grid system. There are many potential wind energy farm sites which could be integrated into the grid. Despite this it seems anomalous that no attempts have been made to demonstrate either a pilot farm or a well-sited single medium-sized wind power generator in New Zealand.

INTRODUCTION
The New Zealand Energy Research and Development Committee was established in 1974 in response to the oil crisis. Its principal function was to fund and manage contracts for energy R & D directed towards both long- and short-term goals. These included energy resource studies in the field of renewable energies. The Wind Energy Resource Survey for New Zealand started in 1974. A Wind Energy Task Force coordinated a cross-disciplinary team of engineers and physicists drawn from universities and government departments. In Phase 1 the team analysed NZ Meteorological Service records [1-4]. Regional wind energy surveys in Otago [4, 5] and Canterbury [6, 17] were carried out under Phase 2. Finally in Phase 3 more detailed measurements were made at prospective wind energy farm sites [7, 8]. These wind energy studies took place in the context of a number of energy scenario studies [9-12] and a broadly based study of potential wind energy utilization in New Zealand [13].

NEW ZEALAND ENERGY RESEARCH AND DEVELOPMENT COMMITTEE
The New Zealand Energy Research and Development Committee (NZERDC), formed in 1974 in response to the 1973 "oil crisis", ceased to exist in 1987 when it handed over responsibility for ongoing contracts to the New Zealand Ministry of Energy. During its 13 years of support for energy research the NZERDC managed over 460 individual research contracts and expended over $10 million. The committee successfully brought together managers, economists, engineers and scientists from industry, commerce, the universities and government in contracted research activity in a manner which has been emulated in Australia and other countries.

Energy research and development projects were funded in the areas of energy use, energy management and market development, technologies for energy production and use, environmental studies and resource assessment. The decision to fund a National Wind Energy Resource Survey, following individual proposals from university research groups, was driven by the perceptions of limited hydro-electric growth, the increasing burden of imported liquid fuels and the consequent attractiveness of "renewable" energy technologies.

While the decision to terminate the NZERDC was not predicted nor the recent decline in the price of oil and gas, there seems no reason to doubt the view that "enthusiasm for wind power comes in gusts, with long calm periods in between" and that the failure to follow-up the New Zealand Wind Energy Resource Survey with significant plant demonstration can be partly attributed to the current cheapness of liquid fuels.
WIND ENERGY RESOURCE SURVEY OF NEW ZEALAND

Individual proposals from university groups for resource surveys and wind power plant development and demonstration were coordinated, managed and funded through a Wind Energy Task Force set up by the NZERDC. University participants comprised Lincoln College (Agricultural Engineering); University of Canterbury (Mechanical Engineering); University of Auckland (Physics) and University of Otago (Physics). The main government participation came from the New Zealand Meteorological Service of the Ministries of Works and Energy. The Task Force did not include representation from private industry. No machine development or demonstration was funded. However, a Wind Energy Resource Survey was initiated under the direction of the Task Force. The Task Force met regularly, receiving interim reports from the regional surveys, approving proposals for transmission to the NZERDC and generally providing technical coordination and direction of the resource survey from 1975 to 1980.

The New Zealand Wind Energy Resource Survey was implemented in three distinct but not consecutive phases:

Phase 1: Analysis of existing meteorological data.
Phase 2: Acquisition and analysis of wind data on a regional basis.
Phase 3: Detailed site investigations.

A fourth study phase was envisaged, but not implemented. This was to include feasibility studies followed by demonstrations of individual machines and wind energy farms. However, several machine demonstrations did in fact take place, funded separately by private industry, university departments and government.

Phase 1: Analysis of existing data

Wind data from the New Zealand Meteorological Service formed the primary data base. These data included mean wind speeds from over 90 anemometer sites and daily wind-runs from an additional 80 sites. As expected, these sites were frequently located in sheltered areas and were often found to provide unrepresentative samples of the surface wind field. Nevertheless, taken in conjunction with upper air data at 900 millibars, and supplemented by the regional surveys, two strong wind regimes of interest were identified [4, 17]. These are: (a) exposed coastal areas and (b) elevated inland areas. Referring to Fig. 1, the wind regime characteristics in New Zealand can be understood from its topography and location in the Southern Pacific, stretching from 34°S to 47°S latitude, in the path of the strong westerlies known as the "roaring forties". The Southern Alps in the South Island and their extension in the North Island are well exposed to the prevailing westerlies. This elevated land is characterized by annual mean wind speeds of at least 6 m/s at 10 m above ground level at elevations greater than about 1000 m. Typically the diurnal and seasonal range is small, perhaps 20%.

The long New Zealand coastline provides a second strong wind regime at exposed sites. These are characterized by larger diurnal "sea-breeze" variations than for the inland sites.

Phase 2: Regional wind energy surveys

Two regional surveys were carried out, both in the South Island. These provided an additional 50 sites generally well exposed to strong winds. The main recording media were mechanical and electronic wind-run counters supplemented by slow-speed magnetic tape recorders [5]. The objectives of the surveys were twofold: (a) to provide more representative sampling and characterization of the surface wind field and (b) to assist in the identification of strong wind sites and regimes [4].

Figure 2, for example, illustrates the relatively small diurnal variation in wind speed and wind power flux at a typical elevated inland site in the province of Otago [8]. Figure 3 illustrates typical wind speed frequencies from another such site to which a two parameter Weibull distribution has been fitted. These regional surveys generally revealed no surprises. The increase in mean wind speed with elevation, the wind speed frequency distribution and the distribution of lulls were all found to be in general agreement with measurements made in similar mid-latitude terrain in the northern hemisphere [14]. Strong wind sites were identified and annual mean wind speeds were estimated from short term regional records by using long term meteorological service data from selected base line sites.

Phase 3: Detailed site investigations

Prospective wind energy farm sites were identified in the first two phases of the survey using wind data in conjunction with topographic information and land-use criteria. Cherry [17] identified 16 prospective wind farm sites in this way. These are shown in Fig. 4.

A typical coastal site is at Waiuku on the Auckland west coast with a mean wind speed of 7 m/s at 30 m above ground level. Detailed measurements, using tower-mounted anemometers and vanes, of the wind speed and direction have been reported by Clegg [7].
Wind energy resource studies in New Zealand

Fig. 1. Relief map of New Zealand.
who also calculated wind speed frequencies, directional wind roses, peak wind gusts, vertical shear in the horizontal wind speed, wind power density and modelled the energy yield of the NASA MOD2 turbine.

Edwards & Dawber [8] made similar measurements at Rocklands, an elevated (600 m) site in inland Otago in the South Island with the same annual mean wind speed. Turbulence measurements were also made and found to be consistent with the wind shear and turbulence expected above smooth, flat terrain [15].

FEASIBILITY STUDIES

Modelling

Wind farm performance has been modelled at a number of these prospective sites [4, 7, 16, 17]. Average power yields of about 0.7 MW km\(^{-2}\) corresponding to a capacity factor of 24% are expected.

This yield is of the same order as for the hydroelectric lake storage areas which, in the South Island, are located at lower elevations inland. It is not surprising then that arguments for wind farming in preference to hydro development have been advanced on the basis of multiple and reversible land use. To date these arguments have not been accepted, presumably in part because they have not been supported by competitive economics and credible technology. Cherry [17] has modelled the output of twelve 250 MW farms which together could provide some 30 PJ annually, about 30% of the present New Zealand system output.

Remote area diesel–wind electric hybrid systems have been modelled and costed by Edwards [15]. Because of the extensive penetration by the national
grid, potential use is probably of most interest on the Chatham Islands and on Stewart Island where individual diesel installations are common. The wind regime on both these islands is favourable to small scale wind power development and suitable sites have been identified in the surveys [1, 4]. A central diesel generating station is planned for Stewart Island which will provide electric power at a price of 40 c/kWh [18]. Estimates of the fuel cost savings achieved by hybrid wind–diesel plant are rather modest at prevailing oil prices [15] and may not be deemed sufficient to warrant installation of wind plant. However, present plans call for the installation of a demonstration aeroenerator on the Chatham Islands to supplement the existing diesel system. Table 1 illustrates comparative costs of a number of notional remote-area power systems operating in New Zealand conditions.

**Aerogenerator demonstrations**

Chasteau [19] successfully developed and operated a grid-connected 5 kW Darrieus wind turbine from 1975 to 1982. The Mechanical Engineering Department of the University of Auckland in association with Fletcher Challenge Ltd also tested a 30 kW Danish induction machine for possible commercial application [20, 21]. This development has not been proceeded with.

The Ministry of Works and Development tested an 11 kW turbine for use on the Chatham Islands in 1980/1981. These tests were not successful. However it is now planned to install a 20 kW machine as a diesel fuel saver. No successful demonstrations of medium scale wind turbines at strong wind sites have yet been made.

**PROSPECTS FOR THE FUTURE**

The Wind Energy Resource Survey of New Zealand has provided a high quality data and knowledge base for the future exploitation of a significant and proven

<table>
<thead>
<tr>
<th>System</th>
<th>Average daily demand (kWh)</th>
<th>Peak load (kW)</th>
<th>Availability (hours/day)</th>
<th>Capital cost [NZ$ (1986)]</th>
<th>Net present worth [NZ$ (1986)]</th>
<th>Annual cost [NZ$ (1986)]</th>
<th>Annual cost of electricity (NZ c/kWh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 kW wind/4kVA diesel/ battery (homestead)</td>
<td>11</td>
<td>4</td>
<td>24</td>
<td>22,650</td>
<td>46,700</td>
<td>3750</td>
<td>91</td>
</tr>
<tr>
<td>1 kW wind/4kVA diesel/ battery (homestead)</td>
<td>11</td>
<td>4</td>
<td>24</td>
<td>20,150</td>
<td>44,200</td>
<td>3550</td>
<td>87</td>
</tr>
<tr>
<td>1 kW wind/4kVA diesel/ battery (homestead)</td>
<td>11</td>
<td>10</td>
<td>24</td>
<td>29,150</td>
<td>58,750</td>
<td>4714</td>
<td>115</td>
</tr>
<tr>
<td>4 kVA diesel battery (homestead)</td>
<td>11</td>
<td>4</td>
<td>24</td>
<td>12,650</td>
<td>56,900</td>
<td>4550</td>
<td>111</td>
</tr>
<tr>
<td>10 kW wind/100 kW twin (small community) diesel</td>
<td>310</td>
<td>100</td>
<td>24</td>
<td>104,000</td>
<td>965,000</td>
<td>77,400</td>
<td>68</td>
</tr>
<tr>
<td>100 kW twin diesel (small community)</td>
<td>310</td>
<td>100</td>
<td>24</td>
<td>75,000</td>
<td>1,087,000</td>
<td>87,200</td>
<td>77</td>
</tr>
<tr>
<td>5 kVA diesel (homestead)</td>
<td>11</td>
<td>10</td>
<td>16</td>
<td>6000</td>
<td>88,800</td>
<td>7120</td>
<td>174</td>
</tr>
<tr>
<td>2 kW wind/petrol/battery (homestead)</td>
<td>11</td>
<td>4</td>
<td>24</td>
<td>23,600</td>
<td>47,200</td>
<td>3780</td>
<td>92</td>
</tr>
</tbody>
</table>
renewable energy resource. It is unfortunate and somewhat anomalous that, despite confirmation and quantification of the size of this generous wind energy resource, its utilization lags well behind other less well endowed countries such as Australia. In the immediate future, remote island power supply will provide the first test of the viability of wind power in New Zealand.

REFERENCES

Upper Limits to the Hard X-Ray Flux from the Quiet Sun and Jupiter

Paul J. Edwards and K. G. McCracken

Southwest Center for Advanced Studies, Dallas, Texas

The results of a survey of celestial sources of X rays with energies in the range 20–60 keV are used to place new upper limits to the hard X-ray flux from the quiet sun. The data, obtained in 1965 with a balloon-borne scintillation detector, allow the upper limits to the hard X-ray flux suggested by OSO 1 measurements to be lowered by nearly two orders of magnitude. Upper limits to the Jovian X-ray flux are also discussed.

1. Introduction

A balloon-borne X-ray telescope has been used to search for celestial sources of X rays with energies above 20 keV. Two flights from Hyderabad, India (17°25'N, 78°35'E), and one flight from Mildura, Australia (34°13'S, 142°05'E), were launched during April 1965. This letter deals with a daylight flight launched from Hyderabad on the morning of April 2, 1965 (day 92). All the conclusions presented here are supported by the later Australian flight.

2. Instrumentation

The geometry of the X-ray telescope is indicated in Figure 1. The field of view of the detector was determined by a parallel set of 1/32-inch brass plates that provided a minimum attenuation of X9 for photons in the energy range 20–60 keV. These brass collimators determine a fan-shaped field of view with the fan passing through the zenith, the axis of the detector making an angle of 22° with the vertical. The angular width of the fan was 19° (FWHM), measured in the plane containing the telescope axis, perpendicular to the collimator plates. The telescope package was connected to the balloon through a 'line twister' that rotated the telescope at a uniform azimuthal rate of about 0.5°/second. A horizontal flux-gate magnetometer provided azimuth information with an accuracy of ±3°. The telescope scanned continuously in azimuth during the period 0830–1430 local mean time at an atmospheric depth of 4.5 g/cm².

The sun (declination 5° N) crossed the balloon meridian at a zenith angle of 12°. Atmospheric absorption limited significant observations to a 5-hour period around noon during which the solar zenith was less than 40°.

---

1 On leave from Physics Department, University of Adelaide, Adelaide, South Australia.
\( \sigma(E) \) specifies the energy resolution of the detector, \( R(\text{FWHM}) = 2.35\sigma(E)/E \).

With \( x = 4.5 \, \text{g/cm}^2 \, \text{air}, x' = 0.3 \, \text{g/cm}^2 \, \text{aluminum}, \theta_a = 22^\circ; \sigma(E) = E^{\alpha} \) (\( E \) in kev) and taking account of the increasing transparency of the collimator with energy, upper limits to a monoenergetic solar X-ray flux have been calculated as a function of energy by setting \( S(H)dH \) equal to 3 \( \sigma \) (three standard errors) in the relevant channel counting rates. The results are plotted in Figure 3 and compared with OSO 1 data obtained by Frost and Peterson [Lindsay, 1963]. Table 1 lists the new limits to the solar and Jovian X-ray line fluxes derived from the present observations. Upper limits to X-ray continuum fluxes may be estimated from these results for most spectra of interest.

The main uncertainty in the calculation concerns the attenuation of the radiation in its passage to the surface of the scintillator. However, since photon scattering into the detector has been neglected in calculating air path and detector mount attenuations, we believe these upper limits should be conservatively regarded.

**THE QUIET SOLAR X-RAY FLUX**

The Zurich sunspot number was zero for the two days prior to the observation and it remained zero for the following week. Three importance 1—flares occurred three days before the flight. Rocket measurements at energies of several kev made by the Lawrence Radiation Laboratory group [Chodil et al., 1965] suggest a thin plasma spectrum (\( T = 4.5 \times 10^6 \, \text{K} \)) for the quiet sun X-ray continuum.

Extrapolation of these data to higher energies yields an integral flux (\( E > 20 \, \text{kev} \)) of \( \sim 10^{-27} \, \text{cm}^{-2} \, \text{sec}^{-1} \), many orders of magnitude below the detection threshold of the balloon-borne telescope for such a steep spectrum (\( 1.5 \times 10^{24} \, \text{cm}^{-2} \, \text{sec}^{-1} \) (Table 1)).

On the other hand, observations of the non-flare sun near solar maximum [Chubb, 1966] do suggest a detectable flux above 20 kev. Extrapolation of these data gives an integral flux (\( E > 20 \, \text{kev} \)) of \( \sim 1 \, \text{cm}^{-2} \, \text{sec}^{-1} \). This flux would have been easily detected in the present data. We conclude therefore that the hard solar X-ray flux varies by at least one order of magnitude over a solar cycle. The present results also suggest that the integral flux at solar minimum is at least two orders of magnitude below the figure of \( 1 \, \text{cm}^{-2} \, \text{sec}^{-1} \) (\( E > 30 \, \text{kev} \)) estimated by Haymes and Juday [1965].

<table>
<thead>
<tr>
<th>Energy of Line, kev</th>
<th>Photon Flux Limit, ( \text{cm}^{-2} , \text{sec}^{-1} ), 95% confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>( 1.5 \times 10^{-1} ) ( \text{Sun} )</td>
</tr>
<tr>
<td>30</td>
<td>( 1.3 \times 10^{-2} )</td>
</tr>
<tr>
<td>40</td>
<td>( 1.0 \times 10^{-2} )</td>
</tr>
<tr>
<td>60</td>
<td>( 1.3 \times 10^{-2} )</td>
</tr>
<tr>
<td>100</td>
<td>( 3.3 \times 10^{-2} )</td>
</tr>
<tr>
<td>150</td>
<td>( 1.0 \times 10^{-1} )</td>
</tr>
</tbody>
</table>
Fig. 2. The counting rate due to energy losses in the range 26–45 kev as a function of geomagnetic azimuth for two periods of observation at 4.5-g/cm² atmospheric depth; Hyderabad, 1965.

1810

EDWARDS AND McCracken

The detector energy resolution was measured as 42% at 30 kev. In the flight package the photomultiplier pulses were subjected to pulse height analysis, the counting rates in five contiguous differential channels between 20 and 58 kev and an integral channel for energy losses above 58 kev being continuously telemetered to ground, together with the magnetometer data.

3. RESULTS

Information concerning discrete X-ray sources was extracted from the data in the following manner. The counting rates in each of the six pulse-height analyzer channels, averaged over 10-second intervals, were labeled with the corresponding magnetic azimuth of the telescope. Mean counting rates were then computed as functions of azimuth for each successive 90-minute period of observation.

Figure 2 presents a section of the data obtained in this way for the period 0830–1130 LMT. Four of the channel counting rates have been summed to provide the counting rate in the energy loss range 26–45 kev. The peak in the lower histogram demonstrates the existence of strong X-ray sources in the constellation Cygnus. [McCracken, 1965; McCracken and Edwards, 1966]. The upper histogram shows a smaller peak shifted to the west as is to be expected for a source that has moved further west of the meridian and to lower elevations.

No significant solar component is apparent, and it is clearly evident that the solar flux was less than that from Cygnus. Since no signal from Jupiter is evident (Figure 2) a similar statement may be made about the X-ray flux from this planet. After subtracting the background counting rate and applying appropriate corrections for atmospheric absorption, the integral photon flux over the energy range 26–45 kev from the Cygnus sources was calculated to be 0.12 ± 0.01 photon/cm² sec on April 2, 1965.

When the data obtained between 1000 and 1430 IST in each of the 6 energy loss channels were examined for a solar component, the counting rates at those azimuths for which the detector was viewing the sun were found to be less than 0.7σ above the background rates in all energy ranges. An upper limit to the solar flux was therefore set by assuming the minimum detectable flux to be that which should result in the accumulation of 3σ counts above background in any channel. The probability of detecting a flux of this magnitude in the presence of background fluctuations therefore exceeds 0.95.

The differential pulse height spectrum $S(H) dH$ (the counting rate in the pulse height interval $dH$ at $H$) was related to the photon energy spectrum $j(E) dE$ of a discrete source by

$$S(H) = \frac{\hat{A}(\theta)}{(2\pi)^{1/2}} \int_{E_s}^{E} \frac{\epsilon(E) \cdot j(E)}{\sigma(E) \cdot \exp \left( - \left( \mu(E)z \sec \theta + \mu'(E)z' \right) \right)} \cdot \sec (\theta - \theta_s) + \frac{(H - E)^2}{2\sigma(E)^2} \; dE$$

where $\hat{A}(\theta)$ is the projected area of the telescope at zenith angle $\theta$, averaged over its azimuthal field of view; $\epsilon(E)$ is the efficiency of the detector (taken as unity); $\mu(E)$, $\mu'(E)$ are the narrow-beam attenuation coefficients of air and the crystal mounting, respectively; $\theta, \theta_s$ are the source and telescope zeniths, respectively; and...
THE JOVIAN X-RAY FLUX

No observations of a Jovian X-ray flux have been published in the literature. The most likely source of X radiation would appear to be bremsstrahlung from the low-energy electrons assumed responsible for the decametric radio emission [Ellis, 1963; Warwick, 1964]. Order-of-magnitude calculations suggest a hard X-ray flux at the earth of $10^{-8}$ photons cm$^{-2}$ sec$^{-1}$ for a mean electron lifetime of $10^8$ sec. This is at least three orders of magnitude below the threshold sensitivity of our equipment. However, considering the present state of knowledge of the Jovian radiation belts, the present results are of value in that they set a limit to certain parameters of the Jovian energetic particle environment through a physical process completely different from that previously open to observation. Thus, for example, writing the energy content of electrons of energy $\lesssim 50$ keV in the radiation belts as $J$ and assuming a mean energy of 100 keV we can set an upper limit to the ratio $J/\tau$, where $\tau$ is the average electron lifetime, of

$$\frac{J}{\tau} \lesssim 10^{24} \text{ erg sec}^{-1}$$

This is then an upper limit to the power input required to maintain an equilibrium-trapped flux ($E > 50$ keV) of electrons in the Jovian belts.

Acknowledgments. The Indian balloon flights were made as part of the EQUEX expedition coordinated and implemented by the National Science Foundation under the able guidance of Dr. B. Stiller and Messrs. R. Kubara, and T. Pappas. The assistance of Miss M. Steinbock and Messrs. J. Corwin and C. Shelton in various phases of the investigation are gratefully acknowledged. The Australian flight was implemented by the Australian Department of Supply launching crew under the direction of Messrs. E. Curwood and D. Scott. This research was supported by National Aeronautics and Space Administration under contract NASr-198. The Australian flight was supported by U.S. AEC and NASA.

REFERENCES


McCacken, K. G., and P. J. Edwards, X-ray emission by the Cygnus XR-1 X-ray emitter in the wavelength range 0.3A to 0.8A, COSPAR, 1966.


(Received August 6, 1966; revised December 22, 1966.)
X-RAY SURVEYS OF THE QUIET SUN, THE MAGELLANIC CLOUDS AND THE CYGNUS REGION

P. J. EDWARDS AND K. G. MCCRACKEN
Department of Physics, University of Adelaide

Balloon borne alt-azimuth mounted X-ray telescopes flown from Hyderabad, India (17°N), Palestine, Texas (32°N), and Mildura, Australia (34°S), have detected hard X-rays (\(h\nu>20\text{KeV}\)) from discrete sources in the Cygnus region.\(^1\,^2\)

Data from the first (April, 1965) of the northern hemisphere flights indicated an X-ray source in the region of Cygnus A and Cygnus XR1. The limited angular resolution of the telescope did not allow confirmation of the NRL rocket observation,\(^3\) also in April, of roughly equal X-ray fluxes from both these objects. A power law, \(j(E) = (E/E_0)^{-1.7}\), for the differential photon energy spectrum of the composite source was derived from the rocket and balloon data.

We flew a second telescope having higher angular resolution six months later in Texas and were able to resolve the flux into two components with different spectra. Examination of the spectral and positional information showed the flux from the direction of Cyg A to be significantly harder than that from the direction of Cyg XR1. Furthermore, the azimuthal intensity profile was appreciably wider than for a single point source, a result consistent with the detection of two separate sources. After correction for atmospheric absorption and telescope geometry, we found the photon flux (\(E>30\text{KeV}\)) to be \(55\pm15\%\) higher than that measured six months earlier in April. The flux measured in a third flight in June 1966 was not significantly different from that in the April flight. Although one cannot definitely rule out instrumental effects, we believe these results provide further evidence for a time variation\(^3\) in the intensity of the Cygnus composite source. If the radiation is of synchrotron origin such variation may reflect fluctuations in the acceleration processes since the lifetime of \(10^{13}\) eV electrons in a \(10^{-2}\) gauss field is only of the order of days.

Upper limits to the hard X-ray flux from a number of celestial objects have also been obtained.\(^4\) The Magellanic Clouds and the radio source Fornax A were surveyed from Mildura in 1965. No radiation was detected from either of these sources. An upper limit of about \(10^{-2}/\text{cm}^2\)-sec-KeV at 35KeV, approximately equal to that from Cygnus A and XR1, may therefore be established for these objects. New upper limits to the flux from the quiet sun and Jupiter deduced from observations made on day 92, 1965, are
The curves show upper limits at the 0.95 confidence level to a line flux at any single energy \(20<E<150\text{KeV}\).

The Zurich sunspot number was zero for the two days prior to the observation and it remained zero for the following week. Three importance 1-flares occurred three days before the flight. Rocket measurements at energies of several KeV suggest a thin plasma spectrum \((T = 4.5 \times 10^6\text{K})\) for the quiet sun X-ray continuum. Extrapolation of these data to higher energies yields an integral flux \((E>20\text{KeV})\), many orders of magnitude below the detection threshold of the balloon borne telescope for such a steep spectrum \((1.5 \times 10^{-1}\text{cm}^{-2}\text{sec}^{-1})\).

On the other hand, observations of the non-flare sun near solar maximum\(^5\) do suggest a detectable flux above 20KeV. Extrapolation of these data gives an integral flux \((E>20\text{KeV})\) of \(\sim 1\text{ cm}^{-2}\text{sec}^{-1}\). This flux would have been easily detected in the present data. We conclude therefore that the hard solar X-ray flux varies by at least one order of magnitude over a solar cycle. This is almost certainly a gross underestimate of the variation.

No observations of a Jovian X-ray flux have been published in the literature. The most likely source of X-radiation would appear to be bremsstrahlung from the low energy electrons assumed responsible for the decametric radio emission.\(^6\) Order of magnitude calculations suggest a hard X-ray flux at the earth of \(10^{-5}\) photons \text{cm}^{-2}\text{sec}^{-1}\) for a mean electron lifetime of \(10^3\text{ sec}\). This is at least three orders of magnitude below the threshold sensitivity of our equipment. However, writing the energy content of electrons of energy \(>50\text{KeV}\) in the radiation belts as \(J\) and assuming a mean energy of \(100\text{KeV}\) we can set an upper limit to the
ratio \( J/\tau \), where \( \tau \) is the average electron lifetime of

\[
\frac{J}{\tau} \leqslant 10^{24} \text{ erg sec}^{-1}
\]

This is an upper limit to the power input required to maintain an equilibrium trapped flux \((E>50\text{KeV})\) of electrons in the Jovian belts.

2 McCracken, K. G., and Edwards, P. J., 'X-ray Emission by the Cygnus XR-1 X-ray Emitter in the Wavelength Range 0.3\AA{} to 0.8\AA{}, *COSPAR* (1966).
Spectral Properties of the X-Ray Objects GX 3+1, GX 354-5 and Sco XR-1

by

G. BUSELLI
M. C. CLANCY
P. J. N. DAVISON
P. J. EDWARDS*
K. G. McCracken
R. M. THOMAS

Department of Physics,
University of Adelaide

The initial flight of a new balloon-borne X-ray observatory was made from Mildura, Australia, on February 29, 1968, to study the properties of a number of celestial objects near the galactic centre. Significant X-ray fluxes were observed from five X-ray objects, and upper limits have been established for four others. It is the purpose of this report to present the spectra observed from three X-ray sources, Sco XR-1, GX 3+1 and GX-5-6; and to present an improved position for the latter object, which we will now designate GX 354-5. Detailed descriptions of the X-ray detectors, and the totality of the results, will be published at a later date.

The X-ray observatory contains two independent X-ray detectors, one similar in principle to the active collimator detector pioneered by Peterson et al., the other basically similar to the graded shield detector developed by Boldt et al. The active collimator telescope comprises a 2 mm thick X-ray detection scintillator, shielded by at least sixteen mean free paths of scintillator (at 100 keV) in all directions other than those within a conical viewing angle of 8°. Considerable pains have been taken to minimize the spurious response of this detector. The pulses from the X-ray detection scintillator are analysed by a 16 channel pulse height analyser, the upper and lower energy limits being 7 and 167 keV and each channel having a width 10 keV. The total sensitive area of this detector is 54-3 cm². The temporal stability of the energy calibration of the complete detection system is monitored throughout the flight by introducing a 152Gd radioactive source into the detector viewing cone for 1 min every 17 min. The graded shield telescope comprises a Pb-Ag-Cu graded shield and a scintillating polyethylene anticoincidence shield, surrounding a 19-0 cm diameter, 3 mm thick CsI(Tl) scintillation detector. The sensitive area is 220 cm². The pulses from this X-ray detector are sorted into four contiguous energy channels set at 30-40, 40-50, 50-60 and 60-80 keV. The graded shield telescope has a rectangular field of view, 10° x 30°.

The two telescopes were inclined to the vertical (by 32° and 43° respectively) and pointed in the same azimuthal direction. Two orthogonal fluxgate magnetometers were used to determine the azimuthal pointing direction of the telescopes during the flight. Pre-flight calibrations, and an in-flight calibration, indicate that all azimuth angles are known to ±1°. By displaying the pointing direction in real time, an operator on the ground was able to control the direction of pointing of the telescope by sending appropriate radio commands to activate a motor which turned the observatory in either direction relative to the balloon. The telescopes were thus scanned back and forth over limited regions of sky known to contain X-ray objects of interest.

X-ray Object Positions

Fig. 1 displays the azimuthal scan obtained by the active collimator telescope for one period of 15 min during which a source previously named Sgr XR-1 (ref. 3) was the nominal object of interest. These data, and similar sets of data accumulated over three other 15 min periods, indicate the presence of two X-ray objects in the general direction reported for Sgr XR-1. The angular separation of these two objects is such that they would not have been
using a least squares technique, in order to obtain the curve was fitted to each set of data as shown in Fig. 1 of the celestial sphere. A theoretical angular response resolved in earlier balloon-borne surveys of this portion of the celestial sphere. A theoretical angular response curve was fitted to each set of data as shown in Fig. 1 using a least squares technique, in order to obtain the azimuth and counting rate of maximum detector response. This procedure was found to provide an adequate fit to the data on the assumption of two X-ray objects only. Furthermore, the motion of the object through the range of zenith angles covered by the detector produced a time curve applicable to the X-ray object 3+1. The best fit curves are from the active collimator results only. In each case, the flux in each energy interval was obtained from a least squares fit of the theoretical response curve to the observed data, the counting rate increment so obtained then being corrected for detector efficiency, atmospheric absorption and foreshortening. Contemporary rocket results for energies in the vicinity of 4-14 keV are also presented. In Figs. 3 and 5, $\lambda$ is the index of the power law spectrum

\[ \frac{dN}{dE} = A E^\lambda \]

and $kT$ is the characteristic energy in the exponential spectrum

\[ E \frac{dN}{dE} = B \exp\left(\frac{-E}{kT}\right) \]

It is significant that we find no appreciable flux from GX 5-1 (ref. 4), an object which is brighter than GX 3+1 at photon energies $\approx 4$ keV. This indicates that GX 5-1 has a markedly softer spectrum than GX 3+1. Fig. 1 also displays the azimuthal plot of counting rate about the nominal direction of Sco XR-1, an X-ray object the position of which is known with high accuracy. These data indicate that a $2^\circ \pm 1^\circ$ correction is necessary in our various azimuth data, and this correction has been applied in all other directional data reported here.

**X-ray Object Spectra**

The spectra for GX 3+1, GX 354-5 and Sco XR-1 are shown in Figs. 3, 4 and 5, respectively. Figs. 3 and 5 include data from the graded shield detector; no results are available from this detector on the spectrum of GX 354-5, which could not be resolved from background. The best fit curves are from the active collimator results only. In each case, the flux in each energy interval was obtained from a least squares fit of the theoretical response curve to the observed data, the counting rate increment so obtained then being corrected for detector efficiency, atmospheric absorption and foreshortening. Contemporary rocket results for energies in the vicinity of 4-14 keV are also presented. In Figs. 3 and 5, $\lambda$ is the index of the power law spectrum

\[ \frac{dN}{dE} = A E^\lambda \]

and $kT$ is the characteristic energy in the exponential spectrum

\[ E \frac{dN}{dE} = B \exp\left(\frac{-E}{kT}\right) \]
Considering GX 3+1 in Fig. 3, the data for \( h\nu \geq 20 \) keV show a good fit to a power law photon number spectrum given by

\[
\frac{dN}{dE} = 3.4 \times 10^{-2} E^{-2.5} \text{ photons cm}^{-2} \text{s}^{-1} \text{keV}^{-1}
\]

which, when extrapolated to lower energies as shown in Fig. 3, is in reasonable agreement with the results of Gursky et al. and Bradt et al. Furthermore, we note that previous workers have also obtained power law photon number spectra for Tau XR-1 and Cyg XR-1; for example, Peterson et al. quotes

\[
\frac{dN}{dE} = \begin{cases} 
3.50 \times 10^{-1.93} & \text{for Tau XR-1} \\
3.58 \times 10^{-1.93} & \text{for Cyg XR-1}
\end{cases}
\]

at low photon energies. The thin plasma spectrum derived by Peterson and Jacobson with \( kT = 14.8 \) keV fails to agree with the high energy portion of the X-ray spectrum, as was, in fact, pointed out by Peterson and Jacobson themselves. We conclude that the best single fit to the observed data over the energy range 5 < \( h\nu < 70 \) keV is obtained using a power law spectrum, of exponent \(-3.79 \pm 0.36\). That is, while the spectrum is much steeper

The spectral exponents in the above three spectra are statistically identical, while there are other objects known to exhibit markedly different spectral characteristics (Sco XR-1, for example, in Fig. 5). This suggests that the three X-ray objects GX 3+1, Tau XR-1 and Cyg XR-1 are physically similar to one another. The fact that Tau XR-1 is known to be a supernova remnant, and the spectral shape itself suggesting that the X-ray emission from Tau XR-1 may be due to magnetic bremsstrahlung, suggests a similar explanation for GX 3+1 and Cyg XR-1. Because of this it seems desirable to determine the radio emission from both these objects, for a magnetic bremsstrahlung origin of the X-rays would imply magnetic bremsstrahlung at radio wavelengths.

The spectral data for Sco XR-1 in the energy range \( h\nu \geq 20 \) keV are shown in Fig. 5, and can be seen to be in good agreement with the power law photon number spectrum

\[
\frac{dN}{dE} = \left(1.08 - 0.21 + 0.27\right) \times 10^{-4} \left(\frac{E}{20}\right)^{-3.79 \pm 0.36} \text{ photons cm}^{-2} \text{s}^{-1} \text{keV}^{-1}
\]

This spectral law is in reasonable agreement with the data for lower energies. By way of contrast, however, an adequate fit to the \( h\nu \geq 20 \) keV data can be obtained by a thin plasma spectrum (\( kT = 14.8 \) keV). Such a spectrum, however, is in error by more than an order of magnitude
than those applicable to Tau XR-1, Cyg XR-1 and GX 3+1, it still suggests a magnetic bremsstrahlung origin for the X-radiation.

This work was supported by NASA, the Australian Research Grants Committee, the Air Force Office of Scientific Research of the United States, and the US Atomic Energy Commission. We thank the Australian Department of Supply balloon launch facility, and, in particular, Messrs E. Curwood and D. Scott for their help.

Received July 17, 1968.


ERRATUM. A mistake occurred in the article "Spectral Properties of the X-Ray Objects GX 3+1, GX 354-5 and Sco XR-1" by G. Buselli et al. (Nature, 219, 1124; 1968), because of an inadvertent alteration made in the Nature office to the original typescript. It was stated in the article that the spectra of GX 3+1, GX 354-5 and Sco XR-1 suggest that these three X-ray objects are similar to one another. The data are at variance with this statement, and in fact the authors' conclusion is that GX 3+1, Cyg XR-1 and Tau XR-1 possess similar spectra, and are probably due to similar celestial objects. One of the authors, K. G. McCracken, has written to stress how this conclusion was reached:

"Our measurements of GX 3+1 above 20 keV show a good fit to a power law photon number spectrum given by

$$\frac{dN}{dE} = 3.4 \times 10^{-2} \pm 0.4 \text{ photons cm}^{-2} \text{s}^{-1} \text{keV}^{-1}$$

which, when extrapolated to lower energies, is in reasonable agreement with the results of Gursky et al. (Ap. J., 150, L75; 1967) and Bradt et al. (Ap. J., 152, 1005; 1968). Furthermore, we note that other workers have obtained power law photon number spectra for Tau XR-1 and Cyg XR-1; for example, Peterson et al. (Proc. Tenth Intern. Conf. Cosmic Rays, Calgary; 1968) quote

$$\frac{dN}{dE} = (3.50 \pm 0.2) \times 10^{-2} \text{ photons cm}^{-2} \text{s}^{-1} \text{keV}^{-1}$$

The spectral exponents in the spectra of GX 3+1, Tau XR-1 and Cyg XR-1 are statistically identical, while there are other objects known to exhibit markedly different spectral characteristics (our paper shows that Sco XR-1, for example, has \( \frac{dN}{dE} \sim \frac{1}{E^{3.6\pm0.4}} \) above 20 keV). This suggests that the three X-ray objects GX 3+1, Tau XR-1 and Cyg XR-1 are physically similar to one another. The fact that Tau XR-1 is known to be a supernova remnant, the spectral shape itself suggesting that the X-ray emission from Tau XR-1 may be due to magnetic bremsstrahlung, suggests a similar explanation for GX 3+1 and Cyg XR-1."
Thermal Models of Centaurus XR-2

P. J. Edwards and J. R. Harries
Department of Physics, University of Adelaide

Recent observations of the Centaurus region by Harries et al., Chodil et al., and Cooke et al. have firmly established the existence of a variable X-ray source, Centaurus XR-2. This source is unique in that its variability has been established beyond doubt.

Francey et al. compare their data from two rocket flights in April 1967 with those of Chodil et al. obtained 44 days later in May. They find the data to be consistent with monotonic cooling of a thin plasma source from $4.2 \pm 0.6 \times 10^7$ K to $1.8 \times 10^7$ K (see Figure 1), accompanied by a fivefold decrease in the (2-8) keV energy flux during this period. Data presented by Cooke et al. suggest a somewhat higher flux decrease.

Figure 1. The measured Centaurus XR-2 temperatures for the two extreme assumptions of an optically thin hot plasma and a black-body source. The plasma cooling curve corresponds to the shock-wave model while the black-body curve is for cooling by radiation alone.

BLACK-BODY MODELS

The assumptions of an optically thin plasma source rest solely on the detailed energy spectrum obtained by the LRL group on May 18 (day 138). Inspection of this spectrum shows it to be reasonably well fitted by a black-body curve with a temperature of $0.85 \pm 0.05 \times 10^7$ K. The X-ray
spectra obtained for the three flights are consistent with a cooling black body of constant dimensions.

The measured black-body colour-temperatures are shown on Figure 1 together with the cooling curve expected for a black body cooling by radiation alone. Bahcall and Wolf have calculated the cooling of a neutron star as a result of the emission of neutrinos produced in nucleon-nucleon processes. In Figure 2 we compare the measured temperatures with the cooling curve calculated from the neutron star model, taking into account the expected gravitational red shift and normalizing the curve to the observed temperature on 1967 April 4. It is evident that, although the X-ray source cooled faster than the Bahcall model, the data do not allow the neutron star hypothesis to be rejected. It seems likely however that optical inspection of the Centaurus region both before and after the X-ray measurements would enable stringent restrictions to be placed on the nature of the X-ray object. Catastrophic origin of a neutron star in a contemporary supernova can obviously be discounted since the flux and temperature measurements place such a source at a distance of rather less than 1 kpc. In passing we notice that the apparent position of the source lies within about 3° of the non-thermal radio source.
13S6A considered to be the remnant of the AD185 supernova, situated, on the basis of its initial luminosity, at about the same distance as the hypothetical neutron star.

**THIN PLASMA MODEL**

The assumption of an isothermal, optically thin plasma as the source of the radiation leads to a model of the cooling plasma which appears to be satisfied by a shock-heated circumstellar gas envelope of a type which might be expected to follow a nova outburst.

At the observed temperatures, the radiation originates principally in electron-proton bremsstrahlung. The spectral intensity $j_E$ at distance $d$ from an isothermal thin plasma of volume $V$ containing $n_e$ electrons/cm$^3$ is then given by

$$j_E d d \sim 1 \times 10^{-28} \left(\frac{n_e^2 V}{d^2}\right) T^{1/2} \left(\frac{e^{-E/kT}}{kT}\right) dE \text{ erg cm}^{-2} \text{ s}^{-1}.$$ 

Using this relation we find that $(\frac{n_e^2 V}{d^2})$, initially $\sim 5 \times 10^{17}$ cm$^{-5}$, had decreased by a factor of $\sim 2$ at the time of the LRL rocket flight, 44 days ($4 \times 10^6$ s) later. At these high temperatures it appears highly probable that the expansion was explosive and that the X-radiation originated in a shell of shock-heated gas.

The gas temperature 1.5 to $4 \times 10^7$°K limits the radial expansion velocity to $v \sim 10^8$ cm/s. Also, the mass conserving $(n_e V = \text{constant})$ volume change sets a radial expansion $\Delta R/R \sim 1$ over a period of $4 \times 10^8$ s. When first seen on April 4, the source was therefore no more than a month or so old, having expanded to a radius $R_1$ of about $5 \times 10^{14}$ cm (30 A.U.). An upper limit to the mean electron density of $n_e \leq 5 \times 10^8$ cm$^{-3}$ then follows from the extreme assumption that the plasma cooled by radiation alone. Taking $n_e = 10^8$ cm$^{-3}$ we find a total mass ($\sim n_e m_H V$) of about $10^{27}$ g ($5 \times 10^{-5} m_0$) and an emission measure $(n_e^2 V) \sim 4 \times 10^{60}$ cm$^{-3}$ on April 4. The thermal energy content, $3n_e kT V \sim 10^{45}$ erg and luminosity $L \sim 10^{-27} T^{1/2} n_e^2 V \sim 10^{37}$ erg/s give a radiative lifetime of about a year and the distance to the source (from $n_e^2 V/d^2 \sim 5 \times 10^{17}$) becomes $d \sim 1$ kpc. This distance may be regarded as an upper limit, since there is no need to invoke significant radiative cooling, and the electron density may well be lower than $10^8$ cm$^{-3}$. 
These parameters are typical of those for recurrent novae deduced by Wallerstein\textsuperscript{6} from coronal line measurements. We may use the blast wave solutions of Rogers\textsuperscript{7} in order to find a more detailed model of the expansion. The cooling data for Centaurus XR-2 show that any such model must conserve the quantity \((n_e^2V/T)\) during the expansion. This condition is satisfied for an adiabatic strong shock which propagated radially outward through a non-uniform gas envelope in which the density,

\[ n \propto r^{-\alpha} \text{ with } \alpha = 2. \]

With this solution, the density behind the shock is a linear function of the radial distance from the origin of the explosion, increasing to \((\gamma + 1)/(\gamma - 1) = 4\) times the pre-shock envelope density at the shock front. The gas velocity similarly increases linearly with radial distance up to the shock boundary. The emission measure \(n_e^2dV\) falls off rapidly behind the shock front and it can be shown that the effective X-radiator is a nearly isothermal shell in which

\[ n^2V \propto T \propto R^{-1}t^{-2/3}. \]

Since the (2-8) keV energy flux suffered a sixfold decrease and the temperature a twofold decrease in \(4 \times 10^6\) s, the shock wave would therefore have been \(\sim 2\) weeks old when first seen by the UAT group. The shock wave temperature at that time determines an expansion velocity \(v \sim 2 \times 10^8\) cm/s. The corresponding shock radius, \(R = \lambda tv \sim 3 \times 10^{14}\) cm. This would have increased to \(\sim 8 \times 10^{14}\) cm when the source was seen by Chodil et al.\textsuperscript{2}

If the source is an average nova at this distance we might expect it to have become brighter than 4m in March 1967. Some additional support for the nova hypothesis comes from the optically observed deceleration of nova envelopes\textsuperscript{8} in which \(v^2 \propto 1/R\) as in the present model of the X-ray source.

Nova Model of Centaurus XR-2

Recent observations of the Centaurus region\(^1\)–\(^3\) have firmly established the existence of a variable X-ray source, Centaurus XR-2. This source is unique in that its variability has been established beyond doubt\(^2\)–\(^4\).

Francey et al.\(^4\) compared their data from two rocket flights in April 1967 with those of Chodil et al.\(^2\) obtained in May, 44 days later. They found the data to be consistent with monotonic cooling of a thin plasma source from \(4.2 \pm 0.6 \times 10^7 \text{ °K}\) to \(1.8 \times 10^7 \text{ °K}\), accompanied by a five-fold decrease in the 2–8 keV energy flux during this period. Data presented by Cooke et al.\(^3\) suggest a somewhat faster decrease.

These observations, together with the assumption (supported by the detailed energy spectrum of Chodil et al.\(^2\)) of an optically thin X-ray source, suggest a model of the cooling X-ray source which is satisfied by the physical conditions expected\(^1\)–\(^6\) in a circumstellar envelope shortly after a nova outburst.

At the observed temperatures, the radiation originates principally in electron proton bremsstrahlung. The spectral intensity, \(j_E\), at distance \(d\) from an isothermal, thin plasma of volume \(V\), containing \(n_e\) electrons/cm\(^2\), is then given by\(^7\)–\(^8\)

\[
j_E \, \text{d}E \sim \left(1 \times 10^{-28}\right) \left(\frac{n_e^2 V}{d^2}\right) T^{\frac{3}{2}} \left(\frac{e^{-E/kT}}{kT}\right) \, \text{d}E \, \text{ergs/cm}^2 \, \text{sec}
\]

where we take the gas to have the normal chemical abundances. Using this relation we find that \((n_e^2 V/d^2)\), initially about \(5 \times 10^{17} \text{ cm}^{-5}\), had decreased by a factor of about 2 at the time of the Lawrence Research Laboratory rocket flight\(^2\) 44 days (4 \times 10^6 sec) later. Assuming that the mass of radiating gas was conserved during this period, this at once implies a polytropic gas expansion; \(TV^{\gamma-1}\) equal to a constant, with exponent \(\gamma\) equal to \(2\)–\(1.5\). Although this is consistent with steady adiabatic expansion (\(\gamma = 5/3\)), isothermal motion (\(\gamma = 1\)), as proposed by Manley\(^9\), is clearly at variance with the observations. Tucker\(^4\) has pointed out the difficulty of maintaining steady gas expansion through gravitational constraint at these high temperatures. It seems highly probable therefore that the expansion was explosive and that the X-radiation originated in a shell of shock heated gas. Before showing that the time decay of the source can be explained naturally in terms of an adiabatic blast wave propagating radially out through a circumstellar gas envelope, we note the following orders of magnitude.
The gas temperature, \((1.5-4) \times 10^7 \, ^\circ \text{K}\), limits the radial expansion velocity to \(v\) about \(10^8 \, \text{cm/sec}\). Also, the mass conserving \((n_e V = \text{constant})\) volume change sets a radial expansion \(\Delta R/R\) of order one over a period of \(4 \times 10^6 \, \text{sec}\). When first seen on April 4, the source was therefore no more than a month or so old, having expanded to a radius \(R_1\) of about \(5 \times 10^{14} \, \text{cm}\) (30 astronomical units). An upper limit to the mean electron density of \(n_e \leq 5 \times 10^8 \, \text{cm}^{-3}\) then follows from the extreme assumption that the plasma cooled by radiation alone. Taking \(n_e = 10^8 \, \text{cm}^{-3}\), we find a total mass \((n_e m_H V)\) of about \(10^{29} \, \text{g} \, (5 \times 10^{-5} \, m_\odot)\) and an emission measure, \(n_e^2 V\), of about \(4 \times 10^{16} \, \text{cm}^{-3}\) on April 4. The thermal energy content, \(3n_e k T V\), circa \(10^{45} \, \text{ergs}\), and luminosity, \(L \sim 10^{-27} n_e^2 V\), circa \(10^{37} \, \text{ergs/sec}\), give a radiative life time of about a year and the distance to the source (from \(n_e^2 V/d^2\), \(\sim 5 \times 10^{17}\)) becomes about 1 kparsec. This distance may be regarded as an upper limit because there is no need to invoke significant radiative cooling, and the electron density may well be considerably lower than \(10^8 \, \text{cm}^{-3}\).

These parameters are typical of those deduced for recurrent novae by Wallerstein from coronal line measurements. We now use the blast wave solutions of Rogers in order to find a more detailed model of the expansion. The cooling data for Centaurus XR-2 show that any such model must conserve the quantity \((n_e^2 V/T)\) during the expansion. This condition is satisfied for an adiabatic strong shock which propagated radially outward through a non-uniform gas envelope in which the density

\[ n \propto r^{-\alpha} \text{ with } \alpha = 2 \]

The shock wave solutions give the radius, \(R\), of the shock front at time \(t\) as

\[ R = kt^{1/\lambda} \]

and the velocity of the shock as

\[ v = R/\lambda t \]

where \(\lambda\) is an expansion parameter which for an adiabatic shock is related to \(\alpha\) by

\[ 2\lambda = 5 - \alpha \]

Because \(n_e^2 \propto R^{-2\alpha}, V \propto R^3\) and \(T \propto v^2 \propto R^{2(1-\alpha)}\), we have \(n_e^2 V \propto T\) as required from the data, providing that

\[ 2\lambda = 2\alpha - 1 \]

Thus

\[ \alpha = 2 \]

and

\[ \lambda = 3/2 \]
With this solution, the density behind the shock is a linear function of the radial distance from the origin of the explosion, increasing to \((\gamma + 1/\gamma - 1) = 4\) times the pre-shock envelope density at the shock front. The gas velocity similarly increases linearly with radial distance up to the shock boundary. The emission measure, \(n^2 dV\), falls off rapidly behind the shock front (as \((r/R)^4\)), and we may take the effective X-radiator to be a nearly isothermal shell of thickness \(\Delta R\), about \(R/4\), and volume \(2.5R^3\), having the density, temperature and radius of the front.

Because the 2–8 keV energy flux has suffered a six-fold decrease and the temperature, \(T\), a 2.3-fold decrease in \(4 \times 10^8\) sec, we find from \(n^2 V \propto T \propto R^{-1}(t^{-2/3})\) that the shock wave would have been about 2 weeks old when first seen by the Universities of Adelaide and Tasmania group. The shock wave temperature at that time determines an expansion velocity \(v\) of about \(2 \times 10^8\) cm/sec and a corresponding shock radius of \(R = \lambda tv\), about \(3 \times 10^{14}\) cm. This would have increased to about \(8 \times 10^{14}\) cm when the source was seen by Chodil et al.\(^2\). The distance to the source becomes \(d\) about 300 parsec for \(n_e = 10^8\) cm\(^{-3}\) and the energy of the blast, \(E\) about \(10^{44}\) ergs.

If the source is indeed an average nova at this distance, we might expect it to have reached a brightness greater than \(4m\) in March or April 1967. Some additional support for the nova hypothesis comes from the optically observed deceleration of nova envelopes\(^11\) in which \(v^2 \propto 1/R\) as in the present model of the X-ray source.

In conclusion we note that the total X-ray luminosity decreases faster than \(t^{-1}\) even without radiative cooling, and that, \(10^7\) sec after the initial observation, the energy flux above 2 keV would have decreased by nearly two orders of magnitude.

I thank J. Harries, R. Francey and Professor K. G. McCracken for useful discussions.

P. J. EDWARDS

Department of Physics,
University of Adelaide,
Australia.

Received November 23, 1967.

Galactic X-Rays from Unresolved Flare Stars

Hudson et al.\textsuperscript{1} have suggested the existence of a class of unresolved low intrinsic luminosity X-ray sources to account for their own observations and those of Cooke et al.\textsuperscript{2} of a galactic disk component of diffuse X-radiation. I shall show that X-ray bursts from red dwarf flare stars may generate a time averaged disk intensity above 1.5 keV, which, conservatively estimated, satisfies the observational requirements, and that current estimates of the luminosity\textsuperscript{3,4} of individual flare X-ray bursts imply a disk intensity substantially greater than that observed.

Lovell\textsuperscript{5} has also suggested that X-rays from flare stars contribute a few per cent of the observed diffuse X-ray flux and has revived Unsold's suggestion\textsuperscript{6} that a substantial part of the galactic cosmic ray flux has its origin in stellar flares. This latter possibility has been re-examined by Edwards and McQueen\textsuperscript{7} who concluded that the cosmic ray contribution from flare stars is negligible.

Grindlay\textsuperscript{3} has calculated a non-thermal bremsstrahlung X-ray luminosity (\(>10\) keV) of \(3.5 \times 10^{31}\) erg s\(^{-1}\) during a typical one magnitude optical flare lasting several minutes on UV Ceti. This is somewhat greater than the optical flare luminosity for such a flare as calculated by Kunkel\textsuperscript{8}. Kunkel\textsuperscript{9} has also shown the existence of a well defined upper limit to the time-averaged optical flare luminosity, \(L_0\), equal to 1\% of the quiescent photospheric luminosity for individual flare stars with a wide age distribution in the solar vicinity. For UV Ceti the average optical flare luminosity \(L_0 \sim 5 \times 10^{28}\) erg s\(^{-1}\) (0.5\% of quiescent luminosity) and is close to the average \(\langle L_0 \rangle\) for all stars in Kunkel's survey. The volume emissivity of active solar neighbourhood flare stars with space density\textsuperscript{10} \(\langle n \rangle = 0.3\) pc\(^{-3}\) is therefore \(\langle n \rangle \langle L_0 \rangle \sim 10^{27}\) erg s\(^{-1}\) pc\(^{-3}\). Consideration of the luminosity, age and mass distributions of fully convective flare stars of spectral type later than KO by Edwards and McQueen\textsuperscript{7} leads to a similar estimate of the optical flare emissivity in the disk.

On the basis of Grindlay's calculation of the individual flare burst luminosity, and using Kunkel's optical data, \(\langle L_0 \rangle \sim \langle L_x \rangle (>10\) keV\) and the total hard X-ray emission from the galactic disk of volume \(V_D\) by \(\langle n \rangle V_D \sim 10^{10}\) stars is therefore \(\langle n \rangle \langle L_x \rangle V_D \sim 5 \times 10^{38}\) erg s\(^{-1}\), a result which
is roughly two orders of magnitude greater than the upper limit deduced from the OSO III data. It is also one order of magnitude greater than the total galactic emission in soft X-rays (<12 keV) required of the hypothetical class of unresolved sources by Hudson et al. to account for the low energy disk component. I conclude, therefore, that the hard X-ray luminosity of flare stars has been substantially overestimated. An upper limit to the volume emissivity in the low energy range (1.5 keV–12 keV) of ~10^{26} erg s^{-1} pc^{-3} may also be set from the data. Since the observations are probably contaminated by the flux from a few strong sources a more realistic upper limit may be 10^{25} erg s^{-1} pc^{-3}. This immediately sets an upper limit to the ratio

\[ \frac{\langle L_x (1.5 \text{ keV}) \rangle}{\langle L_0 \rangle} = \frac{\langle n \rangle \langle L_x (1.5 \text{ keV}) \rangle}{\langle n \rangle \langle L_0 \rangle} = 10^{-2} \]

That is, the energy radiated in soft X-rays is less than 1% of the optical flare energy. Because a value of 10% is quoted for solar flares, 1% seems a conservative figure for stellar flares.

Although uncertainties in the flare star parameters \langle n \rangle and \langle L_x \rangle should not be underestimated, it seems difficult to avoid the conclusion that these objects of the disk and spiral arm populations with a field density of

\[ \frac{\langle n \rangle R^3}{3} \sim 10^{10} \text{ stars - sr}^{-1} \]

(where \( R \) is the line of sight distance through the disk) are likely candidates for the class of unresolved weak sources suggested by Hudson et al. Their time integrated intensity, which one might expect to be related to the surface density of young stars in a given direction, should, from the discussion above, using an average absorption mean free path of \( \lambda_x = 5 \text{ kpc} \), typically be

\[ J_x (> 1.5 \text{ keV}) = \langle n \rangle \langle L_x \rangle \lambda_x / 4\pi \sim 10^{-9} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1} \]

with the steep spectra and Fe emission lines characteristic of solar flares.

This work was supported by the New Zealand University Grants Committee.

P. J. Edwards

*Physics Department,*  
*University of Otago,*  
*Dunedin*

Received September 27; revised November 10, 1971.
Recent observations of solar and stellar flares are used in a reexamination of the proposal due to Unsöld (1957) that flare stars may make a substantial contribution to the galactic cosmic radiation. Flares in young convective stars of spectral type later than KO are considered. Observations of the X-ray brightness of the galactic disc indicate a negligible cosmic ray contribution if the energetic charged particle output relative to the X-ray output, is the same in stellar flares as in solar flares.

1. Historical Introduction. Flare stars were proposed as cosmic ray sources on the basis of observations of radio and energetic charged particle emission from solar flares. Unsöld (1957) suggested that flare stars could account for both the non-thermal galactic radio flux and galactic cosmic rays providing they were generated in the same ratio as in large "white light" solar flares. Lovell (1963) observed radio emission concurrent with optical flares in stars of UV Ceti type and was able to show the time-averaged flare star contribution to the 240 MHz brightness temperature of the anti-centre to be less than 0.5%. This upper limit was obtained by taking a uniform space density of 0.18 stars (pc)$^{-3}$ equal to that then estimated for red dwarfs of the solar neighbourhood, and assuming all these were UV Ceti type flare stars. A count of known flare stars within 5 pc of the sun actually gives an order of magnitude lower space density. However, more recent UV Ceti observations (Petit 1970), indicate a mean occurrence rate of $\sim$ 8 flares/day (with $\Delta$B $> 0.5$) compared with Lovell's estimate of only 0.7/day, so that the temperature estimated in this way is not significantly reduced.

Morrison (1961), Johnson (1961), Ginzburg and Syrovatskii (1964) mention flare stars as possible cosmic ray sources. These authors have generally conjectured that the contribution to the galactic cosmic ray flux is not likely to be significant. This conclusion follows from Unsöld's original hypothesis in view of the minor calculated contribution to the radio sky brightness. However, uncertainties in the disc-averaged space density and in the time-averaged charged particle output from stellar flares have precluded any firm conclusions and suggestions are still made, for example, Lovell (1970), of relativistic cosmic ray origin in these events.
2. Stellar Flare Observations (Optical). We shall use the term "flare star" to describe those eruptive variables which exhibit solar flare-like activity in both radio and optical emission. Flare stars then comprise (1) Solar neighbourhood stars of which the emission red dwarf UV Ceti (dm 6e) is the type star. (2) Eruptive variables in stellar clusters and associations such as the Pleiades, Hyades, Orion nebula, NCG 2264. These stars are all of spectral type later than KO (Haro 1964) consistent with their being young, fully convective stars of the disc and spiral arm populations (Schatzmann 1962, Poveda 1964, Haro 1968).

The observed optical flares are similar in temporal and spectral behaviour to solar white light flares but generate greater optical outputs, in the range $10^{32} - 10^{37}$ erg/flare (Kunkel 1970, Slee and Higgins 1970). Active solar neighbourhood stars flare at rates between 1 - 10/day (Petit 1970). Although flare rates are highest in the faintest stars (Petit 1970), Kunkel (1970) has shown that there exists a well defined upper limit to the time-averaged optical flare output to 1% of the quiescent photospheric luminosity for stars with ages ranging from $10^4 - 10^8$ years in the solar neighbourhood. We calculate the average flare luminosity of UV Ceti $L_o = 4 \pm 1 \times 10^{28}$ erg/s ($\sim 0.5\%$ of quiescent luminosity) to be not significantly different from the average for all stars in Kunkel's 1970 survey. The volume emissivity of active solar neighbourhood flare stars with space density $n \sim 0.03$ (pc)$^{-3}$ is therefore $n < L_o > \sim 10^{27}$ erg/s-pc$^3$. The emissivity in stellar associations is substantially higher. For example the estimates 700 flare stars in the Pleiades (Ambartsumian 1970) of age $2 \times 10^7$ years (Haro 1968) generate $\sim 10^{35}$ erg/s-pc$^3$ in optical flare radiation if the ratio $< L_o (flare)/ L_o (quiescent) > = 5 \times 10^{-2}$ as in the solar vicinity.

3. Stellar Flare Observations (Radio). From early observations (Lovell Whipple and Solomon 1963, 1964) Lovell (1964) compared estimates of the total optical/total radio emission for the sun and several flare stars. Lovell found $E_o/E_R$ of order $10^2 - 10^3$ for flare stars, $10^4 - 10^6$ for the sun. Kunkel (1970) has recalculated the optical luminosities for these and other flares and finds order of magnitude agreement between solar and flare star luminosity ratios. These estimates of the total radio emission have been made by extrapolating the observed metric flux to higher and lower frequencies. However, since most of the flare star measurements have been made in the 100 MHz - 500 MHz range, we feel that it is more reasonable to compare integrated radio outputs for the sun and flare stars between these frequencies. Table I shows the energy outputs for solar and stellar flares and also the ratio of the energy in the optical spectrum to that in the radio spectrum between 100 and 500 MHz. It can be seen that the relative energy output in this frequency range is about four orders of magnitude higher for flare stars than for the sun.
In solar flares several orders of magnitude more energy is emitted in the microwave region than in the metric region. From the null measurements of Higgins et al (1968) at 2650 MHz it would seem that this is not the case in flare stars. However, if stellar flares have any similarity to solar flares a great variation between radio events is to be expected. This question which is clearly relevant to energetic particle acceleration cannot be fully resolved until more microwave observations are made of flare stars.

We find \( \langle E_0/E_R \rangle_{\text{SG}} \sim 10^5 \), well below the spread of values for the sun. We conclude therefore that \( \langle E_0/E_R \rangle_{\text{SG}} \ll \langle E_0/E_R \rangle_{\text{sun}} \) for \( E_0 > 10^{32} \text{ erg} \), a result which indicates more efficient radio mechanisms on flare stars than on the sun. Flare star brightness temperatures as high as \( 10^{20} \text{ K} \) (Slee and Higgins 1970) indicate coherent radio emission.

Grindlay (1970) has calculated the non-thermal bremsstrahlung X-ray flux for UV Ceti flares using the electron energy spectrum (\( \gamma = 4-5 \)) indicated by the radio data. His model predicts an integral (> 10 keV) flux of \( 4 \times 10^{-8} \text{ erg/cm}^2\text{-s} \) during a 1 magnitude flare. This implies a mean X-ray luminosity, \( L_x \sim 10^{23} \text{ erg/s}, \) rather greater than the mean optical flare luminosity. The hard X-ray flux in a large solar flare is only \( \sim 1\% \) of the optical flux, the difference being due, in the context of this model, to the

<table>
<thead>
<tr>
<th>Star</th>
<th>Energy Output (ergs)</th>
<th>Energy Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Optical 100 MHz</td>
<td>Radio 500 MHz</td>
</tr>
<tr>
<td>Orion</td>
<td>( 3 \times 10^{37} )</td>
<td>( 1.3 \times 10^{32} )</td>
</tr>
<tr>
<td>YZ CMi</td>
<td>( 4 \times 10^{34} )</td>
<td>( 7 \times 10^{29} )</td>
</tr>
<tr>
<td>UV Ceti</td>
<td>( 1 \times 10^{32} )</td>
<td>( 2 \times 10^{27} )</td>
</tr>
<tr>
<td>Sun (1)</td>
<td>( 3 \times 10^{32} )</td>
<td>( 5 \times 10^{22} )</td>
</tr>
<tr>
<td>(2)</td>
<td>( 5 \times 10^{31} )</td>
<td>( 5 \times 10^{22} )</td>
</tr>
</tbody>
</table>

Table I: Shows total optical energy and total radio energy between 100 MHz and 500 MHz emitted during flares in: the Orion nebula (Slee and Higgins 1970) and on: YZ CMi (Kunkel 1970, Lovell 1970), UV Ceti (Lovell 1964), the sun (1) the 3b+ flare of February 23 1956 and the sun (2) the 2b flare of March 7 1970. Also shown is the spectral index in the metric wave region and the ratio of optical to radio energies.
flatter solar electron spectrum \((\gamma = 2-3)\). The disc brightness of
\[
J_x(> 10 \text{ KeV}) = \frac{1}{4\pi} \int_0^R <L_x> \cdot n(r) \, dr \sim 10^{-7} \text{ erg/cm}^2 \cdot \text{s-sterad}
\]
(with \( R = 10 \text{ kpc}, \ <n> = 0.03/\text{pc}^3 \))

which this implies is not consistent with observations however, which indicate an upper limit to \(<L_x>\) of about \(10^{24} \text{ erg/s-pc}^3\).

4. Evolution and Lifetimes. The upper limit flare luminosity \(L_0(F) = 10^{-2} L_0(Q)\) (Kunkel 1970) allows an estimate of the upper limit flare output during the convective lifetime of a star. By a well known application of the virial theorem for a contracting star of perfect gas in hydrostatic equilibrium, \(L = -\frac{1}{2} (d\Omega/dt)\) where the gravitational potential energy \(\Omega = -6 GM^2/7R\) for a convective star. The optical flare output generated during contraction to radius \(R\) is then \(W_0(F) \leq 10^{-2} |\Omega|/2\). For UV Ceti \(W_0(F) \sim 3 \times 10^{44}\) erg. The age of UV Ceti, from \(t \sim \Omega/6L\) (Kumar 1963) is \(\sim 10^8\) years so its average flare luminosity over this time is \(\leq 10^{43}\) erg/s, several times the present value (Section 2). According to Kumar (1962) low mass \((0.04\ M_\odot)\) stars like UV Ceti have total convective lifetimes of \((1-5) \times 10^8\) years and eventually become degenerate black dwarfs if their masses lie below \(0.07\ M_\odot\). Above \(0.4\ M_\odot\) radiative equilibrium eventually occurs and the star evolves towards the main sequence. Stars of intermediate mass remain fully convective for their luminous lifetimes near the main sequence (Poveda 1964). In the case of the sun the convective lifetime (to \(3 R_\odot\)) was much shorter, \(\sim 10^6\) y giving \(W_0(F) \sim 10^{46}\) erg and a mean flare luminosity \(\leq 10^{43}\) erg/s. We estimate the total flare energy from contracting convective stars in the following way. The differential mass spectrum of young stars is well described by a power law
\[
F(M) = C \cdot M^{-2.33} \text{ for } 0.2 M_\odot < M < 10^2 M_\odot \text{ (Reddish 1966)} \text{ and } F(M) = C(0.2 M_\odot)^{-2.33} \text{ for } 0.04 M_\odot < M < 0.2 M_\odot \text{ (Belerene 1970).}
\]
\(M\) is the limiting radii \(R(M)\) below which the stars are no longer fully convective are given by Poveda (1964) and Kumar (1963) as functions of mass. Values of \(\Omega(M)\) at this limit show a broad maximum near \(M = M_\odot\) of \(\sim 10^{48}\) erg dropping rapidly below \(0.4\ M_\odot\) to \(\leq 10^{47}\) erg below \(0.1\ M_\odot\). The total loss of gravitational energy during the convective contractions of stars of total mass \(\int F(M) \cdot M \, dM\) is then \(\int F(M) \cdot \Omega(M) \, dM\). Using this mass distribution we find that the main contribution to \(\dot{\Omega}\) comes from stars of mass \(M \sim M_\odot/5\) and \(\dot{\Omega}/M\) totals \(\sim 10^{48}\) erg/M_\odot.

5. Cosmic Ray Production. In the contraction of \(10^{11}\ M_\odot\) we therefore expect less than \(5 \times 10^{-3} (\dot{\Omega}/M) \cdot 10^{11} \sim 5 \times 10^{56}\) erg of optical flare energy. If a fraction \(\epsilon_{cr}\) of this is released as relativistic cosmic rays in the mean stellar lifetime of \(\sim 2 \times 10^8\) y the mean cosmic ray production rate, \(P_{cr} \sim (\epsilon_{cr} \cdot 10^{41})\) erg/s is of the same order as that required to maintain a galactic
cosmic ray proton energy density of $10^{-12}$ erg/cm$^3$ in the disc providing $\varepsilon_{cr} \sim 1$. Since the mass in such young stars is probably less than 5% of the galactic mass (Blaauw 1965), the fractional contribution to the proton energy density is $\lesssim 2 \times 10^{-1} \varepsilon_{cr} \tau_7$, where $\tau_7$ is the effective disc lifetime (in units of 10$^7$ y) of protons. The energy contribution is therefore $< 2 \times 10^{-3}$ for $\tau_7 < 1$, $\varepsilon_{cr} \sim 0.01$. The density of flare stars in the solar neighbourhood is $\sim 25$-50% of all stars by number, an order of magnitude less in mass. The total mass in the disc calculated directly from the local mass density of flare stars is therefore $\sim 1\%$ of the galactic mass. The total number of stars is then $\sim 5 \times 10^3$.

Although the hard X-ray flux from stellar flares is conjectural (Section 3) one may calculate the disc component of soft flare X-radiation by putting $<n> = 0.03$ pc$^{-3}$ and assuming the soft X-ray luminosity to be 10% of the optical flare luminosity as in the sun (Thomas and Teske 1971). We find $J_x > (1 \text{ keV}) \sim 10^{-9}$ erg/cm$^2$-s-sterad, close to that found by Cooke et al. (1969). The X-ray brightness due to flare stars will of course correlate with the columnar density of young stars. In any case the X-ray observations allow an upper limit to $<n L_x(> 1 \text{ keV})>$ of $\sim 10^{24}$ erg/s-pc$^3$ to be established. If the energetic charged particle energy is $\sim 10\%$ of the soft X-ray energy as for solar flares, then $<n L_{cr}> \lesssim 10^{25}$ erg/s-pc$^3$ and the disc production rate $P_{cr} \leq 10^{34}$ erg/s, thereby contributing $< 10^{-4}$ to the density of galactic cosmic radiation.

Acknowledgements. We acknowledge the financial support of the New Zealand University Research Grants Committee.

References.
IONOSPHERIC EFFECT CAUSED BY CELESTIAL X-RAYS

By
P. J. EDWARDS, G. J. BURTT
and
F. KNOX

Ionospheric Effect caused by Celestial X-rays

The study of variable X-ray sources is severely handicapped by the short duration of a single rocket or balloon flight. The advantages of a method of monitoring celestial X-ray fluxes from the ground for prolonged periods are obvious. Two techniques suggest themselves: (1) observation of fluorescent optical and infrared radiation to which the atmosphere is transparent; (2) radiowave observations of those ionospheric parameters (for example, electrical conductivity) known to be affected by X-radiation.

We have detected an ionospheric effect, namely an enhancement of the nocturnal D-region conductivity, which we believe to be caused by X-radiation from the strong X-ray source, Scorpius XR-1, and other weaker sources in the vicinity of the galactic centre.

The influence of solar flares on the propagation of very low frequency (VLF) radio waves in the Earth ionosphere wave guide is well known. It takes the form of a sudden advance in the phase of VLF waves propagated over a sunlit path. The effect is due to reduction of the effective reflexion height of the ionosphere as a result of the increase in the ionization density (and therefore in conductivity) which follows the absorption of solar flare X-radiation in the upper D-region.

During the night the electron density in this region is some two orders of magnitude lower than during the day and is presumably maintained, in quiet conditions, by cosmic rays, celestial X-rays and hydrogen Lyman-α radiation. The background X-rays are approximately isotropic with an intensity for that \( h > 2 \text{ keV} \) of

\[ j_x \sim 10(hv)^{-0.8} \text{ keV cm}^2 \text{ s sr} \]

such that \( 2\pi j_x \), the hemispherical energy flux, is approximately equal to the measured flux\(^1\) from the strongest discrete X-ray source Sco XR-1.

We have calculated the ion production rates in the upper D-region and find that ionization by Sco XR-1 exceeds that produced by either the galactic cosmic radiation\(^3\) or the celestial X-ray background at altitudes between 80 and 90 km, the altitude range within which 20 kHz waves are reflected at night\(^4\). Ionization by the nocturnal Lyman-α radiation may be significant at these altitudes but its contribution is difficult to assess in view of the present uncertainty in the concentration of nitric oxide. For a Lyman-α intensity of \( \sim 1 \text{ krayleigh} \) and a
nitric oxide concentration as high as $5 \times 10^6$ cm$^{-3}$ (an order of magnitude higher than that suggested by Mitra$^7$), the Lyman-α electron production rate, $q \sim 10^{-3}$ cm$^{-3}$ s$^{-1}$ at 82 km, is about the same as the cosmic ray and X-ray background rates and about half the Sco XR-1 production rate.

The perturbation produced by the Scorpius X-radiation can be approximately estimated by computing the altitude dependence $\omega(h)$ of the conductivity parameter,

$$\omega \propto N_e(h)/\nu(h) = q(h)^{1/2}/(\psi(h)\nu^{1/2})(h))$$

where $N_e$ is the electron density, $\nu$ the collision frequency and $\psi$ the electron loss rate$^7$, for both unperturbed and perturbed conditions.

Following illumination by Sco XR-1, the conductivity may increase by as much as 50 per cent between 80 and 85 km and the perturbed conductivity altitude profile may be reproduced by lowering the unperturbed profile by increments ranging from 0.5 km at 80 km to 1.5 km at 85 km. Assuming therefore a 1 km drop in reflection height and using the results of Galejs$^6$ we calculate a consequent phase advance of about one rad (8 s) in the 20 kHz signal from WWVL in Boulder, Colorado, received at Wellington, New Zealand. This is of the order of 10 per cent of the observed diurnal variation shown in Fig. 1 for a 3 yr period.

The magnitude of the diurnal phase change is somewhat less than that computed from the solar minimum ionization profiles of Deeks$^8$ by Galejs’s method and in fact shows a long term decrease presumably associated with the approach to solar sunspot maximum in 1967-68. A number of periodicities are present, some of which have been suppressed by the 52 day averaging period, but the most striking of those remaining is the annual variation with minimum diurnal phase difference occurring in April-May. It is in these months, when the period of full illumination of the Boulder-Wellington path by Sco XR-1 coincides with solar night over the path, that maximum reduction in the nocturnal propagation time is expected. The magnitude of this reduction (6-10 ms) is in good agreement with that approximately calculated for Sco XR-1.

An annual variation of this type might be caused by seasonal effects, although the equatorial symmetry of the great circle propagation path makes this unlikely. The celestial origin of the observed effect is in any case demonstrated in Fig. 2 which shows the sidereal change of 2 h/month in the time of onset of the phase advance. Taking into account the zenith angle dependence of the conductivity increase, the time lag between ground star rise and the onset time is in excellent agreement with that expected for Sco XR-1 acting alone. The integrated X-ray flux from other weak sources near the galactic
Fig. 1. 52 day means (± standard deviation) of the diurnal (night minus day) phase difference (Δφ) in the WWVL signal received at Wellington, 1965-68. Arrows indicate when Sco XR-1 (†) and the Sagittarius cluster (†) should produce a minimum in (Δφ).

Fig. 2. Onset times (UT) of the nocturnal phase advance in the WWVL (Boulder) signal received at Wellington as determined from 10 day means. ○, 1966; □, 1968. Data from 1967 have been excluded because of possible interference from Centaurus XR-2. Times of ground star rise of Sco XR-1 at Boulder (— -- --) and Wellington (--- ---) are also shown.
centre, however, may contribute significantly to the perturbation.

The potential utility of this and similar ground based techniques for the future (and also retrospective) study of strong variable X-ray sources like Sco XR-1 and Centaurus XR-2 (ref. 9) is apparent.

We thank Dr G. F. Preddey for assistance.

P. J. EDWARDS

Department of Physics,
University of Otago,
Dunedin, New Zealand.

G. J. BURTT

Physics and Engineering Laboratory,
Lower Hutt, New Zealand.

F. KNOX

Received January 21, 1969.

A New Radio-Wave Technique in X-Ray Astronomy

P. J. Edwards

Department of Physics, University of Otago, New Zealand

It is evident that the limited duration of a single rocket or balloon flight is a severe handicap to the study of variable X-ray sources like Cen XR-2, the early history of which can only be conjectured from the few available rocket observations. The whole question of the variability of X-ray sources remains in doubt because of the difficulty of relating infrequent high-altitude flux measurements of short duration made with a variety of instruments. There are obvious advantages in a method of monitoring celestial X-ray fluxes from the ground for prolonged periods of time. Such a method has recently been found and, although presently restricted to strong sources like Sco XR-1 and Cen XR-2, is capable of considerable refinement.

The method utilizes standard very-low-frequency (VLF) radio-wave techniques to detect the enhancement in ionospheric D-region conductivity which occurs when the nocturnal ionosphere is illuminated by celestial X-radiation.

The effect of solar flares on the propagation of VLF radio waves travelling in the Earth-ionosphere waveguide is well documented. It takes the form of a sudden advance in the phase of VLF waves propagating over a sunlit path. The phenomenon is caused by transient X-rays generated in the flare. This radiation enhances both the electron density and the electrical conductivity in the upper D-region. The result is that the effective reflection height for radio waves is reduced and their phase velocity increases.

At night the electron density in the D-region is some two orders of magnitude lower than during the day and is presumably maintained, under quiet conditions, by cosmic rays, celestial X-rays and hydrogen Lyman-α radiation. The importance of celestial X-radiation in maintaining the D-region ionization does not appear to have been previously recognized.

The background X-rays are approximately isotropic with an intensity \( j_\nu \sim 10(\nu)^{-0.6} \text{ keV/keV cm}^2 \text{s sterad} \), such that \( 2\pi j_\nu \), the hemispherical energy flux is approximately equal to the measured energy flux from the strongest discrete X-ray source, Sco XR-1.
Calculation of the ion-production rates in the upper D-region reveals that ionization by Sco XR-1 exceeds that produced by either the galactic cosmic radiation\(^7\) or the celestial X-ray background at altitudes between 80 and 90 km, the altitude range within which 20 kHz waves are reflected at night.\(^8\) Ionization by the nocturnal Lyman-\(\alpha\) radiation is significant at these altitudes. For a Lyman-\(\alpha\) intensity of \(\sim 1\) kilorayleigh\(^9\) and a nitric oxide concentration of \(10^7\) cm\(^{-3}\), the Lyman-\(\alpha\) electron production rate, \(q \sim 10^{-3}\) cm\(^{-3}\) s\(^{-1}\) at 82 km, the cosmic ray plus X-ray background rate, and the Sco XR-1 production rate are approximately equal.

The magnitude of the VLF phase anomaly due to an X-ray source may be estimated by computing the altitude dependence \(\omega(h)\) of the conductivity

\[
\omega(h) \propto \frac{N_e(h)}{\nu(h)} \propto \{q(h)\}^{1/4} \nu(h),
\]

where \(N_e = \) electron density, \(\nu = \) collision frequency, \(\psi = \) electron loss rate. The reduction in the height of reflection \(h_r\) is then given by

\[
\Delta h_r \sim \Delta \omega \left(\frac{d\omega}{dh}\right)_{h= h_r}^{-1},
\]

where \(\Delta \omega\) is the increase in conductivity due to the source. For Sco XR-1, \(\Delta \omega / \omega \sim 30\%\) giving \(\Delta h_r \sim 1\) km at altitudes between 80 and 85 km. Although this is small and comparable with the tidal movement of the ionosphere, it is readily detectable using current techniques.

Figure 1 displays the amplitude of the diurnal phase change in the 20 kHz signal from WWVL (Boulder) received at Wellington, New Zealand, over a three-year period. The diurnal phase change of \(\sim 100\) \(\mu\)s (12 rad) indicates that the height of reflection is about 12 km higher during the night than during the day. The effect of Sco XR-1 is evident here as an annual modulation of the nocturnal height of reflection, with the maximum reduction of 1 km in the reflecting height occurring, as expected, in April-May of each year when the period of full illumination of the propagation path by Sco XR-1 coincides with solar night over the path. Analysis of the data in sidereal time confirms the celestial origin of the effect and shows it to be as expected from Sco XR-1. The limited time resolution (~1 h) does not exclude a contribution from other weaker sources in the vicinity of the galactic centre but their integrated effect, on the basis of the known fluxes, should be comparatively small.

Several conclusions of astronomical interest may be drawn from Figure 1.

(a) The amplitude of the X-ray perturbation, when corrected for the secular trend and the biennial component in the phase data, remains constant to within
Figure 1. 52-day means (± standard deviation) of the diurnal (night minus day) phase propagation time ($\Delta\phi/\omega$) in the WWVL signal received at Wellington, New Zealand (1965-68). Dashed line shows the anomaly due to Cen XR-2 as predicted by Chodil et al.\textsuperscript{10}

±15\% from year to year. We conclude therefore that there is no evidence of any change in the Sco XR-1 flux when sampled for two-hourly intervals each night and averaged over 100 nights in each year covering the April-June period in the years 1965-68. The mean deviation of 15\% can be entirely accounted for by the scatter of the nightly points which appear to be the same whether or not the X-ray ionization is present. As the amplitude of the phase shift is proportional to the square root of the electron production rate due to X-rays with energies in the 2-8 keV range, one can set limits of ±30\% to any change in the energy flux in this range.

(b) The X-ray luminosity of Cen XR-2, prior to April 1967, did not change in the manner suggested by Chodil et al.\textsuperscript{10} who predicted a Centaurus flux in excess of five times the Scorpius flux in February 1967. The phase anomaly which would have then occurred is indicated in Figure 1, taking into account the different zenith angles of the two sources. A more detailed analysis of the nightly points does in fact reveal a small effect due to Cen XR-2 which appears to be consistent with the shockwave model proposed by Edwards.\textsuperscript{11, 12}

I wish to thank Mr G. Burtt and Mr F. Knox, Physics and Engineering Laboratory, Lower Hutt, New Zealand, for providing the VLF data, and for valuable discussions.

Simmons Limited, Glebe, Sydney.
Upper Limits to the Prompt X-ray Flux from Supernova 1987a

P. J. Edwards, Department of Electronics Engineering and Applied Physics Canberra College of Advanced Education

Abstract: The literature contains numerous predictions of an X-ray pulse associated with the emergence of the shock-wave at the surface of a star undergoing a supernova explosion. Applied to supernova 1987a, these suggest a small, solar-flare like event lasting about a minute with a terrestrial flux of order $1 \mu$W/m$^2$ and a black-body temperature between $10^8$K and $10^9$K occurring between 0900 and 1000 UT on February 23, 1987. Unfortunately, no direct X-ray observations appear to have been made. However, such an event would be expected to give rise to an observable enhancement of ionization in the upper D-region of the ionosphere. The absence of a detectable perturbation in VLF radio signals propagating in the earth-ionosphere waveguide sets upper limits to the X-ray flux and temperature below those predicted.

1. Introduction

Early supernova models (Colgate 1967) based on compact stellar progenitors led to predictions of a short (~1 s) $\gamma$-ray pulse with total energy of order $10^{50}$erg, representing 0.1% of the initial explosion energy of $E \sim 10^{51}$erg.

Later predictions of a longer (~10 s) soft (<1 keV) X-ray flare of similar total energy have been based on late type supergiant models with extended envelopes (Lasher and Chan 1979; Klein and Chevalier 1978; Falk 1980). In these models, the radiation dominated shock wave is thermalised as it expands into the base of an exponential atmosphere and an X-ray/UV pulse is emitted at low optical depths. Predictions of an additional, harder (>10 keV) X-ray flare from gas which has been piston heated by a viscous shock have been discounted by Epstein (1980) on the grounds that radiation diffusing ahead of the shock front prevents formation of a required velocity jump.

The progenitor star for supernova 1987a has been identified (White and Malin 1987) as SK 202-69, a blue supergiant with an estimated radius (Arnett 1987; Woosley et al. 1987; Dopita et al. 1987) of $3 \sim 3.5 \times 10^{13}$ cm. This radius is consistent with the time delay of <3 h between the neutrino event and the optical brightening, (Arnett 1987) corresponding to the shock travel time to the photospheric surface. The associated shock temperature is then $\sim 5 \times 10^9$ K for $E = 10^{51}$ erg, compared with $T_o = 10^8$ K for the Colgate model and $T_o \sim 5 \times 10^9$ K for the red supergiants of radius $\sim 5 \times 10^{13}$ cm modelled by Lasher and Chan.

If we assume a total X-ray pulse energy of $10^{48}$ erg for 1987a emitted in a time ~1 minute, corresponding to light travel time across the stellar surface, the X-ray flux density at the distance to the LMC is of the order of $10 \mu$W/m$^2$. Such an event would be comparable in magnitude to a class B-C solar X-ray flare and would give rise to a sudden ionospheric disturbance (SID) (Nicolet and Aikin 1960) which would be readily detectable.

Excess ionisation in the nocturnal ionosphere due to the X-ray binary Sco X-1 (Edwards 1969, 1970; Svensson et al. 1972) has been detected at a level of $10^4$ $\mu$W/m$^2$, some five orders of magnitude below the anticipated flux from SN 1987a. The non detection of the effects of an X-ray flare from the supernova in ionospheric records therefore provides an upper limit to the temperature and flux density significantly below that expected.

2. The Anticipated X-Ray Flare

The mode detailed calculations of shock breakout have been made by Lasher and Chan (1978) and Klein and Chevalier (1978). In the following discussion the possible hard X-ray contribution from a gas-viscosity shock front will be ignored following Epstein (1980). Table 1 summarises calculations of the following parameters: shock temperature, $T_o$; minimum flare duration (as set by the stellar radius), $r$; average photon energy $<h\nu>$ (corresponding to temperature, $T_o$); shock front travel time, $\Delta t$ for a range of stellar radii, $r$ bracketing SK 202-69.

The shock temperature has been calculated from the relation (Lasher and Chan 1978),

$$T_o^{(1,2)} = 1.5 \left(\frac{E}{10^{51}}\right)^{1/4} \left(\frac{r}{10^{12}}\right)^{3/4}$$  \hspace{1cm} (1)

The approximate X-ray flare duration has been calculated, assuming negligible shock thickness, from the relation (Lasher and Chan 1978),

$$\tau = 0.8 \frac{r}{u}$$ \hspace{1cm} (2)

The time delay (in seconds) from core collapse to the onset of the flare has been calculated from the relation (Shigeyama 1987),

$$\Delta t = 14. \left(\frac{r}{10^{12}}\right) \left(\frac{M/H_0}{10^5}\right)^{1/2} \left(\frac{E}{10^{51}}\right)^{-1/2}$$ \hspace{1cm} (3)

The upper two entries in Table 1 refer to the red supergiants modelled by Lasher and Chan (1978). The remaining entries refer to models close to those proposed for SN 1987a by Arnett (1987), Shigeyama et al. (1987) and Dopita et al. (1987). The time delays, $\Delta t$ are consistent with the lower limit set by Jones' (1987) visual observation of the LMC, 106.5 m after detection of the neutrino burst.

The temperatures of Table 1 represent upper limits to the spectral temperature of the flare radiation. In the red supergiant calculations of Lasher and Chan the effective blackbody flare temperatures $T_{eff}$ are typically ~20% of the shock temperatures, $T_o$. However, these authors point out that thermalisation by electron scattering is ineffective and that their assumption of blackbody emission at $T_{eff}$ leads to lower limit estimates of spectral temperature. Similar results were obtained by Falk (1978). Klein and Chevalier (1978) compute the flare spectrum under similar conditions and find the spectral temperature to be a factor of 3 higher than the effective temperature. These results collectively suggest that typical spectral temperatures (as...
determined by the mean photon energy) of \(-7T_0/2\) might be expected for red supergiant supernova X-ray flares.

The characteristics of the X-ray/\(\gamma\)-ray supernova flares expected from compact \((R \sim 10^8 \text{ cm})\) stars have been considered by Colgate (1966) and Colgate and White (1967). Their estimates of flare spectra must be regarded as speculative in the absence of a fully developed model of relativistic shock wave thermalisation along the lines followed for slower shocks in red giant envelopes. Nevertheless, there seems general support for a hard X-ray pulse of temperature of order \(T_0\), total energy \(-E/1000\) and duration \(R/C\).

The characteristics of the X-ray/UV flare are strongly influenced by the radius of shock wave breakout (Table 1) so that the flare from SN 1987a can be anticipated to have characteristics intermediate between those attributed to compact stars on the one hand and stars with extended envelopes on the other, as indicated in Table 1.

Following Lasher and Chan (1978), the temperature behind the shock is taken to be that given by the Sedov (1959) similarity solution as expressed in Equation (1). The emergence of this shock into the stellar envelope is governed by two parameters: the ratio of density scale height to shock wave thickness, and the optical thickness of the shock transition.

For the red supergiant models of Weaver et al. (1978) both these quantities are large and the shock velocity, \(V\ll c\). As the shock thickness widens and becomes comparable with the scale height, radiation diffuses to the stellar surface, typically when the front reaches an optical depth of \(-10-100\) below the surface, and the flare begins.

The situation is rather different for the more compact models appropriate to SN 1987a. Thus, the scale height is much smaller \((-10^6 \text{ cm})\) in the \(5.6 M_\odot\) model proposed by Wood (1987), the shock is relativistic, its optical thickness is of order 1 and the transition thickness is of the same order as the photon mean free path \((-10^1\text{ cm at shock break-out in the above model})\). Adopting the method suggested by Lasher and Chan leads to a total flare energy in excess of \(10^{48} \text{ erg}\), and an anticipated terrestrial X-ray flux of \(-10^{-2} \text{ erg/cm}^2\text{s} (10 \mu\text{Wm}^{-2})\) lasting \(-1.5\) minutes with a spectral temperature of \(-5 \times 10^6 \text{ K}\) occurring between 0900 and 1000 UT on February 23, 1987.

3. Detectability of the X-ray Flare
It appears that no direct X-ray flux observations were made at the expected time of shock breakout. However, the influence of solar X-ray flares on radio wave propagation in the earth-ionosphere waveguide is well known as are their geomagnetic effects (Nicolet and Aikin 1960). Both effects are due to the enhanced conductivity which results from the absorption of ionizing X-radiation in the upper D-region of the daytime ionosphere. The threshold X-ray flux required for the production of a sudden ionospheric disturbance (SID) is highly temperature dependent varying from \(-10^{-2} \mu\text{Wm}^{-2}\) at \(T = 10^7 \text{ K}\) to perhaps \(-1 \mu\text{Wm}^{-2}\) at \(T = 10^6 \text{ K}\).

The detection threshold is substantially lower during the night, a number of X-ray sources including SCO X-1, Cen X-2 and Cen X-4 having been detected at a level of \(10^{-4} \mu\text{Wm}^{-2}\) by averaging ionospheric data over several nights (Edwards et al. 1969; Edwards 1969; Svensesson et al. 1972). With these figures in mind it would seem that, with the flare parameters derived in section (1), a detectable SID might be expected, even in daytime. At the expected time of the flare (1900-2000 EAST), eastern Australia, the Pacific and the west coast of South America were in darkness, presenting an ideal opportunity to examine ionospheric data with a low X-ray flux detection threshold.

4. Observations
Phase and amplitude records of very low frequency (10.2, 11.33, 13.6 kHz) signals transmitted from Omega navigation stations in Argentina, Hawaii and Japan, received in Canberra and Sale (Australia) were carefully examined. No effects attributable to SN 1987a were found.

Magnetograms from Eyrewell (New Zealand) and Canberra (Australia) were also examined for evidence of a magnetic crotchet due to possible enhancement of the "quiet day" ionospheric (Sq) current system. Again, no effects were found at the expected times.

### Table 1

<table>
<thead>
<tr>
<th>Energy ((10^{51} \text{ erg}))</th>
<th>Stellar Radius ((10^{12} \text{ cm}))</th>
<th>Shock Temperature ((K))</th>
<th>Mean Photon ((\text{keV}))</th>
<th>X-Ray Flare Duration ((\text{s}))</th>
<th>Stellar Mass ((M_\odot))</th>
<th>X-Ray Flare Energy ((10^{48} \text{ erg}))</th>
<th>Shock Transit Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>67</td>
<td>(6.4 \times 10^5)</td>
<td>0.17</td>
<td>2000</td>
<td>15.5</td>
<td>1.7</td>
<td>3.5 d</td>
</tr>
<tr>
<td>3</td>
<td>67</td>
<td>(8.5 \times 10^5)</td>
<td>0.23</td>
<td>2000</td>
<td>15.5</td>
<td>3.3</td>
<td>2 d</td>
</tr>
<tr>
<td>1</td>
<td>3.5</td>
<td>(6.0 \times 10^6)</td>
<td>1.5</td>
<td>90</td>
<td>4.5</td>
<td>&gt;1</td>
<td>105 m</td>
</tr>
<tr>
<td>1</td>
<td>3.5</td>
<td>(6.0 \times 10^6)</td>
<td>1.5</td>
<td>90</td>
<td>6.5</td>
<td>&gt;1</td>
<td>125 m</td>
</tr>
<tr>
<td>1.5</td>
<td>3.0</td>
<td>(7.3 \times 10^6)</td>
<td>2.0</td>
<td>80</td>
<td>15</td>
<td>&gt;1</td>
<td>130 m</td>
</tr>
</tbody>
</table>
Ionograms for Canberra, Darwin, Hobart, Mundaring and Townsville (Australia) were also examined without success. Unfortunately higher frequency ionospheric propagation records with higher sensitivity through the expected time of shockwave breakout do not appear to be available. However, the absence of detectable effects in the foregoing ionospheric and magnetic data do allow conservative upper limit fluxes to be established with reasonable confidence which are significantly below those expected.

5. Conclusions
The absence of detectable ionospheric and magnetic disturbances attributable to the expected X-ray/UV flare from supernova 1987a places certain restrictions on the characteristics of the flare.

Conservative upper limit fluxes of $10^{-2}$ $\mu$Wm$^{-2}$ at $T = 10^7$ K, $10^{-1}$ $\mu$Wm$^{-2}$ at $T = 5 \times 10^6$ K and $1$ $\mu$Wm$^{-2}$ at $T = 10^6$ K may be established.

These upper limits support the lack of formation of a hard X-ray radiating viscous shock as reported by Epstein and Lasher and Chan. They are not consistent, falling short by several orders of magnitude, with predictions of a short, soft X-ray flare calculated using the method proposed by Lasher and Chan (1978).

This inconsistency may be due to shortcomings in the modelling of the flare, or of the progenitor star—for example, the flare temperature may be over-estimated. Alternatively, the flare radiation might be attenuated by a shell of circumstellar material shed by the progenitor in its evolution from an earlier red supergiant phase.

6. Acknowledgements
I gratefully acknowledge the receipt of inospheric data from K. Mitchell, IPS Radio and Space Services, Australian Department of Science; Omega data from K. Flynn, Australian Department of Transport; magnetograms from L. Tomlinson, DSIR (New Zealand) and P. McFadden, Bureau of Mineral Resources, Canberra.

Arnett, W. D., 1987(a), preprint.
Arnett, W. D., 1987(b), preprint.
IONOSPHERIC EFFECTS OF SUPERNOVA EXPLOSIONS

P. J. Edwards
Dept. of Electronics Engineering & Applied Physics
Canberra College of Advanced Education, Australia

1. INTRODUCTION

Early supernova models (Colgate 1967) based on compact stellar progenitors led to predictions of a short (~1s) $\gamma$-ray pulse with total energy of order $10^{48}$ erg, representing 0.1% of the initial explosion energy of $E=10^{51}$ erg.

Later predictions of a longer (~$10^{4}$s) soft (<1 keV) X-ray flare of similar total energy have been based on late type supergiant models with extended envelopes (Lasher and Chan 1979; Klein and Chevalier 1978; Falk 1978). In these models, the radiation dominated shock wave is thermalised as it expands into the base of an exponential atmosphere and an X-ray/UV pulse is emitted at low optical depths. Predictions of an additional, harder (>10 keV) X-ray flare from gas which has been piston heated by a viscous shock have been discounted by Epstein (1980) on the grounds that radiation diffusing ahead of the shock front prevents formation of the required velocity jump.

The progenitor star for supernova 1987a has been identified (White and Malin, 1987) as SK 202-69, a blue supergiant with an estimated radius (Arnett 1987a; Woosley et al., 1987; Dopita et al., 1987) of 3-3.5 x $10^{12}$ cm. This radius is consistent with the time delay of <3h between the neutrino event and the optical brightening (Arnett 1987), corresponding to the shock travel time to the photospheric surface. The associated shock temperature is then $\geq 5x10^{6}$K for $E=10^{51}$ erg, compared with $T_{0}=10^{9}$K for the Colgate model and $T_{0}=5x10^{5}$K for the red supergiants of radius $\geq 5x10^{13}$ cm modelled by Lasher and Chan.

If we assume a total X-ray pulse energy of $10^{48}$ erg for 1987a emitted in a time ~1 minute, corresponding to a light travel time
across the stellar surface, the X-ray flux density at the distance to the LMC is of the order of 10 $\mu$W/m$^2$. Such an event would be comparable in magnitude to a class B/C solar X-ray flare and would give rise to a sudden ionospheric disturbance (SID) (Nicolet and Aikin, 1960) which would be readily detectable.

Excess ionisation in the nocturnal ionosphere due to the X-ray binary Sco X-1 (Edwards 1969, 1970; Svensson et al., 1972) has been detected at a level of $10^{-4}$ $\mu$W/m$^2$, some five orders of magnitude below the anticipated flux from SN 1987a. The non-detection of the effects of an X-ray flare from the supernova in ionospheric records would therefore provide an upper limit to the temperature and flux density significantly below that expected.

2. THE ANTICIPATED X-RAY FLARE

The most detailed calculations of shock breakout have been made by Lasher and Chan (1978) and Klein and Chevalier (1978). In the following discussion the possible hard X-ray contribution from a gas-viscosity shock front will be ignored following Epstein (1980).

Table 1 below summarises calculations of the following parameters: shock temperature, $T_o$; minimum flare duration (as set by the stellar radius), $\tau$; average photon energy $\langle h\nu \rangle$ (corresponding to temperature, $T_o$); shock front travel time, $\Delta t$, for a range of stellar radii $R_{12}$ bracketing SK 202-69.

<table>
<thead>
<tr>
<th>$(E/10^{51}$ ergs)</th>
<th>$(R/10^{12})$</th>
<th>$(T_o/K)$</th>
<th>$\langle h\nu/\text{keV} \rangle$</th>
<th>$(\tau/s)$</th>
<th>$(M/M_\odot)$</th>
<th>$(EE/10^{48}$ ergs)</th>
<th>$\Delta t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>67</td>
<td>6.4 $\times$ 10$^5$</td>
<td>0.17</td>
<td>2000</td>
<td>15.5</td>
<td>1.7</td>
<td>3.5d</td>
</tr>
<tr>
<td>3</td>
<td>67</td>
<td>8.5 $\times$ 10$^6$</td>
<td>0.23</td>
<td>2000</td>
<td>15.5</td>
<td>3.3</td>
<td>2d</td>
</tr>
<tr>
<td>1</td>
<td>3.5</td>
<td>6.0 $\times$ 10$^5$</td>
<td>1.5</td>
<td>90</td>
<td>4.5</td>
<td>1</td>
<td>105m</td>
</tr>
<tr>
<td>1</td>
<td>3.5</td>
<td>6.0 $\times$ 10$^5$</td>
<td>1.5</td>
<td>90</td>
<td>6.5</td>
<td>1</td>
<td>125m</td>
</tr>
<tr>
<td>1.5</td>
<td>3.0</td>
<td>7.3 $\times$ 10$^5$</td>
<td>2.0</td>
<td>80</td>
<td>15</td>
<td>1</td>
<td>130m</td>
</tr>
</tbody>
</table>

The upper two entries in Table 1 refer to the red supergiants modelled by Lasher and Chan (1978). The remaining entries refer to models close to those proposed for SN 1978a by Arnett (1987b), Shigeyama et al. (1987) and Dopita et al. (1987). The time delays, $\Delta t$ are consistent with the lower limit set by Jones' (1987) visual
observation of the LMC, 106.5m after detection of the neutrino burst.

The temperatures of Table 1 represent upper limits to the spectral temperature of the flare radiation. The characteristics of the X-ray/γ-ray supernova flares expected from compact ($R \sim 10^8$ cm) stars have been considered by Colgate, (1966) and Colgate and White, (1967). There seems to be general support for a hard X-ray pulse of temperature of order $T_o$, total energy $\sim E/1000$ and duration $\sim R/C$.

The characteristics of the X-ray/UV flare are strongly influenced by the radius of shock wave breakout (Table 1) so that the flare from SN 1987a can be anticipated to have characteristics intermediate between those attributed to compact stars on the one hand and stars with extended envelopes on the other, as indicated in Table 1.

For the red supergiant models of Weaver et al. (1978) the shock velocity, $V \ll c$. As the shock thickness widens and becomes comparable with the scale height, radiation diffuses to the stellar surface, typically when the front reaches an optical depth of $\sim 10^{-10}$ below the surface, and the flare begins.

The situation is rather different for the more compact models appropriate to SN 1987a. Thus, the scale height is much smaller ($\sim 10^9$ cm) in the 5.6 $M_\odot$ model proposed by Wood (1987), the shock is relativistic, its optical thickness is of order 1 and the transition thickness is of the same order as the photon mean free path ($\sim 10^{11}$ cm at shock break-out in the above model). Adopting the method suggested by Lasher and Chan leads to a total flare energy in excess of $10^{48}$ erg, and an anticipated terrestrial X-ray flux of $10^{-2}$ erg/cm$^2$-s (10μW/m$^2$) lasting ~1.5 minutes with a spectral temperature of $\lesssim 5 \times 10^6$K occurring between 0900 and 1000 UT on February 23, 1987.

3. DETECTABILITY OF THE X-RAY FLARE

It appears that no direct X-ray flux observations were made at the expected time of shock breakout. However, the influence of a solar X-ray flare on radiowave propagation in the earth-ionosphere waveguide is well known as are their geomagnetic effects (Nicolet and Aikin, 1960). Both effects are due to the enhanced conductivity.
which results from the absorption of ionizing X-radiation in the upper D-region of the daytime ionosphere. The threshold X-ray flux required for the production of a sudden ionospheric disturbance (SID) is highly temperature dependent, varying from $-10^{-2}$ $\mu$W/m$^2$ at $T = 10^7$ K to perhaps $-1$ $\mu$W/m$^2$ at $T = 10^6$ K.

The detection threshold is substantially lower during the night, a number of X-ray sources including Sco X-1, Cen X-2 and Cen X-4 having been detected at a level of $10^{-4}$ $\mu$W/m$^2$ by averaging ionospheric data over several nights (Edwards et al., 1969; Edwards, 1969; Svennesson et al., 1972). With these figures in mind it would seem that with the flare parameters derived in section (1), a detectable SID might be expected, even in daytime. At the expected time of the flare (1900-2000 EST) eastern Australia, the Pacific and the west coast of South America were in darkness, presenting an ideal opportunity to examine ionospheric data with a low X-ray flux detection threshold.

REFERENCES

High Energy Astrophysics

High Energy Transient Ionospheric Monitoring Network

Paul J. Edwards, Department of Electronics, Engineering & Applied Physics, School of Information Sciences & Engineering, Canberra College of Advanced Education, GPO Box 1, Belconnen, ACT 2616

Abstract: Continuous, wide sky coverage is essential for the detection and monitoring of infrequent, short-lived events of astrophysical interest such as supernova and nova explosions, variable X-ray sources, gamma ray bursts, gravity waves and stellar and solar flares. We propose to (1) examine past radio propagation records and (2) develop new computer based radio receivers to monitor and log ionospheric perturbations associated with these events.

1. Introduction

Ionospheric effects due to celestial X-ray sources were first predicted by Edwards, Burtt and Knox (1969) in the context of a search for ground based, wide field-of-view observing techniques. The magnitude of the expected increase in conductivity due to illumination of the nocturnal D-region of the ionosphere by the X-ray source, Scorpius X-1 was calculated and the resulting advance in the phase of a long path, very low frequency (VLF) signal propagating in the earth-ionosphere waveguide was found to be in general agreement with observations of the WWVL 20 kHz signal received in Wellington from Boulder, Colorado (Edwards et al. 1969, Edwards 1969). This work was followed by reports of a reduction in the amplitude of the low frequency (LF) signal from Radio Tashkent received in Ahmedabad during the transits of Sco X-1 and the Crab X-ray sources, 3U0531 +21 (Ananthakrishnan and Ramanathan 1969).

These initial observations received extensive discussion from both observational and theoretical viewpoints. Svennesson et al. (1972, 1979) reviewed and augmented the observational evidence and concluded that nocturnal ionisation by the strong X-ray source Sco X-1, and the strong transient sources Cen X-2 and Cen X-4 was detectable as a perturbation in the phase of VLF signals propagated over long night-time paths at frequencies above 15 kHz.

The original context in which this early work was performed, namely the search for ground-based, wide field-of-view, 'all-sky' observing techniques appears to be still relevant. For example, because of the lack of suitable satellite and spacecraft X-ray observing platforms at the time of shock-wave breakout from Supernova 1987A, no direct observations were made. However, Edwards (1987, 1988a) was able to place an upper limit to the breakout temperature of 2 x 10^6 K based on the absence of observable ionospheric perturbations.

More recently Fishman and Inan (1988) reported a sudden decrease in the amplitude of the 16 kHz signal from Rugby, recorded in the Antarctic at the time of the strong gamma ray burst, GB830801.

These two events, coupled with the imminent launch of the Gamma Ray Observatory containing the Burst and Transient Source Experiment (BATSE; Tuohy 1989), have rekindled interest in this technique.

In the twenty years which have elapsed since these phenomena were first reported, significant advances in communications and data processing technology have taken place. Previous observations of VLF propagation anomalies have been severely restricted by receiver and analogue recording limitations to a small number of frequencies and propagation paths.

It is now possible to develop an extensive network of computer-controlled multichannel wideband radio receivers with digital data logging facilities capable of being interrogated remotely through the public switching network. Thus the ambiguities arising from over-reliance upon a narrow data base can be reduced and, as well, the spatial, temporal and frequency-domain morphologies made accessible. By this means the ionospheric mechanisms may be clarified, the threshold of detection may be lowered and the astrophysical parameters more easily deduced.

2. Detectable transient X-ray phenomena

Transients extra-terrestrial X-ray sources capable of causing detectable enhancements of ionospheric D-region conductivity include solar flares (Mitra 1974), 'X-ray novae' such as Cen X-2, Cen X-4, and A0620-00 (Maraschi et al. 1976), gamma-ray and X-ray bursts (Hartmann and Woosley 1978), and galactic supernova explosions (Lasher and Chan 1978).

For solar flares detected during day-time, the threshold X-ray flux for the production of a detectable sudden ionospheric disturbance (SID) is clearly dependent on both source characteristics (temperature and duration) and on the technique employed. Mitra (1974) quotes a threshold flux, for $E > 1.5$ keV, of $10^{-5} \text{erg cm}^{-2} \text{s}^{-1} (1 \mu \text{W m}^{-2})$ at $T = 10^7 \text{K}$, corresponding to $10^{-5} \text{erg cm}^{-2} \text{s}^{-1}$ for $E > 4$ keV.

The detection threshold is substantially lower during the night because of the greatly reduced level of ambient ionisation due, in turn, to the absence of direct solar X-UV radiation. There now seems ample evidence (eg. Svennesson and Westerlund 1979) for detection of both stationary and transient X-ray sources at a flux level of $10^{-7} \text{erg cm}^{-2} \text{s}^{-1}$ for photon energies above several keV.

3. Detection Techniques

Conventional techniques for monitoring the D-region ionospheric effects of solar flares involve measurements of the complex amplitude of ELF, VLF, LF and HF signals from terrestrial transmitters. The phenomenon of sudden ionospheric disturbances (SID) due to solar flares and their observational techniques are comprehensively treated by Mitra (1974). The propagation of VLF waves in the earth-ionosphere waveguide has been extensively studied (Bracewell & Straker 1949, Galejs 1972, Wait 1964) and VLF and LF phase and amplitude measurements during solar flares have been used to construct ionization height profiles using the mode theory calculations of Wait and Spies (1964). Spectral irradiance of the ionizing flux can then be estimated from the excess ionization profile. The use of VLF/LF propagation data to determine the X-ray spectrum of solar flares has been performed successfully by Miao et al. (1985) using techniques described by Barletti and Tagliaferri (1969) and modified by Lewis et al. (1973).
The extension of these concepts and methodologies to X-ray illumination of the night-time ionosphere is obvious and has been described by Edwards et al. (1969).

(a) Phase Anomalies
Wait and Spies (1964) successfully modelled VLF propagation in the earth-ionosphere waveguide by representing the VLF ionospheric conductivity, \( \sigma(h) \approx N(h)/\nu(h) \) as a two-parameter exponential function of altitude.

The conductivity parameter,

\[
\omega(h) = N(h)/\nu(h)
\]

is assumed to have an exponential profile (Wait and Spies 1964) given by

\[
\omega(h) = \omega_0 \exp(\beta(h - h_0))
\]

with \( N(h, t) = \) electron density
\( \nu(h) = \) collision frequency
\( \beta(h) = \) conductivity gradient
\( h_0 = \) reference altitude at which \( \sigma \) is 2.2 \( \times \) 10\(^{-6} \) S m\(^{-1} \)

For small impulsive ionising events, there is a transient lowering of the height of reflection according to the above model by an amount,

\[
\Delta h = \frac{(\Delta Q/Q)/2\beta}{\alpha(h)}
\]

with \( Q(h, t) = \) ionization rate
\( \alpha(h) = \) effective recombination rate,

providing the duration of the impulse is substantially longer than the 'ionospheric time-constant', \( \tau \approx 1/\alpha N \). This leads to a corresponding advance in the phase of a wave reflected back to a terrestrial observer. Night-time phase advances are well modelled for sources such as Sco X-1 by a drop in reflection height \( \Delta h \approx 1 \) km for \( \Delta Q/Q \approx 1 \), depending on the ambient ionization rate, \( Q \), in the 80-90 km height range. Ambient ionization is largely accounted for by Lyman-\( \alpha \) ionization of nitric oxide, diffuse X-rays and galactic cosmic radiation.

It should be pointed out that transient changes in ionization due to precipitation of charged particles into the ionosphere, solar modulation of the galactic cosmic ray flux and solar generated cosmic ray events all contribute to the noise level which must be recognised and discriminated against.

(b) Amplitude Changes
Sudden enhancements of signal field strength (SES) and sudden field anomalies (SFA) are well known solar flare effects. Kasturirangan et al. (1974) searched unsuccessfully for nocturnal LF signal amplitude changes associated with ten gamma ray bursts recorded by the Vela satellites (Strong et al. 1974) with durations of the order of 10 seconds and peak flux of the order of \( 10^{-6} \) erg cm\(^{-2} \) s\(^{-1} \) above 7 keV. The absence of detectable effects is consistent with their short duration and relatively hard spectrum, the latter leading to significant energy deposition only below the effective height of reflection (Baird 1974).

The gamma ray burst, GB830801 was one of the strongest ever observed with a fluence of \( 2 \times 10^{-2} \) erg cm\(^{-2} \). A coincident decrease occurred in the amplitude of the 16 kHz signal transmitted from Rugby to Palmer Station, Antarctica along an early night-time path (Fishman & Inan 1988).

Calculations presented by Edwards (1988b) confirm that no detectable phase advance would have been expected but suggest that a different mechanism, namely increased non-deviate absorption could account for the observed decrease in field strength. The additional ionization during this event appears to be similar in magnitude to the galactic cosmic ray ionization background.

4. Proposed Monitoring Network
The aims of the proposed ionospheric propagation network are to:
(i) Clarify the physical mechanisms and parameters involved,
(ii) Refine detection and estimation techniques,
(iii) Support international and national 'all-sky' X-ray and gamma-ray transient surveys, and
(iv) Search for transient ionizing events not otherwise detected.

Dedicated monitoring equipment incorporating a multi-band (VLF/LF/HF) computer-based 'smart' scanning receiver, data logger and communications node would be deployed at existing ionospheric, geophysical and astronomical observatories. These would be located to cover a full longitudinal range of near meridional propagation paths in both hemispheres.

These monitoring stations would employ transient signal recognition algorithms to reduce demands on data storage. They would be remotely programmable and would provide data access via the international public communication system.

Phase and amplitude records would be obtainable from any one station with a time resolution of 0.1s from as many as 16 different transmissions with carrier frequencies between 10 kHz and 10 MHz. A larger number of channels could be monitored with lower time resolution.

The provision of redundant channel monitoring over wide geographical and frequency domains should extend the spectral sensitivity to ionizing events from below 1 keV to above 10 keV and significantly lower the detection threshold flux through post detection processing.

5. Conclusion
The proposed network should help to clarify existing ambiguities in the ionization mechanisms, spectral sensitivity and frequency domain characteristics of transient ionizing astrophysical events. Such a network could provide useful supporting data with a wide field of view for direct satellite and spacecraft studies of transient ionizing radiation.

Acknowledgements
I should like to thank Dr J. Crouchley (University of Queensland), Dr R. Sood and Dr R. Manchanda (University College, ADFA) for their encouragement to proceed with the proposal.


Ionospheric Propagation Anomalies due to Celestial X-Ray and Gamma Ray Sources

P.J. Edwards
Centre for Advanced Telecommunications and Quantum Electronics Research
Faculty of Information Sciences and Engineering, University of Canberra, ACT 2601, Australia
Tel:+61 2 6201 2515, Fax:+61 2 6201 5041, e-mail: paule@ise.canberra.edu.au

Introduction

The first reported celestial effect on ionospheric radio propagation [1] was a seasonal variation in the diurnal phase change in the 20kHz signal from WWVL, Boulder received in Wellington associated with the transit of the galactic x-ray source, Sco XR-1. The reality of this effect was initially challenged on both observational [2] and theoretical [3] grounds and subsequent analysis revealed that the sidereal component of the variation, although statistically significant, was smaller than expected. Numerous observations of stellar x-ray induced VLF phase advances were subsequently reported. A comprehensive study [4] involving the evaluation of thirty different VLF propagation paths supported the New Zealand results but concluded that propagation anomalies reported on some other paths could not be reliably attributed to stellar sources. This report reviews the ionospheric parameters revealed by transient and steady celestial sources of ionizing radiation including the recent intense gamma ray burst of August 27 1998.

Transient Celestial Sources of Ionizing Radiation

Ionospheric monitoring of transient x-ray and gamma ray burst sources can provide both useful ionospheric and source information. Thus, the ionospheric detection of the x-ray nova Cen XR-2 [4,5] and the non-(ionospheric) detection [6] of the supernova SN1987A were both of astrophysical interest. Current ionospheric interest attaches to the detection of two large gamma ray bursts which have been associated with large VLF phase and amplitude disturbances [7,8,9].

The Gamma ray burst of August 27 1998

A pulsating radiation burst of five minutes duration from the most recent and the larger of these, on August 27.4321 UT 1998, [10] was monitored at energies greater than 40keV by six spacecraft and at least three terrestrial VLF networks, two of these [8,9] recording near day-time ionization levels in the nocturnal D region. HF effects have also been reported [11]. This event offers a unique opportunity to model the lower D region as well as providing useful burst source information otherwise lost because of the saturation of satellite borne instruments.

References

PART (b):

STUDIES IN FLUCTUATIONS, NOISE

and

QUANTUM ELECTRONICS
The Existence of an Optimum Frequency for Line of Sight Telemetry

P. J. EDWARDS*

Summary
This paper examines the frequency dependent factors which determine the optimum carrier frequency for a communication link. The discussion is limited to line of sight propagation of signals with carrier frequencies in the range 50-1000 Mc/s. The results are particularly applicable to the design of air to ground links such as balloon telemetry systems in which transmitter weight must be minimised.

1. Introduction
A number of papers1, 2 have recently dealt with the choice of optimum frequencies for earth-space links. The frequency dependent factors have been examined by Pratt1 who concluded that the optimum frequency lies between 1000 and 10,000 Mc/s. It was considered desirable to make a similar analysis to include situations in which the presence of the earth's surface influences the system performance. This paper examines the choice of frequency and pays particular attention to airborne telemeters. The existence of an optimum frequency may be anticipated simply from consideration of the total noise in the link. The total link noise may be divided into two components, one generated in the receiver and the other contributed by the receiving antenna. The sum of these components, respectively increasing and decreasing functions of frequency in the range considered, will in general have a minimum value at some frequency. In the absence of other frequency dependent factors, operation at this frequency will ensure a maximum signal to noise ratio. The criterion used here is the minimising of the transmitter input power required to maintain the signal to noise ratio above some threshold value under the specified conditions of link operation.

2. Discussion
The frequency dependent factors to be discussed are:
1. Receiver noise.
2. External noise.
3. Transmission loss.
4. Transmitter efficiency.

The transmitter input power required to give a threshold signal is

\[ P_t = \frac{S_o N' k T_o B}{\epsilon A} \]  

where
- \( S_o \) = Minimum permissible RF signal to rms noise power ratio at the receiver input in the bandwidth (B).
- \( N' \) = Effective receiver noise factor.
- \( kT_o = 10^{-20} \) Joules.
- \( B \) = Receiver noise bandwidth.
- \( \epsilon \) = Transmitter power efficiency.
- \( A \) = Power attenuation factor (\( P_R/P_T \)).
- \( P_R \) = RF power available from matched antenna.
- \( P_T \) = RF power radiated by transmitting antenna.

Both \( A \) and \( N' \) are strongly frequency dependent. In general \( N' \), which contains antenna noise and receiver noise terms, falls to a minimum at metre wavelengths for vacuum tube receivers. The transmission loss, \( 1/A \), is a monotonic function of frequency as are the transmitter efficiency and the noise bandwidth. The functions \( A, B \), therefore determine the difference between the frequency for which \( N' \) is a minimum \( f' \) and the optimum frequency of the link \( f_o \). These two frequencies will be equal only if \( A \) and \( B \) do not depend on frequency.

The optimum frequency might have been found by differentiating Eq. (1) with respect to frequency. However, since the expression obtained in this way is analytically somewhat cumbersome, graphical means will be used instead.

*Department of Physics, The University of Tasmania, Hobart, Tasmania.
Manuscript received by The Institution July 5, 1962.
Revised manuscript received October 29, 1962.

August, 1963


Proceedings of the I.R.E. Australia 639
EDWARDS: The Existence of an Optimum Frequency for Line of Sight Telemetry

3. Noise

The noise in the link determines the minimum usable signal power available at the receiver input terminals. The noise internally generated in the receiver will be specified by the noise figure \( N \) while the total link noise will be described by an "effective" noise figure \( N' \).

Now, by definition

\[
N = \frac{S_i/N_i}{S_o/N_o}
\]

where \( S_i/N_i = \) Input signal to noise ratio
and \( S_o/N_o = \) Output signal to noise ratio.

The receiver input is assumed matched to a resistance at \( T_o = 290^\circ K \). Hence the receiver effectively adds noise power \( (N - 1)kT_oB \) to the signal at the input terminals. Now the external noise presented to the receiver is \( kT_oB \), where \( T_o \) is the noise temperature of the receiving antenna. The total link noise, referred to the receiver input, is therefore \( kB(T_o + (N - 1)T_o) \) and the effective noise figure

\[
N' = (T_o/T_o) + (N - 1)
\]

expresses the total link noise per unit bandwidth, referred to the receiver input terminals in units of \( kT_o \). It should be noted that \( N' \) is greater than or equal to \( N \) only if \( T_o \) is greater than or equal to \( T_o \), and that \( N' \) is a noise figure as previously defined if the signal to noise ratios are measured at the input and output of the link. The input measurement is made on the attenuated signal power \( (AP_T) \) available from a resistance at temperature \( T_o \).

4. Receiver Noise

The contribution of receiver noise to the total link noise may be made smaller by the use of maser and parametric amplifier techniques. These techniques, particularly useful in the microwave range, will not be considered here as they do not yet appear to be sufficiently developed for use in routine telemetry operation.

In assessing typical receiver noise a linear dependence of noise figure on frequency, characteristic of low noise triodes, is assumed. An examination of vacuum tube data shows that the expression

\[
N = 1 + \frac{f(Mc/s)}{150}
\]

is representative of noise figures currently attainable in the range 50 to 1000 Mc/s.

5. Antenna Noise

The noise temperature of the VHF receiving antenna has a diurnal and seasonal variation associated with the motion of the earth relative to the galaxy. The maximum and minimum values of the antenna temperature due to galactic emission as summarised by Kraus and Ko\(^3\) and Shklovski\(^4\) have been used in calculating the effective noise figure for receivers with directional antennae. The dipole noise temperatures used are based on figures obtained by the National Bureau of Standards\(^5\).

In order to take some account of the thermal radiation from the ground absorbed by horizon-looking antennae, the minimum antenna temperature is set at 150° by adding this figure to all cosmic noise temperatures. The above procedure may tend to overestimate the antenna temperature at the lower frequencies where the sky is brighter than the ground but it is considered to be adequate in view of the other uncertainties involved in the calculations. No explicit account will be taken of man made noise. It will be assumed either that this noise is absent or that it has been removed by suitable techniques. In Figs. 1 and 2 the curves (a) and (b) show the effective noise figure \( N' \) as a function of frequency. The curves are plotted for both minimum (a) and maximum (b) sky noise. The dotted curves will be described below.

---

The Existence of an Optimum Frequency for Line of Sight Telemetry

6. Propagation

With isotropic or fixed gain antennae the free space transmission loss increases by 6 db per octave. The propagation of a signal from an airborne transmitter however differs significantly from that in free space. Due to the presence of the earth's surface, which acts as a reflector and absorber, the free space attenuation is modified by diffraction below the horizon and by interference between the direct and ground reflected waves above the horizon.

Consider a transmitter at constant altitude \( h_p \) travelling away from the receiving antenna situated at altitude \( h_r \). As the angle of elevation \( \theta \) decreases the field strength oscillates about the free space value until the transmitter nears the horizon. The free space power transfer between dipole antennae varies inversely as the square of both the frequency \( 1/\lambda \) and distance \( d \).

\[
A = \frac{P_r}{P_t} = (0.13 A/d)^2 \cos^4 \theta
\]

The \( \cos^4 \theta \) term applies to vertical antennae. For horizontal polarization this term should be omitted. The conclusions in this section are independent of polarization.

Interference between the direct and reflected components of the signal will, at most, raise the received power by 6 db but may also cause complete signal drop out if the ground acts as an efficient reflector. By assuming that these interference minima produce gaps in air-ground coverage, Norton\(^7\) has shown that for a given value of \( h_r \) there exists a value of \( h_p \) which permits maximum coverage. The link performance then steadily deteriorates as the frequency is raised since the optimum \( h_p \) value, and therefore the distance to the horizon, varies inversely with frequency. In the present discussion this view is not taken since destructive interference will not drop the signal below that at optical cut off unless horizontally polarized antennae are used over smooth terrain. With vertical polarization or over a "rough" surface the coefficient of reflection does not approach unity except at low elevations.

As the angle of elevation of the transmitter drops, the ground reflection coefficient becomes independent of polarization and tends to the value of \( -1 \). At elevations of the order of several degrees the earth may be considered as a perfectly reflecting plane and the (dipole) transmission loss becomes independent of frequency. At the optical horizon the transmission loss again becomes frequency dependent owing to diffraction effects. It follows from the equation due to Domb and Pryce\(^8\), for the field strength at optical cut off, that the dipole transmission loss increases as \( f^{4/3} \). This result strictly applies only to smooth terrain below a standard atmosphere. Since the propagation loss increases rapidly as the transmitter passes below the horizon the discussion will be limited to optical paths.

7. Antenna Limitations

A directional receiving antenna is useful in reducing fading due to multipath propagation and in reducing the transmission loss. Although the maximum effective noise figure will be raised at the lower frequencies (compare Figs. 1 and 2), this is not serious because of the shortened duration of the daily maximum. For a given type of antenna, the ratio of physical to electrical aperture is a constant in the absence of frequency dependent losses. In the frequency range under discussion the antenna losses are small so that mechanical considerations, in fixing the maximum permissible antenna area, also fix the available capture area. The only other limitation, that of angular resolution, is imposed by tracking difficulties. If the beam width is predetermined by acquisition and tracking considerations then the

---


antenna gain will be constant, regardless of frequency. The use of an area-limited antenna will reduce the transmission loss at the rate of 6 db per octave. Since control over the aspect of the transmitting vehicle is not possible an omnidirectional transmitting antenna is generally required, which will not contribute to the frequency dependence of the transmission loss.

8. Transmitter Efficiency

At the power level of interest (of the order of several hundred milliwatts), the choice of transmitter lies between directly heated vacuum tubes and low power transistors. As the "state of the art" is rapidly changing, it is difficult to generalise the frequency dependence of the power efficiency of these devices. It is reasonable to assume that operation above about 600 Mc/s is prejudiced because of reduced efficiency and increased cost. Another factor associated with the transmitter and prejudicial to UHF operation is that of frequency instability. If the information bandwidth of the link is very much greater than the fluctuations in transmitter frequency, then the noise bandwidth (B) will not depend on frequency. On the other hand, if the noise bandwidth is determined solely by frequency instability then a loss in effective radiated power of 3 db per octave would result from the use of oscillators of fixed stability, that is (B/f = constant). For convenience in curve plotting, a loss of 2 db per octave (B/f = constant) will be assumed in plotting narrow band link performance.

9. Applications

The results of the preceding discussion will now be applied in determining the optimum frequency for a number of practical situations. The frequency dependence of the link performance is determined graphically from Eq. (1). The effective noise figure (N') expressed in decibels is plotted (Curves (a) and (b) in Figs. 1 and 2). The contribution to frequency dependence due to the bandwidth (B) and link attenuation (A) is given by a power law of exponent a. The overall link performance is therefore found by applying a linear bias to the N'(f) curves of magnitude 10 a db per decade. Each curve then expresses the required transmitter power in decibels above an arbitrary level as a function of frequency under the conditions specified for the link in question. The optimum frequency is then found by inspection and is tabulated in Table 1.

For a simple balloon-ground telemetry link in which horizon coverage is not required, the plane earth approximation to the path loss suffices. Dipole antennae and vacuum tube receivers are used in the interests of operating simplicity. For a broad band system the only frequency dependent factor is then the effective noise figure (N'). The frequency at which the required transmitter output power is a minimum is then determined solely by link noise. In Fig. 1 the (N') curves (a) and (b) show the relative power requirement in the range (50 to 1000) Mc/s. The optimum frequency (f_o) lies between 150 and 250 Mc/s, depending on sky noise conditions. The curves, being relatively broad, allow considerable latitude in the choice of frequency. For maximum sky noise the link performance is degraded by no more than 3 db from the optimum by a choice of frequency in the range 100 to 750 Mc/s. For a horizon limited system these curves must be weighted by the propagation loss factor (4 db per octave). In Fig. 1, curves (c) and (d) are derived in this way. Optimum frequencies are lowered by 50 and 100 Mc/s respectively and the system becomes more sensitive to deviations above, than below (f_o). The "3 db bandwidth" is also reduced by a factor of more than two. In practice the tolerance in (f_o) will be less than the figures quoted because of reduced transmitter and antenna efficiencies at the higher frequencies.

In order to take account of frequency instability an additional bias of -2 db per octave has been applied and plotted as curves (e) and (f) in Fig. 1. These curves, as well as applying to the narrow band horizon link also represent the broad-band, free space performance with isotropic (or constant gain) antennae. The system performance remains unchanged from 50 to 100 Mc/s but rapidly deteriorates above this range.

Turning to Fig. 2, N' curves (a) and (b) show the frequency dependent performance of a narrow band telemeter using a fixed receiving aperture for horizon coverage. These curves apply to also a broad band free space link. The optimum frequency lies between 150 and 350 Mc/s, depending on sky noise. The broad band link performance, derived as before, is plotted in Fig. 2 as curves (c) and (d). In this case the full advantage of operation in the optimum region (300 to 1000 Mc/s) is conditional upon the use of high efficiency transmitters. These results are summarised briefly in Table 1.

<table>
<thead>
<tr>
<th>Figure Reference</th>
<th>Optimum Frequency f_o (Mc/s)</th>
<th>System</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (a)</td>
<td>150</td>
<td>A</td>
</tr>
<tr>
<td>1 (b)</td>
<td>250</td>
<td>A</td>
</tr>
<tr>
<td>1 (c)</td>
<td>100</td>
<td>B</td>
</tr>
<tr>
<td>1 (d)</td>
<td>150</td>
<td>B</td>
</tr>
<tr>
<td>1 (e)</td>
<td>70</td>
<td>C, D</td>
</tr>
<tr>
<td>1 (f)</td>
<td>70</td>
<td>C, D</td>
</tr>
<tr>
<td>2 (a)</td>
<td>150</td>
<td>E, F</td>
</tr>
<tr>
<td>2 (b)</td>
<td>350</td>
<td>E, F</td>
</tr>
<tr>
<td>2 (c)</td>
<td>250</td>
<td>G</td>
</tr>
<tr>
<td>2 (d)</td>
<td>800</td>
<td>G</td>
</tr>
</tbody>
</table>

A: Broad band (dipole to dipole) plane earth link.
B: Broad band (dipole to dipole) horizon link.
C: Narrow band (dipole to dipole) horizon link.
D: Broad band (dipole to dipole) free space link.
E: Broad band (dipole to fixed aperture) free space link.
F: Narrow band (dipole to fixed aperture) horizon link.
G: Broad band (dipole to fixed aperture) horizon link.
10. Conclusion
On the basis of these results it is clear that a choice of optimum frequency is indeed possible for line of sight communications systems.

It should be pointed out that with the increasing use of low-noise solid state devices, the optimum frequencies will be raised until transmitter efficiency and stability become the limiting factors.

Acknowledgment
The topic of this paper was suggested by Dr. K. B. Fenton whose helpful criticism and that of other members of the Cosmic Ray Group of the University of Tasmania is gratefully acknowledged. The paper was prepared while the author had tenure of a Commonwealth Post Graduate Scholarship.

CONTRIBUTORS

T. E. Cummins is employed by the University of Sydney as Mechanical Technician in the workshop of the Electrical Engineering School. He was formerly a Laboratory Technician with 25 years' experience in electronics and mechanical and electrical fitting.

J. N. Davison is employed as Workshop Superintendent in the Electrical Engineering School of the University of Sydney. He was formerly a Laboratory Technician with 18 years' experience in electronics and mechanical and electrical fitting.

Paul Julian Edwards was born in Launceston, Tasmania, in 1938. He graduated Bachelor of Science from the University of Tasmania in 1959. Following a year's postgraduate research in cosmic radiation at high altitudes, Mr. Edwards was awarded the degree of Bachelor of Science with Honours (1st Class) in 1960. He is at present at the Physics Department of the University of Tasmania and holds a Commonwealth Post Graduate Scholarship. Mr. Edwards is currently engaged in upper atmosphere studies of ionizing radiation with balloon borne instruments.

A. C. Temby was born in South Australia in 1938. He received his secondary education at the Adelaide Technical High School, and graduated from the University of Adelaide in 1959 with the degree of B.E. with first class honours. He continued as a postgraduate student and completed a M.E. in the field of transistorized active filters, from which this paper is extracted.

Since October 1961, Mr. Temby has been employed in the Commonwealth Acoustic Laboratory, Sydney, N.S.W., doing research into psycho-acoustic problems.

Kenneth B. Whiting was born in the United Kingdom in 1933. He received the B.Sc. in Physics and Mathematics from London University in 1953 and the M.Sc. in Mathematics from the same University in 1961. Mr. Whiting was employed by the General Electric Company at their Stanmore Laboratories from 1954 to 1961. During this period he worked upon various advanced microwave duplexing systems for both C.W. and pulsed radar, with particular emphasis upon the design of ferrite components.

In 1961 he moved to Australia, joining the Weapons Research Establishment, Department of Supply, Salisbury, S.A. He is now engaged upon the development of microwave filters, ferrite devices, and other microwave components, for use in airborne equipment.

Mr. Whiting is an Associate Member of The Institute of Physics (U.K.), and a Graduate Member of The Institution of Electrical Engineers (U.K.).

Summary: The results of determinations of the probability density function of narrow-bandlimited, rectified, RC-filtered, electrical noise are given. As expected from considerations involving the central limit theorem, provided the filter time constant is long compared with the reciprocal pre-detector bandwidth, measurements show the output to closely obey gaussian statistics and the standard deviation is also in accordance with theoretical predictions. The results contradict several published reports that the statistics of the output of a linear rectifier with RC filter are non-gaussian.

Statistics of Rectified, RC-Filtered, Band-Limited Noise
R. Hurst* and P. J. Edwards*, M I R E E

1. Introduction
In a recent discussion of radiometer sensitivity, Christiansen and Högbo9m* imply that the probability density function (pdf) of rectified narrow-band noise after passage through a low-pass RC filter is non-gaussian. In particular, results obtained by Robinson2 are quoted in which the probability of a deviation from the mean power exceeding $6\sigma$ is stated to be 2%. The authors note the apparent superiority of a uniform integrator for which "the fluctuations are distributed according to the normal gaussian error function". Robinson also states that the probability distribution depends on the type of detector and smoothing filter used and quotes values for a linear detector and simple RC filter.

It is the purpose of this communication to point out that, providing the time constant, $\tau$, of the post detector RC filter and the noise bandwidth $B$ of the predetector filter satisfy the inequality $B\tau \gg 1$, the statistics of the output voltage and power are in fact gaussian for all practical purposes. This result is a natural consequence of the central limit theorem of statistics according to which, as $B\tau$ increases and the output filter sums an increasingly large number $(\sim B\tau)$ of independent (but not necessarily gaussian) noise voltages, the resultant distribution tends to the gaussian form.

2. Observations
We have measured the pdf of the RC filtered output voltage of a linear rectifier with high precision using a digital multichannel analyser. A block diagram of this instrumentation is shown in fig. 1. The 30 MHz radiometer was operated in the total power mode. Measured half-power and noise bandwidths were $140 \pm 5$ kHz and $170 \pm 5$ kHz, respectively.

The output of the rectifier, linear within $\pm 3\%$, was passed through an integrating amplifier with time constant $\tau$, variable over the range 0.5 ms to 10s. In order to minimise the effect of long term gain drifts, the integrated signal was passed through a differentiating amplifier with $\tau' = 10s$ prior to F.M. mode tape recording. The tape was read into a HP 5400A analyser operated as a sampled voltage analyser with a sample rate of 100 kHz. Data were typically accumulated for 5 minutes of real time. The measured linearity and precision of the data logging system used is better than $\pm 0.5\%$. Fig. 2 (a) shows a typical pdf. Also shown is the pdf of a 0.5 Hz square wave of known peak to peak amplitude used for voltage calibration. Fig. 2(b) shows the associated distribution function plotted on gaussian probability paper. It is evident that the rectified and filtered output voltages closely follows a gaussian distribution about the mean, $V_m$, in the range

3. Conclusions

Our measured values of $\sigma$ are in excellent agreement with this result (fig. 3) for a range of $\tau$ over more than two decades (1ms to 200ms) limited only by the high frequency response of the data logging system on the one hand and by gain instability and noise generator drift for longer integration.
times on the other. These limitations, which also apply to
a uniform integrator are of a practical rather than theoretical
nature and do not affect our conclusions in any way.

The results of our measurements should not be taken to
imply that the use of a uniform integrator has no advantage
over the use of a simple RC filter of similar noise bandwidth.
If a non-stationary noise signal is present, a uniform inte-
grator (or other more sophisticated filter) will generally give
higher signal to noise ratios, as pointed out by Cooper.6

Acknowledgments

We acknowledge the financial support of the New Zealand
University Research Grants Committee and the assistance
of Mr L. Amon in data processing. We thank Dr W. R.
Webber, Director Space Science Centre, University of New
Hampshire for providing the radiometer.

5. Cooper, B. F. C., "Post-Detector Filtering in Radiometry ",

B. H. Briggs

B. H. Briggs was born in Bradford, York-
shire, in 1923, and educated at Bradford
Grammar School and Clare College, Cam-
bridge. In 1943 he joined the Telecom-
munications Research Establishment, Mal-
vern, and worked for three years on radar
circuit design.

He returned to Cambridge in 1947 to carry
out research on the ionosphere and held
various posts at the Cavendish Laboratory
until 1952, when he was appointed Senior
Lecturer, and later Reader, at the Department
of Physics, University of Adelaide.

Here, he has continued to pursue his
interests in ionospheric research, which relate
to irregularities and movements observed by
various radio techniques.

A. Cantoni

Mr. Cantoni's biography appeared in the
April, 1972 issue of Proceedings, Vol. 33,
No. 4, p. 146.

P. J. Edwards

Paul J. Edwards graduated B.Sc. with First
Class Honours in Physics from the University
of Tasmania in 1960. He received a Ph.D. in
Physics from that University in 1964 and was
then appointed Lecturer in Physics at the
University of Adelaide. Dr. Edwards was
Assistant Professor in the Division of Atmos-
pheric and Space Sciences at South West
Centre for Advanced Studies (Dallas, Texas)
1965-66. He is currently Associate Professor
of Physics at the University of Otago, New
Zealand and is engaged in geophysical and
astronomical research using radio and optical
techniques.

Leslie Germany

Leslie Germany, a Fellow of the Institution
of Electronic and Radio Engineers, and a
Fellow of the Royal Television Society, is a
director of Pye TVT Ltd. in Cambridge, Eng-
land. During the years 1940 to 1945 he
was concerned with the development of
ground radar systems and in 1945/6 assisted
in the development of the first post war
Television receiver. When in 1946 the Pye
company entered the field of television broad-
casting systems he helped to initiate its
broadcast division, and since this time has
been responsible for the development of the
Company's television studio equipment. His
first association with Australian broadcasters
was in 1956 during the inauguration of the
Australian television service.

R. B. Hurst

R. B. Hurst completed B.Sc. (Hons.) degree
in Physics at Otago University, N.Z. in 1971.
At present he is engaged in Ph.D. work in
radio-frequency stochastic phenomena.

C. B. B. Wood

Charles Barrington Wood was commis-
sioned in the R.A.F. in the Second World War
and became Officer Commanding No. 1
Radar Assembly Unit. He joined the B.S.C.
Research Department following his war ser-
vie holding the appointments of Head of the
Television Section and, since 1971, Head of
the Engineering Information Department. He
has been for many years closely involved in
investigations and experimentations in the
television field and more recently in relation
to colour television. He has received a number
of technical prizes and awards for his writings
and more recently was honoured by the Queen
with the award of the M.B.E. He is a Fellow
the BKSTS and SMPTE.
Correlation search for dispersed radio emission from the galactic centre

We report here the results of a search for fluctuations in the intensity of radio emission from the vicinity of the galactic centre using a two-channel cross-correlation technique.

The initial incentive to look for a time variable radio flux from the region of the galactic centre arose primarily from the reports of impulsive gravitational radiation. Searches for impulsive radio events have been reported recently. These observations have been restricted by the necessity to recognize individual pulses in the presence of fluctuation noise. They refer therefore to pulses substantially greater than the r.m.s. level. For example, Hughes and Retallack report pulses of height $> 85$ f.u. at 888 MHz occurring at a rate of $\approx 10^4$ s$^{-1}$. Other observers have utilized the expected delay in pulse arrival times at different frequencies due to velocity dispersion in the interstellar medium in order to reject impulsive noise of local origin.

We utilize the dispersion phenomenon in our search by looking for a delayed, correlated component to the intensity fluctuations at two neighbouring frequencies. Unlike those mentioned above, this method allows the detection of dispersed transients well below the r.m.s. fluctuation noise level. The technique is therefore particularly suited to the detection of low level signals, distinguishable only by their dispersed character. We describe its use in a search for dispersed noise having an arbitrary probability density function, for example pulses with amplitudes below those already reported. The results of a concurrent examination of the data for individual large pulses will be reported elsewhere.

If $x(t)$ and $y(t)$ are the envelope detected outputs of the two telescope channels, their time lagged cross product is

$$R_c(t) = x(t + \tau)y(t - \tau).$$

over the observing time interval. The values of this mean lagged product are the cross-correlation estimates. Plotted as a function of the time lag, $\tau$, they constitute a correlogram which may be examined for evidence of a dispersed fluctuation noise component. Provided the characteristic time scale of the fluctuations is shorter than the time delay due to dispersion in the interstellar medium, a characteristic source signature, distinguishable from local noise, should appear in the correlogram. For non-periodic signals, a random pulse train for example, this signature takes the form of a positive peak $R_c(0, \tau)$ at a time lag, $\tau = \Delta t$, the time delay expected between channels separated in frequency by $\Delta \nu$ (MHz).

$$\Delta t(s) = 8 \times 10^6 DM(\Delta \nu/\nu^2)$$

where $DM$ is the dispersion measure of the variable source.

The height of the peak will be proportional to the mean square value of the correlated flux density variations. Its width indicates their time scale. The peak amplitude may be depressed and the characteristic width may be increased by:

1. dispersion within the non-zero channel bandwidths;
2. a non unique value of dispersion measure;
3. post-detector filtering; and
4. decorrelation between the signals occurring at the source or in the medium. The signal to r.m.s. noise ratio of a correlation estimate, $R_c(\tau)$, increases as the square root of the product of the channel bandwidth and observing time. This indicates the high sensitivity of the correlation method for long observing times and wide bandwidths. In practice the presence of slow correlated instrumental drifts reduces the sensitivity below that expected.

The 150 m meridian transit telescope of the University of Otago, situated at the Invermay Radio Observatory near Dunedin, New Zealand (40°S) was used in the search. The telescope was operated at two frequencies, 120 MHz and 130 MHz as a two element phase switched interferometer in order to minimise interference and reduce instrumental drift. The effective aperture of the telescope is $5 \times 10^3$ m$^2$. The telescope was steered to declination $\delta = -29.0^\circ$S for the two months search duration, from June through July 1973. The area of the search was a (2° x 8°) strip centred on the galactic centre covering essentially the same region as that covered by Hughes and Retallack. It included both Sgr A and the source they reported at $\alpha(1950) = 17h 48.2m$, $\delta(1950) = -28^\circ 58'$. The duration of transit between the half power points of the antenna beam was approximately 50 min. The system temperature reached a maximum of 2,200 K during the transit of the galactic centre. The r.m.s. temperature fluctuation was typically 1 K (6 f.u.) for a post detector integration time constant of 10 s. Calibration was performed periodically by noise injection at the antenna.

The receiver outputs, following envelope detection were low pass filtered with time constant 10 s, this value being approximately the pulse broadening time within the receiver bandwidth of 1 MHz expected for a hypothetical impulsive source having a dispersion measure of $2 \times 10^3$ cm$^{-3}$pc. The integrated signals were then AC coupled with time constant of 100 s (in order to reduce slow drifts) to the data logging system. This consisted of a dual channel chart recorder used for visual detection of dispersed events and two dual channel magnetic tape recorders on which the signals were written in FM mode together with a stable clock reference signal used for synchronizing the tape read-out. The data were sampled at 0.3 Hz. Cross correlation analyses were performed on 30 min data blocks.

For the purposes of envelope correlation only those records of highest quality were used. Any 30 min blocks which contained non dispersed individual pulses greater than about 20 f.u. were rejected from this analysis. Twenty-four individual correlograms, representing 12 h (20% of all data) were computed for delay times, $\tau$ from $-400$ s to $+400$ s. A delay of $\tau = +400$ s corresponds to $DM = 10^4$ pc cm$^{-3}$. Figure 1 shows the computed average of the twenty-four correlograms. The most noticeable feature is the positive peak at zero delay. This peak, $10^{-2}$ square degrees ($4 \times 10^4$ f.u.$^2$) in height indicates the presence, in spite of careful selection of data, of undispersed system temperature fluctuations with a r.m.s. value of $10^3$K. For time lags

![Fig. 1 Average cross-correlogram of galactic centre radio wave intensity fluctuations at frequencies of 120 and 130 MHz. Crosses indicate the signature due to a random sequence of dispersed (DM = 5 x 10^3 pc cm^-3) calibration pulses of variance 1.2(f.u.)^2.](image-url)
\(|\tau| \geq 20 \text{ s} \ (DM \geq 500 \text{ pc-cm}^{-3})\) the cross covariance fluctuations have an r.m.s. value of 0.1(f.u.)^2. Any dispersed noise component with variance exceeding $2 \times 10^{-3}(K)^2$ corresponding to $0.8(f.u.)^2$ for a source near the galactic centre, would be clearly visible.

To illustrate this and to check the computation procedures, a random sequence of dispersed pulses was generated, added to the raw data and processed. The resulting correlogram signature shown in Fig. 1 is consistent in location ($DM = 5 \times 10^{-3}$), shape (full width at half maximum = 20 s) and the amplitude of each pulse was $s = 1.5$ f.u., approximately 94% of the r.m.s. noise level.

We conclude that a conservative upper limit of 0.8(f.u.)^2 may be assigned to the mean square flux density variations at 125 MHz of a variable source located either at the position of the impulsive source previously reported or in the direction of the galactic centre. This upper limit remains less than about 3(f.u.)^2 for sources with $\delta(1950) = 29.0 \pm 1.0^\circ$, and $\alpha(1950) = 17$ h 44 min ± 18 min. In order to attach a physical significance to our results and to relate them to other observations, we consider several sources of variability.

As a possible scenario we assume the previously reported pulses* to be the chance superposition of more frequent, randomly occurring subpulses generated perhaps in stellar events in the region of the galactic centre. It is easy to calculate the variance $\sigma^2(s) = \langle s^2 \rangle - \langle s \rangle^2 = a \sigma^2/T$ and the mean flux density $\langle s \rangle = a e$ for impulse, $e$, and resolving time, $T$. If the 'giant' pulses reported contain $m$ unresolved subpulses in time $T$ we may assume for illustrative purposes that, as reported $me = 8.5 \times 10^{-22} J \text{ m}^{-2} \text{ Hz}^{-1}$ (85 flux unit seconds), $T \sim 1$ s, and the 'm-fold' pulse rate $a P(m - 1, T) \sim 10^{-4}s^{-1}$ where $P(m - 1, T)$ is the Poisson probability of observing $(m - 1)$ pulses in time interval $T$. For $m = 3$, $a \approx 6 \times 10^{-3}s^{-1}$, and $\sigma^2 \approx 50/(f.u.)^2$. A value of $T \approx 20$ s is appropriate to our observations at 125 MHz, taking into account post detector integration and convolution with the receiver bandwidths of dispersed impulses generated by a source or sources having $DM \leq 5 \times 10^{-3}$ pc cm$^{-3}$. The amplitude of the correlogram signature would in this case be $\approx 2.5$ (f.u.)^2, more than twice the amplitude of the facsimile signature shown in Fig. 1 and $\approx 25$ times the r.m.s. correlogram noise. Other values of $m$ do not significantly affect this result which clearly shows the 'giant' pulse model to be unlikely. Raising $m$, will in the limit, simulate a physical situation in which gaussian noise fluctuations are generated at the source. Our results applied, say to Sagittarius A, place a conservative upper limit to its correlated variance over a 10 MHz band centred at 125 MHz of $\sim 1(f.u.)^2$ for periods from 50 to 500 s. This upper limit to the correlated variance of $1(f.u.)^2$ may also be used to place a limit on the flare star contribution to the time averaged brightness temperature of the galactic centre region. If the flare stars are concentrated in a 2 deg^2 region of thickness $\Delta (DM) \approx 10^4$ pc cm$^{-3}$, a realistic distribution^9 of flare event sizes then gives $T(n) \leq 5 K$. This is a fractional contribution of $<10^{-2}$ to the mean temperature of the region.

We acknowledge financial support from the New Zealand University Research Grant Committee.

P. J. Edwards
R. B. Hurst
M. P. C. McQueen

Invermay Radio Observatory,
Physics Department,
University of Otago

Received November 19, 1973.

Cross Correlation Studies of Stellar Variability using two Telescopes

P. J. Edwards, R. B. Hurst and G. Christie

Physics Department, University of Otago

The study of short time scale phenomena of astronomical interest is often handicapped by the presence of wide band noise. In particular, in the study of the optical variability of stars, such noise arises from atmospheric scintillation, extinction variations, sky radiance variations and photon statistics. For example if in order to resolve rapid stellar variability we reduce the length (τ) of the time intervals over which a photometric record is integrated and sampled, we find that the fractional noise fluctuations increase, eventually obscuring any intrinsic variability.

If the variability has a periodic component as do the binary dwarf novae studied by Warner and Nather (1971), the appearance of fundamental and harmonic line structure in the power spectrum of stellar flux variation provides a recognizable signature. The coherence of a periodic signal over the total record length T, where τ/T < 1 allows its recovery even though its mean square amplitude δφ may only be of order δφ^2 ΔT (τ/T) where δφ^2 is the variance (noise power) of the sampled record. Threshold detection of periodic variability of X-ray and dwarf nova binaries at a level of δφ^2 ΔT (τ/T) that is Δm ~ 10^-6 mag, has been reported by Middleditch and Nelson (1973), Warner and others at periods shorter than a few tens of seconds.

The detection of low level stochastic variability (‘flickering’) is less easy. Flickering on these time scales corresponding to bandwidths, B ≥ 0.1 Hz might be expected of a variety of irregular variables including red dwarf flare stars, U Geminorum and X-ray binary stars. The sources 2 U 0900-40, 2 U 1700-37 and Cyg X-1 for example all exhibit ‘shot noise’ X-ray variability. Avni and Bahcall (1974) have drawn attention to the importance of observing any associated optical fluctuations. They suggest these may be no more than 10^-7 mag in size and ‘very difficult if not impossible to observe from ground based telescopes because of atmospheric scintillations’.

Stochastic variability such as that arising from ‘shot noise’ X-ray heating is less easy to detect because its spectrum is likely to be a continuum, not necessarily distinguishable from the spectrum of wide band noise mentioned above. Frequency domain filtering may consequently be of limited value although flickering of the order of 10^-2 mag has been detected from power spectra. In principle, variability of this type could be detected and measured by careful normalizing procedures utilizing observations of suitable comparison stars.

Scintillation noise is characterized approximately by a flat variance spectrum extending over a typical bandwidth of B ~ 100 Hz and having Gaussian statistics when measured in stellar magnitudes. The upper frequency limit is determined by meteorological conditions, and also by the size of the aperture which acts as a low pass filter. If the photon count is sufficiently small shot noise will dominate the noise spectrum over a wider bandwidth. In both these limiting cases, the fluctuation noise above 0.1 Hz can usefully be considered to be band limited Gaussian white noise. A number of well known results from stochastic theory (Bendat and Piersol, 1971) follow directly.

In particular the normalized root mean square error in the variance estimate of the scintillating program star,

\[ \epsilon \approx \frac{\sigma_n^2}{<S>_T^2} / (BT)^{1/2} \]

Now, for B = 1.0 Hz, T = 10^4 s we therefore expect an accuracy of order 1% in the normalized variances of both the program and comparison stars. Providing no systematic errors arose in the subtraction of comparison variance from program variance, we would expect to be able to detect an excess variance of this order. For a small telescope

\[ \frac{\sigma_n^2}{B <S>_T^3} \sim 10^{-4} \text{(mag)^3/Hz} \]

under typical conditions. The corresponding statistical detection limit is \( \sim 10^{-3} \text{mag} \). Systematic errors associated with the normalization and subtraction of variances may raise this to \( \sim 10^{-2} \text{mag} \). For instance, while the scintillation noise variance may be normalized to the square of the mean, \( <S>_T^2 \), photon noise power must be normalized to the mean itself, \( <S>_T \).

The correlation technique proposed in this paper provides a means whereby systematic errors may be largely eliminated and the full statistical threshold limit achieved. Simultaneous observations of a suspected variable by an array of two telescopes separated by a base line exceeding the noise correlation distance are proposed.

The zero lag covariance estimate \( R_{xy}(0) \) will have an expectation value

\[ R_{xy}(0) = <x(t) \cdot y(t)> = \sigma_i^2 \]

equal to the variance intrinsic to the star. The length of the base line required to decorrelate atmospheric noise ranges typically from a few metres for scintillation noise to tens of kilometers for slower extinction and sky radiance changes. The normalized rms error in a zero lag cross covariance estimate is given by

\[ \epsilon \approx \frac{\sigma_n^2}{<S>_T^3 (2BT)^{1/2}} \sim 10^{-6} \text{(mag)^3} \]

for B = 1 Hz, T = 10^4 s. The technique is therefore capable in principle of detecting broadband flickering of the order of \( 10^{-3} \text{mag} \).

We have carried out feasibility studies of this technique using the 40 cm University telescope at Signal Hill Observatory, Dunedin, and the 30 cm Cassegrain telescope operated by the Dunedin Astronomical Society. The two telescopes are separated by a distance of 5.6 km.
The data were logged using the dual channel magnetic tape systems previously described by Edwards (1972), which we have already used in a cross correlation study of galactic radio noise (Edwards et al. 1973). The data records obtained with this system were synchronized to within several milliseconds using VNG time signals. Crystal controlled reference oscillators were used. Figures 1 and 2 show auto correlograms derived from 5 minute observations of the optical counterpart to 2 U 0900-40. Equal integration and sampling intervals of $\tau = 0.9766$ s were used in these runs. The data were high pass filtered by subtraction of a least squares fit trend lines before cross correlation. Figure 3 shows the cross covariance plot. The rms fluctuation in this plot is about $1 \times 10^{-5}$ (mag)$^2$. Hence a peak at zero lag (corresponding either to intrinsic variability or to correlated noise of non stellar origin) exceeding say $5 \times 10^{-5}$ (mag)$^2$ would have been easily detected. With longer observing times and larger apertures a detection threshold below $10^{-6}$ (mag)$^2$ should be feasible. The technique could well be extended to a search for stochastic variability in X-ray flux from the binary and other X-ray sources.

We thank Dr M. Thomas and Mr W. Allen for carrying out observations at the Beverley Begg Observatory. We acknowledge financial support provided by the New Zealand University Grants Committee.

CROSS CORRELATION DETECTION OF RAPID STELLAR VARIABILITY USING TWO TELESCOPES

P.J. Edwards & R.B. Hurst
Physics Department
University of Otago
Dunedin, New Zealand

M. Thomas
Beverly-Begg Observatory
Dunedin, New Zealand

The study of rapid, small amplitude stellar variability is handicapped by scintillation and photon noise. Flickering on time scales corresponding to bandwidths \( B \geq 0.1 \) Hz may be expected of a variety of irregular variables including dwarf novae and X-ray binaries. Stochastic variability is harder to detect than periodic variability because its spectrum is likely to be a continuum, not necessarily distinguishable from the spectrum of the noise due to scintillation, extinction variations, sky radiance variations and photon statistics. The proposed correlation technique provides a means whereby systematic errors may be largely eliminated and the full statistical threshold limit achieved. Simultaneous observations of a suspected variable by an array of two telescopes separated by a base line exceeding the noise correlation distance are proposed.

The zero lag covariance estimate \( R_{xy}(0) \) of the irradiance fluctuations \( x(t), y(t) \) will have an expectation value \( R_{xy}(0) = \langle x(t) \cdot y(t) \rangle = \sigma_g^2 \) equal to the variance intrinsic to the star. The length of the base line required to decorrelate atmospheric noise ranges typically from a few metres for scintillation noise to tens of kilometers for slower extinction and sky radiance changes. The normalised rms error in a zero lag cross covariance estimate is given by \( \varepsilon \approx \sigma_n^2 / \langle S \rangle^2 (2BT)^{1/2} \sim 10^{-6} \text{mag}^2 \) for \( B = 1 \text{ Hz}, T = 10^4 \text{s} \), where the spectral density of the mean square noise, \( \sigma_n^2 / \langle S \rangle^2 \sim 10^{-4} \text{mag}^2 / \text{Hz} \) for a small telescope. The technique is therefore capable in principle of detecting broadband flickering of the order of \( 10^{-3} \) mag. We have confirmed the sensitivity of this technique using two telescopes separated by a distance of 5.6 km.
DETECTION OF STELLAR FLICKERING USING ARRAYS OF TELESCOPES

P.J. Edwards  
W. Allen & M. Thomas

Physics Department  
Beverly-Begg Observatory
University of Otago  
Dunedin, New Zealand

Rapid stellar variability of a stochastic type is characteristic of a variety of irregular variables including dwarf novae, x-ray binaries and nova-like variables. Rapid, small amplitude variability is often difficult to observe because of scintillation and photon statistics. Irregular variability is particularly difficult to observe because the power spectrum of the flickering is often not readily distinguishable from the spectra of atmospheric noise and shot noise. However, simultaneous observations of a variable with an array consisting of two or more telescopes separated by a distance exceeding the noise correlation length allows the intrinsic variability to be distinguished from noise.

On short time-scales, the main sources of noise in photoelectric observations with a small telescope are photon counting statistics and atmospheric scintillation. Photon noise is essentially shot noise, bandlimited by the signal processing system. Atmospheric scintillation also has an approximately flat spectrum, out to a cut-off frequency dependent on atmospheric conditions, and typically of order $10^2$ Hz. The power spectral density of photon noise is proportional to the mean photon flux. For scintillation noise, however, the spectral density is proportional to the square of the mean photon flux. Consequently, for increasing star brightness, there is a transition from photon noise to scintillation noise, the latter being greater with a 40 cm aperture for stars brighter than about $B = 9$ mag. Other sources of noise, usually of less importance, are telescope tracking errors, and extinction and sky radiance changes.

Cross-correlation analysis has been proposed by Edwards et al (1975, 1976) as a technique for measuring short time-scale aperiodic variability ("flickering") in the presence of this noise. To provide two or more channels in which the noise components are statistically independent, an array of telescopes is suggested, separated by large enough distances to decorrelate the atmospheric effects. Multi-colour photometry using a single telescope is not sufficient for two
reasons. Firstly, scintillation noise will remain partially correlated, in channels separated in the optical spectrum by as much as an octave. Secondly, in some cases the variability may be confined to a spectral line or group of lines only.

Photon noise is uncorrelated between the telescopes of the array except for the higher-order Hanbury-Brown/Twiss effect which is entirely negligible for the time-scales and long baselines used. Atmospheric scintillation fluctuations become uncorrelated over a baseline of the order of metres. Other atmospheric effects, such as extinction changes and sky brightness, remain partially correlated over distances of kilometres, but effects correlated over these distances have slow time-scales. For the short time-scales investigated, and at levels of sensitivity attained to date, no systematic correlated component due to atmospheric effects has been detected.

In the following work, signal and noise variances are expressed in squared stellar magnitudes. An r.m.s. fluctuation of 1% therefore corresponds to a variance of approximately $10^{-4}$ (mag)$^2$. For example, scintillation noise typically has a power spectral density of $10^{-4}$ (mag)$^2$ Hz$^{-1}$ for the small telescopes used (Edwards et al, 1975) to verify the assumptions on which this technique is based.

Correlation Technique

The presence of intrinsic variability may, in principle, be detected by comparing the observed variance in the irradiance of the program star with that of a comparison star. Both photon and scintillation noise may usefully be considered to be bandlimited Gaussian white noise when measured in stellar magnitudes. A number of well known results from stochastic theory (Bendat and Piersol, 1971) follow directly.

In particular, the root mean square error in the measurement and subtraction of two variances is

$$
\epsilon \approx \sigma_n^2 \left( \frac{2}{B T} \right)^{1/2}
$$

for noise variances $\sigma_n^2$, noise bandwidths $B$, and observation times $T$. Providing no systematic errors arose we would expect to detect an excess variance of a few times $\epsilon$. However, Clarke and Wyllie (1977) and others have pointed out the difficulties of this procedure as it affects, for example, the detection of rapid fluctuations in the spectra of Be stars.

For a telescope of diameter, $D$/cm, the variance of scintillation noise,

$$
\frac{\sigma_n^2}{B S^2} \sim 2 \times 10^{-2} D^{-4/3} \text{ [in (mag)$^2$/Hz]}
$$
under typical conditions where $<S>$ = mean irradiance.

The correlation technique proposed in this paper provides a means whereby systematic errors may be largely eliminated and the full statistical threshold limit achieved. Simultaneous observations of a suspected variable by an array of two or more telescopes separated by a baseline exceeding the noise correlation distance are proposed.

The zero lag covariance estimate $R_{xy}(0)$ will have an expectation value

$$R_{xy}(0) = <x(t) \cdot y(t)> = \sigma_x^2$$

equal to the variance intrinsic to the star. The normalized r.m.s. error in a zero lag cross covariance estimate is given by

$$\varepsilon \approx \frac{\sigma_n^2}{<S>_2(2BT)^{1/2}} \sim 10^{-6} \text{(mag)}^2$$

for $B = 1 \text{ Hz}$, $T = 10^4 \text{ s}$, $D = 40 \text{ cm}$. The technique is therefore capable in principle of detecting broadband flickering of the order of $10^{-3} \text{ mag.}$ by the use of two small telescopes. The signals from an array of $m$ telescopes may be cross-correlated in pairs to form $m(m-1)/2$ correlograms. It may be shown (Hurst, 1977) that under typical conditions these covariances are independent, gaussian variables. As a consequence, an $m$-fold array lowers the detection threshold by a factor $(m(m-1)/2)^{1/2}$, ($\sim m^{1/2}$, for large $m$) relative to a two telescope array.

We have carried out feasibility studies of this technique using the 40 cm University telescope at Signal Hill Observatory, Dunedin, and the 30 cm Cassegrain telescope operated by the Dunedin Astronomical Society. The observatories are separated by about 5.6 km, in a NE-SW line. In principle, the array could have been extended to include other observatories. In practice, experience with the present array situated under urban skies of poor photometric quality has shown that simultaneous data from three telescopes would have been rarely available.

For the earliest cross-correlation observations, the signals were recorded separately on magnetic tape at each observatory. Synchronization to within a few milliseconds was accomplished by starting the recordings at pre-arranged times. To provide turn-on times and reference frequencies for the recordings, each observatory used a stable clock synchronized to radio time signals from station VNC. Subsequent analysis was by computer. Some of these results have already been published (Edwards et al, 1975).
Since then, a radio telemetry system has been installed allowing direct transmission of data from Beverly Begg Observatory to Signal Hill. The signal is transmitted as a pulse train. This eliminates synchronization difficulties and has allowed us to perform correlations in "real time".

Observational Results

The work so far has been confined to stars which have either been reported as possessing rapid variability, or for which there are good theoretical grounds for expecting such variability. Some examples in both categories are discussed below.

HD 77581 is the optical counterpart to the compact galactic x-ray source 3U 0900-40 or Vela X1 (Hiltner et al, 1972). It is a bright early-type member (spectral type B0.5 Ia), $V = 6.87$ (McClintock et al, 1975), of a binary system. The x-ray emission is believed to result from the heating to extreme temperatures of matter falling into the other member of the binary system, which is a small massive companion. The x-ray flux is likely to interact with matter near the condensed object, or with the surface of the companion star (Avni and Bahcall, 1974) producing flickering fluorescent optical emission of order $10^{-3}$ mag. This star has been observed on several occasions at Dunedin. A correlogram already published (Edwards et al, 1975) implies an upper limit of $5 \times 10^{-5}$ (mag)$^2$ for the variance of any flickering of the B-filtered irradiance on time-scales of $1$ s - $50$s. The most sensitive result obtained to date is shown. This was obtained in a run lasting 28.5 min beginning 1143 UT on 3 March 1976. The lag time scale on this correlogram is centred at zero, where any peak caused by intrinsic variability of the star would be located. An upper limit of $2 \times 10^{-5}$ (mag)$^2$ may be inferred for any flickering with time-scales 20ms - 200ms, again for the B-filtered irradiance.

CD-42° 14462 is a nova-like variable which exhibits both narrow band activity (Warner, 1973) and wide band activity (Hesser, Lasker and Osmer, 1972). It was monitored for high frequency ($f > 0.2$ Hz) variability in B for several nights in August and September, 1978. On September 1, 1978, sustained irregular variability of the order of $5 \times 10^{-3}$ mag. with an e-folding time of $\sim 10$ seconds was detected in contiguous 5 minute correlograms over a period of half an hour. Our observations suggest however that high frequency variability above this threshold was relatively infrequent over the observing period.

In conclusion we believe that this technique when used under good photometric conditions with long integration times and multiple large apertures offers promise in the detection and observation of high frequency stochastic variability from condensed objects such as white dwarfs, neutron stars and black holes with a sensitivity better than $10^{-8}$ (mag)$^2$/Hz.

We thank Mr. Singleton for assistance with observations.
MEASUREMENT OF NON-PERIODIC RAPID STELLAR VARIABILITY using the REAL TIME CROSS-CORRELATION of the TWO TELESCOPE SIGNALS

Figure 1: Block diagram of University of Otago cross-correlation two telescope array.
Fig. 2: Cross-correlogram of B-filtered irradiance fluctuations of HD77581 observed for 28.5 min.

Fig. 3: Cross-correlogram of B-filtered irradiance fluctuations of CD-42° 14462 observed for 10 min. Integration time constant = 1s; differentiating time constant = 20s.


Very Low Brightness Fluctuations in the Mid-Latitude Aurora

G. R. STANLEY and P. J. EDWARDS

Physics Department, University of Otago, Dunedin, New Zealand

From ground-based photometric observations made near Dunedin, New Zealand, at $L = 2.71$, weak mid-latitude electron precipitation is detected through the appearance of a characteristic signature in the cross correlogram of $\lambda 3914$ and $\lambda 5577$ brightness fluctuations. The method is sufficiently sensitive to detect rms brightness fluctuations at 3914 Å, as low as 0.03 R in an observing time of several minutes. During magnetically quiet periods with $Kp \leq 3$ we observe that $\sigma(B_0)$, on a time scale of seconds, falls below this limit, which corresponds to a precipitating electron flux of $\sim 200 \text{ el} \text{ cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$ at $40 \text{ keV}$ providing $\sigma(B_0) \sim \langle B_0 \rangle$. For $Kp \geq 5$, there is an approximate increase of 5.6 in flux magnitude per unit $Kp$.

INTRODUCTION

The characteristics of mid-latitude nightglow emission in the first negative bands of $N_2^+$ remain uncertain, particularly during magnetically quiet nights. The intensities of the $(0, 0)$ band at 3914 Å and the $(0, 1)$ band at 4278 Å are of special interest because the measurement of either permits estimation of the total columnar ionization rate and hence of the energy flux of energetic particles precipitating in the upper atmosphere. Observations of the temporal and spatial properties of the mid-latitude $N_2^+$ emissions may therefore be expected to contribute to the continuing debate concerning maintenance of the night-time ionosphere [Gough and Collin, 1973; Potemra and Rosenberg, 1973] as well as to an understanding of the morphology and mechanisms of charged particle precipitation at middle latitudes.

The question of whether the first negative band system is a permanent feature of the night airglow is still unresolved. Estimates of the mean brightness of the $\lambda 3914$ $(0, 0)$ band vary widely. Hirao et al. [1965] estimate 'a few rayleighs' at $L = 1.26$. Yano [1966] gives 10 R at $L = 1.9$. An upper limit of 1 R at mid-latitudes is given by Broadfoot and Huntley [1966] at $L = 1.65$, and an upper limit of 0.3 R at $L = 2.0$ for $Kp \leq 3$ by Jacka et al. [1970]. Potemra and Zmuda [1970, 1972] reviewed satellite, rocket, and ground-based observations of precipitating electrons ($E > 40 \text{ keV}$) and first negative $N_2^+$ emission. They concluded that the average fluxes of dumped electrons with $E > 40 \text{ keV}$ inferred from mid-latitude satellite observations are responsible for the excitation of a permanent $N_2^+$ airglow with a $\lambda 3914$ brightness range from about 0.2 R to about 3 R under quiet conditions for $2 < L < 3$. Potemra and Zmuda state the view that precipitating electrons are an important source of ionization in the D region. The minimum brightness of 0.2 R corresponds to an ionization rate of $5 \times 10^6 \text{ cm}^{-2} \text{ s}^{-1}$ or, equivalently, an incident electron flux ($E > 40 \text{ keV}$) of $\sim 6 \times 10^9 \text{ el} \text{ cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$ as typically observed by low-altitude satellites [O'Brien, 1964, 1965]. However, O'Brien also observed that for $2 < L < 3.5$ the precipitation ($E \geq 40 \text{ keV}$) sometimes fell below the detection threshold of $\sim 20 \text{ cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$ for the Injun 3 satellite. Such a figure would be consistent with the low electron fluxes ($E > 40 \text{ keV}$), $\sim 7 \text{ cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$ at $L = 2.5$ [O'Brien et al., 1965] and $\sim 4 \text{ cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$ at $L = 3.5$ [Gough and Collin, 1973], that have been measured in mid-latitude rocket flights during very quiet magnetic conditions and would suggest that on occasions the resulting $\lambda 3914$ emission is about 0.01 R or less.

In passing we note the experimental difficulties in unambiguously separating the dumped and trapped energetic electron fluxes in the satellite studies, which have, in the main, used wide-angle detectors. We also note the difficulties, well recognized in the literature, of measuring the low-brightness 3914-Å emission in the presence of background starlight.

The present study is concerned only with those fluctuations in the $N_2^+$ light which take place on a time scale of seconds and which are accompanied by correlated fluctuations in green light having the delay characteristic of the O($S_2$) green line at 5577 Å. We have used a cross-correlation technique to detect very small fluctuations in $\lambda 3914$ emission due to time-varying ionizing particle fluxes.

Genuine particle precipitation gives rise to a peak in the cross-covariance estimate

$$R_{\nu\delta}(\tau) = \langle B_\nu(t) \cdot B_\delta(t + \tau) \rangle$$

at a time lag $\tau \leq 1$ s, where $B_\nu(t)$ and $B_\delta(t)$ are the sky brightness fluctuations in the $\lambda 3914$ and $\lambda 5577$ emissions, respectively. This lagged covariance peak introduces an asymmetry into the cross correlogram of violet and green light fluctuations and so may be readily distinguished from sky brightness changes due to artificial sources or to changing atmosphere extinction. This technique provides a sensitive method of detecting time-varying mid-latitude auroral emissions. Using this method, we are able to detect $(0, 0)$ band brightness fluctuations, on a time scale of seconds, having a rms value greater than 0.03 R in a period of several minutes.

OBSERVATIONS AND RESULTS

Observations of mid-latitude aurora have been made near Dunedin, New Zealand, at Invermay Radio Observatory (geographic latitude, 45.87°S, geographic longitude, 170.41°E; invariant magnetic latitude, 52.60°; and $L = 2.71$) since 1973. Figure 1 shows the dual nitrogen band and oxygen line photometer setup, results from which are reported here. The photometers have been calibrated for absolute brightness measurements with a standard tungsten filament lamp. Operating parameters are summarized in Table 1. Note that the glass filters used enable the acceptance solid angle and therefore the "geometric factor" [O'Brien et al., 1965] to be increased well above that permitted by the use of multilayer interference filters, with a consequent improvement in sensitivity.

Typical photometer traces are shown in Figure 2 for moderately disturbed conditions, in which the 24-hour sum $\sum Kp$ is 39. Figure 3 shows the cross correlogram computed from digitized data taken from this period. Note that the asymmetric shape and positive lag of the peak, of height 0.06 R$^2$, clearly show that the optical fluctuations are due to a low-intensity
aurora. Also the high degree of correlation confirms that most of the signal is due to genuine auroral activity, as is evident from Figure 2. Values of \(\sigma(B_v)\) and \(\sigma(B_o)\) are 0.30 R and 0.26 R, respectively, where \(\sigma(B_v)\) and \(\sigma(B_o)\) are the rms values of the \(\lambda 3914\) and \(\lambda 5577\) signal fluctuations, giving \(\sigma(B_o)/\sigma(B_v) = 0.9 \pm 0.2\). That is, the \(\lambda 3914\) and \(\lambda 5577\) fluctuations are approximately the same magnitude, as is frequently observed in auroral zone studies [e.g., Henriksen, 1973].

For comparison, Figure 4 displays a typical cross correlogram for quiet conditions, \(\sum K_p = 3+\), in which a zero lag covariance of \(1.6 \times 10^{-4} R^2\) is evident but a lagged peak in excess of \(1 \times 10^{-4} R^4\) is clearly absent. We may therefore assign an upper limit \(\lambda 3914\) rms brightness fluctuation of \(-0.03\) R during the 3-min period during which this record was obtained, provided \(\sigma(B_o)/\sigma(B_v) \approx 1\).

In Figure 5 we summarize the results of 15 nights of observation, from 1973 to 1976, where we have plotted the maximum hourly average of \(\sigma(B_v)\) as a function of the maximum \(K_p\) during the 12 hours preceding local midnight. The sensitivity of mid-latitude precipitation to magnetic disturbance is clearly apparent, and we estimate that within an order of magnitude, \(\sigma(B_v)\) can be described by the approximate relation 

\[
\log_{10}(\sigma(B_v)) = -4.25 + 0.75 K_p R, \\
\text{or a 3.6 increase in flux magnitude per unit } K_p.
\]

O'Brien [1965] and Gough and Collin [1973] have found a fivefold increase in flux magnitude per unit \(K_p\) for electrons with \(E > 40\) keV. The variation in upper limits for \(\sigma(B_v)\) when no \(\lambda 3914\) is detected is due to changes in the brightness fluctuation noise level from night to night.

Our results indicate that an upper limit of 0.03 R can be

placed on the rms brightness fluctuation of \(\lambda 3914\) emission from the quiet time, \(K_p \leq 3\), mid-latitude nocturnal upper atmosphere. This places a corresponding upper limit of \(-2.5 \times 10^{-4}\) eV cm\(^{-2}\) s\(^{-1}\) on the rms fluctuation in the total energy flux deposition rate of precipitating particles.

Figure 5 also shows that \(\lambda 3914\) emission has been observed for \(K_p \geq 5\) in the 12-hour period preceding local midnight. We note that for maximum \(K_p < 6+\), Invermay Radio Observatory \((L = 2.71)\) lies equatorward of the equatorial boundary of the southern auroral oval [Bond and Thomas, 1971] and inside the plasmasphere [Carpenter and Park, 1973].

**DISCUSSION**

It must be emphasized that our technique is unable to distinguish between small modulations of a relatively bright emission and large fluctuations in a weak emission. For example, our upper limit to \(\sigma(B_v)\) of 0.03 R for \(K_p \leq 3\) does not preclude a steady emission of brightness \(B_v \gg 0.03\) R. However, observed electron precipitation frequently exhibits considerable temporal modulation [Coroniti and Kennel, 1970, and references therein]. Hall [1974] reports 20-50% modulation in 5577-\(\AA\) intensity, on a time scale of seconds, for a mid-latitude

![Fig. 2. Chart records showing correlated \(\lambda 3914\) and \(\lambda 5577\) activity observed on August 20, 1974, during magnetically disturbed conditions \((\sum K_p = 39)\).](image)

![Fig. 3. Cross correlogram of the \(\lambda 3914\) and \(\lambda 5577\) photometer fluctuations shown in Figure 2 computed from 164 s of data starting at 1301:30 UT and sampled every 0.18 s. Positive time lag means that the \(\lambda 5577\) emission is delayed with respect to the \(\lambda 3914\) emission.](image)

---

**TABLE 1. Photometer Operating Parameters**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Photometer 1</th>
<th>Photometer 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength, nm</td>
<td>391</td>
<td>558</td>
</tr>
<tr>
<td>Optical filter transmittance</td>
<td>0.34</td>
<td>0.68</td>
</tr>
<tr>
<td>Optical filter bandwidth</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>(FWHM)*, nm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opening angle (FWHM)*, deg</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Aperture, cm(^2)</td>
<td>230</td>
<td>230</td>
</tr>
<tr>
<td>Geometric factor, cm(^2) sr</td>
<td>140</td>
<td>140</td>
</tr>
<tr>
<td>Effective geometric factor, cm(^2) sr</td>
<td>12.5</td>
<td>7.5</td>
</tr>
<tr>
<td>Noise level (rms), R/(Hz)(^{1/2})</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>High-pass filter corner</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>frequency, kHz</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Low-pass filter corner frequency, kHz</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

* FWHM is full width at half maximum.
† Corrected for filter transmittance and quantum efficiency of photocathode.
pulsating aurora observed during an intense magnetic storm. For modulation in the range 20-50% and \( \sigma(B_v)/\sigma(B_0) = 0.03 \) R the corresponding range for \( B_v \) is 0.08-0.2 R. With this assumption our results are therefore in agreement with those of Jacka et al. [1970] in placing a conservative upper limit of \( \sim 0.3 \) R on the A3914 mid-latitude night sky brightness for \( Kp = 3 \). Broadfoot and Hunten [1966] have suggested that Yano's [1966] observations of intensities of 10 R may have included some twilight A3914 emission.

As a result of the effective lifetime of the metastable \( O(2^3S) \) atoms which produce the A5577 emission, the value of \( \sigma(B_v)/\sigma(B_0) \) is always less than the ratio \( f = B_0/B_0 \) that would occur for a steady electron flux with the same energy spectrum as that producing \( B_v \) and \( B_0 \). Rees and Luckey [1974] have calculated values for \( f \) assuming an energy spectrum \( N(E) dE = N_0 E \exp(-E/\alpha) dE \) and have found that this ratio depends mainly on the energies of the electrons involved and to a lesser degree on their flux magnitude. Mende and Easter [1975] have found reasonable agreement between these predictions and auroral observations. Now values of both \( \sigma(B_v)/\sigma(B_0) \) and \( f \) may be obtained from the power spectra of the light fluctuations. For example, for the data of Figure 2, \( \sigma(B_v)/\sigma(B_0) = 0.9 \pm 0.2, \) and \( f = 1.0 \pm 0.2 \). Therefore using the curves of Rees and Luckey [1974], we find that \( E = 3.7 \pm 1.0 \) keV \( (E = 2\alpha) \).

It has already been noted that for genuine particle precipitation to be detected by the cross-correlation technique, a detectable positive lagged covariance peak must be present in the correlogram. For example, Figure 3 shows a peak with lag 0.5 \( \pm 0.1 \) s. This corresponds to an effective lifetime of 0.7 \( \pm 0.1 \) s for \( O(2^3S) \) atoms as found by a method similar to that of Paulson and Shepherd [1965]. However, we cannot exclude the possibility that for sufficiently high particle energies the effective lifetime of \( O(2^3S) \) atoms will be sufficiently reduced to give a near-zero lag covariance peak that is indistinguishable from random noise effects. Recent calculations of the lifetime of \( O(2^3S) \) atoms as a function of altitude [Slanger and Black, 1973; Henrisken, 1975] show that a lifetime of less than 0.2 s requires the altitude of maximum ionization to be \( \leq 80 \) km, corresponding to \( E \geq 120 \) keV [Rees, 1969]. A detectable flux of such highly energetic electrons appears unlikely in view of (1) the low \( E \) values of only a few keV that we calculate from our observed values of \( \sigma(B_v)/\sigma(B_0) \), (2) the sizable lags of \( \sim 0.5 \) s observed during detectable precipitation periods, and (3) the lower-limit precipitating electron fluxes \( (E > 40 \text{ keV}) \) of \( \leq 40 \text{ cm}^{-2} \text{ s}^{-1} \) observed during mid-latitude rocket flights [O'Brien et al., 1965; Gough and Collin, 1973].

Unlike the satellite observations and anomalous VLF phase change methods [Potemra and Rosenberg, 1973] the optical correlogram method provides a direct total calorimetric estimate of ionizing energy flux. Therefore it appears to us that the suggestion by Potemra and Zmuda [1970, 1972] of a possible dominant contribution to D region ionization by a quiet time mid-latitude electron flux should be viewed with caution. Our results, interpreted in terms of a quiet time monoenergetic electron flux at 40 keV \( (f \sim 0.1) \), imply an upper limit of \( \sim 2 \times 10^4 \text{ cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1} \) at which \( \sigma(B_v)/\sigma(B_0) \) is the fractional rms fluctuation in the A3914 emission. For \( B_v \sim B_0 \) this is consistent with lower-limit rocket measurements of \( \sim 4 \text{ el cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1} \) at a higher flux of \( > 40 \text{ keV} \) [Gough and Collin, 1973].

Acknowledgments. We acknowledge financial support from the New Zealand University Research Committee.

The Editor thanks G. J. Romick for his assistance in evaluating this report.

REFERENCES


Henrisken, K., The effective lifetime-profiles of the \( O(2^3S) \) and \( N_e^*(2^3P_2^0, \gamma = 0) \) excited species, J. Atmos. Terr. Phys., 37, 1491, 1975.


(Received January 3, 1977; accepted March 9, 1977.)
Sigl et al. (1979) have developed an empirical model for the probability distribution of the duration of episodes (runs) during which the surface wind speed remains above and below fixed reference speeds. Whereas earlier workers (e.g., Hurst et al., 1977; Corotis et al., 1978) used power-law distribution functions to fit the data, these authors use a composite distribution consisting of a power law for durations shorter than a "partition parameter" \( t_1 \), and an exponential for longer times. The following comments may be useful:

1) Analysis of run durations from New Zealand sites obtained with higher resolution (\( t_0 = 50 \) s) than that which the authors have used has shown a power law (Pareto) distribution function, 

\[
F(t) = 1 - \left( \frac{t}{t_0} \right)^{-b},
\]

with \( t > t_0, b > 1 \), to be an excellent fit for durations up to several hours. The values of the exponent \( b \) as a function of \( V/V_m \), the ratio of reference speed to median speed, are similar to those found by the authors. Indeed, the power law representation can often be extended with reasonable accuracy to durations of a day or more. Corotis et al. (1978) have, in fact, done this but were forced to truncate the distribution when calculating mean run lengths. As pointed out by Hurst et al. (1977), the Pareto function leads to an infinite expectation length \( E[t] \) for \( b \leq 2 \) and overestimates the probability of long runs. The composite distribution function introduced in this paper must therefore be regarded primarily as a means of obtaining an expectation value close to the observed mean run length. To say this is not to deny the improved fit consequent upon this procedure. However, it should be emphasized that the partitioning is a quite arbitrary computational device and is without any fundamental significance.

2) The artificial partition at \( t_1 \) may be simply avoided. We first note that the power law implies a conditional "hazard" or failure rate, 

\[
h(t) = \frac{f(t)}{[1 - F(t)]} = \frac{b - 1}{t}.
\]

This is the failure rate of runs which have already lasted for time \( t \). The fact that this decreases as the run length increases is one way of picturing the persistence of surface wind speed. If we take a more realistic view and assume that the memory of its past history is gradually lost as the run length increases we may rewrite \( h(t) = [\lambda + (b - 1)/t] \). In the limit of long durations the failure rate then becomes independent of run length and equal to \( \lambda \).

The corresponding distribution function

\[
F(t) = 1 - \left( \frac{t}{t_0} \right)^{-b} \exp[-\lambda(t - t_0)]
\]

has the same limiting forms as the composite distribution but avoids the artificial partition. The expected value is given by

\[
E[r] = t_0[1 + (\lambda t_0)^{b-1} \Gamma(2 - b, \lambda t_0)]
\]

which simplifies to

\[
E[r] = t_0(\lambda t_0)^{b-2} \Gamma(2 - b)
\]

for sampling intervals short compared with the mean (and with \( 1/\lambda \)). Once \( b \) has been determined from the shorter runs, one of these equations can be solved to yield \( \lambda \) by equating the expected value to the observed mean. Excellent fits have been obtained in this way for a wide range of values of \( V/V_m \).

3) The authors find that the mean run duration depends critically on the sampling rate the effect of which "deserves further study." This dependence is hardly surprising when it is realized that the means refer to all episodes greater than \( t_0 \) (equal to half the sample period). Indeed, this is explicitly expressed in their Eqs. (6) and (13a) and in the corresponding equations in these comments. In fact, referring to the latter we note that log \( E(t) \propto (b - 1) \log t_0 \) in the limit of short sampling intervals. This accounts quantitatively for what the authors refer to as "sampling bias." This can be seen by comparing the difference between the logarithmic means.
given by Eqs. (16) and (15), $\Delta \log E[t]$, with that expected, $(b - 1)\Delta \log t_0 = (b - 1) \log 5$, for example, at the mean speed for which $(b - 1) \approx 0.5$. It should be emphasized that the mean run duration at a site has no meaning unless the minimum run length used in the analysis $(t_0)$ is also specified. Also, the value of $G$ in the power-exponential model depends on the choice of sample period $(2t_0)$.

Since the uncertainty in the true length of a run of apparent length equal to $n$ sample periods $(2nt_0)$ is $\pm t_0$, we note the desirability of high sampling rates in run analyses. Short sampling intervals, of course, are mandatory if one seeks optimum aero-generator operating strategies on a time scale of minutes and hours for which the power law distribution holds.

REFERENCES


Comments on "A Zero Crossing-Based Spectrum Analyzer"

Paul J. Edwards

Reprinted from
IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING
Vol. 37, No. 7, July 1989
Correspondence

Comments on "A Zero Crossing-Based Spectrum Analyzer"

PAUL J. EDWARDS

Abstract—Kay and Sudhaker recently proposed a spectrum analyzer in which DFT coefficients are computed from the zero-crossing times of a preprocessed signal. Their proposed method may be easily modified to provide for 1) invariance of the autocorrelation function to processing and clipping, and 2) spectral estimation by FFT of the zero-crossing intervals.

I. INTRODUCTION

The spectrum analyzer proposed by Kay and Sudhaker utilizes the zero axis crossings of a suitably preprocessed signal instead of uniformly spaced samples of signal amplitude. It therefore dispenses with a conventional A/D converter. The authors claim the consequent advantages of hardware simplicity, extended dynamic range, and higher speed for this form of clipped signal analysis. They conjecture that the minimum computational complexity should be of order \( N \log N \), as for the FFT. It is the purpose of this correspondence to point out that 1) the proposed technique belongs to a class of pulse time modulation techniques in which sampled amplitudes are transformed to zero-crossing interval widths, 2) this class of techniques has been routinely used to extend polarity-coincidence spectroscopy to non-Gaussian signals, and 3) modification of their proposed processing scheme permits spectral estimation by direct fast Fourier transformation of zero-crossing intervals or from clipped autocorrelation function estimates.

II. POLARITY-COINCIDENCE AUTOCORRELATION

A conventional digital spectrum analyzer consists of a low-pass anti-aliasing filter followed by an A/D converter. Uniformly spaced digitized samples of the band-limited signal are input to a software-based discrete Fourier transformer. An alternative indirect implementation frequently used in the fields of radio astronomy and laser spectroscopy where random signals with Gaussian statistics are encountered consists of a "polarity-coincidence" autocorrelator [1] which performs hardware-based operations on clipped (1-bit) signals. Estimates \( \rho_s(\tau) \) of the autocorrelation function and hence the spectral density of the signal can then be obtained from the ACF \( \rho(\tau) \) of the clipped signal by means of the van Vleck relation [2]. Methods of extending the 1-bit autocorrelation technique of spectral analysis to amplitude bounded signals with arbitrary statistics were devised independently by Veliman and Kwakernaak [3], Jesspers et al. [4], and Ikebe and Sato [5]. These methods process the band-limited signal prior to clipping by adding to it an auxiliary signal. A suitable auxiliary signal is the realization \( \eta(t) \) of a random process having a uniform probability density function confined to the interval \((-A, A)\) which also bounds the input signal.

Following the method used by van Vleck [2], it may be shown [6] that the ACF of the clipped composite signal

\[
\rho_s(t_1, t_2) = \int \int \text{sgn} y_1 \cdot \text{sgn} y_2 \cdot dP(y_1, y_2; t_1, t_2)
\]

\[= 1/A^2 \cdot \rho_s(t_1, t_2)\]

where \( P(y_1, y_2; t_1, t_2) \) is the joint probability distribution function of the composite signals \( y_1 = x(t_1) + \eta_1; y_2 = x(t_2) + \eta_2 \).

A later implementation proposed by Peek [7] illustrates the close connection between this modification of polarity-coincidence autocorrelation spectrum analysis and the method proposed by Kay and Sudhaker. Quantized periodic auxiliary signals consisting of linear combinations of Rademacher functions are used. These satisfy the requirements of 1), namely, 1) independence of \( \eta_1, \eta_2 \) and 2) constant (discrete-valued) joint pdf \( p(\eta_1, \eta_2) \) over the \( x(t_1), x(t_2) \) signal amplitude plane. A commercial implementation has been reported [8] to utilize a "voltage-to-duty cycle" conversion.

This can be achieved by using a periodic sawtooth or triangular waveform as auxiliary signal. When followed by zero-axis clipping, this operation constitutes linear pulse duration modulation. It is apparent that the scheme proposed by Kay and Sudhaker is a form of (nonlinear) pulse time modulation in which the zero crossings of the auxiliary waveform are displaced or "jittered" by the signal to be analyzed. The processing technique by which a deterministic or random auxiliary signal is added, prior to clipping, to the signal to be analyzed has evidently been anticipated and implemented in the form of the modified polarity-coincidence autocorrelation [3], [4]. The ACF is derived from amplitude information which has been encoded in the mean value, duty cycle, and probability distribution of the clipped composite signal as in a delta or pulse duration modulated signal. In the Kay and Sudhaker method, the amplitude information encoded in the zero-crossing intervals is subjected to discrete Fourier analysis and the zero-crossing rate \( 2f_c \) of the auxiliary signal must be high enough to satisfy the sampling theorem. The same requirement holds for the mean zero-crossing rate of the deterministic or random reference signals used in the 1-bit ACF method.

III. AN ALTERNATIVE IMPLEMENTATION

Kay and Sudhaker propose to use a sinusoid as an auxiliary signal. Bar-David [9] has shown that signals composed of the sum of a sinusoid and of a bounded band-limited process are uniquely (sample-wise) represented by their zero crossings. The sinusoid amplitude must exceed the bound on the signal and its frequency \( f_s \) must exceed the band limit \( W \). This auxiliary signal served as a zero-crossing interpolator. It ensures that the requirements of the sampling theorem are satisfied in that the composite signal has \( 2f_cT > 2WT \) zero crossings in record length \( T \).

A sinusoid is evidently only one of a large class of auxiliary signals which allow the composite signals to be fully sampled. However, it has two obvious disadvantages. The first, as the authors state, is the computational complexity of transforming zero-crossing times to Fourier coefficients. They were unable to find an algorithm as efficient as the FFT used in conventional multilevel spectrum analysis. The second disadvantage is that the shape of the autocorrelation function of the transformed signal is not invariant to sinusoid addition and clipping.

IV. CONCLUSION

A periodic triangular, staircase, or sawtooth auxiliary waveform, although more difficult to implement at high frequencies than...
a sinusoid, provides two options for spectrum analysis. The first of these, via the clipped ACF, requires a sampling rate many times the Nyquist rate in order to match the performance of a conventional DFT analyzer [10]. The second of these, of which the Kay and Sudhaker analyzer is a particular example, requires a similarly high clock rate.

The Kay and Sudhaker analyzer does not therefore appear to offer any advantage in bandwidth over either the 1 bit ACF analyzer or a conventional analyzer. Addition of a triangular auxiliary waveform, rather than a sinusoid, generates naturally sampled pulse duration and pulse position modulation $r(x_i)$. Direct Fourier transformation of the intervals $r$ will lead to spectral distortion except in the limiting case of small signals. However, various algorithms (e.g., Hostetter [11]) may be used to yield the $2N$ discrete complex Fourier transform coefficients as in a conventional analyzer with lower computational complexity than the Kay and Sudhaker method.

REFERENCES

Level Crossing Statistics of Atmospheric Wind Speeds: Probability Distribution of Episode Lengths

Paul J. Edwards

School of Electronic Engineering and Applied Physics
University of Canberra, ACT 2601, Australia

Abstract. The probability distribution of the duration of episodes ("wind runs") during which the horizontal wind speed in the planetary surface boundary layer remains above or below a given threshold value is of interest in the fields of renewable energy generation and pollutant dispersal. There still appear to be no analytic or conceptual models to explain the remarkable constancy of the power law form of the wind run distribution measured at a variety of sites on the earth’s surface for run lengths ranging from a few minutes to a day or more.

INTRODUCTION

Various empirical models have been developed for the probability distribution of the duration of episodes (runs) during which the horizontal wind speed in the terrestrial surface boundary layer remains above (or stays below) an arbitrary wind speed. Wind power generators generally have a cut-in wind speed, typically a few metres per second, below which they cannot be operated, and a furling speed, typically a few tens of metres per second, above which they must be shut down. In order to formulate optimal operating strategies for aero-generators it is therefore desirable to know the statistical distribution of the lengths of episodes during which the wind speed exceeds or falls below these thresholds. This knowledge is required because there is usually some minimum period of operating time for which it is economic to pay the cost of starting up the generator, given that the wind speed has risen above the cut-in value or fallen below the furling speed.

The characterisation of the episode length distribution of wind runs is one practical example, and the fade length distribution for a radio-communications link is another, of a large class of problems in the theory of level crossings of random variables which are apparently analytically soluble only under restricted conditions [1,8]. The wind run length distribution at a variety of sites has been found [2,3,4,5,6,7] to follow a power law over four decades in run length, from minutes to days.
FIGURE 1. Typical wind run length distributions measured at a site with a mean wind speed of 5 m/s.

MEASUREMENTS

Regional and national wind energy surveys involving synoptic studies of the horizontal wind speed and its variability have been undertaken over the past few decades in most countries of the world. Data from a variety of continental and maritime sites in New Zealand, Europe and the USA [2,3,4] have revealed a remarkable consistency in the form of the statistical episode distribution [5,6,7]. It was found [6,7] that the results of the surveys were relatively non site-specific and that the data could be generalised in power law form. Figure 1 shows typical data obtained from an inland New Zealand site. For wind run data from both maritime and inland sites, the cumulative distribution function $F(t)$ giving the probability of an episode length shorter than time $t$ can be written as

$$F(t) = 1 - (t/t_0)^a,$$

for episodes of wind speed above (or below) threshold $v$, lasting longer than the minimum sampling time resolution $t_0$, of the analysis. The exponent $(a)$ is a relatively weak function of the ratio of the threshold level to the median speed, typically varying from -0.6 to -0.8 for speeds between 4 m/s and 20 m/s. This power law is a good fit to the data over a wide range of episode lengths, from minutes to days, particularly for thresholds close to the median speed.
Similar results with coarser time resolution were obtained by other international workers [2,5] for episode resolution longer than one hour. The value of the exponent for the median speed is the same for episodes above and below threshold, typically between -0.5 and -0.7.

The power law form of the distribution permits a forecast of the duration of a wind run. The conditional episode termination probability, \( h(t) = f(t) / [1 - F(t)] = (-\alpha t), \)

where \( f(t) \) is the density function corresponding to distribution function \( F(t) \). This is the conditional failure rate of episodes which have already lasted for time \( t \). It follows that if the wind speed has exceeded a threshold speed for time \( t_i \), then the conditional probability of it remaining above that threshold for a further time \( t_2 \) is given by

\[
P((t_1 + t_2) \mid t = t_1) = (1 + t_2/t_1)^\alpha,
\]

for the appropriate value of \( \alpha \), within the range of validity of equation (1). For values of \( t \) longer than a few days, the probability deviates significantly below the power law relation. This is a physical limitation to its range of validity since values of the exponent \( \alpha \) less negative than \(-1\) lead to an infinite expectation value for the episode length [7]. Using the physically reasonable assumption that the conditional episode termination rate \( h(t) \) eventually becomes a constant, independent of episode length, a semi-empirical distribution consistent with equation (1) has been derived [6] which covers all episode lengths \( t > t_0 \), and gives a finite expectation length in good agreement with measurement.

**LEVEL CROSSING THEORY**

These results would seem to have some intrinsic interest, quite apart from their utility in the design of alternative energy generation systems. The power law form of the episode distribution implies a power law for the probability density of episode lengths with exponent \( (\alpha - 1) \), typically in the range from -1.5 to -1.7, for level crossings close to the median speed.

Given the remarkable constancy of the exponent over three or four decades of episode (level crossing) intervals, it is tempting to seek some relatively simple conceptual model for this behaviour, ideally supported by the theories of atmospheric turbulence and random process level crossings.

Unfortunately, it seems that level crossing theory can provide only limited help. The analytic derivation of the probability density of the level crossing intervals of random variables from their frequency domain (spectra) and amplitude domain (probability) descriptions remains a major unsolved problem in this field [1,8,9] and simulation appears to be the only practical solution in the general case. Nevertheless, the level crossing behaviour of Gaussian related processes appears to be more tractable. The wind speed probability density functions measured in these surveys were generally close to a Rayleigh (circular Gaussian) distribution. Progress in dealing with Rayleigh fading in radio communications and Riceian-distributed optical scintillation [10] suggests that

525
the related case of (near) Rayleigh distributed surface wind speeds with well defined spectra may also yield to a combination of analysis and physical insight.

CONCLUSIONS AND OPEN QUESTIONS

We have pointed out an example in the field of atmospheric wind turbulence of a major long-standing unsolved problem which the theory of level crossings seems unable to address analytically. Answers to the following questions would be helpful.

(i) Does the power law form of the episode distribution have physical significance? (One is inevitably reminded of the historical preoccupation with “1/f” noise processes).
(ii) Can the law be derived by analytic techniques?
(iii) Does chaos theory or the theory of fractal systems have anything to offer?
(iv) Would software or hardware simulation be fruitful?

ACKNOWLEDGEMENTS

I acknowledge with thanks the invaluable assistance of Dr Robert Hurst and the late K. R. Dawber in analysis and data collection.

REFERENCES

Level-crossing statistics of the horizontal wind speed in the planetary surface boundary layer

Paul J. Edwards
Centre for Advanced Telecommunications and Quantum Electronics Research, University of Canberra, ACT 2601, Australia

Robert B. Hurst
Defence Operational Technology Support Establishment, Private Bag 32901, Devonport Naval Base, Auckland, New Zealand

(Received 17 January 2001; accepted 19 April 2001; published 31 August 2001)

The probability density of the times for which the horizontal wind remains above or below a given threshold speed is of some interest in the fields of renewable energy generation and pollutant dispersal. However there appear to be no analytic or conceptual models which account for the observed power law form of the distribution of these episode lengths over a range of over three decades, from a few tens of seconds to a day or more. We reanalyze high resolution wind data and demonstrate the fractal character of the point process generated by the wind speed level crossings. We simulate the fluctuating wind speed by a Markov process which approximates the characteristics of the real (non-Markovian) wind and successfully generates a power law distribution of episode lengths. However, fundamental questions concerning the physical basis for this behavior and the connection between the properties of a continuous-time stochastic process and the fractal statistics of the point process generated by its level crossings remain unanswered.

I. INTRODUCTION

Following the discovery of a power law distribution for the episode lengths during which the horizontal surface wind speed remains above (and below) fixed wind speed levels, several empirical models were developed to characterize and extend the description of these episode length distributions. In addition to their obvious practical applications to the design and analysis of wind power systems and to wind modeling generally, these level-crossing statistics have significant intrinsic interest. When first discovered 23 years ago it was recognized that their analysis placed them within a class of largely unsolved problems in the theory of stochastic processes, namely the level-crossing behavior of continuous random variables. More recent reviews suggest that the analysis of zero and level-crossing phenomena has advanced significantly in the field of fractal stochastic processes. Since the wind speed level-crossing statistics have the power law form and therefore the self-scaling properties expected of fractal stochastic point processes, it is natural to seek a description in those terms.

Fractal stochastic point processes are well known. Typically, several of their statistics, e.g., the power spectral density, autocorrelation function, Fano factor, and interval distribution exhibit self-similar or "scaling" behavior. This scaling property follows from, and leads inevitably to, power-law forms of these statistics in which the various power-law indices are related to the fractal dimension of the particular process. Self-scaling phenomena are particularly common in the atmospheric and geophysical sciences with, for example, power-law descriptions of mesoscale (10 m–10 km) turbulence-related phenomena. However, the underlying physical processes and their mathematical descriptions in fractal terms are still the subject of active debate.

Figure 1 shows the level-crossing points formed by a...
FIG. 1. Sample functions of (a) a stochastic point process and (b) a stochastic binary (alternating) process formed by (c) the level crossings of a continuous-time random variable such as the horizontal wind speed measured at a fixed point close to the earth's surface.

We are immediately led to ask the question: What is the sample function [Fig. 1(c)] of a continuous time stochastic process, such as the horizontal wind speed.

We are immediately led to ask the question: What is the essential feature of the wind speed fluctuations which causes a Markov process, such as the horizontal wind speed.

We shall then call on the theory of Markov processes to model the wind velocity field and its level-crossing behavior. We shall construct a Markov process which gives first-order conditions, the resulting stochastic point process, the sequence of delta functions \( \delta(t-t_j) \) in Fig. 1(a) at crossing times \( t_j \), can be said to constitute a fractal stochastic point process. Further questions specific to fractal processes such as the derivation of the expected value of the fractal exponent then follow. These questions are relevant to long-standing unsolved problems, both conceptual and analytic, in the field of stochastic level-crossing processes.

In what follows we shall analyze real wind speed data and demonstrate the fractal nature of the level-crossing point process. We shall then call on the theory of Markov processes, in particular the Ornstein-Uhlenbeck process to model the wind velocity field and its level-crossing behavior. We shall construct a Markov process which gives first-order agreement with measurements of the marginal wind speed probability density, its energy spectrum, and its self-scaling level-crossing statistics. We shall pose, but will not address here, several wider questions which arise concerning the necessary conditions for the occurrence of self-scaling level-crossing and related statistics.

II. MEASUREMENT AND ANALYSIS

Regional and national wind energy surveys involving synoptic studies of the horizontal wind speed and its variability have been undertaken over the past few decades in many countries. Data from a variety of continental and maritime sites in New Zealand, Europe, and the USA have revealed a remarkable consistency in the form of the wind speed level-crossing interval distribution. Figure 2 shows an example of this.

For wind run data from both maritime and inland sites, the cumulative distribution function \( F(T) \) of Fig. 2 giving the probability of an episode length \( T \), longer than time \( T \), the “interevent survivor function” (ISF) can be written as

\[
P(t>T)=F(T)=(T/T_0)^\alpha
\]

for episodes of wind speed above (or below) threshold \( \nu \), lasting longer than the minimum sampling time resolution, \( T_0 \), of the analysis. The exponent \( (\alpha) \) is a relatively weak function of the ratio of the threshold level to the median speed, typically varying from \(-0.5\) to \(-0.8\) for speeds between 0.5 and 3 times the mean wind speed. This power law is a good fit to the data over a wide range of episode lengths, from minutes to days, particularly for thresholds close to the median speed where the exponent \( \alpha \sim (0.6\sim0.7) \), is nearly equal for episodes above and below threshold. The differential probability distribution of the episode lengths \( p(t) \), is related to \( F(T) \) by \( F(T)=\int_0^T p(t)dt \). For \( t>T_0 \), and \( \alpha<0 \),

\[
p(t)=(1-\alpha)T_0^{-\alpha}t^{-(1-\alpha)}
\]

Similar results with coarser time resolution were obtained in other surveys for episode resolution longer than 1 h. For values of \( T \) longer than a few days, the probability deviates significantly below the power-law relation. This is a physical limitation to its range of validity since values of the exponent \( \alpha \) less negative than \(-1\) lead to an infinite expectation value for the episode length. Using the physically reasonable assumption that the conditional episode termination rate \( h(T) \) (the failure rate of episodes which have already lasted for time \( T \) eventually becomes a constant, independent of episode length, a semiempirical distribution consis-
tent with Eq. (1) has been derived\(^4\) which covers all episode lengths \(t > T_0\), and gives a finite expectation length in good agreement with measurement.

These results have intrinsic interest, quite apart from their utility in the design of wind power generation systems. The power-law form of the episode distribution implies a power law for the probability density of episode lengths with exponent \(-1 - \alpha\), typically in the range from \(-1.6\) to \(-1.7\), for level crossings close to the median speed.

Given the remarkable constancy of the exponent over three decades of episode (level-crossing) intervals, it is tempting to seek an “explanation” for this behavior, perhaps in terms of some relatively simple conceptual model, ideally supported by the theories of atmospheric turbulence and random process level crossings.

Unfortunately, such an explanation is not readily forthcoming and it appears that classical level-crossing theory can provide only limited help. The analytic derivation of the probability density of the level-crossing intervals of random variables from their frequency domain (spectra) and amplitude domain (probability) descriptions remain a major unsolved problem in this field.

The wind speed probability density functions measured in these surveys were generally close to a Rayleigh (circular Gaussian) distribution. For illustrative purposes Fig. 3 shows a typical histogram of the frequency of hourly wind speeds plotted together with the best fit Weibull wind speed probability density function:

\[
p(V) = \frac{k}{c} \frac{(V/c)^{k-1}}{(k-1)!} \exp\left(\frac{-V}{c}\right)
\]

(3)

In this example the shape factor \(k = 1.67\), rather less than the value of \(k = 2\) for a Rayleigh distribution, frequently quoted for wind speed distributions. The best fit distribution does in fact depend upon height above ground, the time resolution of the measurement, the speed resolution, and the threshold speed of the anemometer. In what follows we shall take horizontal wind speeds measured at a fixed point close to the earth’s surface to be typically distributed according to the Rayleigh law:

\[
p(V) = \frac{2V}{c^2} \exp\left(-\frac{V}{c}\right).
\]

(4)

We should recognize, however, that the Weibull shape factor has been shown to vary considerably, e.g., with topographic class. Justus et al.\(^32\) found \(k\) values in the range 1.2–2.3 over the continental United States. Because the level-crossing properties of Gaussian processes have long been recognized as being more tractable than those of other processes we shall take the Rayleigh distribution as our reference standard. Davenport\(^33\) has suggested a simple two-dimensional model of large-scale turbulent atmospheric motion in which the surface wind is the resultant of two uncorrelated, spatially orthogonal Gaussian processes with equal variances. This idealized model leads naturally to a Rayleigh distribution for the wind speed, and the wind direction is isotropic, that is there is no prevailing wind direction, contrary to the usual situation.

A similar approach to the analysis of fading in radio communications channels\(^34\) in terms of a Rayleigh-distributed envelope was pioneered by Rice.\(^8\)

III. STOCHASTIC LEVEL CROSSINGS

The analytic description of the zero-crossing and the level-crossing behavior of random variables has long been recognized to be one of the more intractable problems in the theory of stochastic processes\(^8\)–19. The topic has been extensively treated in various monographs and reviews.\(^12\)–14,36–39 Despite the absence of general analytic solutions, the theory of zero-crossings and level-crossings has been widely applied in the engineering and environmental sciences, particularly in communications engineering, mechanical engineering, hydrology, and meteorology. Interest in the zero-crossing problem began with the work of Kac\(^9\) in 1943. This was followed by the classic work of Rice,\(^8\) McFadden,\(^10\)11 and others on various aspects of “barrier-crossing” statistics, including the fundamental “first-passage time” problem—the distribution of the time taken by a random process to cross a specified level, given its initial value.

The level-crossing statistics of a stationary random process depend on the spectral moments and, equivalently, the time derivatives of the autocovariance function at zero time lag. This indicates the sensitivity of these statistics to the high frequency spectral content and to the structure of the correlation function at very short time lags. For many processes of interest, e.g., the Ornstein–Uhlenbeck process, the correlation function is not differentiable at the origin and the crossing point density (the crossing rate) becomes infinite. This is one reason for the intractability of the general problem: A sample function having crossed a threshold once may continue to do so again and again in a very short subsequent time interval. This clustering of points is one of the notable characteristics of a fractal stochastic process. For this and other reasons a closed form solution for the probability density of the intervals between successive crossings has not been found in the general case.

Rice\(^8\) defined a suite of joint probability density functions (pdfs) \(P_k(t;b)\) for the intervals between crossing points of the level \(b\) conditioned by the number \(k\) of intervening crossing points and derived an integral equation for the
pdf of the zero-crossing intervals which contained \((k+I)\) joint probability integrals. McFadden\(^{11}\) later discovered an equation relating the \(P_k(t; b)\) to the mean crossing rate and the probabilities \(Q_k(t, b)\) that the process crosses level \(v = b\) exactly \(k\) times in an interval of length, \(t\). Neither of these forms of the solution has proved generally tractable.\(^{38,39}\)

In the present case of wind speed measurements, as in many practical cases, the high frequency control of the real data and the possible nondifferentiability of the model process are not major issues. The measurement process itself involves a convolution over some minimum instrumental or data reduction time interval and so generates a low pass-filtered “local average.” In this way the high-frequency microscale details of the stochastic model and of the data are suppressed for short time scales.\(^{39}\)

The zero- and level-crossing distribution of the nondifferentiable (Brownian) Wiener-Levy process,\(^{39-41}\) the Gaussian Markov diffusion process, is known. It has a power-law probability density function for the intervals between crossings with exponent \(-3/2\). From Eqs. (1) and (2), the exponent of the corresponding interval survivor function (ISF) is \(-1/2\). This interesting result suggests a possible approach to modeling wind speed level crossings through Markov diffusion processes.

### IV. FRACTAL STOCHASTIC PROCESSES

A stochastic process is generally termed fractal if several of its relevant statistics exhibit self-similarity (“scaling”).\(^{20,22}\) The only solution which scales without change of mathematical form is a power law in the argument \(x\) with exponent \(\beta\), \(F(x) = x^\beta\), so that \(F(cx) = c^\beta F(x)\). Power-law forms of the statistics are therefore taken to indicate fractal behavior. A stochastic point process such as that represented in Fig. 1(a) can therefore be described as a fractal stochastic point process (FSPP) when its relevant statistics are described by power laws, although these need not necessarily have the same value of the exponent (“fractal dimension”). Point processes which can be represented on a line as in Fig. 1 have a fractal dimension \(D\) between 0 and 1.\(^{20}\)

Relevant statistics include the interevent interval probability density function \(p(t)\) with exponent \(\beta = -(D+1)\) and its integral, and interevent interval survivor function (ISF) of Eq. (1) with \(\alpha = -D\).

The point process formed by the horizontal wind speed level crossings is therefore a FSPP candidate. Other commonly used statistics include\(^{22}\) the power spectral density (PSD), with \(\beta = -D\); the coincidence rate (CR), with \(\beta = (D-1)\); and the Fano factor (FF), with \(\beta = D\).

The simplest FSPP is the standard fractional renewal or “semi-Markov” process (FRP)\(^{29}\) for which the interevent times are independent random variables drawn from the same fractional (power-law) probability distribution with exponent \(-(D+1)\). Also, the box-counting dimension for the FRP is \(D\) and, since the process is ergodic, all the various fractal dimensions are also equal to \(D\).\(^{42}\) Of course, a real process can be expected to show fractal behavior only over a limited dynamic range in a random variable such as the interevent time. It is evident that the fractal character of both real and simulated processes can be tested by analyzing their sample statistics.

Next we investigate whether a suitably chosen stochastic process can generate the same or similar level-crossing statistics as those exhibited by real wind speed data.

We start by noting that the apparent fractal form of the ISFs for wind speed level crossings is reminiscent of that for the fractional Brownian motion (fBM) family of processes.\(^{20}\) Fractional Brownian motion usually has a Gaussian amplitude distribution and a Levy-stable zero-crossing interval distribution. The ISF power-law exponents we observe in the wind data are close to, but more negative than, \(\alpha = -0.5\). The zero-crossing ISF for fBM is known to have power law exponent \(\alpha = (H - 1)\), where \(H\) is the Hurst exponent.\(^{20,41}\) A value of \(H \approx 0.5\) suggests a particular first-order candidate for a model, namely the classical Weiner–Levy (WL) process, otherwise known as regular Brownian motion which, as already noted, has an ISF power law exponent of \(-0.5\) for its zero-crossing interval density. From a modeling perspective a WL process has an advantage over all other fBMs in that it is a Markov process. This implies that the increments of the process are statistically independent of earlier increments.\(^{31}\)

In practice this makes for very efficient computer simulation, as we describe in the following. However, WL processes (and fBMs) are not suitable candidate stochastic models for wind speed without modification, for two important reasons:

(i) they are Gaussian processes whereas it is clear from Sec. III that the wind speed, although Gaussian related, has near Rayleigh amplitude statistics, and

(ii) they are nonstationary—from initial finite states their variances diverge with time without limit.

To address the first point, we note that wind speed as measured by any conventional anemometer is a one-dimensional measurement of the horizontal wind velocity, which is actually a two-dimensional vector process. For an anemometer situated close to the earth’s surface, the third (vertical) velocity component is usually strongly suppressed on the time scales \((T \gg 100 \text{ s})\) considered here.

As noted earlier, the measured wind speeds are approximately Rayleigh distributed. A Rayleigh distribution can arise naturally when one considers a process that is distributed circularly normally about the origin: It is well known that the random variable consisting of just the displacement from the origin (disregarding direction) is Rayleigh distributed. Hence we can see that it may be possible to model non-Gaussian one-dimensional wind speed measurements by considering an associated two-dimensional (2-D) Gaussian process. One straightforward way to generate such a process is to generate the easterly and northerly (or \(x\) and \(y\)) components as independent Gaussian processes and combine them into a vector. This uses the approach originally suggested by Davenport.\(^{33}\)

To address the second problem of nonstationarity, we note that a WL process can be modified in a standard way into an Ornstein–Uhlenbeck (OU) process\(^{29,30}\) whose displacement from the origin is nondivergent. The modification
involves adding a restoring drift toward the origin, whose magnitude varies linearly with displacement. The OU process remains Gaussian, with finite variance $\sigma^2$, and zero mean.

For the purposes of discussion we take $U(t)$ to be a particular realization of an OU process. Then the effect of the restoring drift is this: If $U(0) (= U_0$ say) is different from zero, the subsequent expectation for $U$ at later time $t$ is given by $\langle U(t) \rangle = U_0 \exp(-t/\tau)$, where $\tau$ is a parameter with dimensions of time. For our purposes the most convenient parameters to describe an OU process are the standard deviation $\sigma$, and the exponential time constant $\tau$.

From the above, we see that the effect of the restoring drift can be small for $U_0$ not too large, and for $t \ll \tau$. Over short times an OU process could therefore be expected to behave rather like a WL process. We therefore have some expectation that the power-law level-crossing statistics might be relatively robust to the transformation into an OU process.

**V. SIMULATION**

We next describe the numerical simulation and subsequent analysis of a single point surface wind speed record based on a Markov stochastic process. For comparison, a one-month real wind speed record acquired in 1981 during a New Zealand wind energy survey\(^2\) was reanalyzed. The simulation was designed to match this real record. We chose to compare the simulated and real winds at an intermediate time resolution of 100 s ($\Delta t = 100$). We generated two (two-dimensional) OU processes, $U_x$ and $U_y$, the $x$ and $y$ components of wind velocity with the same values of $\sigma$ and $\tau$, from which the scalar wind speed followed directly.

For a one-dimensional OU process $U(t)$, the increments $[U(t+\Delta t) - U(t)]$ may be generated as

$$\Delta U = -\beta U(t)\Delta t + G\alpha(\Delta t)^{1/2},$$

where the first term represents the restoring drift, and the second term is Gaussian diffusion with diffusion function $\alpha$. $G$ is a Gaussian random process with zero mean and unity variance.

This formulation is suitable for implementing a computer simulation. First, $\Delta t$ the time increment is chosen together with the required values of $\sigma$ and $\tau$. Provided that $\Delta t \ll \tau$, the parameters $\alpha$ and $\beta$ may be shown to be given by $\beta = \tau^{-1}$ and $\alpha^2 = 2\beta\sigma^2$.

Fitting of the parameter $\tau$ was done iteratively: Two trial one-dimensional OU processes and the resulting Rayleigh process were generated, and the autocovariance function of the Rayleigh process $R_{\text{Ray}}(t')$ was then computed as a function of lag time in the usual manner.\(^4\)

The widths of the autocovariance function (ACFs) were matched in lag time. In this way $\tau$ was assigned a value of 45 600 s. Figure 4 compares the ACFs of the simulated Rayleigh process and the real wind data. In Fig. 4 the ACF of the Rayleigh process has been rescaled to match the height of the real wind ACF at zero lag. As is apparent from an analytic treatment of a two-dimensional OU process used to represent noise fluctuations in laser light,\(^4\) the width of the ACF of the Rayleigh process is approximately half the time constant of the two underlying one-dimensional OU processes.

To fix the remaining parameter $\sigma$ of the component OU processes, we stipulated that the mean of the simulated process be equal to that of the real wind data (6.0 m s\(^{-1}\)). For component $x$ and $y$ processes with standard deviations $\sigma$, it may be shown that the associated Rayleigh process has a mean value of $\sigma\sqrt{\pi/2} = c\sqrt{\pi/2}$, from Eqs. (3) and (4). This and the relationships given previously now completely determine the parameters of the simulation. We generated a simulation of the equivalent of approximately eight years of data, to give good statistics for the analysis.

A histogram of the simulated wind speed samples is shown in Fig. 5, along with a histogram of the real wind data with the same mean (6.0 m s\(^{-1}\)), renormalized so that the areas are equal. Evidently this real wind data depart slightly from Rayleigh statistics as did the earlier examples in Fig. 3.

To provide a graphic comparison between the two data sets, typical segments of data from the real wind and the simulation are shown in Fig. 6, plotted to the same time scale. Despite being subjectively selected, these data are nev-
Nevertheless representative of the larger data sets, although the simulated data and the real data appear similar on time scales of hours and longer, the records clearly differ in character on shorter time scales. The real wind shows more short-term variability. Also this short-term variability in the real wind clearly is greater when the wind speed is high. This latter characteristic appears to be absent from the simulated process. More sophisticated models of simulating the stochastic wind field, such as the successful multiaffine approach to the synthesis of atmospheric turbulence, would presumably give better agreement with real wind characteristics at the cost of greater modeling complexity.

The power spectral densities of both processes were calculated (as the Fourier transformations of the ACFs). To avoid complications related to aliasing the spectral calculations were terminated well before the Nyquist frequency of 5 mHz. The spectra are shown in Fig. 7.

Both processes exhibit power-law spectra for frequencies above $10^{-5}$ Hz. For the simulation this is not unexpected as the underlying OU processes have high frequency spectral exponents of $-2$. The Rayleigh process has a slightly less steep spectrum with exponent $-1.9$. The real wind has a spectral exponent of $-1.6$ and thus has significantly more variance at high frequencies. This accounts for the different appearance of the traces in Fig. 6.

Both processes were analyzed to extract the level-crossing statistics. Some results from this analysis are shown in Fig. 8.

The two plots in Fig. 8 have many features in common. For short durations both plots show the straight lines indicating a power-law dependence of the ISFs on the duration $T$. The real wind shows a steeper power law than the simulation. For each process separately, for threshold equal to the median wind speed, the exponents for excursions above and below the median level are nearly equal ($-0.71$ and $-0.73$, respectively, for the real wind, $-0.53$ and $-0.50$, respectively, for the simulation). These numerical values were obtained from regression analyses over the first decade of duration.

VI. DISCUSSION

The episode distributions for the real and simulated wind data are remarkably similar. We also examined conditional survivor functions for the real wind in which the episodes used in the analysis were selected on the basis of the duration...
of the preceding intervals. These conditional ISFs showed evidence of significant correlation between successive episode lengths for duration of the order of tens of seconds. However, for longer intervals successive episode lengths became nearly independent, indicating an approximate fractal renewal process as for the Rayleigh simulation. The Rayleigh process median crossings evidently constitute a FRP with dimension $D = 0.5$.

Another significant difference between the real and simulated winds is that the simulated crossing rate falls considerably short of the real wind process, by factors of approximately 2 for median crossings and 4 for crossings at twice the mean level. Both processes show the eventual expected fall away from the power law for sufficiently long episode durations of around $10^3$ s.

Finally, both the real wind process and the simulation were analyzed for the Fano factor variation with counting time. Both processes show power law dependence of the Fano factor over two decades from $10^2$ to $10^6$ s, but again with different exponents, $-0.51$ and $-0.39$, respectively, for the real and simulated data.

The power-law form of the simulated interval survivor function ISF for the median crossings and its exponent of $-0.5$, which from our previous discussion found its origin in the WL process, have proven to be remarkably robust. They have survived two transformations, the first to the OU process, the second to the Rayleigh process.

VII. CONCLUSIONS AND OPEN QUESTIONS

We have reanalyzed real wind measurements and confirmed the fractal character of the point process generated by the level crossings of the horizontal wind speed over three decades of intermittent time, from minutes to days. These data suggest that the real wind median level-crossing points approximate a fractal renewal process for intervals longer than a few minutes.

We have constructed a two-dimensional Ornstein–Uhlenbeck (Rayleigh) process as an approximate model of the real wind speed fluctuations. The real and the simulated wind data both show power-law behavior of the spectral densities and the level-crossing interval distribution (ISF) and Fano Factor over a similar range of interval lengths, from minutes to days. Furthermore, the power-law exponent of the simulated median level crossing-interval survivor function ($-0.5$) is close to that for the real wind ($-0.7$ in this case).

Our Markovian wind model departs from the real wind in that it generates an atypical isotropic horizontal wind velocity distribution, underestimates the observed rate of level crossings, does not reproduce the amplitude-dependent short-term variability exhibited by the real wind, and generates a Rayleigh (rather than a Weibull) wind amplitude distribution.

Despite these shortcomings, the simulation reproduces the "standard" Rayleigh probability density function of typical surface wind speeds, the corresponding power-law energy spectrum and the fractal statistics of the point process generated by the median wind speed crossings. It seems likely that a more realistic simulation based on a non-Markovian process with antipersistent Hurst factor would give closer agreement between the power law exponents of the real and simulated data.

What has this simulation achieved? Has it answered the fundamental questions posed at the start of this paper? It seems not. For example:

(1) We have been unable to clarify the physical nature of the turbulent diffusion process which leads to a power-law wind speed median level-crossing interval density function with exponent near $-1.6$. Nor have we arrived at a conceptual model to account for this form of the distribution.

(2) We have certainly demonstrated the robustness and ubiquity of the power-law form of the interval distribution. The level-crossing characteristics of the simulated wind speed fluctuations approximate those of the real wind and closely follows those of the Wiener–Levy process on which the model was based. Presumably the real wind level crossings also closely reflect a fundamental process. What is this process?

(3) We have been unable to identify the essential conditions necessary for the generation of a FSPP from the level crossings of a continuous time process. It may be that, like "1/f" noise, there are several different pathways leading to the same result.

(4) We have been unable to clearly identify a well-defined class of continuous time processes with fractal level-crossing points, or to point to the distinguishing and necessary properties of such a class although the simulations are suggestive in this respect.

These fundamental questions remain unanswered. They indicate the existence of significant unresolved problems in the fields of stochastic level crossings and surface boundary layer wind processes which merit further investigation.

ACKNOWLEDGMENT

One of us (R.B.H.) acknowledges with thanks the assistance of the University of Otago in making available data gathered by the Wind Energy Resource Survey of New Zealand.


J. A. McFadden, Trans. IRE IT-2, 146–150 (1956).


40 J. Feder, Fractals (Plenum, New York, 1988).
41 P. S. Addison, Fractals and Chaos (IOP, Bristol, 1997).
Abstract

HIGH IMPEDANCE QUANTUM NOISE REDUCTION
IN GaAlAs INFRARED EMITTING DIODES

Paul J Edwards
Department of Electronics Engineering
and Applied Physics
UNIVERSITY OF CANBERRA

It is now recognised that quantum optics and electronics provide new insights into the operation of optoelectronic devices and systems. In particular, the realisation that shot-noise due to statistical photo-current fluctuations is not an irreducible limitation to either measurement precision or information capacity of an optical communication link, has stimulated workers at NTT (Japan), IBM (USA) and other laboratories (1, 2) to generate and investigate the properties of sub-Poisson ("quiet") light. Quiet light, associated with non-classical "photon-number-squeezed" quantum states, offers potential advances in optical measurement, computing and measurement technologies.

Noise reduction is possible with current stabilised semiconductor light emitting devices such as light emitting and injection laser diodes driven by high impedance current sources. It may be shown (2) that the electron current noise spectral density, relative to the shot-noise, is then given by $S(f)/2\Delta f = 8kT/eV$ with $T$ the device temperature and $V$ the diode bias voltage. For $T = 300$ K, $V = 2.0$ V, the internal shot noise suppression factor is approximately 10 dB. The photon counting statistics appropriate to a minimum uncertainty number state (Fano Factor = 0), observed with quantum efficiency, $\eta$ are binomial, not Poisson and the observed Fano Factor (normalised power spectral density) is $(1-\eta)$.

Initial measurements made at the University of Canberra with systems in which $\eta \approx 0.13$ generally confirm this expectation. The non-classical Binomial shot-noise statistics were verified at room temperature by varying the geometrical photon coupling factor and hence the overall quantum efficiency, $\eta$. These measurements have been performed with a variety of commercially available GaAlAs infrared (0.8911 μm) emitting diodes, selected for high quantum efficiency. The light emitters were closely coupled to large area (1cm$^2$) reverse biassed silicon PN junction photodiode detectors. Measurements of noise power spectral density over a bandwidth of several megahertz have been made at room temperature and at dry-ice temperature (195K). All IRED measurements were referenced to noise measurements made on a standard tungsten filament lamp. These clearly demonstrate the expected quantum noise quieting relative to the reference (shot-noise) source. Overall quantum efficiency and quantum noise reduction typically rises from about 10% at room temperature to 13% at dry-ice temperature. As anticipated, the noise reduction is restricted to frequencies below 1MHz, the limit set by carrier lifetimes.

Measured Fano Factor

\[ F = \frac{S(f)}{2\langle i \rangle e \langle n \rangle} \]

Typical raw noise spectra showing 0.5 dB reduction (D) below the reference shot-noise level (R) for Shot-noise current \( \langle i \rangle = 2\text{mA} \) at dry-ice temperature (195K). Trace (A) represents amplifier noise level only (\( \langle i \rangle = 0 \)).

Figure 1

Noise Power Spectral Density \( S(f) \) as function of overall quantum efficiency, \( \eta \) measured at \( f = 300\text{kHz}, \langle i \rangle = 1\text{mA} \), normalised to the shot-noise level \( 2\langle i \rangle e \) at room temperature (295K) and dry-ice temperature (195K), [lowest point]. The dashed line \( (1-\eta) \) is the post-detection Fano Factor for a noiseless source showing the transition from Binomial to Poisson statistics expected for number-squeezed light. The solid line is calculated from equation (3) of Reference (2).

Figure 2
Sub-Poisson light from GaAlAs infrared emitting diodes

PAUL J. EDWARDS

Department of Electronics Engineering and Applied Physics,
University of Canberra, PO Box 1, Belconnen ACT 2616, Australia

(Received 15th May 1990)

Abstract. Precise shot-noise measurements on infrared light generated by GaAlAs light emitting diodes have been performed at dry-ice and room temperatures using tightly coupled photodiodes with high quantum efficiency.

Power spectral density measurements confirm the predicted [1,2] suppression of noise below the standard quantum shot-noise level for high efficiency constant-current-driven infrared emitting semiconductor junctions. The measured shot-noise suppression of 0.5 dB at 195 K and 0.4 dB at room temperature is consistent with a measured quantum efficiency of 13% at 195 K and 10% at room temperature. The nonclassical (binomial) photon statistics were demonstrated by the trend towards Poisson (full shot-noise) variance with increasing optical attenuation. As expected, sub-Poisson quantum fluctuation noise, characterised by a Fano factor below unity, extends over a bandwidth of 1 MHz, limited by carrier lifetimes.

These measurements of sub-Poisson 'quiet' light imply current noise suppression of more than 10 dB below the shot-noise level.

1. Introduction

The realisation that shot-noise due to statistical photocurrent fluctuations is not an irreducible limitation to either measurement precision or information capacity has stimulated workers [1,2] to generate and investigate the properties of sub-Poisson ('quiet') light. Quiet light, associated with nonclassical 'photon-number-squeezed' quantum states, offers potential advances in optical measurement, computing and communication technologies.

Noise reduction below the shot-noise level is possible with semiconductor light-emitting devices such as LED and injection laser diodes driven by high-impedance current sources. It may be shown [1] that the electron current noise should ideally then be thermal with spectral density, relative to shot-noise, given by

$$\frac{S(f)}{2Ie} = \frac{2kT}{eV}.$$  \hspace{1cm} (1)

for frequency, $f \ll 1/\tau$ and source resistance, $R \gg$ diode differential resistance with $\tau$ the carrier lifetime, $T$ the device temperature and $V$ the voltage drop across $R$. For $T = 300$ K, $V = 1.0$ V, the internal current-noise is approximately 13 dB below shot-noise. The use of a light-emitting diode to generate 'quiet' recombination radiation by sub-Poisson electron injection through space-charge stabilised current flow was proposed by Teich et al.[2]. Machida and Yamamoto [3] succeeded in demonstrating sub-Poisson light generation over a wide bandwidth using an AlGaAs/GaAs distributed feedback transverse junction stripe injection...
laser operating at 77 K. Tapster, Rarity and Satchell [4] measured sub-Poisson noise below 100 kHz at room and dry-ice temperatures using selected commercially available gallium arsenide LEDs. However, they did not observe the full noise reduction predicted, nor did they find consistency between different LEDs of the same type.

If the excitation current noise were zero, the expected (post-detection) Fano factor (the variance in the photo-electron count, normalised to the mean) would be given by

\[ F_n = (1 - \eta). \]  

For a non-zero excitation Fano factor, \( F_n(e) \), we expect \[ 2 \]

\[ F_n = S(0)/2le = 1 + \eta(F_n(e) - 1). \]  

Initial measurements made at the University of Canberra with GaAlAs LED–silicon pin diode systems in which \( \eta < 0.13 \) generally confirm these expectations. The expected nonclassical binomial shot-noise statistics were verified at room temperature by varying the optical attenuation and hence the overall quantum efficiency, \( \eta \). These measurements have been performed with commercially available emitting diodes.

2. Measurement technique

2.1. Equipment

A schematic diagram of the noise measurement set-up is shown in figure 1. The infrared light-emitting diodes under test are physically located and electrically accessed through a receptacle mounted on the front panel of a diecast aluminium box. This, together with a similar housing for the large area (100 mm\(^2\), 50 ns rise-time, silicon pin diode detector can be mounted on an optical bench for room temperature studies or placed in a cryostat.

The source module contains the capacitive and resistive components required to decouple and drive the device under test and the reference sources from external stabilised power supplies. The detector module provides current input to a low-noise transimpedance amplifier with a 3 dB bandwidth of 2 MHz. Following further amplification the shot-noise signal is input to a swept-frequency spectrum analyser (Advantest 4131 B) at a level of −115 dBm Hz\(^{-1}\), corresponding to a shot-noise current of 2 mA. The low frequency noise power spectral density at this current is 20 dB above the amplifier noise level in the final set-up.

2.2. Procedures

Both swept and single frequency noise measurements have been made. Data obtained in these two modes are available for direct plotting and computer processing by means of GPIB connected peripheral equipment. Single frequency spectral density measurements made with a resolution bandwidth of \( B = 30 \) kHz yield a statistical precision of \((BT)^{-\frac{1}{2}} \sim 0.1\% (0.005 \) dB) when numerically integrated over 20 s intervals.
The procedures were directed towards measurement of the detector photocurrent noise power spectral density, relative to the shot-noise level, within the modulation bandwidth of the light emitting diodes which were selected for high quantum efficiency and closely optically coupled to the pin diode photodetector.

Since the Fano factor was expected to be greater than about 0.9 at room temperature, corresponding to overall quantum efficiencies, $\eta$, less than 10%, precise measurement of noise spectral density relative to a standard shot-noise source were required. Hence, although Tapster et al. [4] found the Fano factor of a tungsten lamp to be within 1% of the calculated shot-noise level, other reports [5] have suggested uncertainties greater than this due to excess noise having a dependence on detector illumination geometry and source spectrum. Considerable effort was therefore expended in a search for such effects. None were found within the systematic and statistical uncertainties of the measurements (0.5%) at frequencies between 200 kHz and 5 MHz and colour temperatures between 2000 and 2700 K. This confidence in establishing the shot-noise reference level was reinforced by measurements made with low efficiency LEDs with similar peak emission wavelengths operating under similar geometrical conditions.

The Fano factor, $F_n$, was obtained as the ratio of the noise power spectral densities for the devices under test, $S(0,T)$, and the reference shot-noise source, measured at the same frequency and the same photocurrent:

$$F_n(T) = \frac{S(0,T)}{2\eta e}.$$  

All measurements were performed in an electromagnetically shielded room.
3. Sub-Poisson noise measurements

Measurements were made on a variety of commercially available infrared light emitting diodes.

Consistent reduction of noise below the standard shot-noise limit was observed for high-power GaAlAs type L2656 and L2690 infrared LEDs manufactured by Hamamatsu Photonics. The higher efficiency type L2656 devices exhibited nearly constant quantum efficiency and low frequency quantum noise reduction, typically 10% (0.4 dB) at room temperature, for drive currents from 10 mA to 45 mA. Figure 2 (a) shows a typical set of real-time spectra obtained at dry-ice temperature in which noise quieting below the shot-noise level is clearly evident.

![Graph showing noise reduction](image)

**Figure 2.** (a) Typical raw noise spectra showing 0.5 dB reduction (D) below the reference shot-noise level (R) for shot-noise current \( <i> = 2 \) mA at dry-ice temperature (195 K). Trace (A) represents amplifier noise level only \( (<i> = 0) \). (b) Noise power spectral density measurements showing high impedance suppression of shot noise from infrared emitting diodes at room temperature. (Hamamatsu type L2656 (LED) closely coupled to the pin diode detector with overall quantum efficiency of 10% and with efficiency reduced to 5% by insertion of a neutral density filter (LED + NDF)).
Figure 2 (b) with a wider scale illustrates the increase in Fano factor with optical attenuation expected for sub-Poisson light with binomial statistics [2]. The same results are obtained irrespective of whether the attenuation is increased geometrically or by using neutral density filters. Figure 3 summarizes the measured noise spectral density, relative to shot-noise, as a function of overall system quantum efficiency. The latter quantity was measured as the ratio of detected photocurrent to LED drive current. This increased typically from 0.1 to 0.13 as the temperature was lowered from room temperature to dry-ice temperature (195 K).

Inspection of figure 3 shows the dependence of the noise reduction on overall quantum efficiency, \( \eta \), to be close to that expected from equations (2) and (3).

4. Conclusions

Quantum noise measurements on high efficiency GaAlAs infrared light emitting diodes indicate noise levels significantly (0.4 dB) below the standard shot-noise level at room temperature. The amount of shot-noise suppression is shown to be proportional to the overall quantum efficiency of the source/detector system in accordance with predictions by Teich et al.[2].

This behaviour is consistent with the nonclassical (binomial) photon statistics associated with number-squeezed, sub-Poisson light. Substantially greater noise

![Figure 3. Noise power spectral density \( S(f) \) as function of overall quantum efficiency, \( \eta \) measured at \( f = 300 \) kHz, \( \langle i \rangle = 1 \) mA, normalized to the shot-noise level \( (2\langle i \rangle \sigma) \) at room temperature (295 K) and dry-ice temperature (195 K). [lowest point]. The dashed line \( (1 - \eta) \) is the post-detection Fano factor for a noiseless source showing the transition from binomial to Poisson statistics expected for number-squeezed light.](image)
suppression should be attainable within current technology with applications to measurement and communications [6].

Acknowledgements
The author wishes to acknowledge technical assistance by Mr J. Matthews and Mr R. Allen and helpful discussions with members of the University of Canberra Optoelectronics Group including Dr L. Barmpoulos, Dr W. Cheung, Dr G. French, Dr G. Pollard, Dr P. Lyons and Dr J. Rayner.

References
Quantum Noise-Correlated Operation of Electrically Coupled Semiconductor Light Emitters

P. J. Edwards and G. H. Pollard

Faculty of Information Sciences and Engineering, University of Canberra, Belconnen, Australian Capital Territory, Australia

(Received 1 May 1992)

Electrically coupled light-emitting diodes have been configured to demonstrate multiple light beams with correlated sub-shot-noise intensity fluctuations. Photocurrent covariance measurements of the two light beams from a pair of series-connected diodes show a positive correlation, similar to that observed with nonlinear crystal down-converters. Shunt-connected diodes are shown to generate negatively correlated light beams.

PACS numbers: 42.50.Dv

Sub-shot-noise "amplitude-squeezed" operation of laser and light-emitting diodes using the high-impedance "quantum-watchdog" noise suppression technique [1] is now well established [1–3]. This technique appears to offer a technically simple way of reducing quantum noise well below the standard quantum limit in tightly coupled optoelectronic systems, constrained only by the magnitude of the attainable quantum transfer efficiency between light emitter and detector. Such a reduction in quantum noise is of great practical interest with potential applications in shot-noise-limited metrology and information transfer.

The high-impedance quantum noise suppression technique, which simply involves driving a laser diode or light-emitting diode from a low-noise high-impedance current source, is also of considerable theoretical interest. Particular interest currently centers on the appropriate quantum-mechanical description of the interaction between the classical electrical pump (macroscopic reservoir) and the quantum-mechanical electron-hole recombination (microscopic) system. Richardson and Yamamoto [1] interpret this interaction, in which the junction voltage of a laser diode "measures" the photon number fluctuation, as an example of a quantum-mechanical-watchdog effect.

We have extensively studied the phenomenology of quantum noise reduction using high-efficiency infrared-emitting diodes [3]. These are simpler to analyze and measure than laser diodes because of the absence of amplified spontaneous emission. Following the work of Teich and Saleh [4], we have found a conceptual and quantitative description in terms of electronic and photonic stochastic point processes to be entirely adequate. In particular, in investigating the Richardson, Machida, and Yamamoto partition noise model of the transverse junction stripe laser, which they used to obtain an 85% reduction below the shot-noise limit [5], we have been led to the realization of a new class of multiple-beam light-emitting devices. These are arrays of laser or light-emitting diodes which generate twin (or multiple) light beams between which the intensity fluctuations are strongly correlated and in which the correlation extends into the quantum (sub-Poissonian photon statistics) regime.

The largest reported (85%) noise reduction, achieved with a transverse junction stripe laser [5], has been the subject of considerable debate. It appears to run counter to the conventional wisdom that the fractional noise reduction, relative to the shot-noise level, is just equal to the quantum efficiency $\eta$ as measured by the current transfer ratio between mean emitter and detector currents.

Referring to the light-emitting junction diode—diode detector noise equivalent circuit of Fig. 1 and utilizing the Burgess variance theorem [4], the mean-square photocurrent noise fluctuation (the photocurrent variance) in bandwidth $B$ in the detector is given by

$$\langle i_d^2 \rangle = \eta^2 \langle i^2 \rangle + \eta (1 - \eta) 2 \langle i \rangle eB,$$

where the mean quantum efficiency $\eta = \langle i_d \rangle / \langle i \rangle$, and $i = \langle i \rangle + i$, $i_d = \langle i_d \rangle + i_d$. We have verified [3,6] Eq. (1) using high-quantum-efficiency infrared-emitting diodes.

FIG. 1. Simplified low-frequency noise equivalent circuit of light-emitting junction diode (with $R =$ series resistance, $\langle v_r^2 \rangle = 4kTBR$, $r =$ differential resistance $= mkT/e\langle i \rangle$, $\langle v_r^2 \rangle = 2mk \times TRB$) coupled to lightwave detector with quantum efficiency $\eta = \langle i_d \rangle / \langle i \rangle$. Detector current and junction diode current variances are related through Eq. (1). For the diodes used, $m = 1.3$, $R = 2.5$ Ω.
MEN Sub-shot noise LED operation

The normalised spectral density

\[ F = \left( \frac{\langle i_2^2 \rangle}{2 \langle i_d \rangle eB} \right) \]

These initial results [3] have now been extended to obtain a noise suppression greater than 1.5 dB of amplitude squeezing using (a) balanced delay line direct detection, and (b) direct detection using an incandescent lamp as a shot-noise reference source. Figure 2 shows the measured spectral density to be a linear function of the system quantum efficiency, as expected from Eq. (2). Measurements by the NTT group of greater quantum noise reduction have been interpreted [5] in terms of the suppression of current partition noise within the diode structure.

Referring to Fig. 3 which represents two shunt-connected light-emitting diodes, configured to simulate the proposed physical laser model [5], it may be shown from Kirchoff's circuit equations that the branch current noise variance may be written in the form

\[ \langle i_1^2 \rangle = \langle i_2^2 \rangle / 4 + \langle i_p^2 \rangle. \]  

The first term in Eq. (3) represents the mean-square current noise, \( \langle i_2^2 \rangle \), split between the two current branches, and the second term, \( \langle i_p^2 \rangle \), represents the current partition noise. The partition noise term for identical branches and symmetric current splitting is

\[ \langle i_p^2 \rangle = kTB(mr + 2R) / (r + R)^2, \]

so that the additional partition noise, together with the junction current noise, should be suppressed as the differential resistance of the diodes

\[ r = \frac{mT}{e} \left( \frac{1}{I_1} \right) \]

is reduced relative to \( R \), the series resistance. In the limit as \( \langle I_1 \rangle \) increases and \( R \gg r \), sub-shot-noise operation is restored, since

\[ \langle i_1^2 \rangle / 2 \langle i_d \rangle eB = r / R \ll 1. \]

The branch current fluctuations \( i_1 \) and \( i_2 \) may be shown by a simple application of the classical Kirchoff circuit laws to be negatively correlated via the current partition noise.

Their covariance may be written

\[ \langle i_1 i_2 \rangle = \langle i_1^2 \rangle / 4 - \langle i_p^2 \rangle. \]

Equation (5) indicates that the partition noise may be directly measured by cross correlation of the two branch fluctuation currents \( i_1 \) and \( i_2 \). We note in passing that Eqs. (3)-(5) are analogous to those describing the random partitioning of a photon beam at an optical beam splitter or an electron beam between charge-collecting

FIG. 2. Noise power spectral density as a function of current transfer (quantum) efficiency (\( \eta \)) measured at 300 kHz, normalized to the shot-noise level for infrared-emitting diode (Hamamatsu L2656) coupled to silicon p-i-n diode detector. The dashed line is given by Eq. (2).

FIG. 3. (a) Shunt-connected infrared-emitting diodes configured to generate negatively correlated lightwave intensity fluctuations. (b) Detector photocurrent covariance plot for shunt-connected L2656 infrared-emitting diodes (\( \eta = 0.12 \)). Horizontal scale, 0.5 ps/bin. \( R_S \approx 1000 \Omega \). Maximum (negative) correlation: \( -0.030 \pm 0.005 \).
electrodes in a vacuum tube. The current partition noise term $(i_2^2)$ is evidently the classical analog of the quantum-mechanical vacuum fluctuation as has been pointed out by Yamamoto [7].

We have verified the physical basis of the NTT model [5] by cross correlating the detected photocurrent fluctuations from two shunt-connected Hamamatsu L2656 light-emitting diodes, as shown in Fig. 3(a). The output currents from two large-area $(100 \text{ mm}^2)$ silicon photodiode detectors were amplified, low-pass filtered $(B = 350 \text{ kHz})$, sampled at $2 \times 10^6/5$, and cross correlated in a single-bit digital correlator (Saicor model 42A). All measurements were performed in an electromagnetically screened room. As in previous measurements on single light-emitting diodes [3], particular attention was given to reducing common-mode power supply and amplifier noise and to minimizing parasitic coupling between correlator channels. These measures were facilitated by the high sensitivity of the digital correlator when operated for integrating periods of the order of minutes. As predicted, a small negative correlation was observed. Figure 3(b) shows a typical result. We find that this correlation, due to partition noise according to Eqs. (4) and (5), disappears in the limit $rIR \ll 1$, as expected. For convenience, the measurements were performed at room temperature with a relatively low value $(12\%)$ of the quantum efficiencies $(\eta_1, \eta_2)$. Operation at and below the shot-noise level was confirmed as in previous experiments [3] to eliminate the possibility of spurious classical noise correlations.

These results directly support the proposed laser model. Also, it is to our knowledge the first demonstration of intensity-correlated operation of a pair of semiconductor light-emitting devices. We believe from the similarity between the noise equivalent circuit representations of laser and light-emitting diodes that the result holds for pairs of shunt-connected laser diodes as well as for shunt-connected light-emitting diodes.

The covariance between the detected photocurrent fluctuations is related to the branch current covariance by

$$\langle i_1i_2 \rangle_d = \eta_1\eta_2\langle i_1i_2 \rangle,$$

and the relation between the variances is given by Eq. (1). Accordingly, we may show that the correlation coefficient for symmetric, current-balanced partition with $\eta_1 = \eta_2 = \eta$ is given by

$$r_{12}(\eta, x) = -\eta F_e(x)/[1 + \eta F_e(x) - 1],$$

where $F_e(x) = x(2 + mx)/m(1 + x)^2$ is the "electron" Fano factor in either branch, taken here to be dominated by the partition noise $(\langle i_2^2 \rangle \gg \langle i_1^2 \rangle)$, and $x = rIR$ is the ratio of differential diode resistance to series resistance in each branch.

It is evident that full negative correlation is expected for $\eta = 1$, and that the covariance vanishes for either $\eta = 0$ (high attenuation) or $F_e(x) = 0 (rIR = 0$, complete partition noise suppression). The maximum negative correlation coefficient

$$r_{12}(\eta, \infty) = (-\eta/2)/(1 - \eta/2)$$

occurs for the limiting case $x \gg 1$, for which case, $F_e = 0.5$.

The situation is similar for series-connected diodes [Fig. 4(a)] in which case the correlation is positive, as shown in Fig. 4(b). Considering the case of $n$ series-connected diodes, and using the equivalent circuit of Fig. 1, we find the electron Fano factor in the current loop,

$$F_e(x) = x(2 + mx)/mn(1 + x)^2.$$

The correlation coefficient between the photocurrent fluctuations due to a pair of series-connected light-emitting diodes is then given by Eq. (7) with a change of sign.

It is evident that full correlation again occurs for $\eta = 1$ and also that maximum quantum noise reduction occurs in the limit as $x$ tends to zero. For the "ideal" case of negligible series resistance $(x \gg 1)$, the electron Fano factor $F_e = 1/n$. The measured correlation of $0.047 \pm 0.005$ shown in Fig. 4(b) for $n = 2, x = 2.6$, and $\eta = 0.12$, is in

![FIG. 4. (a) Series-connected infrared-emitting diodes configured to generate positively correlated intensity fluctuations. (b) Detector photocurrent covariance plot for series-connected L2656 infrared-emitting diodes $(\eta = 0.12)$. Horizontal scale, $0.5 \mu\text{s/bin}. \ R = 5 \Omega$ (internal to diodes). Maximum correlation, $0.047 \pm 0.005$.](image)
reasonable agreement with the predicted value of 0.052.

This demonstration of a strong positive quantum correlation for series-connected light-emitting diodes is of theoretical and practical interest. Providing the correlation is maintained at low photon fluxes, series-connected light-emitting diodes evidently provide a technically simple means of generating quantum noise-correlated light beams similar to the "photon-paired" beams generated in nondegenerate parametric down-converters [8]. Although we have not yet demonstrated this phenomenon with laser diodes, the similarity between the electronic and photonic models of LEDs and laser diodes encourages us to expect similar behavior. Unlike photon-pair generation in nonlinear crystals, there is no momentum, polarization, or energy coupling between the photon pairs. The mean photon energies in each of the beams may be varied independently of those in the other beams by varying the energy band gaps. Also, unlike the situation with nonlinear crystals, one is not limited to photon pairs. Triplets and high-order multiplets may just as easily be generated with no loss of correlation. Of course, from Eqs. (7) and (8), the maximum photocurrent correlation between any pair in an array of $n$ emitters is just

$$P_{2} = \frac{\eta}{[\eta+n(1-\eta)]} . \tag{9}$$

The correlation evidently extends below the shot-noise limit. Indeed, for both series and shunt diodes the individual beams will normally be amplitude squeezed, as we have demonstrated. We see here a clear example of a quantum noise correlation [9], in the form of a classical correlation between sub-Poissonian photon fluxes. No violation of the "two-beam" Cauchy-Schwarz inequality [10] occurs because the magnitude of the correlation coefficient is always less than or equal to unity.

One of the practical applications of the correlation may lie in those optical attenuation measurements in which the noise is dominated by intensity fluctuations in the source [8]. In the present case we might consider the use of dual-beam absorption spectrophotometry of weakly attenuating media as an example. Note that by subtracting from each of the single-beam photocurrents $I_d$ a fraction $\eta$ of the diode array current $I$, we recover the case described by Eq. (2) in which the noise reduction below the shot-noise level is the maximum permitted by statistical fluctuations in the single-beam quantum transfer efficiency.

Another possible application is in the generation, by $n$-fold replication, of robust near-photon-number states as proposed by Yuen [11].

One of us (P.J.E.) would like to thank Y. Yamamoto and S. Machida of NTT Basic Research Laboratories and P. Tapster and J. Rarity of the Royal Signal and Radar Establishment (Malvern, United Kingdom) for useful discussions. This work was supported by the Australian Department of Industry, Technology & Commerce and by Hamamatsu Photonics (KK).

Reduction of Optical Shot Noise from Light-Emitting Diodes

Paul J. Edwards

Abstract—We address the mechanism responsible for the reduction of optical shot noise from high impedance driven light-emitting diodes. Contrary to recent assertions by Kikuchi and Kakui [1], the electronic feedback mechanism is shown to be the same as that for laser diodes with spontaneous recombination providing the underlying Poisson process in place of stimulated recombination.

I. INTRODUCTION

In a recent paper [1], Kikuchi and Kakui report a correlation between the intensity fluctuations in the light beams from a pair of series-connected light-emitting diodes. They report using this correlation to achieve a 0.45 dB noise reduction below the standard shot noise limit. These measurements are similar to those which have previously been reported using electrically coupled light-emitting diodes [2]-[7] following earlier quantum noise reduction achieved with single light-emitting diodes [8]-[12], [13]. However, unlike the authors of this previous work, the authors of reference [1] interpret their results as showing no reduction in the quantum noise level in the light from a constant current driven LED. Indeed, they suggest that such reduction is restricted to laser diodes [1].

It is the purpose of this paper to draw attention to the considerable corpus of experimental and theoretical results which suggests the operation of the same electronic feedback mechanism for quantum noise reduction in light-emitting diodes as in laser diodes. This mechanism was initially proposed and demonstrated by Yamamoto et al [3], for lasers and by Teich and Saleh [14] for light-emitting diodes.

The first reported quantum noise reduction using light-emitting diodes [8], has been confirmed by several authors [2], [10]-[12]. Good agreement between experiment and theory has been demonstrated. The diode noise equivalent circuit of Fig. 1, together with a point-process statistical model of the carrier recombination/photons emission/attenuation/detection process apparently provides an adequate model for determining the current noise and optical noise of single and multiple element LED arrays [6].

II. CURRENT NOISE IN LIGHT-EMITTING DIODES

Referring to the diode noise equivalent circuit of Fig. 1, in which \( \eta = \text{quantum efficiency}, \ C = \text{diode diffusion capacitance}, \ v_i = \text{Nyquist noise voltage associated with series resistance}, \ R \) and \( v_n = \text{shot noise voltage associated with diode differential resistance}, \ r \), we write the bias current as

\[
I_b(t) = \frac{1}{4} + i(t),
\]

the recombination current as

\[
I(t) = I + i(t),
\]

the photodetector current as

\[
i_d(t) = ii + i_d(t),
\]

and the diode junction voltage as

\[
V(t) = V + v(t) = V + q(t)/C,
\]

from which the state equation may then be written in the form

\[
C \frac{dv}{dt} = -v(1/r + 1/R) + v_n/r + v_i/R \quad (1)
\]

or

\[
\frac{dq}{dt} = -q(1/\tau + 1/RC) + v_n/r + v_i/R. \quad (2)
\]

The two independent noise sources evidently constitute a noise term of mean square value

\[
\langle f_i^2 \rangle = \langle v_i^2 \rangle /r^2 + \langle v_i^2 \rangle /R^2
\]

\[
= \langle i_d^2 \rangle + \langle i_t^2 \rangle
\]

with corresponding current noise spectral density

\[
S_i(f) = 2I \cdot e + 4kT/R. \quad (4)
\]

The first term represents the shot noise current spectral density for the recombination current noise, \( i_n = -v_n/r \) which flows when the junction voltage is held constant. The second term represents the Nyquist noise current

Manuscript received October 12, 1992; revised January 25, 1993.
The author is with the University of Canberra, Belconnen, Australia.
IEEE Log Number 9210243.
spectral density for the total junction current noise \( i_s = v_s / R \), which flows in the external resistance when presented with a low impedance (\(< < R\)) by the diode.

By writing (1) in the form

\[
C \frac{dv}{dt} = -i(t) + i_b(t), \text{ from which}
\]

\[
\frac{dv}{dt} = (v_a - v) / \tau + (v_a - v) / RC
\]

it may be seen that the spectral density of the junction voltage noise, \( v(t) \), decreases by 6 dB/octave above corner frequencies of \( 1 / \tau \) (for \( R >> r \)) and \( 1 / RC \) (for \( R << r \)), where \( \tau = rC \), the carrier lifetime. Thus the junction voltage noise is strongly suppressed at high frequencies, \( f >> 1 / \tau \), providing \( R >> r \), and strongly suppressed at all frequencies for \( R << r \) [13]. For such constant junction voltage operating conditions, it is evident from (5) or from Fig. 1 that the recombination current noise \( i(t) = i_b(t) \), the shot noise current. These are the two situations for which we would therefore expect the standard electronic shot noise level with spectral density given by the first term of (3), and for which the standard photon shot noise level is, in fact, observed [2], [7], [11], [12].

However, on time scales \( \Delta t >> \tau \), a negative feedback mechanism operates to suppress the noise below the shot noise level. This mechanism belongs to a class of processes referred to by Teich as “rate compensation by excitation feedback” [14]. Random fluctuations due to the underlying shot noise process are low pass filtered and fed back, via the fluctuating junction voltage, to reduce the variance of the recombination current noise.

If we choose \( R >> r \) then the first term on the right hand side of (5) dominates. Then, at frequencies \( f << 1 / \tau \), junction voltage, \( v(t) \), closely follows the shot noise voltage, \( v_a(t) \). Under these conditions the recombination current noise, \( i = (v - v_a) / r \), is obviously much reduced. It will, in fact, be nearly equal to the Nyquist current noise, \( i_b(t) \), since, at low frequencies, the capacitive current in (1) will be negligible in comparison with the resistive branch currents.

Thus in this limit, \( \langle i^2 \rangle = \langle i_b^2 \rangle \) and the spectral density of the recombination current noise is that of the Nyquist current noise source, \( 4 kT/R \). This results in recombination current noise well below the shot noise level since the ratio of the Nyquist spectral density to the shot noise spectral density [2].

\[
F = (4 kT / R) / 2 I \cdot e
\]

is very much less than one.

The operation of the internal feedback mechanism which is responsible for the current noise reduction is demonstrated in Fig. 2. The state variable \( q(t) \), the stored charge fluctuation, determines the junction voltage fluctuation \( v(t) = q(t) / C \), which provides current noise reduction through the negative feedback path shown. At high frequencies, \( f >> 1 / \tau \), the feedback path is effectively disabled, the high frequency components of \( q \) and \( v \) are both negligible and the full shot noise current, \( v_s / r \), flows.

This result can be seen directly from the noise equivalent circuit of Fig. 1 and state equation (5) represented to show the internal negative feedback mechanism which operates, via the junction voltage \( v(t) \), to suppress the current recombination noise, \( i = (v - v_a) / r \), below the shot-noise level.

This is the basis of optical shot noise suppression in both laser diodes and light-emitting diodes. Assuming for the moment an internal quantum efficiency of unity, it is evident that one to one correspondence between charge carrier injection, radiative recombination, photon emission and photon detection must result in identical electronic and photonic noise at low frequencies, that is, on time scales long compared with the recombination lifetime.

### III. Photonic Noise in Light-Emitting Diode Systems

Spontaneous emission through radiative recombination is a stochastic process with probability \( 1 / \tau \) per carrier per unit time. When the junction voltage \( V(t) = V + v(t) \), the junction charge \( Q(t) = Q + q(t) \) and the charge carrier number, \( N(t) = \langle N \rangle + n(t) \), are all held constant, this reduces to a simple Poisson process with mean rate \( \lambda = \langle N \rangle / \tau \). However, in the presence of feedback the above three quantities are all modulated by the (single-pole) low-pass filtered shot noise (Poisson) process. This constitutes a self-exciting point process, SEPP, with reduced electron number variance and reduced current noise spectral density at frequencies within the feedback loop bandwidth [14], [15]. Note that no direct optical feedback is involved. In the preceding section the junction current noise in a light-emitting junction driven by a high impedance source was shown, by arguments similar to those advanced for laser diodes [13], to be suppressed below the shot noise level at low frequencies. A phenomenological argument was advanced to suggest a corresponding reduction in the photonic noise level seen by a high quantum efficiency detector. Since this reduction has been dis-
puted we now show that this quantum noise suppression is as predicted from the Langevin rate equation for the charge carrier number together with a semiclassical treatment of the photon noise [6].

Following (2), we obtain the rate equation for the charge carrier population number fluctuation, \( n(t) \), in the form

\[
\frac{dn}{dt} = -n(1/\tau + 1/RC) + f_\lambda(t),
\]

(8)

where, from (3), the Langevin noise term for the carrier number

\[
f_\lambda(t) = f(t)/e.
\]

The spectral density of the carrier number fluctuation,

\[
S_n(f) = \frac{2\langle N \rangle/\tau}{[1 + \omega^2\tau^2]} + \frac{4kT/e^2 \cdot R}{[1 + (\omega RC)^2]},
\]

(9)

as has already been remarked upon for the junction voltage noise, has corner frequencies at 1/\( \tau \) and 1/RC respectively.

From the equations (5), (8) the recombination rate,

\[
\lambda(t) = \langle N \rangle/\tau + n(t)/\tau + i_\lambda(t)/e
\]

(10)

is a stochastic variable comprising three terms. The first term is the mean rate expected for a Poisson process. The second is the feedback term which contains the carrier number fluctuation response to the stochastic fluctuation in the third, shot noise, term. The rate fluctuations are suppressed below the shot noise level as the carrier number fluctuation rises at low frequencies, \( \omega < 1/\tau \), to compensate for the shot noise current fluctuation:

If \( R \gg r \), the carrier number response to Nyquist noise can be neglected, \( n(t)/\tau + i_\lambda(t)/e = 0 \), and the recombination photon emission rates become noiseless on time scales long compared with \( \tau \).

For unity quantum transfer efficiency between LED current and photodetector current it is evident that the foregoing statements will also apply to the photocurrent noise. In this case, we may write

\[
I_d(t) = e \cdot \lambda(t) = I(t)
\]

and \( \langle \tilde{I}_d^2 \rangle = \langle \tilde{I}^2 \rangle \). Thus, if the junction current is noiseless then so is the photodetector current. A diagrammatic representation of shot noise suppression by electronic feedback is shown in Fig. 3.

More generally, if the overall quantum transfer efficiency, \( \eta (= I_d/I) < 1 \), we must take account of the additional noise due to random photon deletion. This may occur at any point in the link between carrier injection and photon detection.

We may then write the photocurrent

\[
I_d(t) = \eta \cdot I(t) + f_\lambda(t)
\]

where the photon deletion noise term

\[
\langle \tilde{f}_\lambda^2 \rangle = \eta \cdot (1 - \eta) \langle \tilde{I}_d^2 \rangle.
\]

The detected photon noise is then given by [4], [5], [6]

\[
\langle \tilde{I}_d \rangle = \eta^2 \langle \tilde{I} \rangle + \eta \cdot (1 - \eta) \cdot \langle \tilde{I}_d \rangle. 
\]

(11)

For the case of the high impedance current source in which \( \langle \tilde{I}^2 \rangle \ll \langle \tilde{I}_d^2 \rangle \) as described in the previous section, (11) becomes

\[
F = \langle \tilde{I}_d \rangle /\eta \langle \tilde{I}_d \rangle = \langle \tilde{I}_d \rangle /2 I_d \Delta f = 1 - \eta.
\]

(12)

This relation has been demonstrated for a number of high efficiency light-emitting diodes, in particular the Hamamatsu type 2556 infrared emitting Gallium Arsenide heterojunction diode [2], [7], [11], [12].

**Conclusion**

Quantum noise reduction below the optical shot noise level is not limited to laser diodes and to process involving stimulated emission. It is due to the effect of negative feedback on the recombination rate which occurs when the junction voltage and carrier number are permitted to respond to the underlying (Poisson) stochastic recombination process as in a light-emitting diode or laser diode driven by a high impedance source. Extensive measurements on light-emitting diodes confirm the magnitude of the noise suppression predicted from semiclassical theory.

**Acknowledgments**

The author thanks Dr. J. Rarity and P. Tapster (DRA, Malvern) and Prof. Y. Yamamoto (NTT and Stanford University) for useful discussions and Hamamatsu Photonics KK for support.

**References**


Paul J. Edwards was born in 1938 in Launceston, Tasmania, Australia. He received the B.Sc. (first-class honors) in physics in 1960 from the University of Tasmania. He obtained the Ph.D. in the field of cosmic ray instrumentation, specializing in data telemetry, in 1964 at the same university.

He was appointed Lecturer in Physics at the University of Adelaide ('64-'68), Senior Lecturer in Physics ('68-'70), and Associate Professor of Physics ('70-'82) at the University of Otago, New Zealand. In 1982, he returned to Australia to head up the first civilian undergraduate engineering course (in electronics and communications) in Canberra at the Canberra College of Advanced Education, (now the University of Canberra). Professor Edwards has worked in the fields of signal processing, radiowave and lightwave instrumentation, astronomy, and atmospheric physics with a special interest in stochastic processes. His current interest is in the engineering applications of squeezed light in communications, metrology and computing. He has worked in these areas at the Royal Signals and Radar Establishment, Malvern, U.K., and at the NTT Basic Research Laboratories, Musashino-shi, Tokyo.

Dr. Edwards is a Fellow of the Institution of Radio and Electronics Engineers (Australia), Institution of Engineers (Australia), Australian Institute of Physics, and New Zealand Electronics Institute.
An Experimental Introduction to Non-Classical Light for Second Year Electronics and Communications Engineering Students at the University of Canberra

PAUL J EDWARDS AND HUANG XU

The Second Year Applied Physics Course at the University of Canberra is designed to equip engineering and applied physics students with a theoretical and conceptual basis for a physical understanding of the materials, devices and systems used in modern information, communications and electronic technologies.

We present it in the form of six modules, each of twelve lectures, delivered over two semesters. The associated laboratory and tutorial programme occupies approximately sixty hours. The module titles indicate their content: Statistical and Quantum Physics; Atomic, Molecular and Quantum Physics; Semiconductor Physics; Electromagnetics; Optics; Electronic Materials and Devices. Students are asked to submit two formal reports. These are based on the laboratory work and lecture material and address engineering and other applications. The course emphasises physics as an engineering science: we attempt to show the relevance of modern physics to current, new and emerging technologies.

There is an initial emphasis in the lectures on probability and statistics. We judge physics to provide as good a vehicle as any to deliver this essential material in an engineering course like ours which contains strong communications and instrumentation strands. In the third and fourth years of the BE course, the second year physics context for this material changes to a discussion of more specific device, instrumentation, and communication applications.

The laboratory work includes investigations of radioactive decay, thermal noise, electronic shot noise and photonic shot noise, these serving to illustrate lecture material relating to elementary probability concepts and stochastic processes with Poissonian, binomial and Gaussian statistics.

Here we describe one of these, a laboratory investigation of electronic and photonic shot noise, taken from our research programme in quantum electronics [1], which includes a simple demonstration of non-classical light with sub-Poissonian photon number counting variance. Since we wish to encourage students to form simple conceptual models and to gain some physical insight into these phenomena, we take a semi-classical approach [2] rather than attempt a rigorous quantum mechanical treatment. This investigation has been successfully incorporated into the second year laboratory program for several years.

Prof Paul Edwards and Dr Huang Xu are in the School of Electronics, Engineering and Applied Physics Faculty of Information Sciences and Engineering at the University of Canberra.
Stochastic Phenomena

Our starting point, following an introduction to the elements of probability theory, is the concept of Bernoulli trials. We show that the repetition of a trial or "experiment" in which an event occurs with probability, $p$, generates a random variable – the number of events in $N$ trials – having a binomial distribution. We illustrate this by solving a classical problem: the fluctuations in the number of atoms present in one part of a container of an ideal gas. This leads naturally, through a discussion of number density fluctuations, to the Poisson distribution as a limiting case of the binomial distribution as the number of trials (molecules), $N$, becomes large and the event probability, $p$, (fractional sample volume) becomes small.

We then move to the time domain and introduce the idea of a stochastic point process comprising the set of $N$ event times distributed over time interval, $T$. We again show the transition from binomial to Poisson statistics for the number of events in a subinterval $\Delta T$ as $\Delta T \rightarrow 0$ and $N \rightarrow \infty$.

We discuss the ubiquity of the Poisson process in providing a realistic model for such physically diverse stochastic phenomena as telecommunications traffic, motor traffic, queueing, thermionic emission, photoelectron emission, photodetection, charge carrier injection in semiconductor junctions, atomic spontaneous emission and radioactive decay. We emphasise the unique "memoryless" character of the Poisson process as characterised by the lack of dependence of the number of events occurring in a given time interval on the number of events outside that interval. All students are asked to demonstrate in the laboratory the applicability of the Poisson probability density function to describe the distribution of counts in a radioactive decay experiment.

Shot Noise

The linear superposition of Poisson distributed single electron charge pulses to generate thermionic and photoelectric "shot noise" currents is discussed in lectures and also investigated in the laboratory. In lectures, simple variance algebra is used to relate the mean square current fluctuation, $\langle \Delta I^2 \rangle$ to the variance in the electron number count, $\langle n \rangle$ in time, $T$, and thence to derive the well known shot noise formula for the mean square current noise in frequency bandwidth, $B$:

$$\langle \Delta I^2 \rangle = 4e I \left( \frac{eT}{2} \right)$$

where $e$ is the electron charge, $T$ the time interval, and $I = \langle I \rangle$ is the mean current. The magnitude of the mean square shot noise current and its linear dependence on the DC (mean) current $\langle I \rangle$, (equation (1)) is confirmed using thermal noise as a reference standard. Students are asked to note the reduction in the electronic shot noise below that given by equation (1) when the current is space charge limited, by operating the thermionic diode with low anode voltage. In this mode, as has been well understood for many years [2], an internal negative feedback mechanism operates, via the fluctuating coulomb barrier presented by the electronic space charge within the vacuum diode, to reduce the number fluctuations in the electron stream arriving at the anode. This so called "anti-bunching" effect has become well-known in the field of quantum optics in semi-classical descriptions of sub-Poissonian photon number counting noise [2].

The laboratory investigation is consequently a suitable precursor to the next: "Photonic Shot Noise", in which students demonstrate photon noise reduction below the normal shot noise level, sometimes termed "intensity squeezing". They measure mean square current fluctuation noise some 30% below the standard shot noise level of equation (1) using non-classical light from a light emitting diode of high quantum efficiency driven by a high impedance current source and coupled to a high efficiency detector.

Sub-Poissonian ("Quiet") Light

The generation of non-classical light with photon number variance, $\sigma^2(n) = \langle n^2 \rangle - \langle n \rangle^2$, below the mean, $\langle n \rangle$, using constant-current driven semiconductor laser diodes and light emitting diodes is now well accepted [3,4,5]. It is the basis for a number of proposed new optoelectronic technologies in the field of communications, computing and metrology [6] where performance is often fundamentally limited by the presence of shot noise, previously regarded as god-given and irreducible. Our quantum electronics research group is interested in both the fundamentals and the applications of what we have termed [7] "Quiet Light". At the second year teaching level we describe in a qualitative way the negative feedback mechanism, which as in the case of the space charge-limited thermionic diode, suppresses the random bunching of charge carriers injected into a semiconductor junction from a high impedance source. We describe a simple "leaky charge reservoir" model of a quiet light emitter in which the depletion of the reservoir by radiative recombination leads to a negative correlation between the charge carrier number and the radiated photon number and thence to photon anti-bunching on a time scale long compared with the recombination lifetime. Students are able to contrast this model of a fluctuating population (with mean number $N$) of spontaneously recombining electrons (with mean lifetime, $\tau$) decaying to generate quiet light having suppressed photon number fluctuations, with the canonical model of a fixed (non-fluctuating) population, $N$, of excited atoms, nuclei or electrons with mean lifetime, $\tau$, which spontaneously emit a stream of quanta with Poissonian statistics. This model leads to a simple picture of a more or less regular stream of photons, each photon being emitted from an ideal quiet light source by each recombining electron. We point out that if the detector were also ideal, there would then be a one to one correspondence between each electron recombination event in the light source and each photoelectron generation event in the detector as shown in Figure 1. The recombination current, the photon current and the photoelectron current would then have identical, strongly

\[ \sigma^2(n) = \langle n^2 \rangle - \langle n \rangle^2 \]

Australian & New Zealand Physicist Volume 32, Number 8, August 1995
suppressed statistical fluctuations. The “real life” situation which falls rather short of this ideal is then investigated in the laboratory and the results are discussed in terms of Bernoulli deletion of photons between source and detector.

Non-Classical Light Investigation

Sub-Poissonian Light Generation with an Infrared Emitting Diode

This investigation compares the fluctuation noise measured in the photocurrent of a lightwave detector coupled to a “quiet” light emitting diode with that measured when the detector is irradiated with “ordinary” light from a tungsten lamp. The experimental set up is as shown in Figure 2.

The tungsten lamp is used as a reference light source. It has been shown to generate photocurrent noise within one percent or so of the shot noise level. The “unbunched” character of this light can in fact be easily demonstrated in a Hanbury Brown-Twiss type experiment using a beam splitter and intensity correlator. This reveals the absence of any intensity correlation between the split beams from the lamp while showing a negative intensity correlation, as expected, between the split beams from the quiet light source [8]. We do not ask second year students to carry out these correlation measurements although they are not unduly difficult to perform. In the first part of this investigation, the linearity of equation (1) is confirmed by plotting the mean square current noise at room temperature as a function of the DC photocurrent, using the lamp as source. The mean square current noise can be measured in the same way as for the electronic shot noise and thermal noise. However, more precise and convenient measurements are possible with a swept frequency spectrum analyser and digital plotter. Although their operation requires I>

Simplified semiclassical picture showing the ideal one to one correspondence between electron hole recombinations in a light emitting diode, the resulting photons, and the detected photoelectrons. If the driving impedance, $R_1$, is much greater than the diode impedance then the photon stream is antibunched (has sub-Poisson variance) on a time scale, $T \gg \tau$ the recombination lifetime and the low frequency photocurrent noise is below the normal shot noise level.
Undergraduate Second Year Engineering and Applied Physics Laboratory set up at the University of Canberra to demonstrate non-classical light and measure photon noise reduction.

We have found that our electronics and communications engineering students do not experience any significant difficulty in obtaining good plots of reproducible noise spectra (Figure 3).

In the second part, students are asked to measure the degree of shot noise suppression and relate it to the overall quantum efficiency, \( \eta \), easily shown to be the ratio \( \frac{I_d}{I} \) of the DC detector current, \( I_d \), to the DC light-emitting diode current, \( I \).

The expected form of that relation follows from a consideration of the statistical deletion process which is characterised by the conditional probability, \( \eta \), of a photoelectron hole pair being generated at the detector given an electron hole recombination event at the diode junction.

The photoelectrons are deduced to have a binomial probability distribution since their generation involves a set of \( N \) independent attempts (Bernoulli trials) by a regular stream of \( N \) recombinuing electrons, to each generate a photoelectron with probability of success, \( \eta \).

Since the number variance of the binomial distribution is \( \eta(1-\eta)N \), and the variance of the Poisson distribution is \( \eta N \), the fractional mean square photocurrent noise spectral density from the diode relative to that from the lamp, \( F \), is expected to be just \( (1-\eta) \) at the same DC current:

\[
F = \frac{\langle i_d^2 \rangle}{\langle i_{\text{em}} \rangle^2} = (1-\eta)
\]

The quantum efficiency of the light emitter is temperature dependent for reasons which are not immediately relevant to the main thrust of the investigation. At 77 K the intrinsic quantum efficiency of a Hamamatsu type L2656 infrared emitting diode is in fact approximately double its room temperature value.

Typically the overall transfer efficiency with a large area \((100\text{mm}^2)\) silicon pin diode detector increases from about 15% to 30% as the system is cooled to liquid nitrogen temperature. The expectation is therefore of a maximum shot noise suppression of close to 30%. This is readily measurable. >
Experimental Procedures and Results

The linearity of the shot noise equation is confirmed at room temperature by varying the lamp irradiance to give a range of detector currents up to 10 mA. This procedure serves to build confidence in the use of the equipment and to check the overall linearity of the set up. A series of pairs of noise density measurements at room temperature at specified DC currents using the lamp and the Hamamatsu infrared emitting diode alternately is then obtained and plotted. Since the expected differences are quite small, in the vicinity of 15%, this procedure also helps to hone the experimental technique and remove "finger problems" in the operation of the equipment. The diode current and detector currents are of course both measured to enable the quantum efficiency to be calculated. These initial results typically confirm equation (2) to within 10%, consistent with the statistical noise spectral density sampling error of a percent or so. Next, liquid nitrogen is cautiously added to the cryostat, initially under the supervision of the demonstrator. Temperature is measured with a copper constantan thermocouple with which students have already become familiar in Frank-Hertz, Magnetic Curie Point and similar thermal investigations. A cooling rate of less than two degrees per minute is enforced where possible to limit the failure rate due to differential contraction of the relatively expensive ($100) large area pin diode detectors. The noise measurements are then repeated on the cooled LED down to liquid nitrogen temperature. Figure 3 shows typical noise density plots. Finally, the noise suppression factor, $F$, is plotted as a function of the measured quantum efficiency, $\eta$, and compared with the theoretical expectation as in Figure 4. Students are asked to report, relate and comment on the three shot noise related investigations of electronic, thermal and photonic noise.

Non-classical light phenomena can be quite easily demonstrated in the undergraduate laboratory. A similar investigation to ours has also been introduced into the second year of the optoelectronics degree course at Macquarie University. Both these investigations emphasise the statistical, particle-like properties of photons. Quantum noise reduction experiments of this type, besides foreshadowing an emerging technology to students, also serve to introduce the topics of noise, stochastic processes, uncertainty relations, complementary and wave-particle duality in the lecture course. In the case of our engineering
Typical plot of student results for the photon noise suppression factor (Fano factor), $F$, as a function of the measured quantum efficiency, $\eta$, using the set up of Figure 2, compared with the semiclassical expectation, $F = (1 - \eta)$.

students, the noise topics are taken up again in later years of the course. Our applied physics and engineering graduates will increasingly be called on to use (and develop) new “quantum chips” and quantum physics based systems. At the University of Canberra we believe we must therefore equip them for their work into the next century with a practical and theoretical introduction to the relevant quantum phenomena.

We thank second year Electronics and Communications Engineering students: R Cruse, J Davies, PH Golding, JW Hecker and M Wing for permission to use their results and for their helpful comments and Dr HB Sun for assistance in the laboratory.

**Bibliography**


1. A quantum noise-correlated, multiple beam, lightwave generating and detecting system comprising:

(a) An array of electrically coupled semiconductor junction light emitters of high quantum efficiency.

(b) An array of high quantum efficiency light detectors optically coupled to the array of light-emitters referred to in (a) above.

(c) A means of optically coupling members of the array of light emitters to members of the array of light detectors referred to above with high efficiency.

(d) A low noise, modulated current supply to drive the array of light emitters.

(e) Provision for the placement of one or more optical cells in the optical coupling paths between members of the light emitting array and members of the light detecting array.
A means of combining, or otherwise operating upon, the plurality of detected beam signals in order to enhance the performance of the system by virtue of the quantum noise correlation existing between the beam signals.

MULTIPLE-BEAM QUANTUM NOISE-CORRELATED LIGHTWAVE SYSTEM

This invention relates to a method of generating and utilizing a plurality of lightwave beams which exhibit strongly correlated quantum noise fluctuations.

The present invention provides a multiple lightwave-beam-generating device comprising two or more electrically coupled, high quantum efficiency light-emitting semiconductor junction and/or laser diodes which are optically coupled to one or more lightwave detectors with high overall current transfer efficiencies.

In one form of the invention, two high quantum efficiency laser diodes are electrically connected in series and driven from a common power supply and modulator. Each laser diode is separately optically coupled to a high quantum efficiency p-i-n diode detector with high overall current transfer efficiency.

In another form of the invention the two laser diodes are connected in parallel (rather than in series), optically coupled as before, and driven from a common power supply and modulator.

The invention may be implemented using either visible or infrared light-emitting diodes or laser diodes, providing these are of high quantum efficiency.

For example, high quantum efficiency infrared emitting diodes and photon counting detectors may be substituted for the laser diodes and p-i-n diode detectors previously mentioned.
The Prior Art

Current state-of-the art devices for generating noise-correlated light wave beams are technically complex. They usually employ non-linear crystals operating as two-mode optical parametric oscillators or optical parametric down converters. These devices are usually bulky and inefficient, limited to two beams, and do not permit independent tuning of the wavelengths of the two lightwave beams.

The present invention provides for the more efficient and versatile generation of a plurality of intense noise-correlated light beams by a novel method which may be incorporated into lightwave instrumentation and communications systems.

Background to the Invention

The inventor was the first research worker in Australia (and the second in the world) to demonstrate, in 1990, the generation of quantum noise below the usual shot-noise level using light-emitting diodes driven from a high impedance current source. In the course of the analysis and generalisation of this work, the inventor was led to the realisation that laser diodes and/or light emitting diodes, when suitably electrically coupled generate quantum noise-correlated lightwave beams. The utilisation of noise-correlated beams of this type for precise measurement and secure communication has been foreshadowed by several workers. However, such utilisation has been dependent on the use of optical crystal parametric down converters or similar devices. The present invention provides a new way of generating multiple beams in which the quantum noise in each beam is at, or below, the shot-noise level when detected with high quantum efficiency. This fact taken together with the high quantum correlation between the plurality of beams permits the employment of noise reduction techniques and algorithms to reduce system noise. For example, a dual (or multiple) beam spectrophotometric instrument can utilise the invention to eliminate, or greatly reduce, quantum noise and other noise originating in the lightwave sources. Conventional
instruments can reduce noise down to, but not below, the shot noise level. The inventor has developed and demonstrated a prototype of the invention.

Brief Description of the Drawings

To assist with understanding the invention, reference will now be made to the accompanying diagrams which show two examples of the invention. The drawings are not meant to limit the scope of the invention in any way.

In the diagrams:

Figure (1) shows one example of a dual-beam lightwave system in which the detected lightwave intensities are negatively correlated.

Figure (2) shows an example of a dual-beam lightwave system in which the detected lightwave intensities are positively correlated.

Referring to Figure (1), the current supplied by the modulated high-impedance power supply (1) is split between the two semiconductor junction light sources (2, 3) which are optically coupled to p-i-n diode detectors (4, 5) with high current transfer efficiencies. These diode detectors are connected to a current summing device (8).

Providing the impedance of the modulated current supply in Figure (1) is very much higher than the diode resistance and the quantum transfer efficiencies are both unity, the intensity fluctuations are

(a) reduced below the normal shot-noise level
(b) negatively correlated with correlation coefficient of -1.

This means that, in the ideal case, the sum of the photodetector currents is then free of quantum noise and contains only the modulator signal.
Thus in this case the system could be used to measure differential optical attenuation between the two beams which traverse light paths containing optical cells (6, 7) with an effectively noise-free source leading to a consequent increase in precision and sensitivity.

Referring to Figure (2), the diodes (2, 3) are connected to a common current supply (1). Only one diode need be modulated (9). If both diodes are modulated, they must be modulated out of phase. In this case, referring to the previous discussion, the detected noise fluctuations will again be below the shot-noise level. However, in this case, the correlation coefficient =+1 for the ideal case. The difference current (8) derived from the detectors (4,5) will then contain the modulation and will be quantum noise-free.

The difference current, as shown in Figure (2), is then a measurement of differential absorption which may be made as a consequence of placing an optical sample in one or both of the optical cells (6,7).

In double or multiple beam sources of the type shown in Figure (2), the quantum noise correlation may be ideally maintained at +1, independently of the number of series diode elements.

The invention covers double and multiple beam lightwave systems element which incorporate as their essential feature a double or multiple element array of lightwave emitting semiconductor junction diodes with correlated quantum noise fluctuations, used (for example) in measurement and communications applications.

Current state-of-the-art generators of this type are technically complex. Current experimental dual-beam sources of this type include two-mode optical parametric oscillators (TMOPO) and optical parametric down converters.

The present invention in the form shown in Figure (2) is analogous to the TMOPO.
The claims defining the invention are as follows:

1. A quantum noise-correlated, multiple beam, lightwave generating and detecting system comprising:

   (a) An array of electrically coupled semiconductor junction light emitters of high quantum efficiency.

   (b) An array of high quantum efficiency light detectors optically coupled to the array of light-emitters referred to in (a) above.

   (c) A means of optically coupling members of the array of light emitters to members of the array of light detectors referred to above with high efficiency.

   (d) A low noise, modulated current supply to drive the array of light emitters.

   (e) Provision for the placement of one or more optical cells in the optical coupling paths between members of the light emitting array and members of the light detecting array.

   (f) A means of combining, or otherwise operating upon, the plurality of detected beam signals in order to enhance the performance of the system by virtue of the quantum noise correlation existing between the beam signals.

2. A system as claimed in Claim 1 where the light emitting array comprises laser diodes.

3. A system as claimed in Claim 1 where the light-emitting array comprises light-emitting diodes.
4. A system as claimed in Claims 1-3 where the light-emitting array comprises series-connected light emitters.

5. A system as claimed in Claims 1-3 where the light-emitting array comprises shunt-connected light emitters.

6. A system as claimed in Claims 1-5 where the means of optical coupling referred to in Claim 1(c) includes optical wave guides.

7. A system as claimed in Claims 1-6 where the optical cells referred to in Claim 1(e) include light-modulating devices.

8. A system as claimed in Claims 1-7 where the optical cells referred to in Claim 1(e) contain material samples whose properties are to be measured.

9. A system as claimed in Claims 1-8 where the array of lightwave detectors includes photon counting detectors.

10. A multiple beam correlated lightwave system substantially as herein described with reference to the accompanying drawings.
Abstract

A multiple beam lightwave system is disclosed in which a strong correlation is introduced between the detected quantum noise fluctuations in the beams. An array (2,3) of coupled laser diodes or light-emitting diodes is optically coupled with high quantum transfer efficiency to one or more lightwave detectors (4,5). Because of the quantum noise correlation, signal processing operations such as subtraction of photo detected beam intensities (8) allow the system noise due to quantum noise in the sources to be significantly reduced below the shot noise level. There is a consequent improvement in the precision of lightwave attenuation and other measurements.
LOW-NOISE OPTOELECTRONIC AMPLIFIER USING SUB-SHOT NOISE LIGHT

P. J. Edwards

Indexing terms: Optoelectronics, Amplifiers

A new type of low-noise optoelectronic amplifier is proposed which uses intensity-squeezed light from a tandem array of high quantum efficiency laser diodes or light-emitting diodes. Applications include amplification of intensity-squeezed light, sub-shot noise modulated lightwave generation for free-space and fibre transmission, and low noise small signal RF amplification.

Introduction: Sub-shot noise (i.e. 'intensity-squeezed') operation of laser and light-emitting diodes is now well established. The use of a high impedance current driven, high quantum efficiency emitting junction stripe laser has been reported \[2\]. Other reduction of 8 dB below the usual shot noise level was pioneered by Yamamoto et al. [1]. A reduction of 8 dB below the usual shot noise level was pioneered by Yamamoto et al. [1]. Other work [3] has confirmed the theoretical expectation that, for the limiting case of large electronic noise suppression, \( F_i \ll 1 \), the photonic noise (relative to normal detector shot noise) is given by

\[
F_0 = (1 - \eta) + \eta \cdot F_i \approx 1 - \eta
\]

(1)

for system quantum efficiency \( \eta \).

This equation expresses the well known fragility of intensity-squeezed light to the random deletion processes of attenuation and nonideal photodetection.

The direct-detection signal to noise ratio is then, in terms of input photon number \( n \), or detected photocurrent \( I \), measured in time \( \Delta t \):

\[
SNR = \langle I \rangle / \langle I^2 \rangle
= \langle I \rangle \cdot \Delta t / (1 - \eta) \cdot \epsilon
= \eta \cdot \langle n \rangle / (1 - \eta)
\]

(2)

From eqn. 1, the mean square input detector current noise, relative to shot noise, is

\[
F_s = 1 + \eta_0 (F_i - 1)
\]

(4)

Referring to the noise equivalent circuit of Fig. 2, it may be shown that the mean square current noise at the detected

\[
NF = 1 + \frac{1}{G \eta_1} [G (1 - \eta) + (1 - \eta_0) / \eta_0]
\]

(7)

Thus, for \( \eta_0 = \eta_1 = 0.9 \), \( G = 11 \) and sub-shot noise input light characterised by \( F_i = 0.1 \), the noise figure is 2.2 (3.4 dB). This degradation in SNR is mainly due to noise generated by the lossy \( (\eta = 0.9) \) detector. The optical gain \( G \), may, of course, be raised by introducing positive optical feedback to either the input detector or an auxiliary detector.

As \( \eta \) approaches unity, the input noise drops to zero in the limit \( \eta_1 = 1 \) and the system functions as a low noise photon number amplifier in the way envisaged by Yuen [4] provided \( G \eta_0 \gg 1 \). In this limiting case, setting \( \eta_1 = 1 \), eqn. 7 reduces to eqn. 6. This equation also describes the function of the system as a modulated sub-shot noise light generator. The effective quantum efficiency \( G \eta_0 \) is greater than unity and the photon noise suppression factor \( F_0 = 1 - \eta_0 \). The consequential signal to noise ratio advantage is then \( G (1 - \eta_0) \) over a shot noise limited (coherent) system with the same overall attenuation, \( \eta_0 \), in accordance with eqn. 3.

In this mode, the array may be modulated directly by, for example, a microwave satellite earth station receiving antenna. Sub-shot noise operation of the array [3] results in very low noise preamplification prior to lightwave transmission or direct detection.

In the latter case of direct detection, the system operates as a low noise amplifier with noise figure

\[
NF = 1 + \frac{1}{G \eta_1} [G (1 - \eta) + (1 - \eta_0) / \eta_0]
\]

(8)

Referred to the photonic noise in the input beam the noise figure may be written

\[
NF = 1 + \frac{1}{G \eta_1} [G (1 - \eta) + (1 - \eta_0) / \eta_0]
\]

and with excess system noise temperature,
output of the tandem array, relative to detector output shot noise, is

\[ F_0 = G \cdot \eta_0 \cdot F_d + (1 - \eta_0) \]

The noise figure of the light amplifier, referred to the photon-noise limited output of the input detector, is

\[ NF = (SNR)_o/(SNR)_n \]

\[ = 1 + (1 - \eta_0)/F_d \cdot G \cdot \eta_0 \]

\[ T_s = \frac{290 \cdot (1 - \eta_0) \cdot \langle I \rangle \cdot (R/50)}{\eta_0} \]

for reference temperature \( T_s = 290 \text{ K} \), laser current \( I \) (mA) and system impedance \( R \) (ohms).

Thus, for example, looking ahead to low-threshold laser diode/LED array technology in which large shot noise suppression is achieved at a pump current of \( \langle I \rangle = 1 \text{ mA} \) with a coupling efficiency \( \eta_0 = 0.9 \), we find from eqn. 9 that \( T_s = 3 \text{ K} \) for 20dB gain \( (N = 11) \) in a 50Ω system. Although this level of performance is not presently achievable at room temperature, it appears to be feasible with further development in laser diode array technology.

Conclusion: The imminent development of low threshold, high quantum efficiency laser diode and light-emitting diode arrays will provide a means towards realising:

(a) the Yuen noiseless photon number amplifier
(b) low noise direct modulation of high intensity sub-shot noise light
(c) low noise optoelectronic general purpose amplification with optical isolation.

Acknowledgments: This work has been supported by the University of Canberra, Hamamatsu Photonics KK, and NTT Basic Research Laboratories.

27th November 1992

P. J. Edwards (RF/Lightwave Communications Group, Faculty of Information Sciences & Engineering, University of Canberra, PO Box 1, Belconnen, ACT 2616, Australia)

References
Squeezed Light-emitting Diode Arrays

BY PAUL J. EDWARDS, UNIVERSITY OF CANBERRA, BELCONNEN, ACT, AUSTRALIA

A new type of squeezed light optoelectronic system has been devised that has potential applications in lightwave communications and instrumentation situations where performance is shot-noise limited. An electrically coupled tandem array of light-emitting diodes or laser diodes operates as a modulated photon number amplifier of sub-shot noise light.

Initial theoretical work on the generation of amplitude squeezed (sub-shot noise) light by light-emitting diodes and laser diodes has led to a number of laboratory investigations that demonstrate the ease with which amplitude squeezed light may be generated by semiconductor junctions, particularly light-emitting diodes.4,5

In the course of a recent investigation we were able to demonstrate, for the first time, the generation of quantum-correlated, sub-shot noise, twin-light beams from a pair of electrically-connected light-emitting diodes. This demonstration opens up the possibility of novel squeezed light devices and systems that use the quantum correlations between multiple light wave beams generated by a tandem array of semiconductor light emitters.

For example, if a tandem array of N semiconductor light emitters is driven by a high impedance source, or, alternatively, if the number, N, is sufficiently large, quantum noise is suppressed below the normal shot noise level (SQL) in the individual beams. The resulting composite beam, whether coherently or incoherently formed, will preserve the enhanced signal-to-noise ratio (SNR) characteristic of modulated sub-shot noise light. Moreover, this enhanced SNR can be made robust against attenuation, and thereby overcome the usual degradation of squeezed light in lossy systems.

Such an array may therefore be considered a modulated photon number amplifier of sub-shot noise light. Applications include operation and a sub-shot noise optoelectronic coupler with current gain greater than unity (incoherent operation) and as a generator of robust high SNR lightwave signals suitable for optical fiber input (coherent operation).

If the array is driven by a high quantum efficiency detector, it becomes a low noise photon number amplifier with a limiting noise figure of unity for both classical and non-classical light.4,6

"Wired" optical systems of this type are of theoretical interest because they demonstrate direct coupling between macroscopic, classical electronic circuit noise and non-classical (sub-Poissonian) photonic noise. They are well described semiclassically and consequently easily accessible to engineering analysis and design.

A variety of electronic and photonic manipulations is possible using electrically and optically coupled semiconductors and detectors. These may be more easily implemented than more conventional techniques using gas lasers and nonlinear media.

REFERENCES

Time-dependent Emission Spectra from Molecular Wave Packets

BY THOMAS J. DUNN, JOHN N. SWEETSER, AND IAN A. WALMSLEY, THE INSTITUTE OF OPTICS, UNIVERSITY OF ROCHESTER, ROCHESTER, N.Y., AND CZESLAW RADZEWICZ, INSTITUTE OF EXPERIMENTAL PHYSICS, UNIVERSITY OF WARSAW, WARSAW, POLAND

Wave packet states of atoms and molecules play an important role in studying the boundary between the classical and quantum domains. Such states are essentially nonstationary. Characterizing them requires determining the dynamics of the probability distribution of the particular degree of freedom in which the wave packet is excited. One way to accomplish this goal is to track the wave packet via the time-dependent spectrum of spontaneous emission.7 We have performed such measurements for a wave packet in the nuclear degree of freedom of a sodium dimer.7 This technique permits tracking of the nuclear wave packet in a single excited electronic state over a substantial fraction of its periodic trajectory. It has important applications for studying wave packet excitations in quantum confined electronic microstructures and larger molecules. For example, the quantum control of molecular dynamics is an important goal in chemistry, and the ability to track a wave packet accurately is a significant step toward that objective.

To understand why optics plays such a significant role in this endeavor, consider a wave packet generated by a coherent superposition of vibrational levels in a diatomic molecule. A short optical pulse, resonant with the lowest
Electron-Photon Manipulation
with Sub-Poissonian Semiconductor Junction Arrays

P.J. Edwards

Quantum Electronics Group, Faculty of Information Sciences and Engineering,
University of Canberra, P. O. Box 1, Belconnen, A C T, 2616, Australia.

Abstract. Laboratory investigations have demonstrated the relative ease with
which sub-Poissonian light may be generated by semiconductor junction laser and
light-emitting diodes. A wide variety of electron - photon manipulations is possible
with electrically and optically coupled arrays of sub-Poissonian semiconductor light
emitters and receivers. We discuss systems capable of performing, inter alia,
multiple sub-Poissonian noise-correlated beam generation; low noise photon
number amplification, beam splitting and tapping; "QND-like" photon number
measurement, low noise optoelectronic amplification; and sub-shot noise
interferometry.

1. Introduction

The generation of sub-Poissonian light (SPL), that is, light having photon counting
noise below the normal shot noise limit (SNL) with laser and light emitting diodes
using the high impedance noise suppression method described by Yamamoto et alia
[1], is now well established [2,3,4,5,6,7]. This technique offers a technically
simple way of reducing quantum noise below the SNL in closely coupled
semiconductor junction optoelectronic systems. The amount of noise suppression is
limited mainly by the magnitude of the quantum transfer efficiency which can be
realised between light emitter and detector [8]. Such a reduction in quantum noise
is of practical interest. It has applications in shot noise - limited metrology,
including gravitational wave detection, and in shot noise - limited information
transfer, including optical computing and lightwave communications. Realisation
of SPL based systems in the form of optoelectronic integrated circuits employing
optically interconnected microcavity semiconductor laser and diode detector arrays
appears likely to provide a range of novel electron - photon manipulations and
applications.

2. Sub-Poissonian Light from Semiconductor Junctions

Sub-Poissonian light generation from open loop semiconductor light emitters was
first demonstrated in 1987 by Tapster et alia [2] using infrared emitting diodes and
by Machida et alia [3] using a laser diode. Infrared emitting diodes, cooled to
increase their quantum transfer efficiency, $\eta$, to approximately 11%, exhibited small, up to 4%, noise reduction below the SNL. Machida et alia [3] measured a 7.3% noise reduction using a laser diode with $\eta = 0.3$ at room temperature. This was followed by measurement of a 32% noise reduction for a laser diode cooled to reduce the threshold current and so permit strong pumping to reduce amplified spontaneous noise [4]. Because of incomplete electronic source noise suppression, none of these early measurements gave results in close accordance with the theoretical expectations (Equation (1)) of a fractional noise reduction below the SNL, numerically equal to the quantum transfer efficiency, $\eta$. However, Edwards et alia [5, 6, 7] measured intensity noise reductions of up to 30% with cooled infrared emitting diodes, which were in good agreement with the expression for the Fano factor [8],

$$F_o = \eta \cdot F_i + (1 - \eta) \leq 1 - \eta.$$  

(1)

for negligible source noise as specified by the Fano factor, $F_i$.

The applicability of the above equation has been subsequently confirmed by other workers [9, 10]. It has a simple semiclassical interpretation in terms of the random deletion of photons giving rise to the binomial number variance represented by the second term.

All the above demonstrations utilised the high impedance method of noise suppression [1].

2.1 High Impedance Current Noise Suppression

The photon number (lightwave intensity) fluctuations generated by a semiconductor source of unit quantum efficiency are determined solely by electron hole recombination current fluctuations in the source. Consequently, by raising the impedance of the photon generating current loop, low frequency electronic and photonic current fluctuations can be reduced to generate a near number state with $F_i \ll 1$. Referring to the conventional noise equivalent circuit [11] for a junction diode in Figure (1) we identify the noise current in the branch containing the differential resistance of the diode as the recombination current fluctuation, driven by the voltage generator, $v_n$, representing the internal stochastic processes (spontaneous emission, absorption, stimulated emission, and diffusion) and by Nyquist voltage generator, $v_s$, representing the external thermal pump noise.

This figure serves to illustrate the high impedance noise suppression principle for semiconductor junction light emitters generally, whether or not these exhibit population inversion and stimulated emission. At low frequencies, $\omega \ll 1/\tau$, the current through the "diffusion capacitance", $C$, may be neglected and the junction current and recombination current are consequently identical and easily shown to be negligible compared with the shot noise level providing
We and others [6, 9, 10] have verified electronic noise levels of greater than 15 dB below the shot noise level with infrared emitting diodes driven by high impedance current sources. We have also carried out computer simulations of the systems which closely match the measurements and provide confidence in this simple model [12].

Although the capacitive current may be neglected at low frequencies, the capacitive element provides an important conceptual link with the recombination process. Thus for a light-emitting semiconductor junction, the carrier lifetime is easily shown to be the product of the differential resistance and capacitance. At high frequencies, or with a low impedance pump, the capacitor voltage, and therefore the charge carrier number in the active region of the device, the upper level population, is held constant. Consequently, we expect, and observe "classical" Poissonian photon emission. If, on the other hand, the junction voltage is allowed to fluctuate, which, by inspection of Figure (1) it will do at low frequencies provided $|Z| > r$, the population number will also fluctuate. It will do so in a way which reflects the depletion of the population number by the stochastic emission processes. This can be represented as a classical negative feedback process and can be readily computer modelled [12].

2.3 Semi-Classical Model of SPL Generation

Referring to the light emitting diode equivalent circuit of Figure (1), the state equation for the junction voltage fluctuation, $v$, immediately yields the rate equation for the charge carrier population number fluctuation, $n(t)$ in the form,

$$\frac{dn}{dt} = -n(1/\tau + 1/RC) + f_n(t)$$

$$= [i_b(t)/e - s(t)]$$

where the Langevin noise term, $f_n$, has covariance,

$$<f_n(t) \cdot f_n(t + \tau')> = (2<N>/\tau + 4kT/e^2R) \cdot \delta(t')$$

At low frequencies, $\omega < 1/\tau$, $dn/dt = 0$, and the photon current fluctuation, $a(t)$, and electron current fluctuation, $i_b(t)/e$, are therefore equal. If, in addition, the condition $RC > > \tau$, is satisfied, then $n(t) = f_n(t) \cdot \tau$, and the spectral density of the charge number fluctuation, $S_n(\omega) = 2<N>/\tau$, is as expected for a Poisson variable of mean, $<N>$. The spectral densities of the electron and photon noise currents then both vanish in the limit, $x = r/R = 0$. Consideration of the rate equation shows that, as the pump impedance is raised, the thermal noise in the current pump
eventually determines the low frequency quantum noise level. For this diode model [11], which differs slightly from that discussed by Yamamoto and Machida [1], the photon Fano factor for unit quantum efficiency is equal to $x/(1+x)$.

An equivalent view, following Teich and Saleh, is to treat the electron recombination as a self excited point process exhibiting rate compensation by excitation feedback [8]. Low frequency stochastic fluctuations in the recombination rate are permitted to deplete and enhance the charge carrier population. It is this fluctuating population number (or alternatively, the fluctuating junction voltage) which mediates the negative feedback mechanism responsible for the quantum noise reduction on time scales long compared with the charge carrier lifetime [11].

The final quantum noise level at the detector is then calculable using equation (1). This equation lumps all photon number losses together in the form of a single quantum transfer efficiency, $\eta \leq 1$. This is the product of all the (statistically independent) probabilities which characterise the photon deletion processes in the chain from initial electron injection to final photoelectron excitation. Figure 3(a) shows a simplified representation of a sub-Poissonian light emitter.

3. Arrays of Correlated SPL Emitters

3.1 Generation of Multiple SPL Beams and Photon Number Amplification

It is apparent from the discussion above, that a series connected pair of semiconductor light emitters should exhibit positively correlated, sub-Poissonian photon number fluxes on long time scales. Consideration of classical electrical partition noise at a current branch indicates that shunt connected light emitters should provide negatively correlated sub SNL photocurrents. That such is the case has now been well verified [6, 10] and used to validate a partition model for the largest (8.3 dB) quantum noise reduction observed to date with semiconductor light emitters [13].

The low frequency noise equivalent circuit for a tandem array of diode emitters is shown in Figure (2). It is evident that a different noise reduction mechanism [14] is responsible for the SPL operation of such an array. This new mechanism is analogous to the dynamic pump noise suppression technique proposed for multilevel lasers [15]. It is well known in exponential queuing theory and leads to an N-Erlang recombination event waiting time distribution for any element in an array of N diode emitters. Fano factors and covariances are readily calculated [6]. Again, the physical mechanism responsible for SPL generation is evident in the occurrence of (correlated) low frequency fluctuations in population numbers, as for the high impedance noise suppression method discussed in the previous section.

The light from these individual beams may be separately detected or combined into a single composite beam and then detected. These configurations form the basis for low noise quantum taps (Fig. 3(b)), optoelectronic amplification, QND-like measurement and photon number amplification (Fig. 3(c)) [14, 16].
It is evident from the preceding discussion, that a tandem array of SPL emitters provides the basis for photon number amplification. The tandem diode array consists of an array of \( N \) identical light emitters, each generating a sub-Poissonian beam with Fano factor, \( F_i \). Both the noise and signal modulating currents are evidently the same in each element of the series connected array. Consequently the signal to noise ratio in the incoherently combined beams is the same as that in each of the individual beams while the optical power and the Fano factor are both increased \( N \)-fold. More generally, after attenuation and/or non-ideal detection, the Fano factor becomes

\[
F_o = \eta \cdot N \cdot F_i + (1 - \eta) \tag{6}
\]

In the ideal case, \( F_i = 0 \), the combined beam remains in a number state. This arrangement therefore constitutes a low noise, high power modulated SPL generator. Efficiently coupled to a light detecting array, it constitutes a low noise optoelectronic amplifier. If driven by an ideal detector (Figure 3c) the array functions as a low noise photon number amplifier with a limiting noise figure of unity even for a number state input [14].

We have recently demonstrated the SPL operation of such an array in each of the above modalities and verified the utility of equation (6) for an SPL emitting array of eight infrared emitting diodes with a current gain of 2 and an optical power gain exceeding unity [17].

3.2 Low Noise Photon Number Replication

Yuen [21] has suggested the use of photon number preamplification of sub-Poissonian states as a means of partially compensating for their extreme fragility in a lossy environment. It is well known that an initially noiseless photon number state, with Fano factor, \( F_i = 0 \), rapidly degrades towards the Poissonian level with \( F_i = 1 \), through a succession of intermediate binomial states with binomial number variance equal to \( \eta \cdot (1 - \eta) \cdot < n_i > \) as the overall system quantum efficiency, \( \eta \), decreases towards zero. Following preamplification, the direct detection signal to noise ratio is then, in terms of the input photon number, \( n_i \), or detector photocurrent, \( I \), measured over time interval, \( \Delta t \),

\[
\text{SNR} = \frac{< I >^2}{< i^2 >} = N \cdot < n_i > \cdot \eta \cdot (1 - \eta) \tag{7}
\]

for mean photocurrent, \( < I > \), and mean square photocurrent noise, \( < i^2 > \).

This is a factor of \( N / (1 - \eta) \) higher than for a shot noise limited (Poissonian) signal with the same initial mean photon number, \( < n_i > \) However the fragility remains in the form of the \( \eta / (1 - \eta) \) dependence of the signal to noise ratio.
The tandem diode array can be used to generate a robust form of light signal in which the signal to noise ratio remains enhanced by the factor \( \frac{1}{F_i} \) above that for coherent light with the same initial power as that of a single element, independently of attenuation and overall quantum transfer efficiency, \( \eta \).

In order to achieve this, each element of the array is operated with Fano factor, \( F_i < 1 \), and the composite beam is arranged to be super-Poissonian with Fano factor, \( N \cdot F_i \gg 1 \).

After attenuation (and/or non-ideal detection) with quantum transfer efficiency, \( \eta \) the signal to noise ratio becomes,

\[
SNR = \frac{\langle n_0 \rangle}{\langle F_0 \rangle} = \frac{N \cdot \eta \cdot \langle n_i \rangle}{[1 + \eta \cdot (N \cdot F_i - 1)]} = \frac{\langle n_i \rangle}{\langle F_i \rangle}.
\]

Thus, the signal to noise ratio is independent of \( \eta \), and equal to that at one element of the transmitter, providing the condition \( N \cdot F_i \cdot \eta \gg 1 \) is satisfied.

3.3 Low Noise Optoelectronic Amplification

Consider a tandem array of \( N \) high quantum efficiency light emitters, connected to a modulating signal source, each coupled with overall quantum efficiency, \( \eta \), to a lightwave detector array. Referring to the noise equivalent circuit of Figure (2), it may be shown [14] that the excess system noise temperature,

\[
T_s = \langle I \rangle \cdot e \cdot R \cdot \frac{(F_i + [1 - \eta]/N \cdot \eta)}{2k}.
\]

Thus, for example, looking ahead to low threshold microlaser/LED array structures operating at room temperature with a quantum transfer efficiency of 90\%, and a pump current of 1 mA, a power gain of 20 dB can be achieved in a 50 \( \Omega \) system with an excess temperature of no more than 3 K.

3.4 Noiseless Photon Number Tapping

Figures (3b) and (4) shows how semiconductor junction detector and emitter arrays may be configured to effect a non destructive measurement of photon number. Evidently, non destructive measurement of the intensity of the incident beam can be achieved and the system can be used as a low noise quantum tap [22]. The necessary photon number to electron number conversion gain may also be achieved with a low noise electronic amplifier as described by Roch et alia [23] who show that necessary (although not sufficient) QND criteria are formally satisfied.
The photocurrent from a junction diode detector may also be used to pump a tandem light emitting diode array as in Figure (4). If each emitter is coupled with unit efficiency to a separate detector or information channel then noiseless N-fold replication of the detected signal is achieved. If the input detector is also of unit efficiency, then noiseless splitting of the photonic signal will have been effected. If the emitting array is optically coupled to a single detector, array of detectors or single information channel as in Figure (3c), the configuration constitutes a photon number amplifier with a limiting noise figure of unity [11]. The summed detector array currents, may in turn be used to drive a single light emitter for onward transmission of the regenerated optical signal as shown in Figure (4). Goobar et alia [16], following Yuen [24], describe the possible use of these configurations in transparent optical communication networks.

3.6 Sub-Poissonian Lightwave Interferometry

Holland and Burnett [18] have shown that the minimum level of phase noise permitted by the Heisenberg relation, \( \frac{1}{n} \) radians, is reached in an optical interferometer in which both input ports are driven by number states containing equal photon numbers, \( n \). The intensity difference noise and phase difference noise between the fields at the output ports of the first beam splitter in the interferometer are linked by the Heisenberg relation. Photon interference results in a number variance in the intensity difference (for large \( n \)) of \( n^2/2 \) [19] and hence a variance in the phase noise of \( \frac{1}{n^2} \). The dual Fock state, \( |nn> \), appears to provide advantages similar to those of the two mode squeezed state suggested by Caves [20], and offers potentially very great improvement in sensitivity over a shot noise limited system with coherent input to a single port. It is therefore appropriate to ask whether the dual number state, or some approximation to it, can be realised with "wired optics".

Strongly correlated, sub-Poissonian binomial states can readily be generated with series connected "tandem" arrays of light emitting diodes [6]. Although not yet demonstrated, one may expect this to apply also to arrays of laser diodes. The necessary requirement of first order coherence could be satisfied by a pair or an array of laser diodes enclosed within a common cavity or by a coupled cavity laser diode system.
7. Conclusions

The use of macroscopic electron currents to control, generate and manipulate sub-Poissonian photonic currents, foreshadows a variety of low noise applications.

The systems which have been demonstrated to date have utilised the relatively high electron to photon and photon to electron quantum conversion efficiencies available with semiconductor detectors, and to a lesser degree, with laser diodes and light emitting diodes at low temperatures.

The practicality of these wired optic systems will depend on the ability to achieve high quantum efficiencies at room temperature in arrays of low threshold, high spontaneous emission coefficient microcavity lasers fabricated in optoelectronic integrated circuit form.

References

17. P. Lynam and P. J. Edwards, presented at Sixth International Symposium on Quantum Optics, (University of Auckland, Rotorua, 1994).
Figure 1

Ideal light-emitting semiconductor junction diode noise equivalent circuit. The overall quantum efficiency, $\eta$, is defined as the ratio of detector current $i_d(t)$ to junction current, $i(t)$.
Figure 2

Low frequency noise equivalent circuit for a tandem array of $N$ junction diode light-emitters connected to voltage source, $v_s$. 

\[ v_s \quad R_s \quad i(t) \quad v_n(1) \quad v_n(2) \quad v_n(N) \]
Figure 3(a)

Idealised representation of sub-Poissonian light-emitting semiconductor junction with photon number Fano factor, $F$.

Figure 3(b)

Idealised representation of low noise photon number tap with input photon number Fano factor, $F_i$.

Figure 3(c)

Idealised representation of a photon number amplifier employing an array of sub-Poissonian light-emitting junctions.
Figure 4

Idealised representation of a photon number amplifier/tap employing arrays of SPL emitters and detectors.
Quantum-correlated light from transverse junction stripe laser diodes

YONG-QING Li, PAUL J. EDWARDS, PETER LYNAM and WOOD N. CHEUNG

We report on the generation of quantum-correlated light beams emitted from a pair of electrically coupled transverse junction stripe (TJS) laser diodes. We also report the measurement of sub-shot noise ('quiet') light emitted by each of these laser diodes when driven from a high impedance low noise current source. The measured cross-correlation coefficients between the intensity fluctuations in the two beams and the measured noise reduction of 3 dB below the normal shot noise level are both consistent with the measured quantum transfer efficiencies of approximately 50%. These measurements show that TJS laser diodes provide another method (in addition to the use of optical parametric oscillators and electrically coupled light emitting diodes) for generating twin light beams with intensity difference fluctuations below the normal shot noise level. Potential applications include low noise spectroscopy and interferometry.

1. Introduction

The possibility of suppressing the quantum shot noise ('intensity squeezing') in laser light was first suggested in 1984 [1] and later demonstrated by Machida et al. in 1987 [2] with liquid nitrogen and liquid helium cooled laser diodes using what has now become known as the high impedance pump technique. Pump noise suppression allows stochastic fluctuations in the rate of radiative recombination within the laser cavity to be suppressed by an internal feedback mechanism. Laser intensity noise levels as low as 8·3 dB below the normal shot noise limit have been reported [3] using cooled, free-running TJS laser diodes. More recently, with the use of injection locking and external cavity line-width narrowing techniques to reduce mode-partition noise, intensity noise levels of several dB below the shot noise level have also been reported with QW laser diodes at room temperature [4].

Reduction of the photonic noise in the incoherent light from a light-emitting diode was first demonstrated in 1987 [5]. The high impedance noise suppression mechanism responsible has been shown to be similar to that for laser diodes [6]. Noise reductions of up to 3 dB have been obtained at liquid nitrogen temperature, limited only by the quantum transfer efficiency (current transfer ratio) between emitter and detector [7-9].

The fundamental limit to the degree of quantum noise suppression achievable in practical devices is set by transmission losses. Earlier expectations of communication applications for 'quiet' light have been defeated by its fragility to photon losses. It seems likely therefore that quantum noise suppression techniques will be restricted to specialized systems such as ultra low noise gravity wave interferometers [10] and to optoelectronic logic and linear integrated circuitry in which photon losses can be kept small. Several such quiet light devices based on noise suppressed photon transport between semiconductor junctions have recently been modelled and demonstrated [11, 12].

This paper reports for the first time the generation of 'quantum-correlated' twin beams by laser diodes, a new technique relevant to both these potential applications of quiet light.

2. Photon deletion noise and quantum correlation

2.1. Photon deletion noise

This noise, sometimes referred to as photon 'partition' noise, is a consequence of the lack of a one-to-one correspondence between the individual electron–hole
recombination events in the light emitter and the subsequent photoelectron–hole pair production events at the detector. Any inefficiencies will 'partition' events away from the detection domain. Statistical loss of individual photons from the light beam linking light-emitting and detecting junctions is one source of this partition (deletion) noise. Other sources include photon scattering and non-radiative recombination in the junction and corresponding inefficiencies in the detector. Treating these photon 'deletions' as independent events leads to exact agreement with the predictions of quantum theory [13]. It may then be shown for example that if the detector current $i_d(t)$ is a fraction, $\eta$, of the recombination current $i_r(t)$, this approach yields the following transfer equation [8] relating the corresponding mean square currents:

$$<i_d^2(t) > = \eta^2 <i_r^2 > + (1 - \eta) <i_{r\text{noo}}^2 >$$  \hspace{1cm} (1)

where the first term is the mean square signal (and noise) current transferred from the laser diode circuit to the detector while the second term is the mean square partition noise current. The current fluctuation contributing to this second term is statistically independent of the current flowing in the laser circuit. It vanishes for an ideal system in which the quantum transfer efficiency, $\eta$, is unity. This equation may be divided by the shot noise term $<i_{r\text{noo}}^2 > = 2i_\text{r}e\Delta f$ and then rewritten in the form:

$$F_d = <i_d^2 >/<i_{r\text{noo}}^2 > = \eta F_i + (1 - \eta)$$  \hspace{1cm} (2)

or

$$F_d - 1 = \eta(F_i - 1).$$  \hspace{1cm} (3)

The 'Fano factors' [8] $F_d$ and $F_i$ are the detector and laser mean square current noise levels relative to their respective shot noise levels as defined by the respective DC currents ($i_d = \eta i_r$). From (3), when the laser current noise is completely suppressed ($F_i = 0$) the detector noise, relative to the expected shot noise level is evidently

$$F_d = (1 - \eta).$$  \hspace{1cm} (4)

2.2. Quantum correlation

Correlated intensity fluctuations between two light beams, leading to sub-shot noise fluctuations in their intensity difference ('quantum correlation') are of interest in metrology, offering the prospect of enhanced precision in spectroscopy and interferometry. Quantum correlations have been demonstrated between twin beams generated by optical parametric oscillators [19] and electrically coupled light emitting diodes [8, 14]. Twin beams generated by LEDs are incoherent and therefore cannot be used for narrow band interferometry and spectroscopy. It has been suggested [15, 16] that a pair of laser diodes connected in series might serve as practical sources of quantum-correlated coherent light beams for these purposes. We have demonstrated the generation of quantum-correlated twin beams by two electrically coupled laser diodes. These are modulated by a common pump noise current (with Fano factor $F_i$), and the intensity noise of each laser diode is suppressed by the high impedance method referred to above. It may be shown [14] that after photon deletion the beams should remain quantum correlated and that the measured intensity correlation coefficient between symmetric beams should become:

$$r_{12} = \eta F_i/F_d.$$  \hspace{1cm} (5)

From this equation, a correlation coefficient of 0.5 is expected for $\eta = 0.5$ and $F_i = 1$.

3. Measurements using TJS laser diodes

3.1. Experimental set-up

The experimental arrangement used to measure quantum noise suppression and quantum correlation is shown in figure 1. Two laser diodes (LD$_1$ and LD$_2$) were separately biased by mains powered, heavily filtered constant current sources (DC bias 1 and DC bias 2) through resistances of 500 $\Omega$, which is high compared with the differential resistances needed to suppress pump noise. The two 50 $\Omega$ resistances similarly suppress current noise by raising to 1000 the resistance in the loop containing the two laser diodes. In the absence of externally injected noise, this resistance also suppresses any correlation between the two laser diode currents.

[Diagram of experimental set-up is shown here.]

Figure 1. Experimental set-up. Two TJS laser diodes (LD$_1$ and LD$_2$) are biased separately by two constant current sources (DC bias 1 and DC bias 2). They are modulated in common by a wideband noise source (Mod.) of variable amplitude. The emitted photon beams from the laser diodes are collected by two identical large area pin photodetectors and the output photocurrents are subtracted (or added), amplified, and input to a spectrum analyser.
External noise injection is then required in order to demonstrate quantum noise correlation. This is provided as shown in figure 1 (Mod.) by a wideband gaussian noise generator of variable amplitude, which is connected as a current source to provide, through a 500 Ω resistance, a common mode signal to both lasers. All the resistances were operated at room temperature. The amplitude of this wideband noise source was set to zero when measuring the quantum noise suppression in each individual laser diode. Two large area pin detectors PD1 and PD2 (Hamamatsu S3994) were placed in close proximity at a distance $d \sim 10$ mm from the laser diodes and were tilted to avoid specular reflection into the laser diodes. The deleterious effects of optical feedback [2] were found to be minimized by this configuration.

Two light emitting diodes (Hamamatsu L2656) were used to establish the shot noise reference levels at the same DC detector currents. It has been shown [6, 7] that at frequencies above a few hundred kHz these LEDs can be used as accurate shot noise reference sources. This is a consequence of the limited 'squeezing bandwidth' of these devices. The internal regulatory mechanism responsible for shot noise suppression only operates on time scales comparable with and longer than the radiative recombination lifetime. In the present case this leads to negligible noise suppression at frequencies above 1 MHz.

An ever-present problem with measurements of this type is that of detector saturation. This is a function \textit{inter alia} of optical power density and of optical spot size at the surface of the detector. In particular, in the usual shot noise level calibration procedure, the shot noise level is established by floodlighting the entire detector area using a light emitting diode. The light from the laser diode under test will then usually be restricted to a smaller area spot of higher power density. Under these conditions any saturation effects are likely to lead to an underestimate of the laser fluctuation noise and therefore to an overestimate of the amount of noise reduction. In the present arrangement we established that the shot noise reference level increased linearly (within ±2%) with DC detector currents up to 10 mA. In order to establish a similar linearity regime for the lasers, we modulated the laser diodes with a constant amplitude 34 MHz signal at a fixed DC bias current and then varied both the spot size and the power density at the detector using neutral density filters. We found that this procedure established a similar degree of linearity for detector currents less than 10 mA and laser/detector separations greater than or equal to 10 mm.

### 3.2. Measurements of photon shot noise suppression

Two 'typical' [17] TJS laser diodes were examined. Noise power spectrum and measured photonic Fano factor at 5-0 MHz for the first of these (ML 3308/89-10484) with a front facet reflectivity of 10%, a rear facet reflectivity of 90%, a cavity length of 150 μm, and a room temperature wavelength of 801 nm are shown in figures 2(a) and 3(a), respectively. Upon cooling to 80 K the threshold current drops from its room temperature value of 26 mA to 0.8 mA. At this temperature the noise level initially falls as the pump current increases, reaching the shot noise level at a pump level of $R = (I_L/I_{th} - 1) = 3$. At a pump level of $R = 13.5$ ($I_d = 7.0$ mA at $I_L = 11.6$ mA) the noise suppression is 3.1 dB. After correction for detection and collection losses (approximately 0.95 detection efficiency), this corresponds to 3.4 dB squeezing at the front facet of the laser diode. The measured value is less than the maximum expected (4.0 dB) from the measured quantum transfer efficiency of 0.8, assuming no internal current partition. Figures 2(b) and 3(b) show the results obtained with a second TJS laser of the same type (ML 3308/89-10466). In this case 3.0 dB of
noise reduction was found, again somewhat less than would be expected from equation (4) for the measured quantum efficiency of 0.55.

Since in the present experiment the TJS laser diodes are free-running, the fact that the observed noise reduction is less than the theoretical expectation is probably due to additional photonic mode-partition noise [18]. It is expected that this mode-partition noise would be reduced substantially by the use of either injection locking or external grating feedback techniques.

3.3. Quantum correlation measurement

Figure 4 shows the experimental results. Figure 4(a) shows the measured noise spectral densities of the summed and differenced detector currents. As expected, the fluctuations in the difference current are 3 dB below the shot noise level (SNL) after correcting for the amplifier noise level. We also measured the noise spectral densities of the two separate detector currents, $\langle i_d^2 \rangle$ and $\langle i_d^2 \rangle$. Their cross-covariance is given by $\langle i_{d1} i_{d2} \rangle = (\langle i_{d1}^2 \rangle + \langle i_{d2}^2 \rangle - \langle i_{d0}^2 \rangle)/2$, where $i_{d0} = i_{d1} - i_{d2}$ is the difference between the two detector currents, from which the cross-correlation coefficients $r_{12} = (\langle i_{d1} i_{d2} \rangle - (\langle i_{d1}^2 \rangle)^{1/2})/(\langle i_{d1}^2 \rangle)^{1/2}$ can be calculated. Figure 4(b) shows the correlation coefficients, $r_{12}$, both with and without external noise injection. As expected, with no externally injected noise, the high resistance in the laser diode loop isolates the diodes from each other, suppresses any correlation due to internal laser noise generation, and gives a measured correlation coefficient which is not significantly different from zero (curve ii). If the two laser diodes had been directly connected in parallel a negative correlation [8] would have been expected. The correlation coefficients shown in curve (i) are for the case of external noise injection. In this case, the Fano factor for the injected...
noise varies from $F_i = 2.5$ at frequency 1-5 MHz to 0-85 at 6-5 MHz, leading to a frequency dependent correlation coefficient varying between 0-78 and 0-35. The measured noise varies from $F_i = 2.5$ at frequency 1.5 MHz to 0.85, which is close to but lower than the expected value ($r_{12} = 0.55$ for $\eta \sim 0.55$) calculated from (5). The correlations at other frequencies are also in reasonable agreement with, although lower than, the values expected from (5), especially in the sub-shot noise regime above 4-5 MHz. This decorrelation is due to the presence of modal-partition noise, independently generated in each of the two TJS laser diodes.

### 4. Conclusions

We have demonstrated that TJS laser diodes may be used to generate quantum-correlated light beams with a sub-shot noise intensity difference. The degree of correlation and the amount of shot noise suppression were in good agreement with expectations, being limited by photon deletion noise and modal partition noise. Our measurements, together with others showing significantly lower modal partition noise levels [3] in TJS lasers, suggest their potential for application to low noise interferometry and spectroscopy.

### Acknowledgments

The authors gratefully acknowledge the generosity of Professor Y. Yamamoto (NTT/Stanford University) in providing the laser diodes used in this investigation. We thank Mr. M. Harold (ADFA) for technical help. We also acknowledge financial support from the Australian Research Council and the Department of Education, Employment and Training.

### References


MACROSCOPIC VIOLATION OF THREE CAUCHY-SCHWARZ INEQUALITIES USING CORRELATED LIGHT BEAMS FROM AN INFRA-RED EMITTING SEMICONDUCTOR DIODE ARRAY

P. J. Edwards and X. Huang
Advanced Telecommunications Engineering Centre, Faculty of Information Sciences and Engineering, University of Canberra, PO Box 1, Belconnen, ACT 2616, Australia

Y. Q. Li
Department of Physics, University of Arkansas, Fayetteville, AR 72701, USA

Y.Z. Wang
Joint Laboratory for Quantum Optics, Shanghai Institute of Optics and Fine Mechanics, Academia Sinica, PO Box 800-211, Shanghai 201800, PRC

Abstract

We briefly review quantum mechanical and semi-classical descriptions of experiments which demonstrate the macroscopic violation of the three Cauchy-Schwarz inequalities:

\[ g_{11}^{(2)}(0) \geq 1; g_{11}^{(2)}(0) \geq g_{11}^{(2)}(t), (t \to \infty); |g_{12}^{(2)}(0)|^2 \leq g_{11}^{(2)}(0) g_{22}^{(2)}(0). \]

Our measurements demonstrate the violation, at macroscopic intensities, of each of these inequalities. We show that their violation, although weak, can be demonstrated through photodetector current covariance measurements on correlated sub-Poissonian Poissonian, and super Poissonian light beams. Such beams are readily generated by a tandem array of infrared-emitting semiconductor junction diodes. Our measurements utilise an electrically coupled array of one or more infrared-emitting diodes, optically coupled to a detector array. The emitting array is operated in such a way as to generate highly correlated beams of variable photon Fano Factor. Because the measurements are made on time scales long compared with the first order coherence time and with detector areas large compared with the corresponding coherence areas, first order interference effects are negligible.

The first and second inequalities are violated, as expected, when a sub-Poissonian light beam is split and the intensity fluctuations of the two split beams are measured by two photodetectors and subsequently cross-correlated.

The third inequality is violated by bunched (as well as antibunched) beams of equal intensity provided the measured cross correlation coefficient exceeds \((F - 1)/F\), where \(F\) is the measured Fano Factor of each beam. We also investigate the violation for the case of unequal beams.

1 Theory of The Macroscopic Violation

The first inequality addresses the correlation between the intensity fluctuations in the two beams emerging from a 50/50 optical beam splitter. Loudon\(^1\) gives the standard quantum result for a
single mode beam with mean photon number, \(< n >\):

\[
g_{11}^{(2)}(0) = \frac{< n(n - 1) >}{< n >^2}
\]  

(1)

and Paul\(^{2}\) obtains the same result by treating the photon beam as a beam of classical distinguishable particles, subject to Bernoulli partition. In the absence of interference noise, that is for a broadband, multimode, incoherent source on time scales long compared with the coherence time and with detector areas large compared with the coherence area (Teich\(^{3}\)), this treatment is justified (as it is also for a single mode situation). It is evident from expansion of equation (1) that \(g_{11}\) may be written:

\[
g_{11}^{(2)}(0) = 1 + \frac{(F - 1)}{< n >}
\]  

(2)

This form shows that any violation, (a value less than unity) requires sub-Poissonian variance \((F < 1)\) and must be weak in the macroscopic limit \((n > > 1)\). Nevertheless macroscopic violation can be readily demonstrated in a Hanbury Brown type experiment using a single light emitting diode driven from a high impedance source (Edwards\(^{4}\)). The same configuration serves to show violation of the second inequality. Both these violations can be deduced from the measured covariance between the macroscopic photocurrents \(i_1, i_2\) for the split beams.

As pointed out by Loudon\(^{1}\), violation of this inequality is a fundamental quantum result resulting from the photoelectric detection of either the transmitted or the reflected photon. As we shall see however, Loudon’s assertion that “non-classical effects tend to be most marked for beams with small well defined numbers of photons” is (rather surprisingly) apparently not true for the third inequality:

\[
|g_{12}^{(2)}(0)|^2 \leq g_{11}^{(2)}(0)g_{22}^{(2)}(0)
\]  

(3)

which is violated at macroscopic intensities for bunched and unbunched beams as well as for the antibunched beams for which the first inequality is weakly violated at macroscopic intensities.

It is well known that the Cauchy-Schwarz inequality with time delay \(t\) between the two beams is \(^{3,5,6}\)

\[
|g_{ij}^{(2)}(t)|^2 \leq g_{11}^{(2)}(0)g_{22}^{(2)}(0),
\]  

(4)

where \(g_{ij}^{(2)}\) is the second-order coherence function. For \(t = 0\), we have the following inequality if we use a “classical particle” description\(^{2}\):

\[
\frac{< n_1n_2 >^2}{(n_1 < n_2 >)^2} \leq \frac{< n_1(n_1 - 1) >}{< n_1 >^2} \frac{< n_2(n_2 - 1) >}{< n_2 >^2}.
\]  

(5)

That is

\[
< n_1n_2 >^2 \leq (< n_1^2 > - < n_1 >)(< n_2^2 > - < n_2 >).
\]  

(6)

In order to violate the Cauchy-Schwarz inequality, we should have

\[
< n_1n_2 >^2 > (< n_1^2 > - < n_1 >)(< n_2^2 > - < n_2 >).
\]  

(7)

So that we have

\[
F_1F_2(r^2 - 1) + < n > [2r\sqrt{F_1F_2} + 2 - (F_1 + F_2)] + (F_1 + F_2) - 1 > 0
\]  

(8)
where $r$ is the cross correlation coefficient and $F_1$ and $F_2$ are the Fano Factors for each beam. For macroscopic violation (large $<n>$):

$$[2r\sqrt{F_1F_2} + 2 - (F_1 + F_2)] > 0$$  \hspace{1cm} (9)

If $F_1 = F_2 = F_0$, the measured Fano Factor, we may obtain the following result from inequality (8):

$$(rF_0 - F_0 + 1) > 0,$$  \hspace{1cm} (10)

hence

$$r > \frac{F_0 - 1}{F_0}.$$  \hspace{1cm} (11)

We therefore have the following violation conditions:

1. For $F_0 = 1$ (Poisson), all positive correlations, i.e. $r > 0$;
2. For $F_0 < 1$ (Anti-bunched), all positive correlations, i.e. $r > 0$;
3. For $F_0 > 1$ (Bunched), $r > |(F_0 - 1)/F_0|$.

This case is shown in Figure 1.

![Figure 1: Two different violation regions, I ($F_0 < 1$) and II ($F_0 > 1$).](image-url)
2 Violation For Twin Beams Generated by Coupled LED's

Figure 2 shows the arrangement adapted by Edwards to generate quantum-correlated twin beams.

Figure 2: Series-connected infrared emitting diodes (L2656) configured to generate positively correlated intensity fluctuations.

From Figure 2, we have $i_1 = i_2$ and $\eta_1 = \eta_2$, so $i_{1d} = i_{2d}$ and $<i_{1d}^2> = <i_{2d}^2>$. The correlation coefficient is given by

$$r_{12} = \frac{<i_{1d}i_{2d}>}{\sqrt{<i_{1d}^2><i_{2d}^2>}}$$

This can be easily shown to be

$$r_{12} = \frac{\eta F_i}{F_o}$$

Here $F_i$ is the Fano Factor at the source and

$$F_o = 1 + \eta(F_i - 1)$$

is the Fano Factor measured at the detectors with quantum efficiency, $\eta$. Recall that the macroscopic violation for $F_1 = F_2 = F_o$ was given by

$$r_{12} > \frac{F_o - 1}{F_o}$$

that is

$$\eta \frac{F_i}{F_o} > \frac{F_o - 1}{F_o} = \eta(F_i - 1)$$

$$\frac{F_i}{F_o} > \frac{F_o - 1}{F_o}$$
A violation parameter, $\Delta$ can therefore be written as

$$\Delta = r_{12} - \frac{F_0 - 1}{F_0}$$

$$= \frac{\eta}{F_0} \quad (F_0 > \frac{1}{2})$$

$$= 1 + r_{12} \quad (0 < F_0 < \frac{1}{2})$$

(17)

3 Experimental Results

Referring to Figure 3, these measurements were performed at room temperature using two series connected Hamamatsu type L2656 infrared emitting diodes. The Fano factors were measured as shown with a swept frequency spectrum analyser. Correlations were measured digitally.

![Diagram](image)

Figure 3: The correlated twin beams are generated by light emitting diodes, D1,D2. Tungsten lamps, L1, L2, provide shot noise reference currents in the pin diode detectors P1,P2, and light emitting diodes, respectively. Switch, S provides unbunched (UBN), bunched (BN) and anti-bunched (ABN) twin beams with detected Fano factors measured by the spectrum analyser. The quantum efficiencies are determined directly from the measured DC currents.
Typical results are shown in Figure 4, together with a curve showing the expected values of the violation parameter for a quantum efficiency of 10%, as employed for all measurements with the exception of the left hand point (12%). These results are in good agreement with the theory.

Figure 4: Experimental violation of the Cauchy-Schwarz Inequalities.

4 Conclusions

We have extended theories concerning violation of the Cauchy-Schwarz inequalities (CSI). We have derived a simple condition for the macroscopic violation of a CSI for twin incoherent light beams using a "classical particle" model. We have shown that this CSI is violated for positively correlated, bunched, incoherent twin beams.

References

Violation of a classical Cauchy–Schwarz inequality in photon noise spectra

Yong-qing Li†, Paul J Edwards‡, Xu Huang§ and Yuzhu Wang‖

† Department of Physics, East Carolina University, Greenville, NC 27858-4353, USA
‡ Advanced Telecommunications and Quantum Electronics Research Centre, University of Canberra, PO Box 1, Belconnen, ACT 2616, Australia
§ Physics and Electronics Engineering, University of New England, NSW 2351, Australia
‖ Joint Laboratory for Quantum Optics, Shanghai Institute of Optics and Fine Mechanics, Academia Sinica, PO Box 800-211, Shanghai 201800, People’s Republic of China

Received 10 October 1999, in final form 6 April 2000

Abstract. A Cauchy—Schwarz inequality in photon noise spectra is derived and discussed for macroscopic fields of light. Violation of this Cauchy—Schwarz inequality is demonstrated experimentally in quantum noise-correlated light beams radiated from a pair of electrically coupled light-emitting diodes. Our measurements directly demonstrate that the violation of this Cauchy—Schwarz inequality can be observed in the absence of sub-shot noise fluctuations in either the single-beam intensity or the dual-beam intensity difference. Therefore, the dual-beam quantum correlation condition required for the violation of this Cauchy–Schwarz inequality is not associated with sub-shot noise fluctuations.

Keywords: Quantum correlation, Cauchy—Schwartz inequality, nonclassical light

(Some figures in this article are in colour only in the electronic version; see www.iop.org)

1. Introduction

Cauchy–Schwarz (CS) inequalities are mathematical relations between the cross-correlation function and auto-correlation functions of two intensity-correlated light beams [1]. They are the quantum mechanical equivalents of the mathematical Schwarz relation \(|(i_i(t) i_2(t'))| \leq (i_1^2(t)) (i_2^2(t'))\) between the second moments of two stochastic functions of \(i_1(t)\) and \(i_2(t)\) [2]. Based on the measurement of the second-order coherence functions, a standard CS inequality can be expressed as

\[ |g_{12}^{(2)}(\tau)| \leq g_{11}^{(2)}(0) g_{22}^{(2)}(0). \tag{1} \]

The violation of the CS inequalities is a nonclassical characteristic of light fields. It is associated with the nonclassical correlation phenomenon between the two beams and is associated with the nonexistence of a positive Glauber–Sudarshan P-representation [3].

Since the mathematical Schwarz relation is applicable to any two stochastic functions, the CS inequality of equation (1) can be generalized to any two observable random functions of the light fields, depending on the measurement technique. For weak light fields, photon numbers counted within a time interval (proportional to the intensities of the beams) are convenient quantities measured by photon counting and photon correlation techniques. Experimental violation of CS inequality (1) has been demonstrated in atomic two-photon cascade radiation by Clauser [4] and in parametric fluorescence by Burnham and Weinberg [5].

For intense light fields, convenient measurement quantities are the intensity fluctuations \(i_1(t) = (i_1(t) - \langle i_1 \rangle)\) and \(i_2(t) = (i_2(t) - \langle i_2 \rangle)\) of the light beams. Measurements of the dual-beam photocurrent noise correlation in twin beams emitted from a pair of series-connected light-emitting diodes (LEDs) have revealed quantum correlations [6] and also violation of the CS inequality of equation (1) [7]. Violation of a different form of the CS inequality that is directly related to the intensity fluctuations \(|(i_1(t) i_2(t'))| \leq (i_1^2(t)) (i_2^2(t'))\) has been investigated experimentally [8].

In this paper, we shall derive the CS inequality expressed in terms of the photon noise spectra for macroscopic fields of light

\[ |S_{12}(\Omega)| \leq (S_{11}(\Omega) - \langle i_1 \rangle)(S_{22}(\Omega) - \langle i_2 \rangle), \tag{2} \]

and then demonstrate its experimental violation in two classically fluctuating light beams generated from electrically coupled LEDs. The measurement quantities involved in this CS inequality are the photocurrent noise power spectral densities of the two beams. Just as in the case of the frequency-dependent squeezing spectra of squeezed light, quantum correlation between two real radiation sources is limited by the response time of the system. This means that the violation of the CS inequality may be observed in some frequency ranges but not in others. Therefore, a CS
inexplicable based on the measurement of photon noise spectra is needed in order to study quantitatively the nonclassical correlation phenomenon associated with the violation of the CS inequality.

We shall demonstrate that the violation of CS inequality can be observed in the absence of sub-shot noise fluctuations in either the single-beam intensity or the two-beam intensity difference or sum. This is different from other nonclassical effects [9] observed in macroscopic fields, such as sub-Poissonian photon statistics, amplitude squeezing [10, 11], quadrature squeezing [9] and intense twin-beam correlation [12]. These are characterized by sub-shot noise fluctuations either in the single-beam intensity or in the two-beam intensity difference or sum. This paper is organized as follows. In section 2, we present the theoretical formula of the CS inequality in photon noise spectra. In section 3, experimental results are given to demonstrate the violation of the CS inequality in quantum noise-correlated light beams generated from electrically coupled LEDs. In particular, we are interested in the case in which the CS inequality is violated (i.e. photon noise correlation between two beams is nonclassical), but the noise in the single-beam intensity or in the two-beam intensity difference is still higher than the shot noise level. Section 4 serves as a conclusion.

2. Theory

2.1. CS inequalities in different forms

We consider two beams of light with stationary statistical properties, described by classical intensity $I_1(t)$ and $I_2(t)$ respectively, where $I_1$ and $I_2$ are normalized to be proportional to the photon fluxes of the beams (see appendix A). The mathematical Schwarz inequality is given by [2]

$$\langle |I_1(t)I_2(t')|^2 \rangle \leq \langle I_1^2(t) \rangle \langle I_2^2(t') \rangle.$$  \hfill (3)

This relation leads to the standard classical CS inequality (1) where $g_{12}(\tau) = g_{11}(0)g_{22}(0)$, and $\mathcal{C}(\tau) = (I_1^2(t)I_2(t') - I_1(t)I_2(t'))/\langle I_1 \rangle \langle I_2 \rangle$ is the second-order cross-correlation function between the two beams. Classical inequalities in the single-beam measurements can be derived from equation (1): e.g., $g_{11}(\tau) = g_{11}^{(1)}(0)$ for $I_1(t) = I_2(t)$ [3], and $g_{11}^{(1)}(0) \geq 1$ when $I_2$ is a constant [6].

If the measurement quantities are the intensity fluctuations of the two beams, i.e. $i_1(t) = I_1(t) - \langle I_1 \rangle$ and $i_2(t) = I_2(t) - \langle I_2 \rangle$, another mathematical Schwarz relation can be obtained by replacing $I_1$ and $I_2$ with $i_1$ and $i_2$ respectively, as [2]

$$\langle |i_1(t)i_2(t')|^2 \rangle \leq \langle i_1^2(t) \rangle \langle i_2^2(t') \rangle.$$  \hfill (4)

In the quantum theory of light, the interpretation of the second-order coherence functions is quite different [3, 13]. The beam intensities $I_1(t)$ and $I_2(t)$ are replaced by quantum mechanical operators $\hat{I}_1$ and $\hat{I}_2$, and the second-order coherence functions $g_{11}^2(0) = \langle \hat{I}_1^2 \rangle / \langle \hat{I}_1 \rangle^2$, $g_{12}^2(0) = \langle \hat{I}_1^2 \rangle / \langle \hat{I}_2 \rangle^2$ are defined in the form of normal ordering of the negative-frequency part and positive-frequency part of the electrical field operators. The CS inequalities corresponding to equations (3) and (4) can be expressed as (see appendix A)

$$|\langle \hat{I}_1(t)\hat{I}_2(t') \rangle|^2 \leq \langle \hat{I}_1^2 \rangle \langle \hat{I}_2^2 \rangle$$  \hfill (5)

and,

$$|\langle \hat{I}_1(t)\hat{I}_2(t') \rangle|^2 \leq \langle \hat{I}_1^2 \rangle \langle \hat{I}_2^2 \rangle,$$  \hfill (6)

respectively. Equation (6) can be expressed as

$$|R_{12}|^2 \leq (F_1 - 1)(F_2 - 1),$$  \hfill (7)

where $F_1 = \langle \hat{I}_1^2 \rangle / 2\mathcal{R}_{11}$ and $F_2 = \langle \hat{I}_2^2 \rangle / 2\mathcal{R}_{22}$ are the Fano factors of the two beams, and $R_{12}(r) = \langle \hat{i}_1(t)\hat{i}_2(t') \rangle / 2\mathcal{R}_{11}\mathcal{R}_{22}$ is the cross-correlation function between the two beams, $B$ is the noise bandwidth of the detector filter (appendix B), $\langle \hat{i}_1^2 \rangle_{\text{m}} = 2\mathcal{R}_{11}$ and $\langle \hat{i}_2^2 \rangle_{\text{m}} = 2\mathcal{R}_{22}$ are the shot noise levels of the two beams, respectively.

2.2. CS inequality in photon noise spectral densities

In many cases, the most convenient measurement quantities of macroscopic light beams are photon noise spectral densities. These are directly proportional to the measured photodetector current noise power spectral densities. Therefore, a CS inequality associated with these measurement quantities is desirable in a study of the nonclassical violation of the CS inequality in photon noise spectra.

The photon noise spectrum $S_{11}(\Omega)$ (or $S_{22}(\Omega)$) of a light beam is defined as the Fourier transform of the auto-correlation function $\langle i_1(t)\hat{i}_1(t') \rangle$ (or $\langle i_2(t)\hat{i}_2(t') \rangle$) [13]

$$S_{11}(\Omega) = \int d\omega e^{i\omega t} \langle \hat{i}_1(t)\hat{i}_1(t') \rangle,$$ \hfill (8)

$$S_{22}(\Omega) = \int d\omega e^{i\omega t} \langle \hat{i}_2(t)\hat{i}_2(t') \rangle,$$ \hfill (9)

where $\omega = \omega - \omega$. $S_{11}(\Omega)$ and $S_{22}(\Omega)$ are each proportional to the noise power spectral density of the photocurrent as measured (for example) by a spectrum analyser. Similarly, the cross-correlation noise spectrum $S_{12}(\Omega)$ is defined as the Fourier transform of the cross-correlation function $\langle \hat{i}_1(t)\hat{i}_2(t') \rangle$:

$$S_{12}(\Omega) = \int d\omega e^{i\omega t} \langle \hat{i}_1(t)\hat{i}_2(t') \rangle.$$ \hfill (10)

The CS inequality in photon noise spectra relates the cross-correlation spectrum $S_{12}(\Omega)$ to the auto-correlation spectra $S_{11}(\Omega)$ and $S_{22}(\Omega)$. It can be expressed as (see appendix B)

$$|S_{12}(\Omega)|^2 \leq (S_{11}(\Omega) - \langle \hat{i}_1 \rangle)(S_{22}(\Omega) - \langle \hat{i}_2 \rangle).$$ \hfill (11)

Equation (11) can be expressed in terms of frequency-dependent Fano factors as

$$|R_{12}(\Omega)|^2 \leq (F_1 - 1)(F_2 - 1)$$ \hfill (12)

where

$$F_1 = S_{11}(\Omega)/\langle \hat{i}_1 \rangle,$$ \hfill (13)
From equation (12), the CS inequality in photon noise spectra can be violated in the following cases.

1. If one beam is amplitude-squeezed or sub-shot noise limited \( F_1 < 1 \) and the other beam is classical \( F_2 \geq 1 \), CS inequality equation (12) is violated for all non-zero cross-correlation values. This violation is evidently associated with the nonclassical photon statistics of a single beam.

2. If both beams are identical and nonclassical, i.e. if \( F_1 = F_2 < 1 \), CS inequality equation (12) may not be violated. The violation appears only if \( |r_{12}| > |F_1 - 1|/F_1 \).

3. If both beams are classical \( F_1 \geq 1 \) and \( F_2 \geq 1 \), CS inequality equation (12) can be violated in the absence of sub-shot noise fluctuations in the two-beam intensity difference. For instance, if one beam is nearly shot noise limited \( F_1 \approx 1 \), equation (12) can be violated with any non-zero value in \( S_{12} \) for large values of \( F_2 \). If the other beam is very noisy such that \( F_2 - 1 > 2r_{12}F_2^{1/2} \), it is evident that the fluctuations \( (S_2(\Omega) - 2S_{12}(\Omega)) \) in the two-beam intensity difference \( (I_2 = I_1 - I_2) \) is higher than the shot noise level \( (S_{12}(\Omega)) \) when \( S_{12}(\Omega) = (I_1) + (I_2) \).

Experimentally, we are interested in the case in which both beams are classical and we have demonstrated the violation of CS inequality equation (12) in the absence of sub-shot noise fluctuations in the two-beam intensity difference.

3. Experiment

3.1. Experimental setup

The experimental setup is shown in figure 1. The quantum noise-correlated light beams are generated by two infrared LEDs (LED1 and LED2). These are connected in series and biased by a constant current source through a resistance of 500 \( \Omega \), high compared with the differential resistance of the diodes in order to suppress pump noise in the bias current. The emitted beams from the LEDs (Hamamatsu L2656) are detected by two large-area pin photodiodes PD1 and PD2. The emitted photons from LED1 close proximity (~1 mm) to the LEDs, the overall quantum efficiency (the ratio of the photodetector current to the bias current) is \( \eta_1 = \eta_2 = 16.7\% \) at room temperature. The two photodetector currents are subtracted, amplified and input to a spectrum analyser to measure the noise power spectral density \( S_\omega(\Omega) \) in the dual-beam intensity difference \( (I_d = I_1 - I_2) \). The shot noise spectrum in either of the single beams, \( S_{11}(\Omega) \) or \( S_{22}(\Omega) \), can be measured separately by blocking the other beam. Therefore, the cross-correlation noise spectrum \( S_{12}(\Omega) \) can be calculated by

\[
S_{12}(\Omega) = [S_{11}(\Omega) + S_{22}(\Omega) - S_d(\Omega)]/2. \tag{16}
\]

Typical DC photocurrents are \( \eta_1 I_1 = \eta_2 I_2 = 1.0 \text{ mA} \) for a LED bias current of 6.0 mA. The noise levels \( S_{10}(\Omega) = 2\eta_1 I_1 \Delta \Omega \) and \( S_{20}(\Omega) = 2\eta_2 I_2 \Delta \Omega \) are calibrated using two low-efficiency incandescent filament lamps (while keeping the same photocurrent of 1.0 mA) [6], and by replacing the constant bias current with a shot noise bias current [15] produced by the photodetector current of a large-area pin diode illuminated by an incandescent lamp or a high-power LED. These methods all give consistent results.

In order to change the Fano factors of the two beams separately, the LEDs currents are modulated by wideband Gaussian noise generators NG1 and NG2, as shown in figure 1. Therefore, the excess noise level in the dual-beam intensity difference can be adjusted to be higher than the full shot noise level.

3.2. Experimental results

The experimental results are shown in figures 2–4. When the amplitude of the noise generator NG2 is set to zero and the amplitude of NG1 is set to make the fluctuations in both beams close to the shot noise level (SNL1), the two beams from the LEDs are similar to the conventional twin beams. In this case, the Fano factors in each beam are near unity, \( F_1 = F_2 \approx 1 \). However, the noise in the dual-beam intensity difference is below the shot noise level (SNL4). Figure 2(a) shows the noise power spectral densities for single beams \( S_{11}(\Omega), S_{22}(\Omega) \), dual-beam intensity difference \( S_d(\Omega) \), the
SNL for the single beam (SNL₁), and the SNL for the dual-beam intensity difference (SNL₂). It is evident that the noise reduction in S₂(Ω) below the SNL is frequency dependent, i.e. sub-shot noise is limited to frequencies below 0.8 MHz because of the long radiative recombination time of the LEDs used in this experiment. From equations (13) and (14), we can calculate the Fano factors F₁ and F₂ of the two beams. From equation (16), the cross-correlation noise spectrum S₁₂(Ω) can be calculated and therefore the cross-correlation function R₁₂(Ω) can be calculated from equation (15). Figure 2(b) shows the violation of the CS photon noise spectrum inequality of equation (12). Just as for the case of the sub-shot noise fluctuations in the intensity difference, the violation is frequency dependent and restricted to the low-frequency region.

If the amplitude of NG₂ is set to zero such that F₁ = F₂ and the amplitude of NG₁ is varied, we measure the noise power spectral densities S₁₁(Ω) or S₂₂(Ω) in the single-beam intensities and S₁₂(Ω) in the dual-beam intensity difference at a fixed frequency of Ω = 2π × 0.2 MHz. Thus we can measure a violation parameter

\[ \beta = \frac{|R_{12}(\Omega)|^2}{(F_1 - 1)(F_2 - 1)} \]  

as a function of the measured Fano factor F₁. Figure 3 shows the measurement results which demonstrate that the CS inequality is always violated (β > 1). The theoretical curve in figure 3 is given by

\[ \beta_{\text{theory}} = \frac{|F_{12}|^2}{(F_{12} - 1)^2}, \]

where F₁ = 1 + \eta₁(F₁ - 1). In order to obtain equation (18), we have assumed that the two beams inside the LED sources are fully correlated such that R₁₂ = F₁ = F₂. This is equivalent to assuming 100% internal efficiencies. The source beams are converted to photocurrents in the detectors with effective quantum efficiencies of \( \eta_1 = \eta_2 = 16.7\% \). Therefore, the measured Fano factors and cross-correlation function are expected to be F₁ = F₂ = \( \eta_1(F_{12} - 1) \), F₁ - 1 = \( \eta_2(F_{22} - 1) \) and R₁₂ = \( \eta_1\eta_2 F_{12} \). Equation (18) shows
that the CS inequality is always violated due to the full correlation in LED sources. The experimental measurement results in figure 3 are in good agreement with these theoretical expectations.

In order to demonstrate directly the violation of the CS inequality in the absence of a sub-shot noise intensity difference, (i.e. $S_2(\Omega)$ above the SNL), we set the amplitude of NG$_1$ so that the LED$_1$ beam is close to the SNL ($F_1(\Omega) \approx 1$) and the amplitude of NG$_2$ is set to inject a large excess noise current into LED$_2$, thus generating a super-Poissonian beam ($F_2(\Omega) \gg 1$). In this case, the noise level in the dual-beam intensity difference $S_d(\Omega)$ is well above the SNL, as shown in figure 4(a). However, the cross-correlation function $|R_{12}|^2$ is found to be still larger than $(F_1 - 1)(F_2 - 1)$, as shown in figure 4(b). This means that the CS inequality is still violated. This result follows from the quantum correlation between the LED$_1$ and LED$_2$ beams generated by the noise current of NG$_1$ despite the fact that the noise level in the LED$_2$ beam generated by NG$_2$ is much higher than the SNL.

The quantum noise correlation between the light fields emitted from electrically coupled LEDs in this experiment arises from the correlated electron–hole recombination (photon emission) processes in the LEDs due to the common injection current fluctuations (e.g. from NG1) imposed by the series connection. Following a random fluctuation in the number of charge carriers injected into the active region of LED$_2$, there will be a correlated fluctuation in the number of photons generated through electron–hole recombination on a time scale of the recombination lifetime. As a result of charge conservation, the same fluctuation in charge carrier injection will occur into the active region of LED$_1$ due to the series connection. Consequently, a second correlated photon flux increment will be generated by LED$_1$ on the same time scale. This means that the fluctuations in the photon fluxes are quantum correlated on time scales longer than the recombination time. As previously noted and modelled by Loudon [16], nonclassical intensity correlation and the consequent violation of a CS intensity inequality is a direct consequence of the discrete nature of the light beams arising from the quantized recombination and photodetection processes. This evidently holds true in the present case of quantum correlation and CS violation in the frequency domain.

4. Conclusion

A CS inequality in photon noise spectra has been derived and discussed. The violation of this CS inequality is a nonclassical characteristic of light fields associated with quantum correlation between two light beams arising from field quantization. Two classically fluctuating beams may violate this CS photon spectral inequality as a result of the quantum correlation between them. However, it is not necessary that the mean square noise in their intensity difference be below the standard shot noise limit. The violation of this CS inequality has been observed experimentally in the photon spectra of quantum noise-correlated light beams from electrically coupled LEDs. Unlike the conventional CS intensity inequality, the CS photon spectrum inequality is violated even though the noise in the two-beam intensity difference is well above the SNL. The measured violation parameters are in good agreement with theoretical expectations.

Appendix A. Derivation of equations (5) and (6)

For two separate beams of multimode fields in which each beam has a propagation direction and a polarization direction, the electrical field operator $\hat{E}_i(t)(i = 1, 2)$ can be expressed as

$$\hat{E}_i(t) = \hat{E}_i^{(+)}(t) + \hat{E}_i^{(-)}(t)$$

where $\hat{E}_i^{(+)}(t)$ and $\hat{E}_i^{(-)}(t)$ are the positive- and negative-frequency parts of the electric field, respectively. They can be decomposed into Fourier components [13]

$$\hat{E}_i^{(+)}(t) = \int \frac{d\omega}{2\pi} K_i(\omega) \hat{a}_i(\omega) e^{i\omega t}$$

and

$$\hat{E}_i^{(-)}(t) = \int \frac{d\omega}{2\pi} K_i^*(\omega) \hat{a}_i^*(\omega) e^{i\omega t}$$

where $K_i(\omega)$ is a scaling factor, $\hat{a}_i(\omega)$ and $\hat{a}_i^*(\omega)$ are the annihilation and creation operators, obeying the following commutation relations:

$$[\hat{a}_i(\omega), \hat{a}_k^*(\omega')] = [\hat{a}_i^*(\omega), \hat{a}_k(\omega')] = 0$$

$$[\hat{a}_i(\omega), \hat{a}_k^*(\omega')] = 2\pi \delta_{ik} \delta(\omega - \omega')$$

where $i, k = (1, 2)$ are the labels of the light beam 1 and the light beam 2. The normalized intensity operators (photon flux) can be written as $\hat{I}_i(t) = \hat{E}_i^{(+)}(t) \hat{E}_i^{(-)*}(t)$ and $\hat{I}_2(t) = \hat{E}_2^{(+)}(t) \hat{E}_2^{(-)*}(t)$, when we take $K_1(\omega) = K_2(\omega) = 1$ in equations (A.2) and (A.3) [13]. From equations (A.4) and (A.5), the commutation relations of the electric field operators can be derived as

$$[\hat{E}_i^{(+)}(t), \hat{E}_i^{(+)}(t')] = [\hat{E}_k^{(-)}(t), \hat{E}_k^{(-)}(t')] = 0,$$

$$[\hat{E}_i^{(+)}(t), \hat{E}_k^{(-)}(t')] = \delta_{ik} \delta(t - t').$$

The coherent states of a multimode field can be represented as $|\epsilon_i\rangle = \{|\alpha_i(\omega)\rangle\rangle(\omega = 1, 2)$, satisfying

$$\langle \epsilon_i | \hat{E}_i^{(+)}(t) | \epsilon_i \rangle = \epsilon_i(t) |\epsilon_i\rangle,$$

where

$$\epsilon_i(t) \int \frac{d\omega}{2\pi} K_i(\omega) |\alpha_i(\omega)\rangle e^{-i\omega t}.$$

Thus, in the Glauber–Sudarshan P-representation [14], the density operator for the two beams can be expressed in a diagonal coherent state representation as follows:

$$\hat{\rho} = \int \int P(\epsilon_1, \epsilon_2) |\epsilon_1\rangle \langle \epsilon_1| d\epsilon_1 d\epsilon_2,$$

with

$$P(\epsilon_1, \epsilon_2) = \frac{1}{2\pi} \int P(\epsilon_1, \epsilon_2) |\epsilon_1\rangle \langle \epsilon_1| d\epsilon_1 d\epsilon_2,$$

where $P(\epsilon_1, \epsilon_2)$ is the quasi-probability distribution (P-function).

If we choose the experimental measurable quantities as the ensemble means of the correlation of the two
beam intensities, i.e., $\langle \hat{I}_1(t)\hat{I}_1(t') \rangle$, $\langle \hat{I}_2(t)\hat{I}_2(t') \rangle$, and $\langle \hat{I}_1(t)\hat{I}_2(t') \rangle$, we can check the relation between these three statistical quantities. Consider the error function,

$$\sigma_1(\lambda) = \langle (\lambda\hat{I}_1(t) - \hat{I}_2(t'))^2 \rangle,$$

(A.12)

where the symbol $\langle \rangle$ denotes normal ordering of the electric field operators: that is, creation operators are on the left and annihilation operators on the right, and $\lambda$ is an any real variable number. By using equation (A.11), this error function can be expressed as follows:

$$\sigma_1(\lambda) = \lambda^2 (\langle \hat{I}_1^2(t) \rangle - 2\lambda \langle \hat{I}_1(t)\hat{I}_2(t') \rangle + \langle \hat{I}_2^2(t') \rangle) = \int \int P(e_1, e_2) (\lambda I_1^2(t) - I_2^2(t'))^2 d^2e_1 d^2e_2$$

(A.13)

where $I_1(t) = |e_1(t)|^2$ and $I_2(t') = |e_2(t')|^2$. For classical beams of the field, $P(e_1, e_2)$ is non-negative, thus we have $\sigma_1(\lambda) > 0$ for all real values of $\lambda$. The necessary and sufficient condition for $\sigma_1(\lambda) > 0$ is that the discriminate of the quadratic polynomial $\sigma_1(\lambda)$ is not positive. Therefore, we have the quantum mechanical expression of CS inequality (equation (3)) as

$$|\langle \hat{I}_1(t)\hat{I}_2(t') \rangle|^2 \leq \langle \hat{I}_1^2(t) \rangle \langle \hat{I}_2^2(t') \rangle.$$  

(A.14)

If we choose the experimental measurable quantities as the ensemble means of the correlation of the two-beam intensity fluctuations, i.e., $\langle \hat{I}_1(t)\hat{I}_1(t') \rangle$, $\langle \hat{I}_2(t)\hat{I}_2(t') \rangle$, and $\langle \hat{I}_1(t)\hat{I}_2(t') \rangle$, it is similar that the error function, $\sigma_2(\lambda) = \langle (\lambda\hat{I}_1(t) - \hat{I}_2(t'))^2 \rangle$, is non-negative. This leads to the quantum mechanical expression of CS inequality of equation (4) as

$$|\langle \hat{I}_1(t)\hat{I}_2(t') \rangle|^2 \leq \langle \hat{I}_1^2(t) \rangle \langle \hat{I}_2^2(t') \rangle.$$  

(A.15)

Appendix B. Derivation of equation (11)

Using the commutation relations of equations (A.6) and (A.7), we have:

$$\langle \hat{I}_1(t)\hat{I}_1(t') \rangle = \langle \hat{I}_1 \rangle \delta(t-t') + \langle \hat{I}_1(t)\hat{I}_1(t') \rangle,$$

(B.1)

$$\langle \hat{I}_2(t)\hat{I}_2(t') \rangle = \langle \hat{I}_2 \rangle \delta(t-t') + \langle \hat{I}_2(t)\hat{I}_2(t') \rangle.$$  

(B.2)

Thus equations (8) and (9) can be written as

$$S_{11}(\Omega) = \langle \hat{I}_1 \rangle + \int d\Omega e^{i\Omega t} \langle \hat{I}_1(t)\hat{I}_1(t') \rangle$$

(B.3)

$$S_{22}(\Omega) = \langle \hat{I}_2 \rangle + \int d\Omega e^{i\Omega t} \langle \hat{I}_2(t)\hat{I}_2(t') \rangle$$

(B.4)

where $S_{11}(\Omega)$ and $S_{22}(\Omega)$ are Mandel factors that describe the temporal correlations between photons of each beam, respectively. The first terms in equations (B.3) and (B.4) are the standard shot noise levels which are independent of frequency. The variance of the intensity noise, $\langle \hat{I}_1^2 \rangle$ or $\langle \hat{I}_2^2 \rangle$, can be integrated from the noise spectrum [13],

$$\langle \hat{I}_1^2 \rangle = \int \frac{d\Omega}{2\pi} f_1(\Omega) S_{11}(\Omega),$$

(B.5)

$$\langle \hat{I}_2^2 \rangle = \int \frac{d\Omega}{2\pi} f_2(\Omega) S_{22}(\Omega),$$

(B.6)

where $f_1(\Omega)$ and $f_2(\Omega)$ are the functions associated with the detection filter. Using equations (B.3) and (B.4), one can obtain the usual expressions for the variances

$$\langle \hat{I}_1^2 \rangle = 2B(\hat{I}_1) + \langle \hat{I}_1^2 \rangle,$$

(B.7)

$$\langle \hat{I}_2^2 \rangle = 2B(\hat{I}_2) + \langle \hat{I}_2^2 \rangle,$$

(B.8)

with the noise bandwidth of the detector filters of

$$B = \int_0^\infty d\Omega f(\Omega).$$

(B.9)

where we assume that $f_1(\Omega) = f_2(\Omega) = f(\Omega)$.

Now, consider the difference of intensity noise, $\langle \hat{I}_1(t)\hat{I}_1(t') \rangle = \lambda\hat{I}_1(t) - \hat{I}_2(t)$ for any real values of $\lambda$. The error-function spectrum, $\sigma_3(\Omega, \lambda)$, can be defined as

$$\sigma_3(\Omega, \lambda) = \int d\Omega e^{i\Omega t} \langle \hat{I}_1(t)\hat{I}_1(t') \rangle$$

(B.10)

Using the density operator of the light beams, equation (A.11), we can express $\sigma_3(\Omega, \lambda)$ as

$$\sigma_3(\Omega, \lambda) = \int P(e_1, e_2) S_3(\Omega) d^2e_1 d^2e_2,$$

(B.11)

where

$$S_3(\Omega) = \int d\Omega e^{i\Omega t} \langle \hat{I}_1(t)\hat{I}_1(t') \rangle$$

(B.12)

Physically, $S_3(\Omega)$ is the power spectral density of the stationary function $\hat{I}_1(t)$ and its value is always positive [2], e.g. $S_3(\Omega) > 0$. Therefore, for non-negative distribution $P(e_1, e_2)$, we always have the quadratic polynomial $\sigma_3(\Omega, \lambda) > 0$ for any real values of $\lambda$. This leads to the CS inequality in the form of photon noise spectral densities (equation (6)),

$$|S_{12}(\Omega)|^2 \leq (S_{11}(\Omega) - \langle \hat{I}_1 \rangle)(S_{22}(\Omega) - \langle \hat{I}_2 \rangle).$$

(B.13)

References

Yong-qing Li et al.


Teich M C et al 1990 Phys. Today June p 26

Edwards P J 1991 Int. J. Optoelectron. 6 23


Sub-Shot-Noise laser Doppler Anemometry with Amplitude-Squeezed Light

Yong-qing Li,1 Peter Lynam,2 Min Xiao,3 and Paul J. Edwards1
1Advanced Telecommunications Research Centre, Faculty of Information Sciences and Engineering, University of Canberra, P.O. Box 1, Belconnen, ACT 2616, Australia
2Department of Physics, University College, University of NSW, ADFA, ACT 2600, Australia
3Department of Physics, University of Arkansas, Fayetteville, Arkansas 72701

(Received 21 October 1996)

Amplitude-squeezed light from a quantum-well semiconductor laser with weak optical feedback below the shot-noise level is observed with a feedback factor of $1.5 \times 10^{-4}$. Enhanced sensitivity is demonstrated in the Doppler measurement of a gas flow velocity with an improvement in the signal to noise ratio of 1.0 dB above the shot-noise limit. [S0031-9007(97)02955-4]

PACS numbers: 42.50.Dv, 42.55.Px, 42.62.Fi

In recent years, squeezed light exhibiting fluctuations below the standard quantum limit (SQL) in one quadrature amplitude or in the amplitude of the field has been shown to enable precision measurements with sensitivities beyond the SQL [1–4]. Quadrature squeezed states of light, characterized by reduction of the mean square fluctuation in one quadrature component of the field below that of the vacuum state, have been used to improve the precision of shot-noise limited measurements of weak absorption [2] and in interferometry [3]. A frequency-tunable squeezed light source has been used to demonstrate improvement in the sensitivity of the saturation spectroscopy of atomic cesium [4] and to demonstrate fundamental phenomena in the atom-photon interaction [5]. Amplitude-squeezed states of light, featuring photon number fluctuations below those of Poissonian statistics, have been generated from semiconductor lasers [6–8], and have been recently used for nonlinear spectroscopy and dark fringe interferometry [9]. Low noise amplification has also been demonstrated using sub-Poissonian light generated by semiconductor junction light emitters [10].

In this Letter, we demonstrate the application of amplitude-squeezed light to laser Doppler anemometry with consequent improvement in the sensitivity of velocity measurements above the shot-noise limit. The amplitude-squeezed light was produced from a cooled quantum-well semiconductor laser with weak grating feedback of $-10^{-3}$–$-10^{-4}$. Up to 2 dB of photon number squeezing was observed. We have utilized this squeezed light source to achieve an improvement in sensitivity of 1.0 dB beyond the shot-noise limit for the heterodyne detection of Doppler-shifted light scattered from moving particles. The LDA technique has been employed to measure the velocity distribution of cold atoms trapped in an “optical molasses” [13] and in molecular scattering. The small Doppler frequency shifts in the weakly scattered light may be detected by an optical heterodyne technique using a reference laser beam. The velocity and density of moving particles in a fluid can be obtained from the beat frequency and beat amplitude. For the experimental arrangement shown in Fig. 1, the Doppler shift of the scattered light (dashed line directed to the detector) is given by [12]

$$\nu_D = (n/\lambda)\mathbf{u} \cdot (k_s - k_0) = (2nu/A)\sin(\alpha/2),$$

with $n$ the index of refraction of the flow medium, $\mathbf{u}$ the velocity vector of the flow with amplitude $u$, $\lambda$ the wavelength of the laser beam, $k_0$ and $k_s$ the unit vectors in the directions of the illuminating and scattering (reference) beams, respectively, and $\alpha$ the angle between them.

FIG. 1. Experimental setup. Squeezed source: LD—laser diode; BS—beam splitter with 94% transmission; NDF—neutral density filter; Pol.—polarizer; Iso.—optical isolators; L1, L2, L3—optical lenses; PZT—PZT-controlled mirror; M1, M2—mirrors; A1, A2—apertures. The laser beam from the squeezed source passes through a gas flow twice at an angle of $\alpha$. The scattered light and the transmitted reference beam are collected by a large area pin photodetector (Det.) and the output photocurrent is amplified, and fed into a spectrum analyzer (SA).
The fundamental limitation to sensitivity of this reference beam technique is the shot noise arising from the reference laser beam [12]. The light received by the photodetector comprises a strong local oscillator (reference beam) component and a much weaker Doppler-shifted (scattered light) component. The photocurrent is
\[ i(t) = \eta e[I_0 + 2(n_0 I_0)^{1/2} \cos(2\pi \nu_D t)], \]
where \( \eta \) is the quantum efficiency of the photodetector, and \( I_0 \) and \( I_s \) are the photon number fluxes (intercepted by the detector per unit time) of the reference and scattered beams, respectively. The mean square heterodyne signal current at frequency \( \nu_D \) is then given by \( \langle i_s^2 \rangle = 2e^2 \eta^2 I_s l_{10} \) and the mean square noise current by \( \langle i_n^2 \rangle = 2e^2 \eta I_0 BF_0 \) with \( B \) the noise bandwidth, and \( F_0 \) the Fano factor of the detected reference beam. The signal to noise ratio is then given by
\[ \langle i_s^2 \rangle / \langle i_n^2 \rangle = \eta I_s / BF_0. \] (2)

It is apparent from Eq. (2) that an amplitude-squeezed reference beam (with \( F_0 < 1 \)) leads to enhanced sensitivity of Doppler velocity measurements relative to shot-noise limited measurements (with \( F_0 = 1 \)). The signal to noise ratio of Eq. (2) may be written as an upper limit of \( 2\eta(n_s)/F_0 \), where \( n_s \) is the integrated photon count in a matched filter measurement. In our experiment shown in Fig. 1, both the illumination and reference beams come from an amplitude-squeezed laser source and intersect with a gas flow containing a cloud of smoke particles. The intensity of the weakly scattered light is proportional to the intensity of the illuminating beam, that is, \( I_s = \beta I_{10} \), where \( \beta \) is a scattering parameter. Therefore, for a given \( \beta \), we can improve the signal to noise ratio by employing a photon-number squeezed reference beam as well as by increasing the optical power of the illumination beam. This sub-shot noise, weak scattering laser Doppler technique is therefore potentially valuable in situations where upper limits on permissible illumination intensities exist and where high precision is required, as in medical applications, atomic and molecular scattering, and nondestructive laser measurements generally.

When the total light scattering is very weak (corresponding to an optically thin medium), the optical loss is correspondingly small (in this experiment, less than 2%) and will not significantly degrade the squeezing of the reference beam. This is the basis of this new technique. A similar principle has been demonstrated in absorption spectroscopy with squeezed light [5,9], where the photon losses are also negligibly small. A typical heterodyne spectrum is given in Fig. 2. In the absence of scattering particles, the noise power of the reference beam (from 0.3 to 1.5 MHz) is found to be 1.0 dB below the shot-noise level [Fig. 2(a)], set by a red-filtered white light reference source and balanced detectors [7,8]. The heterodyne signal due to Doppler-shifted light scattered from smoke particles is shown in Fig. 2(b), where the signal to noise ratio is 1.0 dB above the shot-noise limit. The Doppler shift signal at 0.6 MHz corresponds to a flow velocity of 267 mm/s. The signal to noise ratio is improved by 1.0 dB above the usual shot-noise limit.

The experimental arrangement in Fig. 1 is a typical reference beam heterodyning configuration for laser Doppler measurements [12]. The amplitude-squeezed laser beam from the squeezed source (about 25 mW at 770 nm) passes through a polarizer (extinction ratio \( >10^4:1 \)) and two optical isolators (total isolation ratio \( >60 \) dB), and is then focused by lens L1 (with \( f = 75 \) mm) onto the flow; this focused light acts as the illumination beam. The transmitted laser beam, which serves as the reference beam, is collected and focused by lens L2 (with \( f = 100 \) mm). The reference beam is also focused onto the flow but its focus is slightly displaced from that of the illumination beam. This ensures that the effective scattering regions are slightly different for the two beams and so avoids loss of signals by phase decorrelation [12]. The scattered light is focused with the reference beam into an aperture A1 by lens L3 and passes through an aperture A2 in front of a high efficiency pin photodetector (Hamamatsu S3994). The ac part of the photodetector current is amplified and fed into a spectrum analyzer. The aperture A1 is used to ensure that only light coming from or passing through the desired region reaches the detector.
The squeezed laser source used in our experiment was a quantum-well AlGaAs semiconductor laser with weak optical feedback from a grating in a Littman-Metcalf configuration, as shown in Fig. 1. The laser diode (SDL-5411-G1) and collimation lens were cooled down to 80 K inside a liquid nitrogen cryostat. Because of the presence of multiple subthreshold longitudinal side modes in the laser diode [14], some line-narrowing techniques such as injection locking or optical feedback from an external grating must be used in order to obtain amplitude squeezing [7,8]. In our experiment, we used a beam splitter BS (with a transmission of 94%) to reflect a small portion of the laser beam to a grating (1800 lines/mm). The first-order diffraction beam was reflected and fed back to the laser diode by a piezoelectric-translator-controlled mirror. A power meter behind the beam splitter was used to measure the feedback intensity. The maximum feedback intensity was a fraction \(1.2 \times 10^{-3}\) of the output power of the laser diode. Since the feedback beam passes the grating twice, the external cavity in this configuration is highly dispersive and therefore is effective in suppressing longitudinal side modes. This configuration keeps the advantages of low optical loss for the squeezed light output and provides frequency tunability [15]. The threshold current \((3.5 \text{ mA at 80 K})\) was not changed significantly by this weak grating feedback. The linewidth of the single-mode laser with this configuration is estimated to be below 1.0 MHz [15], which corresponds to a coherence length of a few tens of meters, much larger than the optical path difference between illumination and reference beams in Fig. 1.

Curve \(b\) in Fig. 3(a) shows the measured noise power spectral density for the laser beam from the squeezed source (without the polarizer and optical isolators), with the laser diode biased at 26.7 mA giving a corresponding photodetector current of 12.0 mA. Curve \(a\) is the shot-noise level with the same photodetector current. The shot-noise level (SNL) was set by a red-filtered white light source. In order to decrease the effect of detector saturation in calibration of the SNL, the large-area \((10 \text{ mm} \times 10 \text{ mm})\) pin photodiode was placed far away from the focus of the reference beam at A1 so that the laser beam nearly filled the detector aperture. Care was taken to check the consistency between the SNL in two identical detectors set by two white light sources and the noise calibrated by the balanced detectors with a 50/50 nonpolarizing beam splitter [14]. We found that the noise levels agreed to within 5% for dc detector current up to 15.0 mA per detector for the large beam spot described above. As an additional check, as the laser beam was attenuated 50% by using a neutral density filter, the squeezing level \((2.0 \text{ dB})\) was reduced by 50% (curve \(d\) related to the SNL of curve \(c\)) after correction for the amplifier noise level (curve \(e\)). When the polarizer and optical isolators were inserted, the squeezing was degraded to about 1.0 dB as shown in Fig. 2(a) due to the additional optical losses and the interference between the main polarization component and small perpendicular polarization component [7,14].

The gas flow used in the experimental comprised high pressure clean nitrogen gas passing through a small chamber containing a smoke generator. As the smoke passed through the intersection region of the illumination and reference beams, the photodetector current decreased by less than 2%. In order to eliminate possible feedback of scattered light into the squeezed source, we inserted a polarizer and two optical isolators between the squeezed source and scattering region. The optical feedback due to light scattering by the gas flow was then less than \(10^{-8}\). Thus, the optical feedback into the laser diode was dominated by the grating feedback which was set at the maximum feedback intensity \((\sim 1.2 \times 10^{-3})\). Figure 4 shows Doppler shift signals at different nitrogen gas pressures. The flow velocities of 324, 267, and 195 mm/s were inferred from the observed Doppler shift signals in Figs. 4(a)-4(c), respectively. The detection sensitivities for Doppler signal measurements were clearly above the shot-noise level. The linewidth of the Doppler beat signal was limited by transit time, resolution bandwidth of the spectrum analyzer, and turbulence of the gas flow.

![Graph showing noise power spectral densities](image-url)
In summary, we have described and demonstrated a new sub-shot-noise laser Doppler technique, an application of squeezed light quite distinct from the interferometric and spectroscopic applications previously reported. A wideband amplitude-squeezed light source from a quantum-well semiconductor laser with weak grating feedback was employed for laser Doppler velocity measurements. Improvements in measurement sensitivity of 1.0 dB beyond the usual shot-noise limit have been directly observed. By using highly squeezed laser sources [6] or highly correlated twin beams [16,17] in a two-channel reference beam configuration, the measurement sensitivity for laser Doppler measurements could be improved to make this technique useful in practice. This sub-shot-noise laser Doppler technique is also suitable for making velocity distribution measurements of relatively small ensembles of scattering particles, such as cold atoms [13], trapped ions, or biologically active molecules.

The authors acknowledge financial support from the Australian Research Council, the Australia Commonwealth Department of Education, Employment and Training, and the University of Canberra.

A Sub-Shot Noise Optocoupler
with Positive Feedback

P. J. Edwards, W. N. Cheung, H. B. Sun, D.G. O'Brien
Advanced Telecommunications Engineering Centre,
Faculty of Information Sciences and Engineering,
University of Canberra,
PO Box 1, Belconnen, ACT 2616, Australia.
and
P. Lynam
Physics Department
University College, ADFA, University of New South Wales.

Presented at COMAD'94, Macquarie University, Sydney, December 12th - 14th, 1994

ABSTRACT

A novel sub-shot noise optocoupler amplifier has been modelled and realised in prototype IRED form with an open loop current gain of 1.0 at room temperature and 2.0 at 77K. An optical noise figure (for Poissonian light) of 1.4 is achieved at 77 K. Both open and closed photon number and current amplification have been demonstrated with photon partition noise unaffected by the application of positive feedback. The configuration is characteristic of a class of quantum noise reduced systems realised with high quantum efficiency devices which may be fabricated in OEIC form with potential application as low noise optical interconnects, optoelectronic logic elements and low noise amplifiers.

1. Introduction

The use of a tandem array of light emitting diodes or laser diodes for low noise current and light amplification was first proposed by Edwards [1]. High gain and low noise performance may be achieved by closely coupling together series connected arrays of semiconducting light emitting junctions and light detecting elements with high quantum transfer efficiency. High quantum coupling efficiencies provide suppression of the photonic partition noise usually associated with optoelectronic devices. A new class of low noise optocouplers and optoelectronic amplifiers can be realised when high quantum transfer efficiency is combined with high impedance suppression of the normal electronic shot noise present in semiconductor light emitters [2, 3, 4].

We discuss here a further extension of this concept in which positive current feedback is applied from the detector array to the emitting array. This method retains the low noise advantages of the open loop form of the array amplifier in which the light from an array of emitting junctions is combined and detected with high efficiency. The incorporation of positive feedback allows high closed loop current and light gains to be realised. Such an amplifier can be implemented with a single emitter and single detector providing the quantum transfer efficiency between the two elements is greater than one half. The recent development of high efficiency light emitting diodes [5] with external efficiencies of 72% at room temperatures makes this feasible.
2. High Impedance Shot Noise Suppression

Figure 1 shows the equivalent noise circuit of an ideal light emitting diode with diffusion capacitance, $C$, and differential resistance, $r$, driven by impedance, $R_s$, coupled to a detector with overall quantum efficiency, $\eta$.

From this figure it is evident [3] that providing $R_s \gg r \equiv k T/e \langle I \rangle$, the junction voltage, $v(t)$ will fluctuate freely at low frequencies and there will be complete suppression of the shot noise current which would otherwise be driven through differential resistance, $r$, by equivalent shot noise voltage generator, $v_{sn}(t)$. This situation is alternatively expressed in terms of a fluctuating charge carrier population, responding at low frequencies to the stochastic recombination process, in such a way as to suppress the recombination rate fluctuations below the stationary random Poissonian (shot noise) level. For an ideal emitter this results in a photon current with the same sub-Poissonian character as the recombination current in the diode junction. This beam of "quiet light" in which the mean square photon counting noise is reduced to a small fraction $F_i$ of the shot noise value would then generate an essentially shot noise free current in an ideal detector.

In a system with quantum transfer efficiency, $\eta$, the fractional noise is increased above its ideal value because of the random attenuation of photons between emitter and detector.

It can be shown [1] that the detector noise current then comprises two terms: diode shot-noise transferred from the emitter to the detector plus an uncorrelated deletion or "partition" noise term. In the ideal case of unit quantum efficiency, the partition noise term is zero and the first term can be made negligible by using a high impedance modulator.

3. Quantum Noise Suppressed Tandem Diode Array Opto-amplifier

An idealised representation is shown in Figure 2. Here the system is configured as a photon number amplifier and quantum tap, providing both light gain and detection.

In Figure 3, light from each of the $N$ emitting diodes of the LED array is coupled with quantum efficiency, $\eta$, to the output light detector. The emitting array is modulated by current modulator $V_g$ and positive current feedback is provided via impedance, $R_f$. The current modulator may itself be a semiconductor junction lightwave detector in which case the system functions as a photon number amplifier/receiver.

The open loop current gain of the amplifier is evidently just $N.\eta$ and the open loop noise, relative to shot noise, may be shown to be $F_o = N \cdot \eta \cdot F_i + (1-\eta)$. The Noise Figure is then approximately $1 + (1-\eta)/N.\eta$ for a shot noise limited optical signal detected with unit efficiency.

Figure 4 shows the results of typical noise measurements undertaken with a prototype open loop array amplifier. An array of eight Hamamatsu type L2656 AlGaAs DH infrared emitting diodes were coupled to a two element detector with average quantum efficiency, $\eta=0.13$ at room temperature, giving a current gain of close to unity. Upon cooling, the current gain at 300 kHz increased with quantum efficiency to a maximum of $N.\eta = 2.0$ at 77 K. The lowest trace shows the suppressed quantum noise level under high impedance conditions ($F_o = 0.81$). This corresponds to a diode noise of only 3% of the shot noise level. The upper trace shows the 5.5 dB increase expected for a shot noise input ($F_i=1$).
The intermediate trace indicates the noise level for a sub-Poissonian input of \( F_i = 0.77 \), in close agreement with that expected. The corresponding noise figure is 1.4.

4. Optocoupler with Positive Feedback

Providing the open loop gain is greater than one half a closed loop gain exceeding one can be achieved by the application of feedback. This follows from the usual amplifier feedback equations.

This is the basis for the opto-amplifier. This consists of a single laser or light emitting diode optically and electrically coupled to a detector to provide closed loop current gain without degradation of noise performance.

Figure 5 shows a simplified noise equivalent circuit for such an optocoupler driven from a current source with unity feedback from a detector current source. The emitting diode is represented as in Figure 1. The shot noise is completely suppressed and the closed loop output partition noise is at the shot noise level. The signal to partition noise ratio is then identical with that for the open loop but, of course, the signal current is a factor of \( 1/(1-\eta) \) greater. Feedback may also be applied to an \( N \) element array if greater gain bandwidth is required. The application to optical interconnects and OEIC applications generally is clear.

5. Conclusions

The development of high quantum efficiency optoelectronics will enable a new class of optoelectronic devices and systems to be realised in which photonic noise can be largely suppressed. Current, voltage, power and light gain with noise level below the normal quantum noise limit is achievable with a positive feedback optocoupler which can be implemented in OEIC form.

6. References


Figure 1

Ideal light-emitting semiconductor junction diode noise equivalent circuit. The overall quantum efficiency, $\eta$, is defined as the ratio of detector current $i_o(t)$ to junction current, $i(t)$.

Figure 2

Idealised representation of a photon number amplifier/tap employing arrays of sub-Poissonian light emitters and detectors.
Open loop optoamplifier measurements showing: Shot noise suppression (lower trace); Amplified shot noise ($F_i = 1; F_o = 2.75$); sub-shot noise amplification ($F_i = 0.77$). Current gain = 2.0; quantum efficiency = 0.25 at 77K.
Figure 5

Simplified Noise Equivalent Circuit of a single element optocoupler with feedback.

\[ i_o = \eta i_d; \quad i_p = \eta^2 \cdot (1 - \eta) \cdot i_{sn}; \quad \langle i_{sn}^2 \rangle = \frac{\langle u_{sn}^2 \rangle}{r_d^2} \]

Acknowledgements

We acknowledge useful discussions with Ms T. Zhang and the assistance of Ms L. Liu and Ms M. Tatarow.
A photon coupled circuit comprises a high quantum efficiency semiconductor light emitter (TDA) coupled with low photon losses to a high quantum efficiency semiconductor light detector (OLD). Bias current is provided to both the light detector (OLD) and light emitter (TDA). The light output of the light emitter (TDA) is modulated by the signal current flowing therein and the current flowing \( i_0 \) in the detector (OLD) is modulated by the light received from the light emitter (TDA). The quantum transfer efficiency or open loop current gain is greater than 0.5 and a portion of the AC current flowing in the light detector (OLD) is applied as feedback to reinforce or oppose the current flowing in the light emitter (TDA).
FIG. 1

FIG. 2
FIG. 5

![Graph showing Current Gain vs Frequency](image)

FIG. 6

![Graph showing Output Current vs Frequency](image)
LOW NOISE PHOTON COUPLED CIRCUIT

TECHNICAL FIELD

This invention relates to optoelectronic coupling and more particularly to a photon coupled circuit for applications including low noise optocoupling, optical interconnects, current amplifiers, light amplifiers and optical bistable devices. As used herein, the term "photon coupled circuit" refers to a configuration of optoelectronic devices which are photonically and electronically coupled.

BACKGROUND ART

Many prior art devices achieve a current amplification by the use of active electronic circuits. Such devices are not able to offer a low noise operation and in particular are often limited by the shot noise in the system. Australian patent 648682 describes a means of generating multiple noise-correlated lightwave beams in order to suppress shot noise in optical measurement systems.

DISCLOSURE OF INVENTION

It is an object of this invention to provide a sub-shot noise photon coupled amplifier circuit through quantum noise suppression and to provide increased gain or increased bandwidth to the circuit through the application of AC feedback.

Accordingly, in one aspect this invention consists in a sub-shot noise, optoelectronic amplifier circuit comprising a high quantum efficiency semiconductor light emitter; means to provide signal input current and bias current to said light emitter, the light output of said light emitter being controlled by the current flowing therein; a high quantum efficiency semiconductor light detector optically coupled with low photon losses to said light emitter to provide an overall quantum transfer efficiency in excess of 0.5; means to provide a bias voltage to said light detector, the current flowing in said light detector being controlled by the light received from said light emitter and used to provide an output signal; means to isolate a portion of the alternating current back to reinforce or oppose the signal current flowing in the light emitter.

In one form of the invention the feedback is positive to reinforce the signal current flowing in a light emitter and thus to provide increased AC current gain exceeding unity. In another form of the invention the feedback is negative in order to oppose signal current flowing in the light emitter and thus to provide increased bandwidth. As used herein, the term quantum transfer efficiency refers to the efficiency of the transfer of the current modulating the light emitter into current flowing in the light detector. It is therefore equal to the open loop current gain of the circuit.

In one form of the invention the light emitter comprises an array of high quantum efficiency laser diodes or light emitting diodes which are driven by a signal current from a high impedance source.

In one form of the invention using an array of high quantum efficiency semiconductor light emitting or laser diode junctions, the product of the number of elements in the array and the value of the quantum transfer efficiency between each element and the light detector is much greater than 1. In this case the quantum partition noise introduced in the coupling between the light emitting array and the detecting system is negligible and negative feedback may be applied to widen the bandwidth of the system.

In another form of the invention, employing a single light emitter the quantum transfer efficiency preferably exceeds 0.75 and ideally is as close to unity as possible.

Preferably, the light detector comprises a back biased PIN photo diode operating in photovoltaic mode.

The high quantum transfer efficiency between the light emitter and the light detector is preferably achieved by low loss photon coupling through close physical positioning as for example, in an optoelectronic integrated circuit (OEIC) together with selection of high quantum efficiency light emitters and light detectors. Low loss photon coupling can also involve the use of an optical waveguide or a low loss light pipe.

In one form of the invention the current in the light emitter is preferably modulated by the current output of an additional light detector. The device then functions as a light receiver. In an alternative arrangement an input light signal from an external source is received by the light detector and the current flowing in the light detector is modulated by the input light signal.

In another form of the invention the current flowing in the light detector is used to drive individual or arrays of semiconductor laser or light emitting junctions.

The photon coupled circuit of this invention can be used as the basis of an optocoupler, optical interconnector, optical amplifier, electronic amplifier, optoelectronic oscillator, optoelectronic multivibrator or optoelectronic bistable device.

BRIEF DESCRIPTION OF DRAWINGS

One embodiment of the invention will now be described, by way of example only, with reference to the accompanying drawings in which:

FIG. 1 shows a photon coupled current amplifier according to the invention;

FIG. 2 shows an AC equivalent circuit of the amplifier shown in FIG. 1;

FIG. 3 shows (a) current gain, (b) voltage gain and (c) power gain of the photon coupled amplifier shown in FIG. 1;

FIG. 4 is an AC equivalent circuit of the photon coupled amplifier used for noise calculations;

FIG. 5 is a plot of current gain of the photon coupled amplifier with feedback resistance equal to (a) 30, (b) 15 and (c) 7.5 k ohms.

FIG. 6 is a plot of the frequency spectra of the photon coupled amplifier output currents of (a) signal and (b) noise;

FIG. 7 shows a plot of a noise figure variation with a source resistance at different quantum efficiencies;

FIG. 8 shows a second embodiment of a photon coupled amplifier according to this invention; and

FIG. 9 shows a third embodiment of a photon coupled amplifier according to this invention.

BEST MODE FOR CARRYING OUT THE INVENTION

FIG. 1 shows the circuit of an optocoupler amplifier. An array of series connected semiconductor junction light emitting diodes (or laser diodes) TDA each of high quantum efficiency is shown modulated by a voltage source Vp with internal impedance R. It is connected to the array through a blocking capacitor Cr and the DC bias current for the array is provided through resistance R. The light from this array is efficiently collected by an array of PIN diodes OLD which is biased through resistance R.
A fraction of the current output from the detector is fed back from point B through the feedback resistance $R_f$ and the blocking capacitor $C$, to the input of the light emitting array at point A to provide positive current feedback. The amplified current output is available between points A and B or between B and G. The noise behaviour of the positive feedback amplifier will depend on the quantum efficiency of the optoelectronic conversion and the noise figure is expected to improve as the quantum efficiency is increased. In the following analysis, a LED model which has been found useful for both small signal and noise performance evaluations is used.

Let

$$N = \text{number of LEDs in series}$$

$$\eta = \text{differential quantum efficiency per device}$$

$$r_s = \text{differential resistance of the LED}$$

$$R_s = \text{bulk resistance of the LED}$$

$$r_{sd} = r_s + R_s$$

$$R_1 = \text{bias resistance for the LED array}$$

$$R_p = \text{includes the bias and load resistance at the output}$$

$$R_f = \text{feedback resistance (or impedance in general).}$$

The low-frequency small signal equivalent circuit for the system is shown in FIG. 2. It is assumed that the LEDs are identical and that the detector generates an output current $i_d$ given by $i_d = N(i_d)$. where $i_d$ is the signal current flowing through the LED array. The impedance of the detector is assumed to be high compared with $R_1$ or $R_p$. With $i_d$ equal to the current from the signal source, it can readily be shown that the closed-loop current gain of the amplifier is given by

$$A_i = \frac{i_b}{i_d} = \frac{A_0}{1 + \frac{R_f}{R_1}}$$

(1)

where $A_0$ is the open-loop current gain given by

$$A_0 = \frac{A_0}{\eta} \left( \frac{R_1}{1 + \frac{R_f}{R_1}} \right)$$

(2)

and $\beta$ is the feedback factor given by

$$\beta = \frac{R_2}{R_f + R_2} \left( 1 - \frac{R_f}{R_1} \right)$$

(3)

For stability reasons, the amplifier must operate such that the loop gain $A_0 \beta$ is less than unity. If $R_1$ and $R_2$ are large compared with the device resistance $r_s$, the open-loop gain becomes $N_0 \eta$ and the feedback factor simplifies to $R_f/(R_f + R_2)$.

The input impedance of the opto-amplifier is given by

$$R_{inp} = \frac{v_i}{i_i} = \frac{A_{inp}}{\eta}$$

(4)

The effect of the LED bias resistance $R_1$ on the input resistance is already included in the expression for the current gain.

It can also be shown that the voltage gain of the feedback amplifier is given by

$$A_v = \frac{v_o}{v_i} = 1 + \frac{R_2}{R_f + R_2} \left( \frac{N}{r_{sd}} - \frac{1}{R_1} \right)$$

(5)

It will be noted that while the current gain is proportional to the number of LEDs in the array, the voltage gain is independent of $N$. When $R_f = 0$, $A_v$ reduces the open-loop value of $\eta R_{inp}$. For a given circuit, the voltage gain increases as the LED bias current increases. This is because the device resistance decreases as the bias current goes up. The apparent power gain of the feedback amplifier is given by

$$P_i/P_o = A_o A_v$$

where $P_i = v_i i_i$ is the input signal power and $P_o = v_o v_0$ is the output signal power from the detector. FIG. 3 shows typical variations of current gain, voltage gain and power gain of the amplifier as a function of the feedback resistance.

At high frequencies, the LED junction can be represented by a parallel RC circuit. The rate of photons produced, $n_r$, is related to the diode injection current $i_2$ by

$$n_r = \frac{i_2}{1 + \mu \tau_c}$$

(6)

where $\tau_c$ is the charge recombination constant and $q$ is the electronic charge. The detector current is given by

$$i_o = N\eta n_r q = N\eta (1 + j\omega \tau_c)$$

(7)

where $\eta (\omega) = \eta (1 + j \omega \tau_c)$. The high-frequency gain of the amplifier can be found using (2) (3) and (5) with $\eta$ replaced by $\eta (\omega)$. Assuming the LED resistance to be small then

$$\beta = R_f/(R_f + R_2)$$

and consequently

$$A_0 = \frac{N\eta (\omega)}{1 - N\eta (\omega) R_f / R_1} = \frac{N\eta (\omega)}{1 - N\eta (\omega) R_f / R_1}$$

(8)

The 3 dB bandwidth is thus given by

$$1 + 2 \pi R_f / \tau_c$$

and the gain-bandwidth product is $N\eta \tau_c$. This result is consistent with positive feedback, i.e. an increase in low-frequency gain is accompanied by a corresponding reduction in bandwidth. While the GB product is independent of feedback, it is proportional to $N$ which can be increased by using more LEDs in the array.

If wide bandwidth is of primary concern, negative feedback can be used. To implement negative feedback in the photon coupled amplifier of FIG. 1, the polarities of the detectors are reversed and the negative bias supply (V−) is replaced by a positive bias supply. This causes the detector current ($i_b$) to reverse in direction. The analysis of the negative feedback amplifier follows the same lines as for positive feedback except that $i_b$ is now given by $-N\eta_i$, indicating a change of direction of current flow. In comparison with the positive feedback structure, $-\eta$ is assigned instead of $\eta$ to equations (1) (3) to give the current gain of the negative feedback amplifier, i.e.

$$A_i' = \frac{i_b}{i_d} = \frac{A_0'}{1 - \frac{R_f}{R_1}}$$

(9)

where $A_0'$ is the open-loop current gain given by

$$A_0' = -N\eta \left( \frac{R_1}{1 + \frac{R_f}{R_1}} \right)$$

(10)

and $\beta'$ is the feedback factor given by

$$\beta' = \frac{R_2}{R_f + R_2} \left( 1 - \frac{R_f}{R_1} \right)$$

(11)

For large values of $R_1$ and $R_2$, the open-loop gain reduces to $A_0' = N\eta$ while the feedback factor becomes $\beta' = R_f/(R_f + R_2)$ which is equal to $\beta$ in the case of positive feedback. Therefore the current gain is

$$A_i' = \frac{i_b}{i_d} = \frac{-N\eta}{1 - \frac{R_f}{R_1}}$$

(12)

It can readily be shown that the magnitude of the amplifier gain will drop to $N\eta/(1 + N\eta B)$ while the bandwidth increases to $(1 + N\eta B)\tau_c$, with GB again given by $N\eta \tau_c$. By stacking a number of high quantum efficiency LEDs in a negative feedback array amplifier, the bandwidth of the...
amplifier can be made to substantially exceed the intrinsic bandwidth (1/\(f_T\)) of a single LED.

The noise spectral density of the detector current consists of two components: the transmitted diode array current noise and the photon partition noise. The partition noise arises from the Bernoulli photon detection process due to nonideal quantum efficiency. The photon current noise, on the other hand, is the sum of all the random events which independently modulate the injection current of the LED. It consists of the thermal noise of the circuit resistances, the short noise to bandwidth (1/\(T_c\)) of a single LED.

The photon current noise, on the other hand, is the sum of all the random events which independently modulate the injection current of the LED. It consists of the thermal noise of the circuit resistances, the short noise to bandwidth (1/\(T_c\)) of a single LED.

The partition noise is given by

\[
<\xi_p^2> = N(n-1)\tau d n + \frac{N(n-1)}{q} \phi_{n, q}^2
\]

where \(q\) is the electronic charge. The LED bias current, and

\[
<\xi_p^2> = (N(n-1)\tau d n + \frac{N(n-1)}{q} \phi_{n, q}^2)
\]

is the mean square shot noise current spectral density. Standard noise analysis of the circuit of FIG. 4 gives the transmitted photon current noise due to thermal and shot noise sources:

\[
<\xi_T^2> = \left( \frac{N(n-1)}{1-Nn} \right) \left( \frac{2W}{\mu} \right) <\xi_p^2>
\]

where

\[
a = \frac{n_1 R_1 - \tau d n}{n_1 (R_1 + R_2)} = \frac{\tau d n}{n_1 R_1}
\]

\[
W = \frac{\tau d n}{n_1 R_1} + N \left( \frac{\tau d n}{n_1 R_1} \right)^2 \left( \frac{\tau d n}{n_1 R_1} + \frac{\mu}{2} \right)
\]

The total output noise current is given by

\[
<\xi_{out}^2> = <\xi_T^2> + <\xi_p^2>
\]

where

\[
U = 1 + \frac{N(n-1)}{1-Nn} \left( \frac{R_2}{R_1 + R_2} \right)
\]

The second term in (17) represents the total effect of the partition noise, i.e., it consists of the open-loop partition noise and the returned partition noise through the feedback loop. Similar circuit analysis yields the output signal current

\[
I_{out} = \frac{V_j}{R_j} \left( 1 - \frac{N(n-1)}{1-Nn} \right)
\]

where \(V_j\) is the input signal voltage. The output signal-to-noise ratio is therefore given by

\[
SNR = \frac{<\xi_T^2>}{<\xi_p^2> + \left( \frac{2W}{\mu} \right) + \frac{1-n}{Nn} \left( 1 - Nn \phi_{n, q}^2 \right)^{-1}}
\]

The input signal and noise are defined as the LED modulation currents due to the signal source \(V_j\) and the thermal noise of the source resistance \(R_s\), respectively. Thus the input SNR is given by

\[
SNR = \frac{<\xi_T^2>}{<\xi_p^2> + \left( \frac{2W}{\mu} \right) + \frac{1-n}{Nn} \left( 1 - Nn \phi_{n, q}^2 \right)^{-1}}
\]

where \(\phi_{n, q} = \frac{q}{2GL}\), \(\mu\) being the ideality factor.

Finally, the noise figure of the feedback amplifier is given by

\[
F = \frac{SNR}{SNR_{in}} = \left( \frac{\mu R_s}{2 \phi_{n, q}} \right) \left( \frac{1-n}{Nn} \left( 1 - Nn \phi_{n, q}^2 \right)^{-1} \right)
\]

To compare the noise figure of the feedback amplifier with the open-loop case \(R_f\) and \(R_s\) are assumed for simplicity to be large compared with \(N\phi_{n, q}\) which leads to a \(\alpha = R_f/(R_f + R_s)\).

\[
F = 1 + \frac{R_s}{\phi_{n, q}} \left( \frac{\tau d n}{n_1 R_1} + \frac{\mu}{2} \right)
\]

That is, under ideal bias conditions, the noise figure of the amplifier is independent of positive feedback. This is equivalent to saying that the output SNR of the amplifier is not affected by the applied feedback (as the input SNR may be fixed at a constant).

When \(\eta = 1\), the last term of (24) vanishes and the noise figure decreases as \(R_s\) increases. When the quantum efficiency is less than unity, the same term will contribute to the rise of the noise figure for large source resistance. Differentiating (24) with respect to \(R_s\) gives the optimum source resistance for minimum noise figure.

\[
R_s = N\phi_{n, q} \left( \frac{2}{1-\eta} \right) \left( \frac{\tau d n}{n_1 R_1} + \frac{\mu}{2} \right)
\]

The corresponding minimum noise figure is then given by

\[
F_{(min)} = 1 + 2 \left( \frac{1-\eta}{\eta} \right) \left( \frac{\tau d n}{n_1 R_1} + \frac{\mu}{2} \right)^2
\]

which is independent of the number of light emitters used. In practice, finite bias resistances will increase \(F_{(min)}\) slightly.

For the purposes of verification of the operation of the photon coupled amplifier of FIG. 1, a photon coupled amplifier based on a Hamamatsu type L2656 AlGaAs/GaAs Heterojunction LED was modelled in Pspice using known techniques. The parameters used for the LED were:

- \(\tau = 34,400\) ns (at 1 mA), \(\tau = 3,600\) ns, \(\mu = 1,323\) and \(\tau = 9,297\) μs.
- With \(V_j = 1\) μV, \(R_1 = 160\) Ω, \(R_2 = 100\) kΩ, \(N = 1\) and \(\eta = 0.95\), the current gain of the amplifier is obtained as shown in FIG. 5. Curves (a), (b) and (c) are the current gains with the feedback resistance equal to 30, 15 and 7.5 kΩ respectively. It is clearly demonstrated that, as positive feedback is increased, the current gain goes up and the bandwidth decreases correspondingly. FIG. 6 shows the frequency spectra of the output signal and noise currents of the amplifier with the same parameters used to obtain curve (b) in FIG. 5. With constant input voltage, the 3-dB bandwidth of the output current spectrum is wider than that of the current gain response because the input current \(i_s\) increases at high frequencies due to the shunting effect of the LED capacitance. The output SNR at low frequencies can be
obtained directly from the spectral plots which is 54.25 dB and the input SNR obtained from (21) is 55.77 dB, thus giving a noise figure of 1.52 dB which agrees with calculation by (22).

FIG. 7 shows the noise figure of the positive feedback amplifier with a single LED in the array (N=1). The results are calculated as a function of the source resistance using the general expression (22) with R1=R2=100 kΩ. Curve (a) is for the ideal case of η=1. The noise figure decreases monotonically to 0 dB as the source resistance increases. This is due to high impedance suppression of the shot noise in the LED. Approximately the same noise figure is obtained when the feedback factor is varied between 0 and 1.

Curves (b) and (c) are the noise figures obtained for η=0.95 and 0.9, respectively. In both cases, increasing the source resistance no longer reduces the amplifier noise figure monotonically. As the source resistance increases, the shot noise is suppressed as before but the partition noise which increases as η decreases, becomes dominant hence causing an increase in noise figure. As the source resistance becomes very small, the shot noise current through the LED gradually increases, again causing an increase in noise figure. A minimum noise figure occurs at some intermediate value of the source resistance. For example in curve (b), the minimum noise figure is 1.15 dB and the optimum value of R1 is 160Ω which agrees with calculation using (25). FIG. 7 also shows that the noise figure of the feedback amplifier improves as η increases. This is obviously due to the reduction of partition noise in the amplifier at high quantum transfer efficiency.

Referring to FIG. 8, the current modulator formed by Vp and R2 in FIG. 1 is replaced by an input light detecting system, ILD. The current generated by ILD when light falls on it is used to modulate the array of light emitters. This current is amplified in the same way as in the form of the invention shown in FIG. 1, with the TDA bias and OLD bias currents provided in the same way. Resistance R3 and voltage supply Vb provide DC bias current for the output light emitting array, OTDA.

The output current from OLD is used to modulate the current in the laser diode or light emitting array, OTDA. Capacitor C3 couples the signal from OLD to OTDA. Impedances Z1 and Z2 together determine the feedback factor β. The circuit functions as a low noise light amplifier.

Another embodiment of the invention is shown in FIG. 9. In FIG. 9 the LED array labelled TDA shown in FIG. 1 consists of a single light emitting diode or laser diode and the detecting system consists of a single detector.

Providing the quantum transfer efficiency, otherwise called the open-loop current gain, between the light emitting element and the detector is greater than one half, the closed-loop current gain of this configuration can exceed unity. In this form of the invention a current gain exceeding unity can thus be obtained by applying positive current feedback from a single detector to a single emitter.

Referring to FIG. 9, voltage source Vp, with resistance Rg generates current i0 to drive the light emitter LE through coupling capacitor C. The DC bias current for the semiconductor light emitting junction is provided through resistance R1 from the positive voltage power supply V+. Semiconductor light detector LD is optically coupled to light emitter LE to provide open-loop current transfer ratio (ηp=i1/i0) of at least one half. DC bias current for the semiconductor detector junction is obtained by R1 from power supply V-. Positive current feedback is provided via capacitor C. The output signal may be measured in the form of a voltage drop across impedance Z0.

The use of high impedances for R1, R2 and Z1 is necessary to obtain large suppression of the diode shot noise current which may otherwise flow through the light emitting element and be amplified. Almost all the output current i0 then flows back via C through the light emitter LE and the closed loop current gain is then approximately equal to 1/η(1−η). Photon partition noise is minimised by arranging the open-loop current transfer ratio, otherwise called the quantum transfer efficiency, to be as close to unity as possible.

The circuit of FIG. 9 can also be used as the basis of a lightwave amplifier and receiver by inputting light directly to detector LD.

It will be apparent that the best performance of the photon coupled circuit according to this invention is achieved using a single mode laser diode, resonant cavity LED, or micro cavity laser diode or LED with the highest possible quantum efficiency. Currently available LEDs have an external efficiency at room temperature in excess of 85%.

We claim:

1. A sub-shot noise, optoelectronic amplifier circuit comprising a high quantum efficiency semiconductor light emitter: means to provide signal input current and bias current to said light emitter, the light output of said light emitter being controlled by the current flowing therein; a high quantum efficiency semiconductor light detector optically coupled with low photon losses to said light emitter to provide an overall quantum transfer efficiency in excess of 0.5; means to provide a bias voltage to said light detector, the current flowing in said light detector being controlled by the light received from said light emitter and used to provide an output signal; means to isolate a portion of the alternating current flowing in said detector and to feed this portion of the alternating current back to reinforce or oppose the signal current flowing in the light emitter.

2. A circuit as claimed in claim 1 wherein said feedback is positive to reinforce the signal current flowing in the light emitter and thus to provide increased current gain.

3. A circuit as claimed in claim 1 wherein said feedback is negative to oppose the signal current flowing in the light emitter and thus to provide stable gain with increased bandwidth.

4. A circuit as claimed in claim 1 wherein the light emitter comprises an array of semiconductor light emitters.

5. A circuit as claimed in claim 4 wherein said array comprises a plurality of semiconductor laser diode or light emitting diode junctions electrically connected in series and/or in parallel, driven by a high impedance signal current source.

6. A circuit as claimed in claim 4, wherein the circuit has an open loop gain that is much greater than 1 to which negative feedback is applied to increase the bandwidth.

7. A circuit as claimed in claim 1 wherein the light emitter is a single semiconductor device.

8. A circuit as claimed in claim 7 wherein said semiconductor device comprising the light emitter is a single light emitting diode, single light emitting junction, single laser diode or single laser diode junction.

9. A circuit as claimed in claim 7 wherein the quantum transfer efficiency of said coupling is at least 75%.

10. A circuit as claimed in claim 1 wherein the light detector comprises one or more back biased photodetectors operating in photoconductive mode.

11. A circuit as claimed in claim 1 wherein the lowest optical coupling is achieved by close physical positioning.

12. A circuit as claimed in claim 1 wherein the least optical coupling is achieved by using an optical wave guide or low loss light pipe.
13. A circuit as claimed in claim 11 wherein said circuit is in the form of an optoelectronic integrated circuit.

14. A circuit as claimed in claim 1 further comprising a second semiconductor light detector, the output current of said second light detector modulating the current flowing in said light emitter.

15. A circuit as claimed in claim 14 wherein said second light detector is adapted to receive an input light signal from an external source and the current flowing in the light detector is modulated by the input light signal.

16. A circuit as claimed in claim 15 further comprising a light output device driven by the amplified current from said light detector.

17. A circuit as claimed in claim 16 wherein said light output device comprises one or more semiconductor laser diode or light emitting junctions.
Abstract - This paper discusses a new method of optoelectronic amplification which utilises an array of one or more series connected light emitters from which the light is collected by closely coupled PIN diode detectors with high quantum efficiency. Positive current feedback is applied to the opto-coupler in order to increase the current gain of the system. It is shown that the gain-bandwidth product of the system is proportional to the number of light emitters used. Current, voltage and power gains of the opto-amplifier are derived and typical results shown. The system noise becomes very small if devices of high current transfer efficiency are used. Further, the noise figure is found to be independent of the positive feedback applied. For a given system, there is a minimum noise figure and it is independent of the number of light emitters used in the array. With a transfer efficiency of 0.95, a single-light emitter opto-amplifier can provide a current gain of 14.7 dB with a noise figure of about 1.5 dB.

I. INTRODUCTION

The use of a tandem array of light emitting semiconducting junctions for current and lightwave amplification has been proposed by Edwards [1]. Noise and gain measurements of a prototype open loop LED array amplifier using this principle have been reported in [2,3]. This paper discusses a new method of optoelectronic amplification which utilises an array of light emitters and light detectors coupled together with high quantum efficiency to which positive current feedback is applied.

The new method has several advantages. One of these is that the shot noise in the system is suppressed below the normal level by the use of high impedances and high quantum efficiency devices. Another is that the amplification is accomplished without the use of active electronic elements such as transistors. The incorporation of positive feedback allows high closed loop gains to be realised while retaining the low quantum noise characteristics of the open loop form of the array amplifier. The proposed opto-coupler amplifier, or opto-amplifier in short, can also be implemented with a single emitter and detector providing the current transfer ratio (quantum efficiency) is greater than one half. The technique forms the basis of a number of functional devices such as low noise current amplifiers, low noise light amplifiers and optical bistable devices which may be implemented in optoelectronic integrated circuit form.

Fig. 1 shows the circuit of an opto-amplifier. An array of series connected semiconductor junction LEDs each of high quantum efficiency is shown modulated by a voltage source \( v_g \) with internal impedance \( R_g \). It is connected to the array through a blocking capacitor \( C_i \) and...
the DC bias current for the array is provided through resistance $R_1$. The light from this array is efficiently collected by the PIN diode detector which is biased through resistance $R_2$.

A fraction of the current output from the detector is fed back from point $B$ through the feedback resistance $R_f$ and the blocking capacitor $C_2$ to the input of the light emitting array at point $A$ to provide positive current feedback. The amplified current output is available between points $A$ and $B$ or between $B$ and $G$.

II. LOW FREQUENCY CHARACTERISTICS

In applying positive feedback to the opto-amplifier we expect to find that the gain in the system can be increased significantly whereas the gain-bandwidth product remains the same as that of the open-loop opto-coupler. The noise behaviour of the positive feedback amplifier will depend on the quantum efficiency of the optoelectronic conversion and the noise figure is expected to improve as the quantum efficiency is increased. In the following analysis, we assume the LED model used by [4,5] which has been found useful for both small signal and noise performance evaluations.

Let $N =$ number of LEDs in series

- $\eta =$ differential quantum efficiency per device
- $r_d =$ differential resistance of the LED
- $r_s =$ bulk resistance of the LED
- $r_{sd} = r_s + r_d$
- $R_1 =$ bias resistance for the LED array
- $R_2 =$ includes the bias and load resistance at the output
- $R_f =$ feedback resistance (or impedance in general)

The low-frequency small signal equivalent circuit for the system is as shown in Fig. 2. We have assumed identical LEDs and that the detector generates an output current given by $i_o = N(\eta i_2)$, where $i_2$ is the signal current flowing through the LED array. The impedance of the detector is assumed to be high compared with $R_f$ or $R_2$. With $i_1 =$ the current from the signal source, it can readily be shown that the closed-loop current gain of the opto-amplifier is given by

$$A_i = \frac{i_o}{i_1} = \frac{A_o}{1 - A_o \beta} \quad (1)$$

where $A_o =$ the open-loop current gain given by

$$A_o = N\eta \left( \frac{R_i}{R_1 + N r_{sd}} \right) \quad (2)$$

and $\beta$ is the feedback factor given by
For stability reasons, we must operate the amplifier such that the loop gain \( A_0 \beta \) is less than unity. If \( R_1 \) and \( R_2 \) are large compared with the device resistance \( r_{sd} \), the open loop gain becomes \( N \eta \) and the feedback factor simplifies to \( R_2/(R_f+R_2) \).

The input impedance of the opto-amplifier is given by

\[
R_{in} = \frac{v_i}{i_i} = \frac{A_i r_{sd}}{\eta} \tag{4}
\]

Note that the effect of the LED bias resistance \( R_1 \) on the input resistance is already included in the expression for the current gain.

It can also be shown that the voltage gain of the feedback amplifier is given by

\[
A_v = \frac{v_o}{v_i} = 1 + \frac{R_f R_2}{R_f + R_2} \left( \frac{\eta}{r_{sd}} \right) \tag{5}
\]

It is of interest to note that while the current gain is proportional to the number of LEDs in the array, the voltage gain is independent of \( N \). When \( R_f = \infty \), \( A_v \) reduces to the open loop value of \( \eta R_2/r_{sd} \). For a given circuit, the voltage gain increases as the LED bias current increases. This is because the device resistance decreases as the bias current goes up. The apparent power gain of the feedback amplifier is given by \( P_o/P_i = A_v A_i \), where \( P_i = v_i i_i \) is the input signal power and \( P_o = v_o i_o \) is the output signal power from the detector. Fig. 3 shows typical variations of current gain, voltage gain and power gain of the amplifier as a function of the feedback resistance.

III. **BANDWIDTH**

At high frequencies, the LED junction can be represented by a parallel RC circuit. The rate of photons produced, \( n_p \), is related to the diode injection current \( i_2 \) by

\[
n_p = \frac{i_2/q}{1 + j \omega \tau_c} \tag{6}
\]

where \( \tau_c \) is the charge recombination constant and \( q \) is the electronic charge. The detector current is given by

\[
i_o = N \eta n_p q = N \eta (\omega) i_2 \tag{7}
\]
where \( \eta(\omega) = \eta/(1+j\omega \tau_c) \). The high frequency gain of the amplifier can be found using (2), (3) and (5) with \( \eta \) replaced by \( \eta(\omega) \). Assuming the LED resistance to be small we have \( \beta = R_2/(R_f+R_2) \) and consequently

\[
A_i = \frac{N\eta(\omega)}{1 - N\eta(\omega) \beta} = \frac{N\eta}{1 - N\eta \beta + j\omega \tau_c}
\]  

(8)

The 3 dB bandwidth is thus given by \((1-N\eta \beta)\tau_c\) and the gain-bandwidth product is \(N\eta/\tau_c\). This result is consistent with positive feedback, ie, an increase in low frequency gain is accompanied by a corresponding reduction in bandwidth. While the GB product is independent of feedback, it is worth noting that this figure of merit is proportional to \(N\) which can be increased by using more LEDs in the array. If wide bandwidth is of primary concern, negative feedback will be more appropriate. It can readily be shown that, with negative feedback, the amplifier gain will drop to \(N\eta/(1+N\eta \beta)\) while the bandwidth increases to \((1+N\eta \beta)/\tau_c\), with GB again given by \(N\eta/\tau_c\). By stacking a number of high quantum efficiency LEDs in a negative feedback array amplifier, the bandwidth of the amplifier can be made to substantially exceed the intrinsic bandwidth \((1/\tau_c)\) of a single LED. The general characteristics of the negative feedback structure will be given in a separate paper.

IV. NOISE FIGURE

The noise spectral density of the detector current consists of two components: the transmitted diode array current noise and the photon partition noise [1]. The partition noise arises from the Bernoulli photon deletion process due to nonideal quantum efficiency. The photon current noise, on the other hand, is the sum of all the random events which independently modulate the injection current of the LED. It consists of the thermal noise of the circuit resistances, the shot noise in the LEDs and the feedback of any current fluctuations in the detector, including partition noise. A low-frequency noise equivalent circuit of the positive feedback opto-amplifier is as shown in Fig. 4. \(R_m\) represents the parallel combination of \(R_g\) and \(R_f\), and the equivalent input voltage is given by \(V_m = V_g(R_m/R_g)\). The RMS shot noise voltage of each LED is represented by \(v_{sn} = i_s r_d\) where \(i_s\) is the RMS shot noise current. The thermal noise voltage associated with each of the resistances \(R_m, R_f, R_2\) and \(N_r s\) has been omitted from the figure for the sake of clarity.

It has been shown that the partition noise is given by [1]

\[
<i_p^2> = N\eta(1-\eta)2qI = N\eta(1-\eta)<i_s^2>
\]  

(8)

where \(q\) is the electronic charge, \(I\) the LED bias current, and \(<i_s^2>(=2qI)\) is the mean square shot noise current spectral density. Standard noise analysis of the circuit of Fig. 4 gives the transmitted photon current noise due to thermal and shot noise sources:
\[
<i_a^2> = \left( \frac{N\eta}{1-N\eta\alpha} \right)^2 \frac{2W}{\mu} <i_s^2>
\]

where \( \alpha = \frac{nR_2 - rd \frac{r_d}{nR_m}}{n(R_f + R_2)} \)

\[
W = \frac{rd}{R_m} + \frac{rd}{R_f + R_2} + N \left( \frac{rd}{R_p} \left( \frac{r_s}{rd} + \frac{\mu}{2} \right) \right)
\]

The total output noise current is given by

\[
<i_{no}^2> = <i_a^2> + U^2 <i_p^2>
\]

where \( U = 1 + \frac{n\eta}{1-N\eta\alpha} \frac{R_2}{R_f + R_2} \)

The second term in (12) represents the total effect of the partition noise, i.e., it consists of the open-loop partition noise and the returned partition noise through the feedback loop.

Similar circuit analysis yields the output signal current

\[
i_{so} = \frac{V_g}{R_g} \frac{N\eta}{1-N\eta\alpha}
\]

where \( V_g \) is the input signal voltage. The output signal-to-noise ratio is therefore given by

\[
SNR_o = \left( \frac{V_g/R_g}{<i_s^2>} \right)^2 \frac{2W}{\mu} + \frac{1-N\eta\alpha}{N\eta} (1-N\eta\alpha)^2 U^2 \]

We define the input signal and noise as the LED modulation currents due to the signal source \( V_g \) and the thermal noise of the source resistance \( R_g \), respectively. Thus the input SNR is given by

\[
SNR_i = \frac{V_g^2}{4kTR_g} = \frac{(V_g/R_g)^2}{<i_s^2>} \left( \frac{\mu R_g}{2rd} \right)
\]

where \( rd = \mu kT/qI, \mu \) being the ideality factor.

Finally, the noise figure of the feedback amplifier is given by

\[
F = \frac{SNR_i}{SNR_o} = \left( \frac{\mu R_g}{2rd} \right) \left[ \frac{2W}{\mu} + \frac{1-N\eta}{N\eta} (1-N\eta\alpha)^2 U^2 \right]
\]
It is of interest to compare the noise figure of the feedback amplifier with the open loop case. To simplify discussion, we assume $R_1$ and $R_2$ to be large compared with $N_{rd}$ which leads to 

$$a = \frac{R_2}{R_f + R_2}, U = \frac{1}{1 - N_\eta \alpha}$$

and

$$W = \frac{r_d}{R_g} + N \left( \frac{r_d}{R_g} \right)^2 \left( \frac{r_s}{r_d} + \frac{\mu}{2} \right)$$

For the open loop amplifier, $R_f \gg R_2$ so that $a = 0$ and $U = 1$. When feedback is used, $R_f \ll R_2$ so that $a = 1$ and $U = \frac{1}{1 - N_\eta}$. In both cases (17) simplifies to

$$F = 1 + \frac{N_{rd}}{R_g} \left( \frac{r_s}{r_d} + \frac{\mu}{2} \right) + \left( \frac{1 - \eta}{\eta} \right) \frac{(\mu R_g)}{2N_{rd}}$$

That is, under ideal bias conditions, the noise figure of the amplifier is independent of positive feedback. This is equivalent to saying that the output SNR of the amplifier is not affected by the applied feedback (as the input SNR may be fixed at a constant).

When $\eta = 1$, the last term of (18) vanishes and the noise figure decreases as $R_g$ increases. When the quantum efficiency is less than unity, the same term will contribute to the rise of the noise figure for large source resistance. Differentiating (18) gives the optimum source resistance for minimum noise figure:

$$R_g = N_{rd} \sqrt{\frac{2\eta}{(1-\eta)\mu} \left( \frac{r_s}{r_d} + \frac{\mu}{2} \right)}$$

The corresponding minimum noise figure is then given by

$$F_{(\text{min})} = 1 + 2 \sqrt{\left( \frac{1 - \eta}{\eta} \right) \left( \frac{r_s}{r_d} + \frac{\mu}{2} \right)}$$

which is independent of the number of light emitters used. In practice, finite bias resistances will increase $F_{(\text{min})}$ slightly.

V. RESULTS

The parameters of the LED used in this study are: $r_d = 34.4 \ \Omega$ (at 1 mA), $r_s = 3.6 \ \Omega$, $\mu = 1.323$ and $\tau_c = 0.297 \ \mu s$. The opto-amplifier is modelled in PSpice using the method of [5]. With $V_g = 1 \ \mu V$, $R_g = 160 \ \Omega$, $R_1 = R_2 = 100 \ \text{k}\Omega$, $N = 1$ and $\eta = 0.95$, the current gain of the amplifier is obtained as shown in Fig. 5. Curves (a), (b) and (c) are the current gains with the feedback resistance equal to 30, 15 and 7.5 k\Omega, respectively. It is clearly demonstrated that, as positive feedback is increased, the current gain goes up and the bandwidth decreases correspondingly. Fig. 6 shows the frequency spectra of the output signal and noise currents of the amplifier with the same parameters used to obtain curve (b) in Fig 5. Note that, with
constant input voltage, the 3-dB bandwidth of the output current spectrum is wider than that of the current gain response because the input current $i_1$ increases at high frequencies due to the shunting effect of the LED capacitance. The output SNR at low frequencies can be obtained directly from the spectral plots which is 54.25 dB and the input SNR obtained from (16) is 55.77 dB, thus giving a noise figure of 1.52 dB which agrees with calculation by (17).

Fig. 7 shows the noise figure of the positive feedback amplifier with a single LED in the array ($N=1$). The results are calculated as a function of the source resistance using the general expression (17) with $R_1=R_2=100$ kΩ. Curve (a) is for the ideal case of $\eta=1$. The noise figure decreases monotonically to 0 dB as the source resistance increases. This is due to high impedance suppression of the shot noise in the LED. Approximately the same noise figure is obtained when the feedback factor is varied between 0 and 1.

Curves (b) and (c) are the noise figures obtained for $\eta=0.95$ and 0.9, respectively. In both cases, increasing the source resistance no longer reduces the amplifier noise figure monotonically. As the source resistance increases, the shot noise is suppressed as before but the partition noise, which increases as $\eta$ decreases, becomes dominant hence causing an increase in noise figure. As the source resistance becomes very small, the shot noise current through the LED gradually increases, again causing an increase in noise figure. A minimum noise figure occurs at some intermediate value of the source resistance. For example, in curve (b), the minimum noise figure is 1.5 dB and the optimum value of $R_g$ is 160 Ω which agrees with calculation using (19). Fig. 7 also shows that the noise figure of the feedback amplifier improves as $\eta$ increases. This is obviously due to the reduction of partition noise in the amplifier at high quantum transfer efficiency.

VI. CONCLUSIONS
A novel method of optoelectronic amplification with positive feedback has been discussed. The gain, bandwidth and the noise characteristics of the amplifier have been analysed. The gain-bandwidth product of the amplifier is given by $A/1/T_c$ which is significant in that it can be controlled by the number of light diodes used in the array. If it is implemented in the form of optoelectronic integrated circuit, it is possible to have an opto-amplifier with very high GB product. We have noted that the opto-amplifier with positive feedback provides increased current and power gains with the corresponding decrease in bandwidth. If an increase in bandwidth is required, then negative feedback can be applied as in conventional electronic amplifiers.

The noise performance of the opto-amplifier has been studied in terms of its noise figure which is commonly used in communication systems. Significantly, we have shown that the noise figure is independent of the positive feedback applied provided that the diode bias resistances are large compared with the differential resistance of the diode array.
Furthermore, low noise figure can be obtained with high quantum transfer efficiency. For each value of $\eta<1$, we have shown that there is an optimum source resistance at which the noise figure is minimum. This optimum condition is easily realized in practice with the use of simple impedance transformation. Since such high efficiency LEDs and laser diodes are feasible, the proposed low-noise high gain optoelectronic amplifier is achievable in the foreseeable future.

**ACKNOWLEDGMENT**

This work is supported by the Australian DEET Targeted Institutional Link and ARC Grants. Assistance by H B Sun and T T Zhang in LED parameter measurements is also acknowledged.

**REFERENCES**


Captions to figures

Fig. 1 LED array opto-coupler amplifier with positive feedback.

Fig. 2 AC equivalent circuit of the opto-amplifier with each LED represented by $r_s$ and $r_d$.

Fig. 3 (a) Current gain, (b) voltage gain, and (c) power gain of the positive feedback amplifier with $N=8$, $\eta=0.15$, and $R_f=R_2=2 \, k\Omega$.

Fig. 4 AC equivalent circuit of the opto-amplifier for noise calculation.

Fig. 5 Current gain of the opto-amplifier with feedback resistance equal to (a) 30, (b) 15 and (c) 7.5 $k\Omega$.

Fig. 6 Frequency spectra of the amplifier output currents in dB$\mu$A: (a) signal and (b) noise, with $N=1$, $\eta=0.95$, $V_g=1 \, \mu V$, $R_g=160 \, \Omega$, $R_1=R_2=100 \, k\Omega$ and $R_f=15 \, k\Omega$.

Fig. 7 Noise figure variation with the source resistance at different quantum efficiencies: (a) $\eta=1$, (b) $\eta=0.95$ and (c) $\eta=0.9$. 
Fig. 1
Fig. 2
Fig. 3
Fig. 4
Fig. 5
Fig. 6
Modelling of a DH LED using one-port impedance measurements

W. N. CHEUNG, G. N. FRENCH and P. J. EDWARDS
Advanced Telecommunications and Quantum Electronics Research Centre,
University of Canberra, P.O. Box 1, Belconnen, ACT 2616, Australia

(Received 7 July 1998; accepted 13 October 1998)

We review the linkage between the rate equation of the LED and its equivalent circuit. We describe a procedure to extract parameter information from the input impedance of an LED that is measured over a broad frequency range using a network analyser. We confirm that the method, which utilizes both magnitude and phase information of the device impedance, enables us to determine the equivalent circuit of the LED with good reliability. The device model is shown to be valid over a frequency range of up to 100 MHz.

1. Introduction

Differential carrier lifetimes in semiconductor light-emitting diodes and laser diodes may be obtained from single port impedance measurements [1–3]. Electrical characterization involving DC parameter and RF impedance measurement also yields useful information about the small signal equivalent circuit topology and parameter values [4–6]. It is well known that there is close correspondence between the parameters of the equivalent circuit and the rate equation description of the LED. Electrical characterization can therefore yield information on both the small signal modulation response and the intrinsic noise of the device. One of the important applications of an accurate equivalent circuit of an LED is in determining the noise and modulation characteristics of photon coupled circuits. An accurate physical model of the LED is also useful for predicting the transfer performance in optocoupler amplifiers [7].

While direct optical transfer measurement will yield important characteristics, such as quantum transfer efficiency as well as the modulation bandwidth, one-port impedance measurement is simple and reliable. This paper describes a procedure for deriving the model parameters from the one-port impedance measurements of the LED made using a network analyser. The salient feature of our procedure is the incorporation of the phase information as well as the magnitude of the device impedance in the modelling process.

2. Equivalent circuits

In the LED, since electrons and holes are injected in pairs and recombine in pairs, it is enough to consider the rate equation for only one type of charge carrier. Let $N$ be the number of electrons, say. The rate of increase of the number of electrons in the active region is given by [8]:

$$\frac{dN}{dt} = \frac{1}{q} \left( \frac{N}{T} - \frac{N_m}{\tau_e} \right),$$

(1)

where $q$ is the electronic charge, $I$ is the injected current, and $\tau_e$ is the recombination time constant. The first term on the right-hand side of (1) is the rate of electrons injected into the active region and the last term is the rate of electron–hole recombination, including radiative and non-radiative recombination.

Consider an injected current consisting of a DC bias and a sinusoidal modulation: $I = I_b + I_m \exp(j \omega_m t)$. Substituting this current into (1) and solving for the steady-state carrier number gives:

$$N(t) = N_b + N_m \exp(j \omega_m t),$$

(2)

where $N_b = I_b \tau_e / q$, and

$$N_m = \left( \frac{\tau_e}{q} \right) \frac{I_m}{I + j \omega \tau_e}.$$

(3)

Note that $N_b$ is the carrier number due to the bias current, and $N_m$ is the amplitude of the modulated carrier number. It is clear from (3) that the modulated amplitude decreases as the modulation frequency increases. Physically, this may be interpreted as that, at high modulation frequencies, the pumping current changes so rapidly that there is insufficient time for some of the charge carriers to recombine.

Rewriting (3) as

$$\frac{q N_m}{\tau_e} = \frac{I_m}{I + j \omega \tau_e}$$

(4)
makes it evident that the left-hand side of (4) is a current, which is directly responsible for carrier recombination. If we assume that the quantum efficiency is 100% so that recombination is entirely radiative, then this current may be regarded as the AC photon generating current, $I_p$. Further, (4) also expresses the frequency response of the device relating the photon current $I_p$ into the sinusoidal current modulation $I_m$. We note that the frequency response is identical to that of a single-pole parallel RC circuit, as shown in figure 1 (a). Writing (4) explicitly as a current we get

$$I_p = \sum \left( \frac{I}{R} \right) \frac{I_m R}{I + j\omega \tau_e} = \frac{V_p}{R}, \quad (5)$$

where $V_p$ is the voltage across the RC circuit. Hence the photon current is represented by the resistive current of the equivalent circuit. In terms of the complex frequency $s$, the photon current can be expressed formally as

$$I_p(s) = \frac{I_m(s)}{1 + s \tau_e}. \quad (6)$$

Next, we must find appropriate representation for $R$ and $C$ of the equivalent circuit. So far, all we know is that the RC time constant is equal to $\tau_e$. From junction current theory, we note that the differential resistance for small signals at low frequencies is given by $R_d = mV_T/I_b$, where $V_T$ is the thermal voltage, $m$ is the diode ideality factor and $I_b$ the bias current. This is applicable to the LED under the condition of low-frequency, small-amplitude modulation. Therefore, we have $R = R_d$ and, consequently, the parallel capacitance is given by $C_d = \tau_e/R_d$.

In addition to the ideal light emitting junction, a practical LED will also have a contact and cladding series resistance $R_s$. The single port equivalent circuit of a light-emitting diode or a sub-threshold laser diode is thus as shown in figure 1 (b). The contact and cladding series resistance $R_s$ and the differential resistance $R_d$ can be separately extracted from the DC current voltage device characteristic or, as shown below, from the low-frequency input impedance measurements of the LED.

The device time constant $\tau_e = R_s C_d$ will determine the operating bandwidth. The modulation and noise characteristics can be determined from the equivalent circuit parameters.

3. Modelling

The input impedance of the LED as modelled by the equivalent circuit is given by

$$Z = \frac{R_d}{I + s R_s C_d} = \frac{(R_s + R_d) I + s \omega \tau_e}{I + s \tau_e}, \quad (7)$$

where $\alpha = R_s/(R_s + R_d)$. It is clear that the input impedance function consists of a pole and a zero. The magnitude and phase of the impedance are given respectively by

$$|Z(j\omega)| = \sqrt{\frac{1 + (\omega \tau_e)^2}{1 + (\omega \tau_e)^2}} \tan^{-1}(\omega \tau_e) - \tan^{-1}(\omega \tau_e). \quad (8)$$

Essentially we require three independent equations to determine the three parameters of the equivalent circuit. (a) The sum of $R_s + R_d$ is obviously given by the impedance at low frequencies. (b) The carrier lifetime $\tau_e$ is related to the pole frequency of the impedance, which may be closely approximated by the 3 dB frequency of the magnitude response. The approximation is valid if $\alpha$ is small so that the zero of the impedance function has little effect on the 3 dB frequency. (c) The ratio of $R_s$ and $R_d$ can be determined from the phase response in the following way.

Taking the tangent of the phase angle $\phi$ and solving for $\alpha$ gives

$$\alpha = \frac{\theta + \tan \phi}{\theta - \theta^2 \tan \phi}, \quad (10)$$

where $\theta = \omega \tau_e$.

At a given frequency, we have the value of $\phi$ and $\theta$ so that we can calculate the constant $\alpha$. In practice, however, $\alpha$ may not be constant over the entire range of frequencies of interest due to measurement uncertainties. In the present case, we average the calculated value of $\alpha$ and use the average $\alpha$ to determine the ratio of $R_s/(R_s + R_d)$.

4. Measurement

Figure 2 shows the schematic of the set-up for the measurement of the input impedance of the LED with appropriate bias. Before making measurements, the network analyser is calibrated with a 50 $\Omega$ standard load, a short circuit, and an open circuit over a frequency range
Modelling of a DH LED using one-port impedance measurements

5. Parameter estimation

As an illustration of the modelling method, we describe in more detail the case with a bias current of 5 mA. From the magnitude curve we find \( (R_s + R_d) = 8 \, \Omega \) at low frequencies. The 3 dB frequency of the magnitude response is estimated to be 21 MHz, hence giving \( \tau_c = R_d C_d = 7.58 \, \text{ns} \). With the help of the phase response and the value of \( \tau_c \), we find the value of \( \alpha \) from (10) to be 0.24 which leads to \( R_s = \alpha (R_s + R_d) = 1.92 \, \Omega \). Finally, we find the remaining two parameters to be \( R_d = 6.08 \, \Omega \) and \( C_d = 1.25 \, \text{nF} \). The ideality factor is deduced to be 1.17.

Table 1 summarizes the results of parameter estimation for the LED sample for all three currents.

Using the estimated parameters, we calculated the input impedance of the device and plotted the magnitude and phase alongside the measured data, as shown in figures 5 and 6 for the case of \( I_b = 1 \) mA. Similar pairs of graphs are shown in figures 7–10 for bias currents of 5 and 10 mA, respectively. In each case, good agreement

<table>
<thead>
<tr>
<th>( I_b ) (mA)</th>
<th>( R_s + R_d ) (Ω)</th>
<th>( \tau_c ) (ns)</th>
<th>( \alpha )</th>
<th>( R_s ) (Ω)</th>
<th>( R_d ) (Ω)</th>
<th>( C_d ) (nF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30.5</td>
<td>10.3</td>
<td>0.058</td>
<td>1.77</td>
<td>28.7</td>
<td>0.359</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>7.58</td>
<td>0.24</td>
<td>1.92</td>
<td>6.08</td>
<td>1.25</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>7.58</td>
<td>0.4</td>
<td>2.0</td>
<td>3.0</td>
<td>2.53</td>
</tr>
</tbody>
</table>
between the model and measurement is observed over a broad frequency range up to the 3 dB frequency. As the frequency increases, the difference between the two sets of data becomes more evident. This discrepancy appears to be the result of high reflection from a low impedance. At high frequencies, the magnitude of the LED impedance is small so that it causes a large reflection coefficient to exist in the measuring system. As outlined in the Appendix, it is a feature of the Smith Chart that the error in the determination of impedance increases with the magnitude of the reflection coefficient.

6. Conclusions
We have described a modelling procedure for determining the small-signal parameters of a DH LED based on broad-band impedance measurements. These
Modelling of a DH LED using one-port impedance measurements

Figure 6. Measured and modelled impedance phase of LED at 1 mA.

Figure 7. Measured and modelled impedance magnitude of LED at 5 mA.

parameters satisfy the fundamental rate equation governing the recombination process in the active region of the semiconductor junction. The validity of the equivalent circuit is confirmed by the close agreement between the modelled and measured input impedance of the LED.

Appendix

Network analysers are essentially vector reflectometers, so that impedance estimates are based on transformations of the reflection coefficient (Γ) measured at the input port of the device. In general, the normalized...
Differentiating the above equation yields the following fractional change in $z$ as a function of the fractional change in $\Gamma$.

\[
\frac{dz}{z} = \left( \frac{2\Gamma}{1 - \Gamma^2} \right) \frac{d\Gamma}{\Gamma}.
\]  

(A2)
Equation (A2) clearly shows that when $F$ is real and large (i.e. when $F$ approaches $\pm 1$), a small error in the measurement of the reflection coefficient will give rise to a large error in the impedance of the device under test.

Acknowledgments

The project is supported by research grants from the Australian Research Council, DEETYA Provision of measuring facilities by the Commonwealth Department of Communication and the Arts, and assistance in measurements by Mr D Brooke are also acknowledged. We thank Mr Takemura of Hamamatsu Photonics KK for providing light-emitting diodes for evaluation purposes.

References


Quantum noise-suppressed operation of TJS and SQW ridge-waveguide laser diodes in photon transistor configurations


1Centre for Advanced Telecommunications and Quantum Electronics Research, University of Canberra ACT 2601 Australia; email: paule@ise.canberra.edu.au
2Hamamatsu Photonics K K, Hamamatsu, Shizuoka, Japan;
3Edward L. Ginzton Laboratory, Stanford University, Stanford CA 94305 USA.

Abstract - We report the first successful incorporation of shot noise-suppressed laser diodes in a closed-loop photon transistor amplifier configuration. We have measured and modelled the current gain, bandwidth and noise level of transverse junction stripe and quantum well laser diodes in open loop (optocoupler) and closed loop (common-emitter) photon transistor configurations. At liquid nitrogen temperatures we have directly measured 4.5 dB suppression of photonic noise below the full shot noise level.

A. Introduction

Recent work [1] has shown that cooled transverse junction stripe GaAs-AlGaAs lasers grown with high As:Ga ratios are 4-5 dB quieter than the standard quantum limit (full shot noise level) when strongly pumped. These devices may be used with advantage in photon coupled circuits such as open loop and closed loop optocouplers where critical applications require that circuit noise be suppressed below the full shot noise level [2,3].

Figure 1 shows the basic open- and closed-loop (photon coupled transistor) optocoupler configurations.

![Fig. 1: Basic optocoupler configurations showing the relation between the open loop (LHS) and the photon-coupled transistor (PCT) closed loop (RHS) configurations.](image)

The open loop (LHS) circuit is analogous to a common base transistor circuit with emitter to collector current gain \( \alpha = i_C/i_E = \eta \), the net quantum transfer efficiency between the laser diode-emitter and p-i-n diode collector circuits. With the anodes (or the cathodes) of the emitter and collector diodes electrically connected together, as in the RHS of Fig 1, we obtain the three terminal transistor configuration. It is evident that the resulting photon coupled transistor containing \( N \) series-connected light-emitting junctions will have current gain,

\[
\beta = i_C/i_B = N\eta/(1-N\eta) \tag{1}
\]
when operated in the common emitter mode with input to the base, B. Inspection of Fig. 1 and Eqn. 1, shows that the common base and common emitter modes may be regarded as open-loop and closed-loop configurations respectively, with unity positive current feedback.

One of the fundamental disadvantages of the photon coupled transistor (PCT) is that, unlike a conventional bipolar junction transistor in which sub-shot noise streams of holes or electrons are transported from emitter to collector, laser emission usually exhibits photonic noise well in excess of the full shot noise level. In order to achieve similarly quiet operation of a PCT it is necessary to suppress both the photonic shot noise and, as well, the photonic partition noise arising from photon losses between emitter and collector.

The collector current noise of a common base-connected (ie open loop), photon or bipolar transistor is the sum of the transmitted emitter noise (1st term) plus the photonic partition noise (2nd term):

$$\langle i_c(t)^2 \rangle = \alpha^2 \langle i_e^2(t) \rangle + \alpha (1-\alpha) \langle i_{sn}^2 \rangle \quad (2)$$

where the mean square emitter current noise $\langle i_e^2(t) \rangle$ in a laser-based optocoupler or PCT, usually exceeds the full emitter shot noise level $\langle i_{sn}^2 \rangle$ because of modal competition and other noise sources. Also, since the current coupling efficiency $\alpha$ is usually much less than unity, significant partition noise is introduced, unlike a bipolar transistor in which $\alpha \approx 1$.

We have succeeded for the first time in incorporating a shot noise-suppressed laser diode in a closed loop PCT configuration. We have lowered photonic noise by more than 4 dB below the full shot noise level with a corresponding increase in signal to noise ratio.

**B. Gain Measurements**

![Test jig for open and closed PCT amplifier measurements. Laser bias current is applied through resistance $R_b$ and detector (collector) bias through $R_c$. Shot noise reference not shown. Positive ac feedback is applied via $C_2$ with switch $S$ in the upper position. Signal currents $i_1$ and $i_2$ are deduced from voltage measurements made with the spectrum analyser and a calibrated oscilloscope.](image)

Fig. 3 shows the detector (collector) current with dc feedback applied from a single high efficiency pin diode detector (Hamamatsu S3994) to an experimental Hamamatsu single transverse junction stripe (TJS) laser diode cooled to liquid nitrogen temperature. The differential dc current gain is increased from $\alpha = 0.56$ (the slope of the OL curve) to $\beta = 1.26$
(the slope of the CL curve) as expected from Eqn. (1). Fig.4 shows the open and closed loop ac gain at frequency 5 MHz as a function of laser bias current at room temperature for an array of two series-connected single quantum well lasers grown and fabricated at the Australian National University. The 6 dB increase in gain following application of feedback is consistent with Eqn. 1 and a measured open loop current gain $A_o = 0.51$.  

![Graph](image1)

![Graph](image2)

Fig. 3: Detected photocurrent at 80 K from squeezed TJS laser in normal open loop (O/L) and close loop (C/L) modes

Fig.4: Current gain of PCT at closed loop (C/L) and open loop (O/L) condition.

Fig. 5 compares measured and Spice-modelled current gains of another SQW laser based PCT. The measured and modelled bandwidths are in satisfactory agreement.

![Graph](image3)

Fig.5: Current gain of PCT with frequency in close loop (C/L) and open-loop (O/L) condition.

C. Noise Measurements

Noise measurements were made with the test set up of Fig. 2 using a spectrum analyser and a LED array as a shot noise reference. We have not succeeded in suppressing the SQW laser noise below the full shot noise level. However, we have observed [2] as much
as 4.5 dB of shot noise suppression with liquid nitrogen-cooled Hamamatsu TJS lasers in open loop configurations and corresponding enhancement of signal to noise ratio. This is the highest noise suppression directly measured to date with these laser diodes [1,2]. Fig. 6 (2 dB/vertical scale division) shows typical 3dB shot noise suppression and signal to noise ratio enhancement in both open and closed loop PCT amplifier configurations and Fig. 7 shows a noise equivalent circuit which provides a satisfactory small signal and noise model.

D. Discussion

We have demonstrated shot noise-suppressed operation of photon coupled transistor current amplifiers both with and without positive current feedback. Their gain, bandwidth and noise levels are close to those predicted. However, before PCTs can rival the wide bandwidths and low noise levels of BJTs their quantum coupling efficiency must be raised close to unity.

Fig. 6: Spectra of closed loop (top); open loop (bottom) photon-transistor noise and 5 MHz signal levels; Full shot noise reference level: (middle trace).

Fig. 7: Proposed noise equivalent circuit of photon transistor, showing: emitter noise (\(i_d\)); partition noise (\(i_p\)); and thermal noise (\(v_t\)) sources.

Acknowledgements

We acknowledge financial support from the Australian Research Council and we thank Prof C. Jagadish and colleagues for growing and preparing SQW laser diodes for this project.

References

Sub-Poissonian Electronic and Photonic Noise Generation in Semiconductor Junctions*

Paul J. Edwards
Centre for Advanced Telecommunications and Quantum Electronics Research, University of Canberra, Canberra, ACT 2601, Australia.

Abstract
This paper addresses sub-Poissonian electronic and photonic noise generation in semiconductor junctions. Recent theoretical and technical advances in the understanding and generation of quantum noise-suppressed ('quiet') light have emphasised the links between photonic and electronic shot noise. Shot-noise suppression and single electron–photon control through the operation of the collective and single-electron Coulomb blockade mechanisms are described.

1. Introduction
As new fabrication technologies are developed to provide the increased switching speed, bandwidth and information packing density required of integrated electronic circuitry, the quantum effects associated with the quantisation of electronic charge and energy have become increasingly apparent. One of the consequences of this is the occurrence of 'granularity noise' (shot noise) in optical interconnects and mesoscopic electronic circuits.

Over the last decade, the quantum (non-classical) properties of light beams have been closely investigated and photonic noise suppression techniques with potential applications in metrology, communications and computing have been developed. Similar attention is now being paid to electronic 'quantum shot noise' and its minimisation in mesoscale and nanoscale circuits. It has now become apparent that close similarities exist between photonic shot noise, ballistic electron shot noise and diffusive electron shot noise, particularly when these three phenomena are viewed as the outcomes of stochastic point processes.

The so-called 'collective Coulomb blockade' effect (Imamoglu and Yamamoto 1994) has been utilised to suppress optical shot noise fluctuations in the light emitted by macroscopic semiconductor junctions (Yamamoto and Machida 1987; Yamamoto et al. 1993; Tapster et al. 1987; Edwards 1990, 1991; Edwards and Pollard 1932; Kim and Yamamoto 1997). Such quantum noise-suppressed devices have obvious application in those optical measurement, communications and computing systems where performance is limited by photonic shot noise.

Macroscopic Coulomb space charge suppression of electronic shot noise was an essential feature of the low noise vacuum electronic devices belonging to

* Refereed paper based on a talk presented to the Workshop on Nanostructures and Quantum Confinements, held at the Australian National University, Canberra, in December 1998.
an earlier era (Schottky 1918). It is not generally recognised that macroscopic space charge smoothing of electronic shot noise is also an essential feature of low noise semiconductor junction diodes and transistors (Edwards et al. 1999), as well as of sub-Poissonian laser and light-emitting diodes (Yamamoto et al. 1987). New mesoscopic and reduced dimension semiconductor devices such as the high electron mobility transistor, resonant tunnelling diode and single electron transistor, employ both classical (Coulomb) and quantum (Pauli) effects to suppress electronic transport noise (Milburn and Sun 1998).

Nanoscale devices such as 'electron–photon turnstiles' have also been envisaged (Imamoglu and Yamamoto 1994) and partially realised (Kim et al. 1999). In these the single particle Coulomb blockade effect (Averin and Likharev 1986) potentially enables the manipulation and control of individual photons and electrons with applications in the efficient transfer and processing of information.

In what follows we present a simple semiclassical discussion of the electronic and photonic shot noise generated in semiconductor junctions. The processes of charge injection, transport, storage and recombination are represented in Fig. 1a for a semiconductor diode and in Fig. 1b for a light-emitting junction diode with the additional process of radiative recombination. Radiative recombination provides a means of directly probing the recombination process in semiconductor junctions.

This paper is organised in five sections. Section 2 introduces electronic shot noise in the form of thermionic shot noise and its suppression by Coulomb interactions. Electronic noise in ideal macroscopic semiconductor junctions is discussed in Section 3. Partition noise and the second order counting statistics of electronic and photonic streams are discussed in Section 4. Section 5 examines photonic noise from semiconductor junctions, stochastic charge injection, the collective and single electron Coulomb blockade effects and concludes with a discussion of heralded photon states in mesoscopic junctions.

2. Electronic Shot Noise

(2a) Thermionic Emission

Shot noise, the random fluctuation in electric current arising from the discrete character of electronic charge, the 'Schroteffekt', was first identified (Schottky 1918) in the thermionic current flowing in a temperature-limited vacuum thermionic diode. Electrons are, in the main, emitted randomly and independently from heated metallic cathodes and thermionic emission therefore constitutes a naturally occurring Poisson point process (Teich and Saleh 1988). Providing this description also adequately describes the charge collection process, the resulting thermionic current fluctuations will have the Markovian correlation function,

\[ \langle i(t)i(t') \rangle = 2Ie\delta(t - t'), \]  

and will show the full single-sided shot noise current spectral density, the Fourier transform of the correlation function, \( S_i(\omega) = 2I e \), associated with the transport of a mean current \( I \) of independent particles, each of charge \( e \). In general, however, the emission, transport and charge collection processes will lead to a non-zero correlation function for \( |t - t'| > 0 \) and will consequently modify the white noise spectrum of this idealised process. Non-zero charge collection time
will reduce the spectral density at high frequencies so that the spectrum must be written more generally to take these effects into account.

In these cases the spectral density will be reduced below the full shot noise level and the Fourier transformed autocorrelation function \( \langle i(t)i(t') \rangle \), the current spectral density \( S_i(\omega) \), will contain a frequency dependent Fano factor \( F_i(\omega) \leq 1 \), and thus,

\[
S_i(\omega) = 2F_i(\omega)Ie. \tag{2}
\]

Here \( F_i(\omega) \) is readily measurable as the frequency dependent ratio of the spectral density of the current noise to the spectral density of the full shot noise current fluctuation.
On the other hand, the related time-domain Fano factor $F_n(\tau)$ expresses the total number counting (integrated current) variance relative to the Poissonian (full shot noise) value. It is formally defined as the variance $\sigma^2_n$ of the number count $n(\tau)$ in time interval $\tau$, normalised to the Poissonian variance $\langle n(\tau) \rangle$ where

$$F_n(\tau) = \frac{\sigma^2_n(\tau)}{\langle n(\tau) \rangle} = \frac{\langle n^2(\tau) - n(\tau)^2 \rangle}{\langle n(\tau) \rangle}. \quad (3)$$

Thermionically generated shot noise tends to be super-Poissonian ($F_n > 1$) at low frequencies, due to positively correlated emission variations on long time scales, and sub-Poissonian at higher frequencies, due to the combined effects of space charge smoothing and finite charge collection time. Similar comments apply to thermionic emission at Schottky junctions and heterostructure junctions (Imamoglu and Yamamoto 1993; Kim and Yamamoto 1997; Kobayashi et al. 1999).

Fig. 2. Representation of space charge smoothing of thermionic vacuum electron emission: a model for the macroscopic Coulomb blockade in semiconductor junctions.

(2b) Space Charge Suppression of Shot Noise

Space charge smoothing (Rack 1938) in thermionic diodes operated in the space charge-limited regime introduces anti-correlated density fluctuations into the electron stream. Fig. 2 illustrates this concept of shot noise suppression by a fluctuating potential barrier. The height of the barrier fluctuates with the electronic space charge population which itself fluctuates in response to random variations in thermionic emission. Low frequency shot noise suppression by Coulomb-moderated negative feedback mechanisms of this type are a common feature of classical and quantum electronic systems and devices.

The space charge-limited thermionic diode provides a conceptual model for photonic shot noise suppression in macroscopic systems (Teich and Saleh 1985;
Yamamoto and Machida 1987) and in mesoscopic systems involving the ballistic transport of electrons where Pauli exclusion provides an additional noise suppression mechanism (Lesovik 1989; Buttiker 1990). Space charge-limited thermionic noise has a thermal character, reflecting its thermal origin, with Fano factor derived in the classical limit as (Teich et al. 1987)

\[ F_n(\omega) = \frac{8kT}{eV}, \]

for cathode temperature \( T \) and applied voltage \( V \) (\( \gg kT/e \)). A similar result applies to the external noise current in a semiconductor junction diode operated under constant current conditions. More recent discussions of shot noise suppression (e.g. Blanter and Buttiker 1999) show how shot noise suppression can be explicitly accounted for in terms of electron–electron interactions.

We shall illustrate shot noise suppression for diffusive transport associated with semiconductor junctions and discuss the transition from collective Coulomb interactions to single electron ‘Coulomb blockade’ in low capacitance mesoscopic junctions.

3. Macroscopic Semiconductor Junction Noise

The first theoretical treatment of shot noise generated in semiconductor junctions was given by van der Ziel (1955, 1957), who initially attributed it to the random transport of charge carriers across the depletion layer. The results of van der Ziel’s analysis were confirmed for both diodes and transistors with the measurement of the full shot noise level for diodes operated with a fixed potential difference. However, the physical basis of the analysis was challenged by Buckingham and Faulkner (1974) and Buckingham (1983) who attributed the shot noise to two independent mechanisms operating in the neutral (bulk) regions of the structure: thermal fluctuations in minority charge carrier diffusion and fluctuations in the rates of generation and recombination. Their model, unlike the van der Ziel model, is consistent with the generally accepted diffusion model of charge carrier transport in a forward biased junction.

This diffusion model was adopted by Yamamoto and Machida (1987) to describe the generation of sub-Poissonian light by semiconductor diode lasers and by Edwards (1993) to describe the generation of sub-Poissonian light in light-emitting diodes. These authors also drew attention to a long-standing confusion in the literature which had led to the erroneous conclusion that the electron–hole recombination noise in a semiconductor junction is characterised by full shot noise, contrary to their observations of suppressed photonic shot noise in the light generated by laser and light-emitting diodes when driven from high impedance constant current sources (Machida et al. 1987; Machida and Yamamoto 1988; Tapster et al. 1987; Edwards 1990).

The charge carrier number present in the vicinity of a semiconductor junction, \( N(t) = N + n(t) \), is determined by the processes of charge injection, diffusion, and recombination which are taken into account in the following rate equation:

\[ \delta N(t)/\delta t = I(t)/e - N(t)/\tau + f_n(t). \]
In the first term on the right-hand side \( I(t) = I + i(t) \) represents the current supplied from an external circuit. It is taken to equal the net (fluctuating) rate at which charge is injected across the space charge layer at the junction into the 'active' (recombination) region. The second term, in which \( \tau \) is the mean lifetime of the diffusing carriers, comprises the mean recombination rate \( N/\tau \) plus a fluctuating term \( n(t)/\tau \). This latter represents the response of the reservoir population, \( N(t) \), and hence of the (population number-dependent) recombination rate to external 'pump' noise, \( i(t)/e \), and to the stochastic charge recombination process itself as represented by the third term, the Langevin noise term \( f_n(t) \). The second and third terms together then represent the fluctuating recombination rate: a population-dependent time varying rate plus an intrinsic stochastic (Poissonian) fluctuation.

Linearisation of equation (5) yields

\[
\delta n(t)/\delta t = i(t)/e - n(t)/\tau + f_n(t). \tag{6}
\]

If the stored electron population \( N \) is fixed, for example, by pinning the junction voltage, then \( \delta n(t)/\delta t = n(t) = 0 \) and the charge carriers will recombine randomly with mean lifetime \( \tau \). This constitutes a Poisson point process with correlation function \( \langle f_n(t)f_n(t') \rangle = 2N\delta(t-t')/\tau \), rate \( N/\tau \), mean square value \( \langle f_n^2 \rangle = (N/\tau)\Delta f \), and double-sided spectral density equal to the mean recombination rate \( N/\tau \). From equation (6) the recombination current noise and the current noise in the external circuit, \( i_\text{sn}(t) = -e f_n(t) \), are both at the full shot noise level. The single-sided mean square shot noise current spectral density is then, as expected,

\[
S_i(\omega) = 2\langle f_n^2 \rangle/\Delta f = 2e^2\langle f_n^2 \rangle/\Delta f = 2ie. \tag{7}
\]

If \( N(t) \) is allowed to freely fluctuate and in addition the injection current noise is suppressed, then \( i(t) = 0 \), and so \( \delta n(t)/\delta t = -n(t)/\tau + f_n(t) \). Taking Laplace transforms, the recombination current noise spectrum in this case becomes

\[
S_i(\omega) = 2\omega^2\tau^2\langle f_n^2 \rangle/\Delta f(1 + \omega^2\tau^2) = 2ie\omega^2\tau^2/(1 + \omega^2\tau^2), \tag{8}
\]

and has the character of single pole high-pass filtered shot noise and vanishes in the low frequency limit of \( \omega\tau \ll 1 \). This illustrates low frequency shot noise suppression according to the 'leaky reservoir' model (Edwards 1993). In passing we note that the electron reservoir number fluctuation spectrum has a complementary low pass character with a total mean square fluctuation of \( \langle N \rangle/2 \), just one-half the Poissonian value.

Writing equation (6) in the form of a state equation using the fluctuations in the junction potential \( \eta_{jn}(t) \), the injected current \( i(t) \), and the charge \( q(t) \) stored in the so-called 'diffusion capacitance' \( C \) as state variables, gives (Edwards 1993)

\[
C\delta \eta_{jn}/\delta t = i(t) - \nu_{jn}(t)/\tau - i_{sn}(t), \tag{9}
\]

and thus

\[
\delta q/\delta t = i(t) - q(t)/\tau C - i_{sn}(t). \tag{10}
\]
Referring to the corresponding noise equivalent circuit of Fig. 3, in which the stored charge fluctuation \( q(t) = n(t)e = Cyjn(t) \) shows that in this model the junction voltage fluctuation is a direct measure of the carrier number fluctuation \( n(t) \). Thus, shot noise suppression at frequencies \( \omega \ll 1/\tau = 1/\tau C \) is evidently achieved by making the external impedance \( R_s \) much greater than \( r \), the internal differential resistance of the junction, so that \( i(t) = \frac{\nu_s(t) - \nu_jn(t)}{R_s} \) vanishes in that limit. The voltage source \( \nu_s(t) \) in Fig. 3 represents the thermal Nyquist noise voltage associated with resistance \( R_s \). Then the mean square injected noise current \( \langle i(t)^2 \rangle \) can be reduced to negligible proportions by raising the value of the resistance \( R_s \). Noiseless charge injection into the reservoir is thus assumed for a high impedance current source. A detailed physical analysis of this noiseless injection process has been recently given (Kim and Yamamoto 1997) and refined by Kobayashi et al. (1999) in connection with a generalised theory of photon noise suppression in both macroscopic and mesoscopic junctions. In more detailed microscopic treatments (Buckingham and Faulkner 1974; Kim and Yamamoto 1997; Kobayashi et al. 1999) this current can be written to explicitly contain two stochastic Langevin terms representing forward and backward charge carrier injection noise.

The junction diode diffusion noise model above adequately describes electronic noise generation in macroscopic circuits and devices. It has also been used as the conceptual basis for models of sub-Poissonian light generation in light-emitting diodes (Edwards 1993) and diode lasers (Yamamoto and Machida 1987). These
account for many of the measured characteristics of macroscopic optoelectronic devices.

4. Partition Noise

The term ‘bunching’ is used in semi-classical quantum optics (Teich and Saleh 1988) to describe departures from a Poissonian photon stream having uncorrelated density fluctuations. Reduced (sub-shot noise) photo-current fluctuations, generally at low frequencies, are associated with ‘anti-bunching’, formally defined in terms of anti-correlated photo-current fluctuations in the two components of a split light beam.

If a stream of countable particles (electrons or non-interfering photons) is subject to Bernoulli partition and randomly partitioned into two beams of equal intensity, $I_1$ and $I_2$, the normalised intensity correlation (Loudon 1980; Oliver et al. 1999) between the two beams is

$$G_{11}^{(2)}(0) = \frac{\langle I_1(t)I_2(t)\rangle}{\langle I_1 \rangle^2} = \frac{(n(n-1))}{(n)^2} = 1 + \frac{(F_n - 1)}{(n)} , \tag{11}$$

where the currents $I(t) = en(t)/\Delta t$; $\langle I_1 \rangle = \langle I_2 \rangle = \langle I \rangle/2$ and $\langle n \rangle = \langle I \rangle \Delta t/e$. This function evidently deviates significantly from unity only in the limit of low mean count $\langle n \rangle$, that is, for weak currents and short integration times $\Delta t$. It is therefore a useful parameter in cases where small controlled numbers of electrons or photons are required. In the macroscopic case it is more useful to rewrite the covariance

$$\langle i_1(t)i_2(t) \rangle = [G_{11}^{(2)}(0) - 1]\langle I_1 \rangle^2 = \frac{1}{2}\langle [i_1^2(t)] - \langle i_1^2(t) \rangle \rangle , \tag{12}$$

as a correlation coefficient

$$R_{12} = \frac{\langle i_1(t)i_2(t) \rangle}{\sqrt{\langle [i_1^2(t)] \rangle \langle [i_2^2(t)] \rangle}} \tag{13}$$

and to express this as a conventional (frequency dependent) correlation coefficient relating the fluctuations in the partitioned streams

$$R_{12}(\omega) = \frac{\langle F_n(\omega) - 1 \rangle(1 - T)/[TF_n(\omega) + (1 - T)]}{(1 - T)\langle n \rangle} \tag{14}$$

The denominator in equation (14) expresses the shot noise as the sum of two independent variance terms, a transmitted noise term and an additional partition noise term. For transmission probabilities $T$ and $(1 - T)$, the ‘partition noise’ (van der Ziel 1970) introduced into particle streams has binomial statistics with variance given by

$$\langle n^2 \rangle - \langle n \rangle^2 = T(1 - T)\langle n \rangle \, \text{ and } \, \langle i_p^2 \rangle = T(1 - T)\langle i_{np}^2 \rangle . \tag{15}$$

From above, the spectral density of the total noise following random loss of particles from the beam can then be written in terms of the transmission factor $T$, and input and output Fano factors $F_i$ and $F_o$, as
This equation which also follows from the cascade variance formula (Burgess 1959) is well known in the quantum optics and mesoscopic literature (Teich 1988; Shimizu and Sakaki 1991). Conceptually, it illustrates the binomial statistics which arise from the operation of independent Bernoulli selection with fixed probability \( T \). It can be generalised (Lesovik 1989; Buttiker 1990) to describe the additive partition noise generated in multiple channel mesoscopic systems. Historically, it describes the shot noise in the anode current of a multielectrode space charge-limited thermionic vacuum device (van der Ziel 1970). The thermionic current fluctuations are initially suppressed below the full shot noise level, typically to \( F_i < 0.05 \), by the fluctuating space charge barrier (Fig. 2) between cathode and anode and then raised by the additional noise introduced by current partition as given by equation (16). Here \( T \) is the conditional probability of an electron being counted, given its emission at the source of the particle stream.

Partition noise is present in bipolar junction transistors, as 'quantum shot noise' in mesoscopic electron transport and as 'quantum vacuum fluctuations' in lossy photonic systems.

5. Photonic Shot Noise

Measurement of the photonic shot noise in the light emitted from a semiconductor junction is represented in Fig. 3. As mentioned in Section 2, photonic noise measurements provide a direct probe of the intrinsic fluctuations in the radiative recombination rate. These are characterised by the Fano factor \( F_i \) in Fig. 3 and equation (16) with \( T = \eta \), the overall quantum transfer efficiency between radiative recombination events in the junction and their subsequent photodetection.

The partition noise is a consequence of the lack of a one to one correspondence between the individual recombination events in the junction and their subsequent photodetection in the form of electron–hole pairs generated at the detector. Providing these photon deletion losses are statistically independent Bernoulli events, they can be treated like those arising at an optical beam splitter with overall transmission probability \( \eta \), as in Fig. 3.

Fig. 4 shows the equivalent circuit of an ideal light emitting diode coupled with quantum efficiency \( \eta \) to a photon detector as in Fig. 3. A fraction \( \eta \) of the recombination current appears at the detector. It therefore becomes possible in principle to measure externally the radiative recombination current in a light-emitting diode and to check the validity of the diode circuit noise model previously discussed.

When the recombination noise is completely suppressed \( (F_i = 0) \), the detector noise relative to the expected shot noise level is evidently \( F_o = (1 - \eta) \). This is an expression of the extreme fragility of sub-Poissonian 'quiet' light to attenuation of the light beam and reveals why it cannot provide any significant advantages in metrology or communications in the presence of even moderate losses.

(5a) Sub-Poissonian Photonic Noise

Fluctuations in photoelectron emission from a photocathode illuminated by a steady light source were for many years regarded as conceptually similar to thermionic emission fluctuations, being assumed to show the full shot noise
due to Poissonian statistics. The direct detection of sub-Poissonian light from semiconductor lasers (Yamamoto and Machida 1987) and light-emitting diodes (Tapster et al. 1987; Edwards 1990) has emphasised the similarities between photonic and electronic shot noise processes. These processes have a common description when viewed as stochastic point processes (Teich and Saleh 1988). For example, Edwards et al. (1999) noted that the equation describing the transfer of quantum noise between a photon emitter and photo-detector is identical with that describing the transfer of electronic shot noise between the emitter and collector of a semiconductor bipolar junction transistor and the generation and propagation of 'quantum shot noise' in a mesoscopic circuit (Shimizu and Sakaki 1991).

Fig. 4. Noise equivalent circuit representation of the Langevin rate equation (6) showing a light emitting junction with junction capacitance $C$, differential resistance $r_m$, and equivalent shot noise voltage source $\nu_{sn}$, driven through external resistance $R_s$, coupled with quantum efficiency $\eta$ to a photon detector with partition noise current $i_p$.

The first measurements of sub shot noise light from light-emitting diodes driven from high impedance circuits were made by Tapster et al. (1987). These showed 4% noise reduction ($F_d = 0.96$), below the full shot noise level. Edwards (1990, 1991, 1992) subsequently confirmed the validity of equation (16) for quantum efficiencies $\eta \leq 0.3$ and $F_i \leq 0.05$. The maximum, quantum efficiency-limited noise reduction reported to date of 50%, obtained by Shinozaki et al. (1997) is in agreement with that expected from equation (15) with complete noise suppression ($F_i = 0$) at the junction. These measurements support the shot noise suppression model based on the simple rate equations (5) and (6). In particular, the spectral dependence of the suppressed recombination noise has been measured and is evidently (Fig. 5) the same as that of the corresponding external current modulation characteristic, as expected for the leaky reservoir recombination model (Zhang et al. 1995). However, it is also evident from Fig. 5 that the bandwidth of the suppressed noise varies with the junction current and
does not equal $B = \frac{1}{2\pi r} = \frac{1}{2\pi rC}$ as would be expected from the simple reservoir model. The model ignores the stochastic injection of minority charge carriers across the space charge region of the junction into the active region where radiative recombination takes place. This is a significant omission, particularly as the size of the junction is reduced to mesoscopic dimensions and collective Coulomb effects give way to the single electron Coulomb blockade phenomenon (Imamoglu and Yamamoto 1993).

Measurements (Kim et al. 1995; Zhang et al. 1995; Kobayashi et al. 1999) on macroscopic and microscopic light-emitting junctions show that the bandwidth over which noise reduction occurs is more accurately predicted by a stochastic injection model in which transport across the depletion (space charge) layer and its depletion capacitance play an important part.

(5b) Stochastic Injection Model of Light-emitting Junctions

These recent measurements on light-emitting microjunctions and mesojunctions have led to a clarification of the diffusion based models of junction noise in favour of stochastic injection models. In these recent models, stochastic injection across the space charge layer into the active recombination region of the junction is accounted for. These models employ a space charge smoothing mechanism similar to that proposed by Teich and Saleh (1985) on the basis of the space charge model of a thermionic vacuum diode (Section 2a).

The charging energy $N_e e^2/2C_{dep}$ at the junction associated with the injection of $N_e$ electrons into the active layer raises the Coulomb barrier against subsequent charge injection. This occurs providing $N_e e^2/2C_{dep} \gg kT$, the thermal energy. It results in a sub-Poissonian stream of anti-bunched electrons on a characteristic
time scale $\tau_i = \tau_{C_{\text{dep}}} = kT/C_{\text{dep}}/eI$. As the injection current is lowered, this time scale lengthens and exceeds the recombination time. The corresponding bandwidth over which noise suppression occurs then becomes $B = eI/2\pi kT/C_{\text{dep}}$. With low injection, this will be less than the recombination bandwidth as is evident from Fig. 5. In this situation the suppression is due entirely to the 'macroscopic Coulomb blockade' process (Imamoglu and Yamamoto 1993) since there can be no significant storage of diffusing charge in the reservoir. The corresponding rate equation is then equation (10) and the corresponding equivalent circuit is Fig. 4, with $C$ replaced by $C_{\text{dep}}$, the depletion capacitance. In general the macroscopic junction capacitance required to correctly model the noise suppression measurements will evidently be the sum of the depletion (collective Coulomb blockade) capacitance and the diffusion (leaky reservoir) capacitance.

(5c) Single Photon Generation

The successive injection of $N_i$ electrons (typically $10^8$ or more) perturbs the junction potential by an amount $Ne/C_{\text{dep}} \gg kT/e$, the thermal voltage fluctuation, and thus introduces antibunching and consequent shot noise suppression. If the electron count or current integration time interval exceeds the characteristic time $\tau_i$, then sub-Poissonian variance and sub shot noise current spectral densities will be observed, although the traditional measure of anti-bunching, the second order coherence function $G^{(2)}(0)$ from equation (11) will not be significantly different from unity since $(\langle n \rangle)$ is large, being of the order of $N_i$. Moreover, as the injection current is reduced the bandwidth of the suppressed noise will continue to contract, requiring ever longer integration times in order to maintain low noise levels. Full shot noise ($F(\omega) = 1$) will result if the integration time is made much shorter than $\tau_i$.

However, if $C_{\text{dep}}$ and $T$ are reduced to make $e^2/kT/C_{\text{dep}} \gg 1$, the potential drop due to the passage of single electrons may become sufficiently large relative to the thermal fluctuations to regulate the flow of individual electrons via the single electron Coulomb blockade effect (Averin and Likharev 1986). This requires a sub-micron sized semiconductor nanojunction with junction capacitance of order $10^{-16}$ F operated at liquid helium temperatures.

If, in addition, the spontaneous recombination time can be made much shorter than the electron inter-arrival time $e/I$ (1 ns for $I = 0.16$ nA), then control over the emission of single photons becomes possible. This is the basis of proposals for realising 'electron photon turnstiles' utilising both single electron Coulomb blockade and quantum confinement effects in mesoscopic semiconductor junctions (Imamoglu and Yamamoto 1993, 1994). Quantum control of photon emission would enable the generation of well-defined photon numbers (photon number states) at specified times. Such 'heralded' photon number states have applications in quantum communications and cryptography. For example, from equation (11), an ideal single photon state source for quantum key distribution purposes would be characterised by $(\langle n \rangle) = n = 1$, $F_n = 0$ and $G^{(2)}(0) = 0$. The inefficient weak Poissonian sources currently used in quantum cryptographic systems are typically characterised by $(\langle n \rangle) \approx 0.1$, $F_n = 1$ and $G^{(2)}(0) = 1$.

Proposed single-photon and single electron devices also have applications in quantum computing, quantum communications and quantum metrology, as well as in fundamental tests of quantum mechanical theory (Kim et al. 1999).
6. Conclusions

Electronic and photonic shot noise remain significant impediments to the performance of optoelectronic and mesoscopic systems. However, investigations of sub-Poissonian light from macroscopic light-emitting semiconductor junctions have highlighted the close parallels between electronic and photonic shot noise, have resulted in a better understanding of the physical processes involved, and have pointed the way to new noise suppression technologies on macroscopic and mesoscopic scales.

Acknowledgments

I acknowledge with thanks the financial support of the Australian Research Council under its Large Grants Program and the encouragement and support of the late Rolf Landauer.

References


Manuscript received 13 August, accepted 26 October 1999
Semiconductor Junction Noise Revisited

P. J. Edwards, R. McDonald and W. N. Cheung

Centre for Advanced Telecommunications and Quantum Electronics, University of Canberra, ACT 2601, Australia

Abstract. We review the conventional wisdom concerning the physical origin and phenomenology of noise in semiconductor junction diodes and transistors following recent developments in the field of sub-shot noise ("quiet") light generation. We identify several unsolved problems in the modelling of semiconductor junction noise in the macroscopic and mesoscopic domains.

INTRODUCTION

Recent interest in the generation of "quiet" light, light in which the quantum noise fluctuations have been suppressed below the normal shot noise level [1,2,3,4], has revealed a number of common misconceptions about the physical origin and the physical modelling of noise in semiconductor junction devices [1,5]. These misconceptions are all the more remarkable when we realise that more than 40 years have passed since the first models of semiconductor junction noise were described by van der Ziel and others [6,7,8]. More recently still, attempts to link shot noise related processes in macroscopic and mesoscopic junctions have revealed additional shortcomings in the conceptual framework [9,10].

Four examples serve to illustrate problems still remaining in our understanding of semiconductor junction noise:

1) For many years the incorrect Langevin equations were used to describe noise generation in laser diodes [1]. Their use led to the erroneous belief [11] that sub-shot noise operation of laser diodes was not possible because of shot noise introduced in the pumping process. This misconception apparently arose from the van der Ziel models [6,7] of junction current noise in a p-n junction diode. These were interpreted to imply shot noise-limited junction current noise.

2) This assumption of a shot noise limit to semiconductor junction diode current, independent of the driving source impedance, lead Kikuchi and Kikui [12] to incorrectly attribute full shot noise to the light from light-emitting junction diodes, contrary to that observed with a high impedance pump [3].

3) Revision of the Langevin equations and the corresponding noise equivalent circuits to accommodate the observations of quiet (sub-shot noise) light along the
above lines, while reproducing the macroscopic features of the observations, fail in at least one important respect. The revised equations do not take into account the stochastic nature of the charge carrier injection into the active region of the junction. This failure becomes particularly acute when attempts are made to generalise the theory of Coulomb blockade in mesoscopic junctions to include collective effects in macroscopic junctions [9,10].

4) Finally, at the pedagogic level, fundamental misconceptions remain enshrined in the undergraduate physics and engineering texts dealing with p-n junction noise. Thus we read [14] that "in a bipolar transistor, minority carriers diffuse and drift across the base region to be collected at the collector base junction. The collector current consists of a series of random current pulses. Consequently, collector current shows full shot noise and this is represented by a shot noise current generator...". This statement about the character of the collector current noise is physically incorrect as is the final sentence in the following statement [14] that "base current $I_b$ in a transistor is due to recombination...and also to carrier injection from the base into the emitter. All of these are independent random processes and thus $I_b$ also shows full shot noise".

This confusion between physical noise sources and equivalent noise sources is not limited to undergraduate texts. It has become customary to refer to the "collector current shot noise" and the "base current shot noise" in the professional literature, eg [15], although neither of these fictitious noise sources has the direct physical meaning implied by their nomenclature.

**SUB-SHOT NOISE LIGHT GENERATION**

There is now an overwhelming body of evidence to show that a light emitting diode with high quantum efficiency emits light with sub-shot noise intensity fluctuations when driven from a high impedance source [3,4]. This has led to a reexamination of models of shot noise generation in semiconductor junctions. The reason for this is that the intensity fluctuations in the emitted light provide a direct probe of the statistical fluctuations in the electron-hole recombination rate. These are thus evidently revealed to be sub-Poissonian in character when a light-emitting diode is driven by a high impedance source.

The reality of this phenomenon was initially strongly contested on the grounds that the noise should be fixed at the full shot noise level by the random carrier transport and recombination processes, independently of the external circuit configuration [12]. The basis for this incorrect assertion can probably be traced back to van der Ziel's original junction noise theory [6,7]. According to this, the shot noise in an ideal diode originates in the random diffusion of carriers across the space charge region. This view was challenged by Buckingham and Faulkner [8] who derived a diffusion model of junction noise based on Shockley's model of an ideal junction. These authors showed that the shot noise current associated with charge transport across the depletion layer should be vanishingly small and quite undetectable in the external circuit. Compensating relaxation currents were postulated to flow as a result of charge redistribution across the depletion layer. Two noise generation mechanisms were
identified, both in the neutral bulk regions of an ideal diode: (i) thermal fluctuations in minority carrier diffusion and (ii) random fluctuations in the spontaneous recombination rate.

Both of these noise mechanisms are influenced by external conditions and can be suppressed by placing the junction in a high impedance environment. This is contrary to the conventional wisdom received prior to the first optical sub-shot noise measurements and their subsequent modelling by Yamamoto and others [1,2,3,13], based on the Buckingham bulk diffusion noise model [8].

This is illustrated in the diode noise equivalent circuit of Fig.1(a) which suggests a shot noise limited recombination current, \( i_{m}(t) \) [12]. Referring to the equivalent circuit of Fig.1(b), negative feedback between recombination rate fluctuations, \( i_{m}(t) \), and fluctuations \( n(t) \) in the stored charge number, occurs when the junction voltage, \( v_{n}(t) = q(t) / C_{e} = n(t) e / C_{e} \), is allowed to fluctuate freely by a high driving impedance. This regulatory mechanism serves to suppress stochastic fluctuations in the recombination (photon emission) rate on time scales long compared with the carrier lifetime, \( \tau = r_{e} C_{e} \), where \( r_{e} \) is the differential resistance and \( C_{e} \) is the junction diffusion capacitance. Thus, the Langevin rate equation derived from this macroscopic model of the junction can be written

\[
\frac{dn(t)}{dt} = \frac{i_{e}(t)}{e} - n(t)/\tau + f_{n}(t),
\]

where the Langevin force \( f_{n}(t) = v_{sn}(t)/r_{e} e, \) has mean square value \( <i_{sn}^{2}>/e^{2} \), and \( i_{sn}(t) \) is the shot noise limited recombination current which flows when the junction voltage \( v_{n}(t) \) is pinned by the capacitor at high frequencies or else when the diode is driven from a low impedance source. On the other hand the recombination current noise \( i_{en}(t) = [v_{n}(t) - v_{sn}(t)]/r_{e} \), and consequently the photodetector current noise in an ideal detector (\( \eta = 1 \)), both vanish in the low frequency, high impedance limit for which \( v_{n}(t) = v_{sn}(t) \). This macroscopic model predicts complete suppression of the shot noise in this limit. Measurements generally support this model although limited to moderate degrees of noise suppression by the partition noise arising from non-ideal detection.

\[
\begin{align*}
\frac{dn(t)}{dt} &= \frac{i_{e}(t)}{e} - n(t)/\tau + f_{n}(t),
\end{align*}
\]

FIGURE 1. Alternative semiconductor bipolar junction diode noise equivalent circuits.
SEMICONDUCTOR JUNCTION NOISE

A more serious limitation to this macroscopic model of bipolar junction noise, concerns its microscopic formulation. It is clear that for both constant voltage and constant current operation of a semiconductor junction, the diffusion and injection of carriers into the active region of the junction are stochastic processes. The model outlined above fails to address this issue although attempts [9,10] have been made to remedy this omission by extending the Buckingham model [8] to explicitly take the depletion capacitance into account in the charge injection process.

An important conceptual advance has arisen from this work concerning the relation between shot noise suppression in macroscopic junctions and single electron Coulomb blockade in mesoscopic, low capacitance junctions. Injection noise suppression in macroscopic junctions has been interpreted as a collective Coulomb blockade phenomenon. The successive injection of \( N \), (typically \( 10^8 \) or more electrons) perturbs the junction potential by an amount \( Ne/C_{dep} >> kT/e \), the thermal voltage fluctuation, and thus introduces anti-bunching and consequent shot noise suppression. In the mesoscopic case the potential fluctuation due to single electrons may be sufficient to regulate the flow of individual electrons providing \( e^2/kT C_{dep} >> 1 \). If, in addition, the spontaneous recombination time is much shorter than the electron charging time \( e\ell \), then there will be no charge storage at the junction and the sub-Poissonian statistics of the injected electron stream will be translated directly into sub-Poissonian photon statistics. Where significant charge accumulation occurs, as in macroscopic diodes, the statistics of the fluctuating stored charge population then determines the recombination noise [13].

The correct microscopic modelling of this situation is of particular practical interest in the development of single photon emitting devices. In this case the aim is to generate a controlled stream of single photons with applications in optical computing and quantum cryptosystems. However, although charge transport across a mesoscopic junction has been quantitatively modelled [9], the fabrication of effective “electron-photon turnstile” devices [16] and the full experimental verification of the transport models remain major unsolved problems.

PARTITION NOISE

The predictions of the macroscopic theory of bipolar junction recombination noise have not yet been completely tested since it has not so far proven possible to implement a system incorporating a light-emitting diode system in which the quantum transfer efficiency, \( \eta \) is much greater than 50\%. In consequence the measured noise contains a partition noise component which sets a lower limit to the photo-current noise of about one half the full shot noise level.

Photo-detection can be modelled as a Bernoulli selection of \( n_i \) recombination events with conditional probability, \( \eta \). The result is additive partition noise having a binomial distribution with number variance \( \eta (1-\eta) <n_i> \). Consequently, the detected noise relative to the shot noise level, the Fano factor \( F_\alpha \), can be written,
The first term is the input noise transferred to the detector output. The second term is the partition noise. This equation, well known in quantum optics, is identical with that used to model the transfer of noise from the emitter-base junction to the collector-base junction of bipolar and photon coupled transistors [17,18].

TRANSISTOR NOISE REVISITED

Not surprisingly we might therefore expect noise in a common base–connected bipolar junction transistor (BJT) to be successfully modelled in the same way as noise in an open loop optical coupler with \( \eta = \alpha \), the collector/emitter current transfer ratio. Likewise there is a close correspondence between physical noise models of the BJT and the photon coupled transistor (PCT). These have been verified by laboratory measurement [17,18]. The models require three independent physical noise sources: a base-emitter junction noise source; a partition noise source in the collector circuit, and a thermal base resistance noise source. These are shown in Fig.2. Of course, when noise currents are calculated, these will be found to be correlated in a manner determined by the circuit topology.

\[
F_o = \eta F_i + (1-\eta)
\]  

(2)

**FIGURE 2.** Bipolar junction transistor noise equivalent circuit incorporating the diode noise equivalent circuit of Fig. 1(b).
FIGURE 3. Measurements of shot noise suppression in a common base connected BJT with high impedance in the emitter lead.

Although partition noise in a BJT is much smaller than in a PCT, since $\alpha = 1$, high impedance suppression of the base–emitter noise current occurs in a BJT just as in a laser diode or light-emitting diode [1,2,3]. Fig.3 shows the measured suppression of low frequency shot noise for the common base configuration with a high impedance in the emitter lead. Inspection of Fig.1(b) and the use of Equation 1 shows that at very low frequencies only partition noise should be present. The full shot noise level is approached only at high frequencies as the base–emitter current rises towards full shot noise. This trend is shown in the measurements shown in Fig.3.

This approach to the modelling of the BJT was first used by van der Ziel [6,7]. However subsequent transistor noise models, by replacing physical noise sources with more convenient equivalent noise generators, appear to have introduced conceptual difficulties and errors in the literature.

Two problems are noteworthy [14,15]. The first of the equivalent noise sources, the so-called “collector shot noise current” is a fictitious entity which has a direct physical significance only in the case when the full base-emitter shot noise current flows and is transferred to the collector together with additive partition noise. Equation 2 shows that the collector noise is then at the shot noise level ($F_o = 1$).

The other problem is that of the so called “base current shot noise”. It will be obvious from the preceding discussion and from inspection of the diode equivalent circuits (Fig.1) that the base current will show full shot noise only under very restricted conditions, if at all. The usual high impedance base bias network will usually limit the base current noise to thermal values. The fact that it is convenient to use an equivalent noise current generator at the input to a common emitter BJT amplifier of value

$$<i_{in}^2> = 2 I_b e \Delta f / \alpha = 2 I_b e \Delta f,$$  \hspace{1cm} (3)
CONCLUSIONS AND OPEN QUESTIONS

In summary, there remain conceptual difficulties in the modelling of bipolar junction device noise. Recent work on shot noise suppression in optoelectronic and mesoscopic devices has highlighted these problems. They are due in part to shortcomings in early junction noise models. Several tasks remain. The revision of "diffusion" models of junction noise to provide electron counting and arrival time statistics in the mesoscopic limit may be possible using the theory of stochastic point processes. Revision and generalisation of the junction noise equivalent circuits also must be completed, and recent work on amplitude-squeezed, sub-Poissonian optical noise incorporated into the mainstream of electronic noise theory and practice.

ACKNOWLEDGEMENTS

We acknowledge the support of the Australian Research Council through its Large and Small Grant programmes and the encouragement of the late R. Landauer.

REFERENCES

The distinction between “physical” and “equivalent” noise sources in bipolar junction transistors and other semiconductor devices has become blurred in the noise literature and in current engineering textbooks. An unfortunate consequence of this is the emergence in the literature of fictitious noise sources such as the “collector-current shot noise” and the “base-current shot noise”. These are often assigned a physical reality and incorrectly treated as real physical noise sources, independent of circuit topology. Text-books have encouraged successive generations of students in this belief. Non-physical noise sources such as these are convenient and legitimate, even essential, for the effective modeling and simulation of circuit noise. However their naïve use in teaching and research is likely to continue to give rise to fallacious concepts and misleading conclusions. The physical modeling of the light-emitting diode, the photon transport transistor and the bipolar junction transistor are briefly discussed to illustrate this view.

**Keywords:** Semiconductor noise; shot noise; photonic noise; partition noise; transistor noise; equivalent noise generators; physical noise generators.

1. **Introduction**

Recent advances in quantum optics and the successful generation of sub-shot noise (“quiet”) light by semiconductor junction light-emitting diodes and laser diodes have revealed a number of fundamental misconceptions about the physical origin of shot noise in semiconductor devices such as bipolar junction diodes and transistors [1–3]. That these misconceptions should persist for more than forty years after the seemingly definitive work on diode and transistor noise by van der Ziel and others is remarkable [4]. It provides an interesting commentary on the way we conceptualise and teach the topic of fluctuations and noise. A glance at the transistor noise literature cited by current textbooks is informative [5]. We are told that “the theory of the low frequency noise performance of the bipolar junction transistor has remained essentially unchanged since its inception” and we are shown a common emitter-connected transistor with a shot noise current generator in parallel with its base-emitter junction and another between the collector and emitter terminals [6]. Most textbooks make similar presentations, omitting to mention that these
are equivalent and not physical noise sources. There are attempts to provide a physical basis for the two shot noise generators in terms of “independent random processes” at the base and collector junctions. We read for example that “the collector current consists of a series of random current pulses (and that) consequently, collector current shows full shot noise and this is represented by a shot noise current generator…” [5,7]. Alas, measurement and simple calculation show this statement to be untrue.

The accompanying statement that because base current \( I_b \) in a transistor is due to recombination and carrier injection from the base into the emitter, “these are independent random processes and thus \( I_b \) also shows full shot noise” is also demonstrably incorrect. The noise present in the external base current is determined by the base circuit impedance.

This confusion between physical noise sources and equivalent noise sources is not limited to undergraduate texts. It has become customary to refer to the “collector current shot noise” and the “base current shot noise” in the professional literature as if these were real physical sources. There have even been attempts to measure these fictitious entities directly [8]. Thus are the myths of the so-called “base current shot noise” and “collector current shot noise” entrenched and perpetuated. These misconceptions have their origins in the original van der Ziel models of junction shot noise in p-n homojunctions [4].

In fact, the shot noise current generators conventionally located at the base-emitter diode light junctions of transistors, at the junctions of light-emitting diodes, between the collector and emitter terminals of transistors and across the terminals of reverse-biased semiconductor detectors are misleading (but remarkably robust and popular) fictions.

2. Measurements

Figure 1 shows the collector current noise, relative to full shot noise, for a low frequency common base transistor with a high impedance (\( \gg r_e \)) connected from emitter to ground.
The current noise spectral density was measured with a swept frequency spectrum analyser. The Fano factor was calculated as the collector noise current spectral density relative to that from a cathode temperature-limited thermionic diode shot noise generator operated with the same dc current and injected at the collector terminal.

Rather than exhibiting full shot noise as might be expected from the conventional wisdom, the collector noise is suppressed below the full shot noise level (Fano factor = 1) by a factor of some fifty times at low frequencies.

This is a direct demonstration that the full collector shot noise generator conventionally shown in common-emitter configurations is specific to that configuration. It is an equivalent noise generator which, together with another (equivalent) base current noise generator, correctly accounts for common-emitter circuit noise. It has no “real” existence independent of its two physical components—the shot noise generated in the base-emitter diode and the partition noise generated at the base-collector current branch.

3. Physical models: Transistor noise revisited

There is a close correspondence between the physical noise models of bipolar transistors and photon transport transistors [3]. Only three independent physical noise sources are needed to characterise the ideal transistor: the base-emitter shot noise voltage, the base-collector branch partition noise current and the thermal base noise voltage.

![Universal noise equivalent circuit for ideal bipolar junction and photon transport transistors showing three independent physical noise sources: shot noise base-emitter voltage generator $v_{sn}$; collector current partition noise generator $i_p$; and thermal base voltage noise generator $v_{bn}$. For a photon transistor, the base resistance $r_b$ and its thermal noise are negligible. In a bipolar transistor $\alpha=1$, therefore the partition noise is negligible.](image-url)

**Fig. 2.** Universal noise equivalent circuit for ideal bipolar junction and photon transport transistors showing three independent physical noise sources: shot noise base-emitter voltage generator $v_{sn}$; collector current partition noise generator $i_p$; and thermal base voltage noise generator $v_{bn}$. For a photon transistor, the base resistance $r_b$ and its thermal noise are negligible. In a bipolar transistor $\alpha=1$, therefore the partition noise is negligible.
Since the probability \(1-\alpha\) of an emitted charge carrier not being collected is small, the Bernoulli partition noise \(<i_b^2> = \alpha(1-\alpha) <i_n^2>\) is much less in a BJT than in a photon transistor where photon losses are much greater. Thus, high impedance suppression of base-emitter shot noise in a common base transistor leaves only low frequency partition noise as seen in Fig. 1.

Figures 1 and 2 together show that at higher frequencies the collector noise approaches the full shot noise level since then

\[
<i_{en}^2> = \alpha^2 <i_n^2> + <i_b^2> = \alpha^2 <i_n^2> + \alpha(1-\alpha) <i_n^2> = \alpha <v_{en}^2>/r_e^2 = 2\alpha I_e e \Delta f = 2 I_e e \Delta f.
\]

Comparison between Fig. 2 and the physical noise equivalent circuit of a light-emitting diode [1-3] reveals that this shot noise suppression is the same phenomenon observed in the recombination light emitted from light-emitting diodes and laser diodes when these are operated in a high impedance loop [1]. Intensity fluctuations in the photon flux emitted from a light-emitting junction provide a direct probe of the statistical fluctuations in the electron-hole recombination rate. It is not surprising therefore that recent advances in the understanding of semiconductor junction noise have come via quantum optics [1-3, 9-11].

For many years the incorrect Langevin equations (and corresponding noise sources) were used to describe noise generation in laser diodes [1]. Their use led to the erroneous belief that sub-shot noise ("quiet") light generation with laser diodes was not possible.

Similar incorrect assertions were made concerning the photon emission from light-emitting diodes. The first measurements of "quiet" light from LEDs were discounted on the grounds that the noise must be fixed at the full shot noise level by random carrier transport and recombination processes, independently of the external circuit.

The basis for these misconceptions can again be traced back to van der Ziel's original junction diode noise model which incorporated a full shot noise current generator driving the external circuit current noise. This model implicitly assumed constant voltage operation and therefore constant stored charge-carrier number, with consequent Poissonian recombination statistics.

The correct physical recombination noise source is that shown in Fig. 2 for the base-emitter junction. This representation clearly shows how the second order statistics of the photon stream emitted from a light-emitting junction and the charge carrier stream injected into the collector-base region of a BJT are both determined by the impedance environment in which the devices are embedded. Full shot noise, \(<i_{en}^2> = <v_{sn}^2>/r_e^2\) is generated only when the junction voltage is pinned as it is at high frequencies and also when the base-emitter loop impedance is low compared with \(r_e\).

In 1974, Buckingham and Faulkner derived a diffusion model of junction noise based on Shockley's model of an ideal junction [12]. Yamamoto and Machida successfully modeled their high impedance laser "quiet" light generation using the Buckingham bulk diffusion noise theory [1].

4. Conclusions

Our approach to the modeling of the BJT is similar to that first used by van der Ziel. However, the replacement of physical noise sources by more convenient but "circuit specific" equivalent noise generators in subsequent engineering models has given rise to conceptual difficulties and errors in the electronics and photonics literature.
The so-called "collector shot noise current" has a direct physical significance in the case when the full base-emitter shot noise current flows and is transferred to the collector together with additive partition noise. This occurs in a common base transistor at high frequencies when the full emitter shot noise current is generated. This is a common situation in photonics. Ironically, for bipolar transistors in which the collector base partition noise is small, the "collector shot noise" is almost entirely due to shot noise generated in the base-emitter junction! Similarly, the so-called "base current shot noise" is nothing of the kind. It is actually the collector partition noise referred back to the input terminal!

Finally, it should be emphasised that the conventional BJT noise equivalent sources are correct and convenient tools for noise simulation. Using physical noise sources in lieu of the conventional equivalent noise sources will not improve the accuracy of the BJT noise simulation. When properly used they both provide identical results.

Why then should we concern ourselves with physical noise sources? The answer must surely be that by narrowing the distance between device noise theory and engineering practice we become better equipped to avoid the gross errors referred to earlier and we gain the physical insight needed to make conceptual and practical advances. This is becoming increasingly important for low noise quantum electronic and quantum optical devices such as the high electron mobility transistor [13].

References

CHAPTER 16

Sub-Poissonian Recombination Noise in Macroscopic and Mesoscopic Semiconductor Junctions

Paul J. Edwards
University of Canberra, Canberra, Australia

CONTENTS
1. Introduction ................................................. 2
2. Electronic Shot Noise ........................................ 3
  2.1. Thermionic Emission Noise .................. 3
  2.2. Space Charge Suppression ................. 3
  2.3. Semiconductor Junction Shot Noise ....... 4
  2.4. Recombination Current Noise Suppression ... 5
3. Photonic Shot Noise ....................................... 6
  3.1. Photonic Recombination Noise Measurements ... 6
  3.2. Photon Counting Statistics .............. 7
  3.3. Sub-Poissonian Photonic Noise ............ 7
4. Macroscopic-Junction Terminal Current Noise ........... 8
  4.1. Historical Background ...................... 8
  4.2. Van der Ziel's Corpuscular Model ........ 8
  4.3. Buckingham's Diffusion Model .......... 8
5. Macroscopic-Junction Recombination Noise ........... 11
  5.1. Macroscopic Coulomb Blockade ............ 11
  5.2. The Stanford (Macroscopic Coulomb Blockade) Model ... 12
  5.3. The Canberra (Fluctuating Reservoir) Model ... 15
  5.4. The Hiroshima (Backward Pump) Model .... 15
6. Mesoscopic-Junction Recombination Noise ........... 17
  6.1. Transition from Collective to Single-Electron Coulomb Blockade ... 17
  6.2. Single-Electron Coulomb Blockade ....... 18
  6.3. Heralded-Photon Generation ............ 18
7. Summary ................................................. 20
  7.1. Current State of the Art .............. 20
  7.2. Future Developments ...................... 20
Acknowledgments .................................. 20
References ........................................ 20
1. INTRODUCTION

Prior to the invention of semiconductor laser and light-emitting diodes, models of the internal noise processes in semiconductor junction devices could not be directly tested. Noise measurements were restricted to the external terminal current and the terminal voltage. The noise generators in the noise equivalent circuits were only required to reproduce the current and voltage noise external to the device. The availability of light-emitting junctions in which radiative recombination dominates the recombination process has provided a diagnostic tool which enables the internal (radiative) recombination and other relevant processes to be examined directly.

We shall use the term “recombination current noise” in an operational sense to refer to fluctuations in the radiative recombination rate as seen in the photocurrent of an ideal (100% efficient) photon detector capable of collecting and detecting all of the emitted photons from a light-emitting junction. Since every electron-hole recombination event in an ideal junction maps to a corresponding electron-hole generation event in an ideal photon detector, this provides direct access to the time-varying recombination current and recombination rate.

The recombination current noise defined in this way is strongly influenced by the electrical environment in which the junction is placed, that is, by the junction bias circuit configuration, although it has its physical origin in the carrier generation, transport, and recombination processes within the junction. Its measurement has provided significant and unexpected new insights into semiconductor junction dynamics as well as foreshadowing a range of quantum information technologies and applications.

One of the unexpected outcomes of such optical recombination current measurements on semiconductor light-emitting junctions has been the discovery of sub-Poissonian "quiet" light. We shall use the terms "quiet," "sub-Poissonian," and "sub-shot noise" light interchangeably to refer to a photon beam in which the variance of the fluctuating photon count is less than for the random (Poissonian) case, corresponding to an integrated mean square current fluctuation below the normal shot noise level. A noiseless light beam will therefore be construed as a uniform stream of photons of constant number density giving rise to a constant, nonfluctuating rate of photoelectron hole-pair generation in an ideal detector.

This chapter addresses the principles of sub-Poissonian electronic and photonic noise generation in semiconductor junctions. We shall discuss recombination shot noise suppression, the transport of single electrons, and the controlled emission of single photons through the operation of the collective and single-electron Coulomb-blockade mechanisms.

External regulation of the carrier injection and recombination processes in semiconductor junctions suppresses the shot noise current fluctuations which have their origin in the thermal diffusion, thermionic emission, quantum tunneling, and carrier generation-recombination processes. In the low current limit, the Coulomb-regulated and Pauli-regulated transport of single electrons in mesoscopic structures leads to the controlled ("heralded") emission of single photons.

We shall review the concepts, the mechanisms, and the techniques for suppressing the shot noise usually associated with the recombination process in semiconductor junction devices, as revealed by measurements of intensity fluctuations in the emitted light. Potential applications of quiet-light and heralded photon emission technologies in metrology, telecommunications, computing, and cryptography are briefly discussed. The aim is to describe the phenomenology and the development of the physical concepts underlying the suppression of recombination photon intensity noise in optoelectronic devices and systems; to describe these in semiclassical terms, and to outline the realization of these suppression techniques in the form of novel optoelectronic devices and systems. Analytic discussion will be restricted to "long" diodes and quantum-confined systems in which the terminal current is due entirely to recombination, so that the recombination current dynamics will be treated in the well-known charge control approximation. This is not a severe restriction, since the basic principles of sub-Poissonian noise generated in the thermal transport and recombination of charge carriers are to be addressed.

The possibility of quantum noise suppression ("intensity squeezing") in laser light was suggested in 1984 [1] and first demonstrated in 1987 by Machida et al. [2] with strongly pumped laser diodes and by Tapster et al. [3] with light-emitting diodes. The laser and light-emitting diodes were operated with high-impedance biasing circuits. Not only does this mode of operation suppress the external terminal current ("pump") fluctuation; it also suppresses the internal fluctuation in the radiative recombination rate by an internal negative feedback mechanism [4] in a manner to be described. The result is that both the external circuit current noise and the electron-hole recombination current noise present in the emitted light are suppressed below the full shot noise level.

In current state-of-the-art demonstrations of quiet light generated in macroscopic semiconductor light-emitting junctions, "photon anti-bunching" reduces the number variance of bunches of photoelectrons below the Poissonian value when measured on time scales longer than some characteristic time, usually the radiative recombination lifetime of minority carriers in the bulk regions of the devices.

Wide-bandwidth photonic noise suppression of 3 dB below the full shot noise level has been measured in the light from light-emitting diodes [5] and up to 5 dB in the light from laser diodes [6–8], limited by photon detection efficiency. These measurements imply an electron-hole...
recombination current noise suppression of more than 20 dB. The fundamental limit to the degree of photon noise suppression achievable in practical devices is set, in fact, by transmission losses. It seems likely therefore that most applications of photon noise suppression techniques will be limited to optoelectronic and mesoscopic integrated circuit devices and interconnects in which the photon transmission losses can be kept small.

A notable exception to this is in the field of quantum cryptography, where the suppression of binary signals containing more than a single photon is essential for secure key transmission, even in lossy situations. In the single-photon limit information can be carried by the individual members of such noiseless photon streams, controlled by "electron-photon turnstiles" and "quantum repeaters" [9, 10] and "heralded" by electronic means. The electron and photon anti-bunching phenomena responsible for sub-Poissonian number variance in the light from macroscopic junctions can be extended to provide heralded emission of single photons in mesoscopic junctions in which the Coulomb blockade and Pauli (quantum confinement) effects operate. The controlled emission of single photons is of potential importance in the development of quantum computer logic gates and quantum cryptographic key generators.

The chapter is organized into six sections. Following this Introduction, Section 2 briefly reviews electronic shot noise concepts in terms of stochastic point processes and spectral densities. Section 3 deals with photonic shot noise in the context of optical measurements of electron-hole recombination noise. In Section 4 models of macroscopic semiconductor junction noise are presented in a historical context. Section 5 addresses physical models of recombination noise in macroscopic heterostructures, including the macroscopic Coulomb blockade process and the nonlinear backward pump model. Finally, Section 6 discusses physical carrier transport in mesoscopic heterojunctions, the controlled tunneling of single-carrier processes, and the consequent regulated emission of single photons.

2. ELECTRONIC SHOT NOISE

2.1. Thermionic Emission Noise

Electronic shot noise, the random fluctuation in electric current arising from the discrete character of electronic charge, was first identified by Schottky [11] in the thermionic current flowing in a temperature-limited vacuum thermionic diode. Electrons, in the main, emitted randomly and independently from heated metallic cathodes, and thermionic emission therefore constitutes a naturally occurring Poisson point process [12]. Providing this description also adequately describes the charge collection process, the resulting thermionic current fluctuations will have a Markovian correlation function,

\[ \langle i(t) \cdot i(t') \rangle = 2ie\delta(t-t') \quad (1) \]

and will show the full single-sided shot noise current spectral density, \( S_i(\omega) = 2le \), associated with the transport of a mean current \( I \) of independent particles, each of charge \( e \). In general however, the emission, transport and charge collection processes will lead to a nonzero correlation function for \( |t-t'| > 0 \) and will consequently modify the white noise spectrum of this idealized process. Nonzero charge collection time will reduce the spectral density at high frequencies, so that the spectrum must be written more generally to take these effects into account.

In these cases the spectral density will be reduced below the full shot noise level, and the Fourier-transformed correlation function \( \langle i(t) \cdot i(t') \rangle \), the current spectral density \( S_i(\omega) \), will contain a frequency-dependent Fano factor \( F_i(\omega) \leq 1 \), and thus,

\[ S_i(\omega) = 2F_i(\omega)le \quad (2) \]

\( F_i(\omega) \) is readily measurable as the frequency-dependent ratio of the spectral density of the current noise to the spectral density of the full shot noise current fluctuation.

The related time-domain Fano factor \( F_n(\tau) \) expresses the total number counting (integrated current) variance relative to the Poissonian (full shot noise) value. It is formally defined as the variance \( \sigma_n^2 \) in the number count \( n(\Delta t) \), in time interval \( \Delta t \), normalized to the Poissonian variance \( \langle n(\Delta t) \rangle \), where

\[ F_n(\Delta t) = \frac{\sigma_n^2(\Delta t)}{\langle n(\Delta t) \rangle} \]

\[ = \frac{\langle (n(\Delta t) - \langle n(\Delta t) \rangle)^2 \rangle}{\langle n(\Delta t) \rangle} \quad (3) \]

This is a more fundamental and appropriate measure of sub-Poissonian electron and photon counting variance in the small number regime encountered in mesoscopic systems discussed in Section 6. Thermionic emission noise and its suppression in semiconductor heterostructure junctions have been discussed extensively by Imamoglu and Yamamoto [13-15] and Kobayashi et al. [16].

2.2. Space Charge Suppression

When thermionic vacuum diodes are operated in the low-voltage space charge-limited regime, anticorrelated density fluctuations are introduced into the electron stream [17], with the result that the noise is suppressed on integration time scales longer than the corresponding correlation time. Figure 1 illustrates this concept of shot noise suppression by a fluctuating potential barrier. The height of the barrier fluctuates with the electronic space charge population, which itself fluctuates.
in response to random variations in thermionic emission. Low-frequency shot noise suppression by Coulomb-modulated negative feedback mechanisms of this type is a common feature of classical and quantum electronic systems and devices.

The space charge-limited thermionic diode provides a conceptual model for photonic shot noise suppression in macroscopic diodes [18, 19] and in mesoscopic systems involving the ballistic transport of electrons where Pauli exclusion can provide an additional anti-bunching mechanism [20, 21]. Space charge-limited thermionic noise has a thermal character, reflecting its thermal origin, as does the external noise current in a semiconductor junction diode operated under constant current conditions.

We shall illustrate shot noise suppression for diffusive transport associated with semiconductor junctions and discuss the transition from collective Coulomb interactions in macroscopic junctions to single-electron “Coulomb blockade” in low-capacitance mesoscopic junctions.

2.3. Semiconductor Junction Shot Noise

2.3.1. Measurable Noise Variables

In this section we shall discuss the origin of shot noise in macroscopic junctions. There are two noise currents and one noise voltage accessible to measurement which any model must account for. These are

(i) circuit or terminal current noise (the noise in the current flowing into the terminals of the semiconductor junction device);
(ii) recombination current noise (the noise in the electron-hole recombination rate as deduced from noise measurements made on the emitted light);
(iii) junction terminal voltage noise (the noise in the voltage drop measured across the terminals of the junction device).

In what follows we shall assume either a long-base homojunction diode, that is, a diode in which the width of the bulk material recombination layer is much longer than the minority carrier diffusion length, or, alternatively, a heterojunction diode in which transport is inhibited beyond the active region. Although these assumptions are not essential, they simplify the discussion considerably, since minority carriers cannot then diffuse as far as the ohmic electrodes and all of the terminal current must then be due to recombination. This situation is represented in Figure 2 [22].

2.3.2. Terminal Current Noise

The charge carrier number fluctuations in the vicinity of a macroscopic semiconductor junction \( N(t) = N + n(t) \) originate in the processes of charge injection, transport, and recombination. Figure 2a illustrates the net effect of these processes, summarized in the following rate equation [4]:

\[
\frac{dN(t)}{dt} = I(t)/e - N(t)/\tau + f(t)
\]  

In the first term on the right-hand side, \( I(t) = I + i(t) \) represents the current supplied from an external circuit. It is taken to equal the net (fluctuating) rate at which charge is injected across the space charge layer at the
Recombination Noise in Semiconductor Junctions

Figure 3. Reservoir-model noise equivalent circuits derived from Langevin rate equations (6) and (7) for an ideal junction diode with differential resistance \( R \), plus a fluctuating term \( n(t)/\tau \). Linearization of Eq. (4) yields

\[
dn(t)/dt = i_e(t)/e - n(t)/\tau + f_n(t) \tag{5}
\]

Writting Eq. 5 in the form of a state equation, using the fluctuations in junction potential \( v_m(t) \), the injected current \( i_e(t) \), and the charge \( q(t) = en(t) \) stored in the junction capacitance \( C \), as state variables [4] gives

\[
dq/dt = i_e(t) - q(t)/rC + f_i(t) = i_e(t) - q(t)/rC - i_m(t) \tag{6}
\]

and\[
Cd v_m/dt = i_e(t) - v_m(t)/r - i_m(t)
\]

where \( ef_n(t) = f_i(t) = -i_m(t) = v_m(t)/r \).

Referring to the corresponding noise equivalent circuits of Figure 3, in which the stored charge fluctuation \( q(t) = n(t)e = C v_m(t) \) and full shot noise voltage generator \( v_{m0}(t) = -i_{m0}(t)/r \) show that in this model the junction voltage fluctuation is assumed to be a direct measure of the carrier number fluctuation \( n(t) \).

From Figure 3, if the bias resistance \( R_\| \) is set to zero the full recombination shot noise current \( i_{m0}(t) = i_e(t) \) then evidently flows in the external circuit at low frequencies, in agreement with circuit noise measurements. If the bias resistance is raised, reference to either of the two equivalent circuits of Figure 3 shows that the terminal current noise is reduced accordingly, again in agreement with circuit noise measurements. From Eqs. (5) through (7) it is evident that the recombination current noise and the current noise in the external circuit, \( i_m(t) = ef_n(t) \), are both at the full shot noise level. The low-frequency mean square terminal current spectral density [4] is then, as expected,

\[
S_i(\omega) = 2(f_i^2)/\Delta f = 2e^2 \cdot (f_i^2)/\Delta f = 21e \tag{8}
\]

2.3.3. Recombination Current Noise

Setting \( R_\| \) is zero pins the junction voltage and fixes the stored electron population \( N \) so that \( dn(t)/dt = n(t) = 0 \) and the charge carriers recombine randomly with fixed mean lifetime, \( \tau \). This constitutes a Poisson point process with correlation function \( \langle f_n(t)f_n(t') \rangle = 2N \delta(t - t')/\tau \), mean rate \( N/\tau \), mean square value \( \langle f_n^2 \rangle = (N/\tau)\Delta f \), and double-sided spectral density equal to the mean recombination rate, \( N/\tau \).

Referring again to Eqs. (4) and (5), the second and third terms on the right-hand side together represent the fluctuating recombination rate; a population dependent, time-varying rate \( -\dot{N}(t)/\tau \); plus an intrinsic stochastic (Poissonian) fluctuation \( f_n(t) \). The former term represents the response of the reservoir population, \( N(t) \), and hence of the (population number-dependent) recombination rate to two sources of noise: (i) external source "pump" noise \( i_e(t)/e \) and (ii) the stochastic charge recombination process itself, as represented by the third term, the Langevin noise term \( f_n(t) \). A long-standing erroneous view held that recombination noise was represented solely by the latter (Langevin) term and therefore could not be reduced below full shot noise level, irrespective of the external loop impedance.

2.4. Recombination Current Noise Suppression

The voltage source \( v_m(t) \) in Figure 3 represents the thermal Nyquist noise voltage associated with resistance \( R_\| \). The mean square pump noise current, \( \langle i_e(t)^2 \rangle \), can be reduced to negligible proportions, and the terminal voltage can be unclamped by raising the value of the resistance \( R_\| \). If the terminal voltage is allowed to freely fluctuate, then \( N(t) \) will also fluctuate freely. If in addition the injection current noise is suppressed, then \( i_e(t) = 0 \). From Eqs. (5) and (6) and Figure 3 the recombination current fluctuation is then

\[
i_e(t) = -dn(t)/dt = en(t)/\tau - ef_n(t)
\]

This is the basis of the high-impedance method of recombination noise suppression in which both the pump noise and the recombination noise currents are suppressed [4, 23]. The detailed mechanisms for this suppression are discussed in Section 5.
2.4.1. Noise Spectra

Taking Fourier transforms, the recombination current noise spectrum in this case becomes

$$S_r(\omega) = 2\omega^2 r^2 (f_1^2) / \Delta f (1 + \omega^2 r^2)$$

$$= 2 \epsilon_0 \omega^2 / (1 + \omega^2 r^2)$$  \hspace{1cm} (10)

This has the character of single-pole high-pass filtered shot noise and vanishes in the low-frequency limit of $\omega \tau \ll 1$. Thus, shot noise suppression at frequencies $\omega \ll 1/\tau = 1/RC$ is evidently achieved by making the external impedance $R_e$ much greater than $r$, the internal differential resistance of the junction, so that $i(t) = (v(t) - v_m(t))/R_e$ vanishes in that limit. This illustrates low-frequency shot noise suppression according to the fluctuating charge reservoir model [4]. The electron reservoir number fluctuation spectrum has a complementary low-pass character with a total mean square fluctuation of $\langle N \rangle/2$, just one-half the Poissonian value.

The macroscopic junction diode diffusion noise model above adequately describes electronic noise generation in macroscopic junction devices. It has also been used as the conceptual basis for circuit models of sub-Poissonian light generation in light-emitting diodes [4] and diode lasers [23]. These account for many of the measured characteristics of macroscopic optoelectronic devices.

Noiseless charge injection into the reservoir has been assumed for a high-impedance current source. A detailed physical analysis of this noiseless injection process was first given by Buckingham and Faulkner in 1974 [24]. Kim and Yamamoto applied that model to light-emitting diodes in 1997 [26], and Kobayashi et al. in 1999 [16] refined the model further with a generalized model of photon noise suppression in commercial heterojunction LEDs. In these more detailed microscopic treatments the injection current noise contains two stochastic Langevin terms representing forward and backward charge carrier injection noise. These models will be discussed in Sections 4 and 5.

3. PHOTONIC SHOT NOISE

3.1. Photonic Recombination Noise Measurements

Measurement of the photonic shot noise in the light emitted from a semiconductor junction is represented in Figure 4. This shows the noise equivalent circuits of an ideal light-emitting diode coupled with quantum efficiency $\eta$ to a photon detector [22]. A fraction $\eta$ of the internal recombination current $i_r$ appears at the detector together with additional photonic current partition noise $i_p$. It therefore becomes possible in principle to measure the internal radiative recombination current noise in a light-emitting diode externally and to check the validity of recombination noise models. Photonic noise measurements evidently provide a direct probe of the intrinsic fluctuations in the radiative recombination rate.

Figure 4. Noise equivalent circuit representation of ideal light-emitting junction diode with junction capacitance $C$, differential resistance $r_m$, and equivalent shot noise current source $i_s$ driven from external resistance $R_e$ and coupled with quantum efficiency $\eta$ to a photon detector with partition noise current $i_p$. Reprinted with permission from [22, Fig. 4]. P. J. Edwards, Aust. J. Phys. 53, 179 (2000). © 2000 CSIRO Publishing.

3.1.1. Photonic Partition Noise

Partition noise [25] is a consequence of the lack of a one-to-one correspondence between the individual recombination events in the junction and their subsequent photon detection in the form of electron-hole pairs generated at the detector. Provided these photon deletion losses are statistically independent Bernoulli events, they can be treated like those arising at an optical beam splitter with overall transmission probability $\eta$, as in Figure 5. For photon detection probabilities $\eta, (1 - \eta)$, the additional "partition noise" introduced into particle streams has binomial statistics with variance given by $\sigma_e^2 = \eta(1 - \eta)\langle n_i \rangle$. That is,

$$\sigma_e^2 = \langle n_e^2 \rangle - \langle n_e \rangle^2 = \eta \cdot \sigma_i^2 + \eta(1 - \eta)\langle n_i \rangle$$  \hspace{1cm} (11)

Figure 5. Representation of the optical measurement of the recombination fluctuations in a semiconductor light-emitting junction. Non-ideal detection ($\eta < 1$) is represented by a notional optical beam splitter with transmission $\eta$. Reprinted with permission from [26, Fig. 3]. J. Kim and Y. Yamamoto, Phys. Rev. B: Solid State 55, 9949 (1997). © 1997, American Institute of Physics.
Equation (11) expresses the shot noise as the sum of two independent variance terms, a transmitted noise term and an additional partition noise term, which gives rise to an additional mean square partition noise current of \( \langle i^2 \rangle = \eta(1 - \eta)(\mu^2) / \nu^2 \). The spectral density of the total noise following random loss of particles from the beam can then be written in terms of the transmission factor \( \eta \), and input and output spectral Fano factors \( F_i(\omega) \) and \( F_o(\omega) \), as

\[
F_o(\omega) = \eta F_i(\omega) + (1 - \eta)
\]

(12)

This illustrates the binomial statistics which arise from the operation of independent Bernoulli selection with fixed probability \( \eta \) and which can be generalized [20, 27] to describe the additive partition noise generated in multiple-channel mesoscopic systems. Historically, it describes the shot noise in the anode current of a multi-electrode space charge-limited thermionic vacuum device as analyzed by van der Ziel [25]. Partition noise is present in bipolar junction transistors, as “quantum shot noise” [27] in mesoscopic electron transport, and as “quantum vacuum fluctuations” in lossy photonic systems [28].

When the recombination noise is completely suppressed (\( F_i = 0 \)) the detector noise, relative to the expected shot noise level, is evidently \( F_o = (1 - \eta) \). This is an expression of the extreme fragility of sub-Poissonian “quiet” light to attenuation of the light beam and reveals why it cannot provide any significant advantages in metrology or communications in the presence of even moderate losses. From Eq. (11), knowledge of the transmission term \( \eta \) allows the original electron-hole recombination noise to be recovered from nonideal optical measurements.

### 3.2. Photon Counting Statistics

The term “bunching” is used in semiclassical quantum optics [12, 28] to describe correlated density fluctuations in a photon stream, unlike a Poissonian stream having uncorrelated density fluctuations. Reduced (sub-shot noise) photocurrent fluctuations, generally at low frequencies, are associated with “anti-bunching.”

The traditional measure of particle anti-bunching is the second-order coherence function [28, 29]

\[
S^{(2)}(\omega) = \langle I_i(t) \cdot I_i(t) \rangle / \langle I_i \rangle^2 = \langle n(n - 1) \rangle / \langle n \rangle^2 = 1 + (F_o - 1) / \langle n \rangle
\]

(13)

derived from a Hanbury Brown–Twiss cross-correlation [30] or coincidence-counting measurement [28] in which a beam of particles with intensity \( I(t) = I + i(t) = I + n(t) / \Delta t \) is randomly partitioned into two beams of equal intensity \( I_1, I_2 \), with \( \langle I_1 \rangle = \langle I_2 \rangle = \langle I \rangle / 2 \), and \( \langle n \rangle = \langle I \rangle \Delta t \).

A more appropriate Hanbury Brown–Twiss measure of macroscopic sub-Poissonian fluxes is the cross-correlation function:

\[
\rho(\omega) = \langle i_i(t) \cdot i_i(t) \rangle / \langle i_i(t)^2 \rangle = (F - 1)/(F + 1)
\]

(14)

which becomes \( \rho(\omega) = -1 \) in the limit of complete noise suppression.

\( S^{(2)}(\omega) \) evidently deviates significantly from unity only in the limit of low mean count \( \langle n \rangle \), that is, for weak currents or short integration times, \( \Delta t \). It is therefore a useful parameter in cases where the statistics of small numbers of electrons or photons are to be characterized as in the case of single-electron photon turnstile emissions. Another parameter used [31] to characterize single-photon emission is the ratio of the photon number probabilities, \( L/S = P(n > 1)/P(n = 1) \). For a Poissonian source this multiple photon “leakage ratio,” ideally zero, is \( \approx \langle n \rangle / 2 \). An ideal single-photon source for quantum key distribution purposes would be characterized by \( P(n = 1) = 1, P(n > 1) = 0, \langle n \rangle = 1, F_o = 0, \) and \( g^{(2)}(0) = 0 \). The weak Poissonian sources currently used in quantum cryptographic systems are typically characterized by \( \langle n \rangle = 0.1, P(n = 1) = \langle n \rangle \exp(-\langle n \rangle) \approx 0.1, F_o = 1, \) and \( g^{(2)}(0) = 1 \).

### 3.3. Sub-Poissonian Photonic Noise

Fluctuations in photoelectron emission from a photocathode illuminated by a steady light source were for many years regarded as conceptually similar to thermionic emission fluctuations, being assumed to show the full shot noise due to Poissonian statistics. The direct detection of sub-Poissonian light from semiconductor lasers and light-emitting diodes in many laboratory experiments has emphasized the similarities between photonic and electronic shot noise processes. These processes have a common description when viewed as stochastic point processes [12]. For example, Edwards et al. [32] noted that the equation describing the transfer of quantum noise between a photon emitter and photo-detector is identical with that describing the transfer of electronic shot noise between the emitter and collector of a semiconductor bipolar junction transistor and the generation and propagation of “quantum shot noise” in a mesoscopic circuit [27].

The first measurements of sub-shot noise light from light-emitting diodes driven from high-impedance circuits were made by Tapster et al. in 1987 [3]. These showed 4% noise reduction (\( F_o = 0.96 \)), below the full shot noise level. Edwards et al. [33–35] subsequently measured 30% noise reduction (\( F_o \leq 0.7 \), implying \( F_i \leq 0.05 \)) with double-heterostructure light-emitting diodes coupled to high-efficiency large-area silicon p-i-n detectors with overall quantum detection efficiencies of
3.3.1. Noise Suppression Bandwidth

The spectral dependence of the suppressed recombination noise in the light from macroscopic and microscopic junctions has been measured by several groups [5, 16, 36-38]. The noise-suppression bandwidth was found [37] to be the same as that for external current modulation (Fig. 6), as expected for the fluctuating reservoir model. However, as first pointed out by Kim Kan and Yamamoto [38] and as shown in Figures 6 and 10, the bandwidth of the suppressed noise varies with the dc current. It approaches the recombination lifetime limited bandwidth \( B = \frac{1}{2\pi\tau_e} = \frac{1}{2\pi\tau_e C_{\text{diff}} } \) only in the limit of high pump current. The reservoir model does not account for this variation and ignores the noise due to the stochastic injection of minority charge carriers across the depletion layer into the active region. The space charge-regulated injection of bunches of carriers across the junction and consequent suppression of pump noise were identified as a "macroscopic Coulomb blockade" process [38], by analogy with single-electron Coulomb blockade [13, 14, 39]. The current dependence of the noise suppression bandwidth is accounted for by these models, which take proper account of the injection noise (Section 5).

4. MACROSCOPIC-JUNCTION TERMINAL CURRENT NOISE

4.1. Historical Background

The first theoretical treatment of terminal current shot noise generated in semiconductor junctions was given by van der Ziel in 1955 [40-42], who initially attributed its origin to the random transport of charge carriers across the depletion layer. The circuit current noise predictions of van der Ziel's analysis were confirmed for both diodes and transistors. For example, the full shot noise terminal current flows when ideal diodes are operated with a fixed terminal voltage, consistent with the noise equivalent circuits of Figure 3. However, the physical basis of the analysis was challenged by Buckingham and Faulkner in 1974 [24, 43, 44] and was shown to be inconsistent with the well-accepted Shockley diffusion model of an ideal homojunction. They attributed the shot noise to two other mechanisms operating in the neutral (bulk) regions of the structure: thermal fluctuations in minority charge carrier diffusion and fluctuations in the rates of generation and recombination. Their model, unlike van der Ziel's, is consistent with the Shockley diffusion model [45] of charge carrier transport in a forward biased homojunction.

4.2. Van der Ziel's Corpuscular Model

Van der Ziel's original corpuscular theory was developed in the context of the successful theory of shot noise in thermionic vacuum diodes. Essentially it proposed semiconductor shot noise to originate in the stochastic transport of minority charge carriers across the depletion layer. Although the physical basis for the theory was challenged, it remains an accurate empirical model of diode and bipolar transistor terminal current shot noise. Edwards and Cheung [32, 46, 47] have recently developed a "neo-corpuscular" model of semiconductor junction shot noise suppression based on the van der Ziel model which accurately models sub-Poissonian circuit shot noise and recombination shot noise in junction diodes, bipolar junction transistors, and photon-coupled transistors, using the equivalent circuit and rate equation descriptions of Section 2. In their model a shot noise-suppressed minority carrier/(photon) flux generated in a high-impedance environment is transported from the charge carrier/(photon) emitter to the charge carrier/(photon) collector, subject to stochastic partition between the base and collector terminals. This successful extension of van der Ziel's circuit-level models to the sub-Poissonian regime should not be surprising since it is consistent with the macroscopic Coulomb blockade models of diffusive and thermionic charge injection across semiconductor junctions. Indeed, it is now clear (Section 5) that in the so-called thermionic limit of weak injection, junction diode shot noise does in fact originate in transport across the depletion layer, as originally proposed by van der Ziel.

4.3. Buckingham's Diffusion Model

Buckingham and Faulkner [24, 43] and Robinson [44] developed their diffusion theory to provide a physical origin for the full shot noise observed in the terminal
current of a long-homojunction diode operated with a fixed terminal voltage. Their analysis demonstrated that homojunction terminal current shot noise did not originate in the depletion layer, but had its origin in the bulk diffusion regions of the junction.

They examined three sources of terminal current noise:

1. Depletion layer noise
2. Thermal diffusion noise
3. Generation-recombination noise

The Buckingham diffusion model replaced the "corpuscular" model of diode shot noise due to van der Ziel with a physically based model.

4.3.1. Depletion Layer Noise

According to the Shockley theory [45], extremely large forward ($I_f$) and backward ($I_b$) diffusion currents flow across the depletion layer. These are both much greater than the external terminal current, by a factor equal to the ratio of the diffusion length to the mean free path. As pointed out by Buckingham [43], these large currents are caused to maintain the equilibrium between the majority carrier population on one side of the junction and the minority carriers on the other.

The terminal current, $I = (I_f - I_b)$, is the difference between these two large currents. It is relatively so small that it can be neglected in conventional calculations of the charge carrier distribution across the junction which assume that the quasi-Fermi levels are constant across the depletion layer and are separated by the applied terminal voltage. The flux of majority carriers (electrons from the n-side of an abrupt n⁺-p homojunction, for example) which constitutes the forward diffusion current $I_f$, consists of those electrons in the bulk n⁺ material incident on the junction which have sufficient energy to surmount the junction potential barrier:

$$I_f(V_j) = (en_p D_n A/l_t) \exp(e V_j(t)/kT) = I_s(L_n/l_t) \exp(e V_j(t)/kT) = I_s A/l_t$$  \hspace{1cm} (15)

for equilibrium electron density $n_p$ at the p-type edge of the depletion layer, Einstein diffusion constant $D_n$, junction area $A$, electron mean free path $l_t$, diffusion length $L_n$, and junction voltage $V_j$.

The flux of minority carriers which constitutes the backward current then comprises all those electrons incident on the junction from the p-side and

$$I_b(t) = e n_e D_p A/l_t = e(L_n/l_t)N/r_f = I_s A/l_t$$  \hspace{1cm} (16)

According to the Shockley boundary condition, the so-called "law of the junction" [45], the minority electron concentration $n_e$ at the p-side boundary exactly follows the junction voltage $V_j$ and

$$n_e = n_p \exp(e V_j/kT)$$  \hspace{1cm} (17)

However, since the net terminal current $I = I_f - I_b$ is nonzero, Eqs. (15)–(17) cannot all be exactly correct. Buckingham [43] modified Eq. (17) by replacing $V_j$ with $(V_j - AV)$, where $AV \ll kT/e$, to take account of the (very) small voltage drop across the junction, and thereby introduced a differential depletion layer resistance

$$r_d = kT/e I_f$$  \hspace{1cm} (18)

For strongly biased homojunctions this resistance is negligible in comparison with the diffusion resistance $r_d = kT/e I$. However, it provides the relaxation mechanism whereby equilibrium is quickly restored following the passage of a carrier across the junction. The passage of an electron across the junction raises the electron concentration at the p-side edge of the junction. This departure from the steady state results in two relaxation current flows: one back across the junction, causing an increase in $I_b$, the other through the p-region, causing an increase in the forward current $I_f$. The source of the current noise due to the $(I_f + I_b)/e$ junction crossings per second, each generating transient current $\pm e\delta(t)$, is the depletion conductance $e I_f/kT \approx I_b/kT$. This is so much larger than the diffusion conductance that the external current noise is negligible. Buckingham and Faulkner showed [24] that it contributes thermal equilibrium voltage noise density $4kT r_f$ to the external current noise that is evidently completely negligible in comparison with the total shot noise voltage $2kT r_d$ due to thermal diffusion and generation-recombination noise in the bulk material.

The noise equivalent circuit of Figure 7 shows the two additional voltage noise generators $v_i(t)$ and $v_b(t)$, which represent the forward and backward diffusion noise at the junction. From Eq. (15), taking $L_n = 1 \mu m$ and $l_t = 10 nm$, the forward and back diffusion currents will each be approximately 1 A for a typical silicon homojunction.

![Noise equivalent circuit of a moderately forward-biased homojunction diode showing the depletion ($C_{dep}$) and diffusion ($C_s$) capacitances, the recombination shot noise voltage generator $v_{re}(t)$, and the depletion-layer shot noise voltage generators $v_i(t)$ and $v_b(t)$ associated with forward and backward diffusion across the depletion layer.](image-url)
carrying a current of 10 mA. The corresponding diffusion resistance \( r_d \) in Figure 7 will be several ohms, and the forward and back resistances \( r_f \) will only be on the order of \( 10^{-2} \) ohms. The independent mean square shot noise voltages \( \langle v^2 \rangle = \langle i^2 \rangle / \langle r^2 \rangle \) will then both be a factor \( l_i/L_n \) (two orders of magnitude) less than the recombination noise voltage \( \langle v^2 \rangle \) and will therefore contribute no more than 1% to the terminal shot noise current. This clearly shows that van der Ziel's corpuscular concept, that diode shot noise "originates" in random crossings of the depletion layer, is not valid in the macroscopic diffusion limit.

The Buckingham theory was extended by Yamamoto and Machida [23], who assumed a long n⁺-p single heterojunction diode with negligible bulk resistance. In their treatment the terminal voltage and therefore the junction voltage are initially fixed by the external bias source. Consequently the electron density on the p-side of the junction and at the p-side terminal are also fixed. Subject to these boundary conditions, the current noise can then be calculated from the fluctuations in the electron distribution in the p-side bulk material.

4.3.2. Thermal Diffusion Noise

Thermal diffusion leads to fluctuations in the spatial distribution of minority charge carriers. This in turn leads to relaxation currents across the junction and within the bulk material, which act to restore the unperturbed steady-state distribution. In Buckingham's treatment the perturbation due to the transport of a single electron over the mean free path distance \( l_i \ll L_n \) between collisions is equivalent to a current \( e\delta(t) \) over path length \( l_i \). The minority carrier relaxation currents which follow such an event are found by solving the time-dependent diffusion (continuity) equation for such a single "event," Fourier transforming the relaxation currents, and then randomly superposing independent "events," using Carson's theorem to obtain the resultant noise power spectral density in the terminal current. The result [23, 43] for the low-frequency, constant-voltage terminal current noise spectral density is

\[
S_i(\Omega) = 4Ae^2(D_n/L_n)[(n_p - n_{po})/3 + n_{po}/2] / (1 + R_s/R_d)^2
\]

For the case where the external bias current supply has resistance \( R_s \neq 0 \), the voltage at the p-type edge of the depletion layer and the electron density there will both fluctuate. According to the Shockley "law of the junction" [45], the electron concentration is a unique function of the depletion layer and the electron density there will both fluctuate. According to the Shockley "law of the junction" [45], the electron concentration is a unique function of the depletion layer and the electron density there will both fluctuate. Consequently the electron density on the p-side of the junction and at the p-side terminal are also fixed. Subject to these boundary conditions, the current noise can then be calculated from the fluctuations in the electron distribution in the p-side bulk material. \[ (19) \]

4.3.3. Generation-Recombination Noise

In the event of the generation or recombination of an electron-hole pair, there will be a perturbation in the minority carrier distribution, and minority carriers will flow to restore the equilibrium. A recombination-generation event at a point is modeled as the disappearance/appearance of an electron. This corresponds to a transient current \( \pm e\delta(t) \) at that point. Again, with the carrier concentration fixed at the terminal of the p-type region, solving the continuity equation for the relaxation currents, Fourier transforming, and randomly superposing the g-r events yields the terminal g-r current,

\[
S_i(\Omega) = 2Ae^2(D_n/L_n)[(n_p - n_{po})/3 + n_{po}/2] / (1 + R_s/R_d)^2
\]

4.3.4. Total Terminal Current Noise

The total current noise in the external circuit is the sum of these diffusion and g-r noise currents,

\[
S_i(\Omega) = 2Ae^2(D_n/L_n)[n_p + n_{po}]/(1 + R_s/R_d)^2
\]

The corresponding voltage noise spectral density is then

\[
S_v(\Omega) = S_i(\Omega) \cdot R_i^2 = 2Ae^2(D_n/L_n)[n_p + n_{po}]/(1/R_s + 1/r_d)^2
\]

These two results provide a physical basis for the conventional representation of the terminal noise equivalent circuits of Figure 3. The current noise spectral density is

\[
S_i(\Omega) = 2e(l + 2L_n)/(1 + R_s/R_d)^2
\]

and the terminal voltage spectrum is

\[
S_v(\Omega) \approx 2elR_d^2/(R_s + r_d)^2
\]

The forward dc current is

\[
I = (eAD_n/L_n)(n_p - n_{po}) = N\epsilon/\tau_c
\]

with excess minority electron population \( N = (n_p - n_{po})L_n \), and reverse saturation current,

\[
I_s = (eAD_n/L_n)n_{po}
\]

The dc junction current is due entirely to recombination of the excess minority charge in the p-type region, and the full short-circuit terminal shot noise spectral density,

\[
S_i(\Omega) = 2e^2L_n \cdot (n_p + n_{po})/\tau_c = 2e(l + 2L_n)
\]

is entirely accounted for by the Poissonian recombination of a fixed population of \( N + 2N_0 \) \( \approx N \) excess electrons with fixed lifetime \( \tau_c \), consistent with the heuristic model of Section 2.
The correlation function for the terminal current noise is given by

\[ \langle F(t)F(t') \rangle = 2(\Delta e) \cdot \delta(t-t')/(1 + R_f r_d)^2 \]  

which shows the full shot noise current for \( R_f \ll r_d \), as expected.

The net result of this analysis is that all of the external current noise can be accounted for in terms of diffusion and g-r noise in the bulk recombination region, where recombination provides the randomly fluctuating sink for all of the minority carriers which have crossed the depletion layer. When the external bias resistance in Figure 7 is set to zero, it is evident that at low frequencies the shot noise voltage generator \( v_{\text{sh}}(t) \) will drive the terminal current at the full shot noise level \( \langle i_{\text{sh}}^2 \rangle = \langle v_{\text{sh}}^2 \rangle / r_{d}^2 \). Two-thirds of this noise is due to thermal diffusion in the bulk p-type region, and one-third to generation-recombination noise in the same region. The noise contributed by diffusion across the depletion layer is negligible.

This directly contradicts van der Ziel's original corpuscular concept. Although the net depletion layer current (the injection current) evidently does carry full shot noise, it originates elsewhere. The Buckingham model successfully accounts for the terminal current noise for a voltage-biased diode in terms of thermal diffusion noise and generation-recombination noise components. Together these constitute the shot noise-limited "pump" current noise which flows when the diode terminals are short-circuited by a low-impedance bias circuit. Nevertheless, the "bottom line" (Poissonian) fluctuation noise in the spontaneous electron-hole recombination rate ensures that the low-frequency terminal current contains full shot noise. Spontaneous recombination therefore has prior claim as the "origin" of shot noise in a strongly forward-biased homojunction.

5. MACROSCOPIC-JUNCTION RECOMBINATION NOISE

5.1. Macroscopic Coulomb Blockade

To obtain "quiet" light (in which the intensity fluctuations are suppressed below the shot noise level), the electron-hole recombination rate fluctuations must be suppressed. For this to occur, the external current noise must also be suppressed because any current variation will modulate the population number and hence the recombination rate. This must be so since the low-frequency terminal current \( I(t) = Q(t)/\tau = eN(t)/\tau \) in the long-diode model. From a circuit-level viewpoint, the shot noise-limited pump current modulates the light output of a light-emitting diode operated at constant voltage and generates the full recombination shot noise in the emitted photon stream. Suppression of the external terminal current noise is therefore a necessary condition for recombination noise suppression.

Initial discussions of recombination noise suppression in light-emitting diodes and laser diodes [2-4] assumed that the use of a high-impedance, "constant current" bias source would automatically suppress the pump noise associated with the injection of charge carriers into the active recombination region of the diode. From Section 4 it is clear that in the case of moderately strongly pumped homojunctions, the forward/back diffusion mechanism identified by Buckingham and Faulkner acts to suppress depletion layer transport noise independently of the bias impedance. However, Imamoglu and Yamamoto [15] pointed out that for weak injection, corresponding to situations in which the diffusion capacitance becomes less than the depletion capacitance, it is not immediately clear which physical mechanisms operate to suppress the noise associated with the injection of carriers across the junction into the active region in the high-impedance regime. They identified the importance of pump noise suppression in the generation of sub-Poissonian photon fluxes and asked the question: How is the stochastic injection noise suppressed? They concluded that injection current-induced voltage fluctuations at the junction provide a negative feedback mechanism to smooth out the carrier injection rate, as in the case of space charge-limited thermionic electron current noise in vacuum diodes [17]. It was pointed out [38, 48] that the macroscopic junction voltage can only respond significantly to large number fluctuations, not to individual injection events.

This recognition of a macroscopic or collective Coulomb blockade effect therefore bridged the conceptual gap between the conjectured single-electron Coulomb blockade suppression of thermionic emission noise in mesoscopic junctions and noise suppression in macroscopic junctions.

The macroscopic junction voltage drop due to the injection of \( N_i \) carriers across a macroscopic junction is \( \Delta V_{\text{i}} = N_e/C_{\text{dep}} \). As a result the forward injection current will on average initially decrease to a fraction \( \exp[-e\Delta V_{\text{i}}/kT] \approx \exp[-N_i/\beta] \) of its initial value, where \( \beta = e^2/kTC_{\text{dep}} \). Individual carrier injection events will therefore not affect subsequent injections because the ratio of the depletion layer charging energy to the thermal energy \( \beta = e^2/kT \ll 1 \) in a macroscopic junction. The injection process might then be expected to be a random (Poissonian) point process despite the fact that the current noise in the external circuit is suppressed to thermal levels by a high-bias resistance. However, collective effects will evidently be significant when \( N_i/\beta \approx 1 \), that is, when the number of injected carriers \( N_i \) is on the order of \( 1/\beta = kTC_{\text{dep}}/e^2 \). Since the mean injection rate is \( I/e \), this establishes a characteristic time scale \( \tau_i \) such that \( \tau_i/e = N_i = 1/\beta \), on which the junction voltage responds and provides negative feedback to regulate the injection rate where

\[ \tau_i = e/\beta I = \tau/\beta = kTC_{\text{dep}}/le = r_d C_{\text{dep}} \]  

(29)
The charging energy $N_e e^2/2C_{dep}$ at the junction associated with the injection of $N_e$ electrons into the active layer drops the junction potential and raises the Coulomb barrier against subsequent charge injection. This occurs if $N_e e^2/2C_{dep} > kT$. It results in a sub-Poissonian stream of anti-bunched electrons on a characteristic time scale $\tau_e = r_{C_{dep}} = kT C_{dep}/e$. As the injection current is lowered, this time scale lengths and exceeds the recombination time. The corresponding bandwidth over which noise suppression occurs is then $B = 1/2\pi T C_{dep}$. For low injection this will be less than the measured recombination bandwidth, as shown in Figure 6. In this situation the shot noise originates in the forward transport of electrons across the depletion layer, as in van der ZieJ's original model. With high-impedance bias it can be suppressed by the space-charge-induced junction voltage fluctuations, since recombination is relatively rapid, so that there can be no significant carrier storage and associated spontaneous noise.

### 5.1.1. Junction Voltage Dynamics

In their formal treatment of noise in $p^n$ heterojunction light emitters, Kim and Yamamoto [26] decompose the diode pump current noise into two components: a Markovian carrier injection current and a band-limited current fluctuation regulated by charging effects at the junction. It is this latter current which partially cancels the low-frequency components of the shot noise arising from the Markovian carrier injection. The junction current is carried by electrons injected from the n layer across the depletion layer into an active p-type layer with subsequent radiative recombination. As in the Shockley model [45], no carrier recombination is assumed to take place within the depletion region.

In the absence of thermionic emission or diffusion across the junction, spatial integration of Poisson's equation across the junction leads to a relation between the depletion capacitance $C_{dep} = [dQ/dV] = eA/s_n$, the external loop current $I_{ext}(t)$, and the junction voltage $V_j(t) = \varphi_n(t) - \varphi_p(t)$, where $s_n$ is the width of the depletion layer in the n$^+$ region and $\varphi_n$ and $\varphi_p$ are the quasi-Fermi levels in the two layers,

$$dV_j/dt = I_{ext}(t)/C_{dep} \tag{30}$$

The current can then be written in terms of the rate of change of depletion layer width as

$$I_{ext}(t) = -eN_d A dx_n/dt \tag{31}$$

The authors point to three mechanisms which modulate the width of the depletion layer and thus vary the junction voltage. These are the external current (which compresses the layer); the forward injection of carriers (which increases the space charge and so increases the width); and the backward injection of electrons from the active region across the depletion layer into the n-type region (which decreases the layer width). The resulting fluctuations in $C_{dep}$ have a negligible effect on the charging current compared with the resulting junction voltage changes $V_j(x_n)$ and are neglected in their treatment.

### 5.2. The Stanford (Macroscopic Coulomb Blockade) Model

The forward injection of carriers across the junction is modeled [26] as the thermal diffusion of electrons from the n-type layer to the active p-type region, although the authors point out that a thermionic emission transport model gives similar results. A large flux of electrons will also diffuse back in the opposite direction, and the net diffusion current is given by the difference between these forward and backward injection currents, as in the Shockley homojunction model discussed by Buckingham and Faulkner [24].

#### 5.2.1. Injection Noise

The noise associated with the injection process will therefore depend on the dynamics of these two currents, both of which are usually very much higher than the net diffusion current in a strongly biased homojunction. When extended to include the transport of charge carriers across the junction, Eq. (30) becomes

$$dV_j/dt = I_{ext}(t)/C_{dep} - (I_1(t) - I_b(t))/C_{dep} \tag{32}$$

The external current $I_{ext}(t)$ therefore consists of the charging current $C_{dep} dV_j/dt$ plus the net charge transport current, $I_1(t) - I_b(t)$. For a standard macroscopic junction LED operating at a temperatures above a few kelvins, the factor $\beta = e^2/kT C_{dep}$ will be much less than unity, and single-injection events will not have any significant effect on the forward current. In this limit the forward diffusion current will then comprise a stochastic injection current plus a slowly varying current averaged over a large number of injection events $N_e$, typically on the order of $10^4$, which varies with the time-dependent junction voltage.

The stochastic injection component can be represented by a Langevin noise term $F_I$ in the equation for the time-varying forward injection current,

$$I_1(t) = (eD_n A n_{ph}/L_n) \exp[eV_j(t)/kT] + eF_I \tag{33}$$

with corresponding Markovian rate correlation function

$$\langle F_I(t) \cdot F_I(t') \rangle = 2\langle I_1(t) \rangle \delta(t - t')/e \tag{34}$$

The correlation function for the backward current,

$$I_b(t) = (eD_n A/L_n) [n_{ph} + (N/A L_n) \cdot \exp(-t/L_n)] + eF_{b1}(t) \tag{35}$$
is written in the same way. The external current from bias voltage $V_e$ with resistance $R_e$ is written as $I_{ext}(t) = (V - V_f)/R_e + eF_{ext}$, with

$$ (F_{ext}(t)F_{ext}(t')) = (4kT/e^2R_e)\delta(t - t') $$

(36)

5.2.2. Langevin Equations

Equation (32) then becomes

$$ dV_f/dt = (V - V_f)/R_eC_{dep} - I_1(t)/C_{dep} + I_b(t)/C_{dep} $$

$$ + (e/C_{dep})[F_{ext} - F_t + F_0] $$

(37)

This equation then allows the suppression of injection noise to be calculated in the “diffusion” regime appropriate to macroscopic junctions. The corresponding equivalent circuit model shown in Figure 2 of [26] illustrates this but omits the important depletion resistance introduced by Buckingham [43] and shown here in Figure 7.

Three simultaneous Langevin rate equations have to be solved to calculate the total recombination noise. The first of these is Eq. (37): when linearized about the steady-state values $N_{av}, V_{av}$, with $n = N - N_{av}$ and $V_f = V_f - V_{av}$, it addresses the relations between the charging current fluctuations and the current noise in the external circuit, the forward and back diffusion noise, the Langevin thermal terminal current noise, and the forward-diffusion and back-diffusion noise terms, respectively:

$$ C_{dep} dV_f/dt = -V_f/R_e - C_{dep}V_f/T + n/n_o + e[F_s - F_i + F_b] $$

(38)

The second linearized noise equation describes the relation between fluctuations in the rate of change of the reservoir population, the recombination current fluctuations, the forward- and back-diffusion current fluctuations, and the Langevin recombination current, forward and back noise terms, respectively:

$$ e dn/dt = -en/T + C_{dep}V_f/T + n/n_o - e[F_i - F_t + F_b] $$

(39)

The third linearized equation expresses the fluctuating photon current noise in terms of the recombination rate fluctuation and the Langevin recombination noise and partition noise terms, respectively, as

$$ s(t) = \eta[n(t)/\tau_t + F_i^2] + F_{ph} $$

(40)

Since it is relatively straightforward to correct the photon noise measurements to take account of nonideal photon collection and detection, this last equation can be simplified by taking the overall photon collection efficiency $\eta = 1$, so removing the Langevin partition noise term $F_{ph}$. Also, one of the most accessible measurable model parameters, the noise suppression bandwidth (NSB), is independent of the detection efficiency $\eta$. This last equation then becomes an equation for the fluctuating electron hole recombination rate,

$$ s(t) = n(t)/\tau_t + F_i $$

(41)

with $(F_s(t)F_s(t')) = 2(N + 2N_{ph})\delta(t - t')/\tau_t$

These equations contain four characteristic times. These are the external current charging time $\tau_{RC} = R_eC_{dep}$, the minority carrier recombination lifetime $\tau_{r}$, and two characteristic depletion layer diffusion times, $\tau_d$ and $\tau_{in}$, defined by

$$ 1/\tau_d = (1/C_{dep})(dI_1(V_f)/dV_f)|_{V_f=V_o} = eI_1(V_o)/kTC_{dep} $$

$$ = 1/\tau C_{dep} $$

(42)

$$ 1/\tau_{in} = (1/e)dI_b(N)/dN|_{N=N_o} = eI_b(N_o)/(N_o + N_{ph}) $$

$$ \approx eI_b(N_o)/N_o $$

(43)

These last two time constants characterize the time scales for thermal fluctuations in the diffusion process. They are both much shorter than either the external charging time constant, $R_e(C_{dep} + C_{diff})$, or the recombination lifetime, $\tau_r$.

5.2.3. Recombination Noise Spectra

Fourier transforming these equations [26] yields the spectral density of the recombination rate fluctuations, the recombination current noise. The frequency-dependent Fano factor is

$$ F_s(\Omega) = S_s/2I/e = 1 - \chi(\Omega)(1 - [R_d/(R_d + R_e)]^2) $$

(44)

where $\chi(\Omega) = 1/[(1 - \Omega^2\tau_d^2 + \Omega^2(\tau_r + \tau_t)^2)$, and the characteristic injection time,

$$ \tau_i = e\beta I = kTC_{dep}/eI $$

(45)

As before, the Fano factor is unity at all frequencies for the constant voltage case ($R_i = 0$). For the high-impedance case,

$$ F_s(\Omega) = S_s/2I/e = 1 - \chi(\Omega) $$

(46)

indicating complete noise suppression in the low-frequency limit and a trend toward full shot noise at high frequencies, in agreement with the measurements.

Figure 8 shows recombination noise spectral densities calculated [26] for a moderately biased LED with the bias current arbitrarily chosen to make $\tau_i = \tau_r(C_{dep} = C_{diff})$. Injection noise and spontaneous emission (recombination) noise are then both present. The bias is varied from constant-voltage to constant-current conditions by varying the bias resistance $R_e$. As expected, the low-frequency noise is most strongly suppressed when the junction voltage is free to fluctuate freely under high-impedance conditions. The NSB is independent of the value of the bias resistance but strongly current dependent.
junction voltage fluctuation, and terminal current noise in a macroscopic junction in which the recombination lifetime is much shorter than the injection time scale and in which the thermal noise is neglected: 

\[ \tau_i = \frac{kT}{e} \]

Figure 8. Suppression of macroscopic junction recombination noise: spectral density calculated for a forward bias current chosen to make the depletion and diffusion capacitances equal. The bias is varied from the constant voltage to the constant current regimes by raising the bias resistance. Reprinted with permission from [26, Fig. 4a]. J. Kim and Y. Yamamoto, Phys. Rev. B: Solid State 55, 9949 (1997). © 1997, American Institute of Physics.

Figure 9 from [26] sketches the junction behavior for constant-voltage operation and constant-current operation. For low-impedance constant-voltage bias Figure 9a shows individual recombination events signaled by transient junction voltage drops and terminal current spikes on a time scale of \( \tau_{pe} = R C_{dep} \), corresponding to a situation in which the recombination time \( \tau_r = \tau_{sp} \) is evidently much shorter than the injection time scale \( \tau_i \).

The junction voltage recovers quickly by relaxation through the external bias circuit, injection events remain independent, and the full terminal and photon recombination shot noise results. Figure 9b is intended to indicate the junction voltage response to macroscopic injection current fluctuations with consequent electron anti-bunching, sub-Poissonian variance in the photon flux, and sub-shot noise terminal current for high-impedance bias. (A realistic representation of the high-current macroscopic junction reservoir dynamics in which \( C_{dep} \gg C_{diff} \) would show the thermal junction noise voltage \( kT/e > e/c \), with reduced photon count fluctuations and reduced terminal-current shot noise over times \( \Delta t > \tau_r = \tau_{sp} \). A sketch including the thermal noise can be found in [9].)

When the bias current is very low, \( \tau_i \gg \tau_r = \tau_{sp} \), \( C_{dep} \gg C_{diff} \), and the NSB, \( B_{3dB} = \Omega_{3dB}/2\pi = 1/2\pi\tau_i \), is well below the limit set by the carrier lifetime. In this so-called thermionic limit, there is little accumulation of minority charge in the diffusion reservoir, so that the back current and its noise vanish, as does the spontaneous emission noise. Equations (38) and (39) can then be simplified to exclude these terms, and the diode noise equivalent circuit (Fig. 7) then contains only the depletion capacitance \( C_{dep} \), forward injection resistance \( \tau_i = kT/eF = kT/eF = \tau_{sp} \), and forward injection shot noise voltage \( \eta_i(t) \).

As the current increases the NSB approaches the recombination limit \( \tau_i^{-1} = \tau_{sp}^{-1} \), and

\[ \Omega_i^2 \tau_i \approx 1 \quad \text{and} \quad \chi(\Omega) = 1/\Omega^2(\tau_i + \tau_r)^2 \]  (47)

The corresponding 3-dB bandwidth is then

\[ B_{3dB} = \Omega_{3dB}/2\pi = 1/2\pi(\tau_i + \tau_r) = 1/2\pi(\tau_r + kT C_{dep}/eF) \]  (48)

The corresponding noise equivalent circuit comprises the differential resistance \( kT/eF \) and the junction and diffusion capacitances connected in parallel as in Figures 3 and 4.

The noise suppression bandwidth is proportional to current in the low-current regime and increases with current to a limiting bandwidth set by the effective recombination lifetime. The shot noise is then entirely due to spontaneous recombination noise in the p-type bulk region. The NSB calculated above is in general agreement with heterojunction LED measurements. The macroscopic junction capacitance required to correctly model the circuit and photon noise measurements will evidently be the depletion ("injection") capacitance at low currents and the diffusion ("reservoir") capacitance at high currents. Injection noise will dominate at low currents, as in van der Ziel's model.

For a typical heterostructure LED operating with a current of, say, 1 mA at room temperature, \( \tau_i \) is on the order of 10 ns. If this were the longest characteristic time scale in the system, that is, if the recombination lifetime were much shorter than 10 ns, then we would expect the 3-dB bandwidth over which the recombination photon noise is suppressed to be given by \( B = 1/2\pi\tau_i \approx 17 \text{ MHz} \). This is a typical situation for weakly pumped
standard heterojunction LEDs: the injection of minority carriers into the active region is quickly followed by recombination.

5.3. The Canberra (Fluctuating Reservoir) Model

As the pump current is increased, the injection time shortens until the fixed recombination lifetime becomes the bandwidth-limiting factor [4] and

\[ B_{16B} = 1/2\pi\tau_i \] (49)

Reservoir noise will then be the main noise source at high currents, as in the Buckingham, Stanford, and Canberra models [4, 24, 26]. The correlation between junction voltage and carrier number fluctuations can be calculated from the linearized Langevin equations for the Stanford model. For high-bias currents, the junction relaxation times \( \tau_f, \tau_p \ll \tau_i, \tau_r \). The result is a perfect correlation between the junction voltage and the carrier number in the active region of a heterojunction and the bulk diffusion region of a homojunction, independently of the value of the bias resistance. According to Kim and Yamamoto [26] the physical reason for this is that the fast forward- and back-injection currents maintain a one-to-one correspondence between the junction voltage and the carrier number. From an equivalent circuit viewpoint, as seen in Section 4, the depletion region is effectively short-circuited by the very low depletion resistance. In the equivalent circuit of Figure 7, \( r_f, u_f \), and \( u_b \) may then all be neglected (as in Fig. 3), and the injection noise then makes no contribution to either the internal recombination noise or the terminal current shot noise, independently of the value of the bias impedance.

For high-bias currents the conceptual noise suppression model [4] is a reservoir of fluctuating minority charge from which there is a reservoir level-dependent leakage flow of photons exhibiting high-frequency "turbulence" at the shot noise level but which at low frequencies shows smooth "laminar" flow characteristics. The low-frequency turbulence is suppressed as a result of fluctuations in the reservoir level which take place on a relatively long time scale compared with the spontaneous recombination "leakage." The level of the reservoir is kept constant over time intervals that are short compared with the recombination time. To place the macroscopic Coulomb effect in perspective, it should be noted that it is the recombination rate fluctuations which, in the end, directly determine the optical noise. The measured recombination current noise is just the fluctuation in the recombination rate. In the limit of strong pumping when the reservoir dynamics dominate, it is the fluctuation in the quantity \( N(t)/\tau_i \). In the limiting case of weak injection, the reservoir of stored minority charge remains empty because of relatively fast recombination, and the recombination dynamics are completely determined by the injection noise.

![Figure 10. Measured dependence of the noise suppression bandwidth of a high-speed heterojunction LED on bias current. Experimental results are compared with the theoretical predictions of the Stanford model [26]. Reprinted with permission from [16, Fig. 6], M. Kobayashi et al., Phys. Rev. 60, 16686 (1999). ©1999, American Institute of Physics.](image)

The Stanford model has been extended by Kobayashi and co-workers [16], who have presented a more detailed model appropriate to heterojunction LEDs.

5.4. The Hiroshima (Backward Pump) Model

5.4.1. Heterojunction NSB Measurements

The publication of the Stanford analysis by Kim and Yamamoto [26] stimulated further theoretical and experimental work on sub-Poissonian light emission from macroscopic heterojunctions. Kobayashi et al. [16], working at Hiroshima University, noticed that their measurements of the noise suppression bandwidth obtained with commercially available heterojunction LEDs were not well fitted by the Stanford model.

Figure 10 shows a typical comparison between the measurements of the NSB current dependence and the predictions of the Stanford model as summarized in Eq. (48). The Stanford model appears to be a good fit in the low and the high, but not the intermediate current ranges. The Hiroshima group questioned whether a macroscopic diffusion model was appropriate for heterojunction light emitters. They examined the transition from thermionic emission at low pump currents to the macroscopic diffusion limit at high currents. They were able to account for their experimental results over a wide range of pump currents with a unified model of injection and recombination based on a detailed reexamination of the backward pump process.

Figure 11 shows a typical energy-band diagram for a commercial p+-N heterojunction light emitter. The diagram takes into account grading of the bandgap at the interfaces between the N-type wide gap and p-type active regions and between the active region and the P-type wide gap region on the right-hand side. The LED used in the Hiroshima work behaved essentially as a single p-n junction device: the active region layer and the adjacent P-type layer were much more highly

...
ward (FP) mean injection currents. The forward current so that the fluctuation in the back current becomes

5.4.2. Backward Pump Modelling

fusion limit previously discussed in which the pump and recombination processes are strongly coupled and

\[ C_{\text{Langevin}} = \text{function of} \quad \frac{1}{1 - a_{\text{d}}}, \]

dep., so that most of the junction voltage drop fell across the p-N junction. As in the Stanford analysis, the device was therefore characterized by electron transport from the N-type region to the p-type active layer and by the recombination dynamics within that active region. Unlike the situation envisaged in the Stanford model, which was based on the classical Shockley/Buckingham diffusion model devised for homojunctions, Kobayashi et al. [16] suggest that the injected electrons may not completely thermalize to the conduction band edge because of hot carrier effects and band-tail filling. As a result, electrons with energies greater than the band edge because of hot carrier effects and band-tail filling. The jump voltage was estimated to be less than 5 \times 10^{-16} \text{cm}^{-3} at the highest pump current (20 mA), much lower than the hole concentration, permitting a linear recombination model.

5.4.2. Backward Pump Modelling

Kobayashi et al. [16] define a parameter \( a_{\text{d}} = I_{\text{BP}}/I_{\text{FP}} \) to characterize the ratio of the backward (BP) and forward (FP) mean injection currents. The forward current is then \( I_{\text{FP}}/(1 - a_{\text{d}}) \), and the back current is \( eN_{\text{o}}(1 - a_{\text{d}}) \). They also define a differential ratio \( a_{\text{d}} = dI_{\text{BP}}/dI_{\text{FP}} \), so that the fluctuation in the back current becomes \( \Delta I_{\text{BP}}(t) = e \cdot \Delta N_{\text{o}}(t)/\alpha_{\text{d}}(1 - a_{\text{d}}) \). The spectral Fano factor for the recombination rate, found by solving the Langevin equations, is then a function of \( \alpha_{\text{d}}, \alpha_{\text{e}}, \) and \( \tau = C_{\text{dep}}/\alpha_{\text{d}}, \) the junction response time.

The case \( a_{\text{d}} = a_{\text{e}} = 1 \) recovers the macroscopic diffusion limit previously discussed in which the pump and recombination processes are strongly coupled and for which \( f_{\text{diff}} = 1/2\pi(\tau_{\text{d}} + \tau_{\text{e}}) \) (Eq. (48)). For the low-injection case, \( \alpha_{\text{d}} = \alpha_{\text{e}} = 0 \), the back current is zero and the forward injection and recombination processes are viewed as separate cascaded processes, as in the thermionic emission limit treated by Imamoglu and Yamamoto [14].

Figure 12 shows a comparison between their NSB data and the predictions of the Stanford model for \( \alpha_{\text{d}} = \alpha_{\text{e}} = 0, 0.5, 1 \). This comparison led the Hiroshima group to suggest a model in which the backward pump rate increased monotonically but in a nonlinear fashion from zero at low currents to equality with the forward injection rate at high currents.

The pump dependence of the backward to forward (BP/FP) injection ratios needed to account for the measured suppression bandwidths was independently found by measuring the static current/voltage characteristic of the diode from

\[ I \approx (eD_{\text{m}}\Delta n_{\text{p}}/L_{\text{m}})\left(\exp[V_{\text{p}}/nkT] - 1\right) \]  \( \alpha_{\text{d}}(V_{\text{p}}) \)

(50)

where \( n \), the ideality factor, is slightly greater than 1. The differential resistance can be similarly rewritten as

\[ r_{\text{d}} = (nkT/e)\left[1 - \alpha_{\text{d}}(V_{\text{p}})/1 - \alpha_{\text{e}}(V_{\text{p}})\right] \]  \( \alpha_{\text{d}}(V_{\text{p}}) \)

(51)

The Hiroshima group found that \( \alpha_{\text{d}}(V_{\text{p}}) \) was greater than \( \alpha_{\text{e}}(V_{\text{p}}) \), consistent with the expectation of a backward injection component increasing with current. They successfully modeled the current dependence of the noise suppression bandwidth, using the BP parameters (\( \alpha_{\text{d}}, \alpha_{\text{e}} \)), determined from the static \( I-V \) characteristics of Eqs. (50) and (51). Figure 13 shows the good agreement obtained between the measured and BP model results for values of \( \alpha \approx 0 \) (the thermionic limit) and \( \alpha \approx 1 \) (the diffusion limit). This work conclusively confirmed the importance of the backward injection process. The Hiroshima group has subsequently demonstrated noise suppression over bandwidths exceeding 300 MHz.
Recombination Noise in Semiconductor Junctions

5.4.3. Constant-Voltage Noise Suppression

An interesting feature of the work of the Hiroshima group is their measurement of a small degree of shot noise suppression for a constant-voltage heterojunction LED. They have presented a theoretical model of this phenomenon and a detailed account of the forward and backward injection process in a heterojunction light-emitting diode. From the preceding discussions, neither the injection noise nor the recombination noise can be suppressed by Coulomb effects unless the junction voltage is allowed to fluctuate freely. The mechanism for constant voltage noise suppression is therefore of theoretical interest and may also be of practical interest because of the difficulty of achieving high-impedance noise suppression of mesoscopic junction recombination noise in the small photon number regime [48].

The essence of the proposed constant-voltage model is the existence of a small nonlinear backward injection dependence on the pumping current. A nonlinear relation between the forward and backward processes results in a nonlinear relation between the forward injection current and the photon emission rate, a tendency toward saturation. The Fano factor will then be greater than or less than unity, the Poissonian (full shot noise) value, depending on whether the emission rate/injection rate relation is super- or sublinear with

$$F_0 = 1 - 2(\alpha_d - \alpha_e)(1 - \alpha_d)/(1 - \alpha_e)$$  \hspace{1cm} (52)

yielding sub-Poissonian noise for $\alpha_d > \alpha_e$.

Room temperature measurements indicate [49] a small noise reduction of several percent below the full shot noise, in reasonable agreement with that predicted from independent measurements of the BP parameters. Two possible mechanisms for the nonlinearity have been suggested. These both operate in the heavily p-doped active region: the "hot electron" effect, in which significant electron heating occurs at currents greater than several milliamps, and conduction band state filling, which accelerates the rise of the quasi-Fermi level with pump current density. Unlike the macroscopic and mesoscopic Coulomb blockade effects, this new constant-voltage non-Coulombic recombination noise reduction mechanism has the potential to provide weak sub-Poissonian photon fluxes on time scales shorter than the recombination lifetime limit.

6. MESOSCOPIC-JUNCTION RECOMBINATION NOISE

6.1. Transition from Collective to Single-Electron Coulomb Blockade

From the discussion in Section 5, the passage of a single electron across the depletion layer of a high-impedance-biased semiconductor junction will drop the junction voltage by an amount $\Delta V_j = e/C_{dep}$. As a result the forward injection current will on average initially decrease to a fraction $\exp[-e\Delta V_j/kT] = \exp[-\beta]$ of its initial value, where $\beta = e^2/kTC_{dep}$. If $C_{dep}$ and $T$ are reduced to make $\beta > 1$, the potential drop should eventually become sufficiently large relative to the thermal fluctuations to block the subsequent passage of individual electrons, as in tunneling junctions [39]. This requires a sub-micron-sized semiconductor nanojunction with a junction capacitance on the order of $10^{-16}$ F cooled to liquid helium temperatures.

Unlike the case of the macroscopic junctions discussed in Section 5, the rate equations describing charge transport across mesoscopic junctions generally cannot be linearized. It is therefore necessary to simulate the injection dynamics by Monte Carlo methods. Imamoğlu et al. [13–15] simulated a p-i-n heterojunction driven by an ideal constant current source. They also showed analytically that the variance $\sigma^2_a$ in the number of thermionically injected carriers measured in time $\Delta t$ would equal $1/\beta$ and be independent of both the measurement time $\Delta t$ and the average number of injected carriers $\langle N \rangle = I\Delta t/e$. Consequently, from Eq. 3, the Fano factor,

$$F = \sigma^2_a/\langle N \rangle = e/(\beta I\Delta t) = \tau_e/\Delta t$$  \hspace{1cm} (53)

with $\tau_e = r_0C_{dep}$, decreases indefinitely as the measurement time $\Delta t$ increases. The spectrum of the carrier injection noise is therefore strongly suppressed, and sub-shot noise current spectral densities will be observed at frequencies $f \ll 1/2\pi\tau_e$, as already described. The second-order correlation $g_2^{(2)}(0)$ will not be significantly different.

Figure 13. Measured noise suppression (squeezing) bandwidths of high-impedance biased DH LED controlled by the addition of external capacitances $C_p$. For two different pump currents, plotted against the reciprocal of the effective junction response time $\tau = r_0(C_{dep} + C_p)$. Reprinted with permission from [16, Fig. 13], M. Kobayashi et al., Phys. Rev. Lett. 60, 16086 (1999). © 1999, American Institute of Physics.
from unity for a macroscopic junction since \( \langle n \rangle \) is large, being on the order of \( N_i \approx 10^9 \). Moreover, as the injection current is reduced, the bandwidth of the suppressed noise \( (1/2\pi\tau_1) \) will continue to contract, implying ever longer integration times to maintain low noise levels. Full shot noise \( (F_n(\Delta t) = 1) \) will result if the counting time \( \Delta t \) is made much shorter than \( \tau_1 \).

Mesoscopic single-electron blockade effects might be expected to appear when \( \beta \geq 1 \) in the form of a peak in the current noise spectral density at frequency \( f = 1/e = \tau^{-1} \), corresponding to a peak in the arrival time interval distribution. When the blockade ratio \( \beta \gg 1 \), the strong junction voltage response to individual charge injection events can result in a more or less regular sequence of injection events separated by time interval \( \tau = 1/e \). This implies that the junction response time \( \tau_1 \ll e/1 \), the mean interval between successive injections. These expectations were confirmed in the Monte Carlo simulations.

6.2. Single-Electron Coulomb Blockade

6.2.1. Thermionic Emission Blockade

Although the Monte Carlo simulations reveal a smooth transition between the macroscopic collective-electron and the mesoscopic single-electron blockade effects in \( p-i-n \) heterojunctions, attempts to implement the single-electron blockade in mesoscopic junctions have not been successful. The reason for this failure appears [48] to be the requirement for high-impedance biasing of the mesojunction with \( R_s \gg e/IE_{dep} \) and \( R_s \gg kT/eI \).

In 1992 Imamoglu et al. [13] proposed the use of a high-impedance biased \( p-i-n \) mesojunction to generate a regulated single-photon stream. Assuming thermionic electron emission, they set down the conditions which needed to be satisfied if the single-electron Coulomb blockade process was to operate successfully to produce a regular photon stream. These were (i) that the depletion capacitance and the junction temperature should both be small enough to ensure that \( \beta = e^2/kT = \tau/\tau_1 \gg 1 \); (ii) that the bias circuit impedance should be large enough to make the junction voltage recovery time long compared with the single-electron charging time, \( R_sE_{dep} \gg e/1 \), as shown in Figure 9b; (iii) that the bias circuit impedance should be large compared with the quantum channel resistance, \( R_s \gg h/2e^2 \); and (iv) that the recombination time should be short compared with both the single-electron charging time \( \tau = e/1 \) and the characteristic junction fluctuation time \( r_0E_{dep} = kT/eIE_{dep} \). Taken together, these conditions also mandate a high circuit impedance relative to the diffusion resistance of the junction, \( R_s \gg kT/eI \).

For \( \beta = 5 \) \( (T = 4K) \), a mean electron emission rate of less than \( 10^{10} \text{ s}^{-1} \) would imply a current of less than \( 1.6 \text{nA} \), an unrealistically high \( R_s \gg 10^{12} \Omega \), and an unrealistically low parasitic capacitance compared with the junction capacitance, \( C_{dep} < 10^{-16} \text{ F} \). There has therefore been some effort to develop single-electron/photon systems which can be operated at constant voltage.

6.2.2. Resonant Tunneling Blockade

A double-barrier resonant tunneling \( p-n \) junction diode typically consists of a GaAs quantum well (QW) sandwiched between an \( n \)-type AlGaAs electron emitter and a \( p \)-type AlGaAs collector but separated from them by quantum-tunneling barriers. Because of quantum confinement, quasi-bound energy levels are formed in the central QW. With the application of a forward bias voltage to the diode, the Fermi level in the emitter can be matched to a bound state level in the QW to produce resonantly enhanced tunneling.

Constant-voltage shot noise suppression is due to several different mechanisms. Following a barrier crossing, Pauli exclusion will inhibit further tunneling to the same sublevel for the duration of its occupancy. In addition, a Coulomb blockade effect operates because of the build up of space charge in the well. Following the ingress of \( N_e \) electrons, the barrier height drops by \( \Delta V = N_e(e/C_p + C_p) \) on the \( n \)-side and rises by the same amount on the \( p \)-side, where \( C_p \) and \( C_n \) are the two barrier capacitances. This is equivalent to an upward shift of the QW energy levels of \( Ne^2/(C_p + C_n) \). This shift will be sufficient to detune the system off resonance following the ingress of a single electron or hole, providing two obvious conditions are satisfied. These are (i) the usual single-electron blockade condition: \( \beta = e^2/kT \gg 1 \); and (ii) the single-electron charging energy must greatly exceed the sub-band level width, \( e^2/kT \gg h\Delta\omega/2\pi \).

Unlike thermionic electron blockade in mesoscopic heterojunctions, tunneling blockade can be realized with low-impedance, voltage-driven bias circuits and is therefore not subject to the practical problems preventing the realization of single-electron thermionic and diffusion blockade.

6.3. Heralded-Photon Generation

Control over individual photon emission would permit the generation of well-defined photon numbers (photon number states) at specified times. Such "heralded" photon number state emissions in which single photons or multiple photons are generated on demand have applications in quantum communications and cryptography. Single-photon devices also have applications in quantum computing, quantum metrology, and fundamental tests of quantum mechanical theory.

In 1992 Imamoglu and Yamamoto [14] discussed both the high impedance and voltage bias of a mesoscopic double-barrier resonant tunneling junction. They showed that, provided the single-electron blockade conditions were satisfied, constant-current high-impedance biasing could generate a periodic sawtooth junction voltage waveform, indicating periodic electron tunneling at the
single-electron charging frequency of $r^{-1} = I/e$ into the active recombination layer, leading to periodic emission of single recombination photons. Recognizing the difficulties associated with high source impedance operation, they examined constant voltage tunneling of electron and (preferential) hole tunneling into a central well in which the hole concentration was maintained at a high level by modulation doping to ensure fast radiative recombination. They found that if the electron and hole resonance voltage bias conditions could be separated by $\Delta V = e/(C_0 + C_p)$, a bias voltage could be applied to import a single electron into the active QW through the n-side tunnel, so raising the well energy by $e^2/C$ to block further electron tunneling but enhancing the resonant entry of a hole. With the entry of the hole, the QW would return to its original state, recombination would occur, and a sub-Poissonian photon stream would be generated because of the dead time between successive photon emissions.

It was later realized [10] that the application of an appropriately chosen alternating "clock" bias voltage, superimposed on a steady DC level, should enable the periodic transport of electrons and holes into the well to produce a regulated single-photon stream.


Since the double-barrier p-i-n mesoscopic junction can accommodate the separately clocked blockade of electrons and holes, provided the electron-hole recombination time can be made much shorter than the clock period, then control over the emission times of single photons becomes possible, resulting in heralded photon emission. Kim et al. [10] successfully demonstrated the first electron-hole-photon turnstile in 1999, utilizing sequential single-electron and single-hole blockade in a cooled GaAs/AlGaAs mesoscopic double-barrier p-n junction.

The central quantum well of this turnstile (Fig. 14) is connected on each side to p-type and n-type wells by tunnel barriers. The lateral size of the junction post structure is kept small ($\approx 100 \text{ nm}$) by electron beam lithography and plasma etching techniques to raise the single-carrier charging energies $e^2/C$ sufficiently to permit blockade operation at a temperature of 50 mK. At a given voltage bias level the resonant tunneling condition is satisfied for an electron from the n-side QW. The Coulomb interaction between this electron and the other (=10) electrons in the central QW raises the sub-band energy above the Fermi level of the n-side QW, thereby blocking further electron tunneling. At this bias level resonant-hole-tunneling cannot occur because the Fermi level in the p-side QW is higher than the central well hole sub-band energy. However, when the bias voltage is increased further, a single hole can resonantly tunnel into the central well, and hole blockade prevents further hole ingress since the Coulomb interactions between the hole and the electrons in the central well detune the resonant hole tunnel from the p-side and reopen the n-side tunnel. By modulating the turnstile with a square wave voltage with amplitude chosen to alternately open each of the two tunnels, a single electron and a single hole can be periodically injected into the central well, where the recombination takes place with a 25-ns lifetime, which is short compared with the modulation period. Consequently no more than one photon is generated in each modulation cycle.

**6.3.2. Semiconductor Quantum Dot Emission**

A single quantum dot can also serve as a source of heralded single photons. Pelton et al. [51] grew InAs in a GaAs matrix with the use of molecular beam epitaxy to form strain-induced nanometer-scale islands which were then etched to make submicron mesa structures containing on average less than one dot each. The quantum dots were excited by laser pulses to produce excitons, the wavelength of the resulting recombination photons being uniquely determined by the number of carriers present on a dot. It was therefore possible to isolate the single exciton photon by spectral filtering. The enhanced emission of single photons was verified by Hanbury Brown-Twiss measurement.

**6.3.3. Photon-Multiplet and Conditioned Single-Photon Generation**

Edwards and Pollard [35] demonstrated the generation of correlated beams of quiet light with series-connected arrays of light-emitting diodes driven from a low-impedance bias source. Li et al. [52] demonstrated the correlation between the light beams generated by semiconductor laser diodes connected in tandem. Edwards et al. [31] suggested that this correlation should be extendable to the mesoscopic single-photon domain.
and could be used as the basis for conditioned single-photon generation. Sumitomo et al. [53] have proposed a series-coupled mesoscopic dual LED system driven by a low-impedance clock voltage to regulate the emission of photon pairs. Their proposed device is similar to the single-photon turnstile demonstrated by Kim et al. [10] but does not require quantum confinement of holes. It seems likely that their concept could be extended to triplet and higher order photon-multiplet generation, opening up the prospect of both conditioned sub-Poissonian and regulated single-photon, photon-pair, and photon-multiplet generation.

7. SUMMARY

7.1. Current State of the Art

The physical processes responsible for the generation of shot noise in semiconductor junctions are now better understood, largely as a consequence of theoretical and experimental studies of the intensity noise in the recombination light from light-emitting junctions.

Optical measurement and analysis of recombination-photon fluctuations have provided valuable insights into the internal dynamics of semiconductor junctions not otherwise available from electronic circuit measurements. This work has clarified the sources of shot noise. It has shown that shot noise is generated in transport across the depletion layer as originally proposed by van der Ziel, and through tunneling junctions. This mechanism is significant for mesoscopic junctions and for macroscopic junctions in the case of weak injection. For moderate to strong injection in macroscopic homojunctions and heterojunctions, shot noise is mainly generated in the recombination regions.

The microscopic physics of six electronic and photonic shot noise suppression mechanisms which operate in these junctions have been identified and are now reasonably well understood. These are Buckingham's backward/forward diffusion mechanism, the fluctuating charge-reservoir mechanism; macroscopic Coulomb blockade, the nonlinear backward pump, single-electron Coulomb blockade, and Pauli blockade. These processes have all been shown to suppress electronic and photonic noise in the laboratory with potential applications in cryptography, communications, computing, and metrology. Nevertheless, electronic and photonic shot noise remain significant impediments to the performance of macroscopic semiconductor junction devices in optoelectronic systems and to the application of mesoscopic devices in next-generation information systems.

Rather modest reductions in bright-light shot noise have been reproducibly achieved with laser and light-emitting diodes in the laboratory, limited mainly by partition noise due to inefficiencies in photon collection and detection. Single-electron/photon turnstile light emitters with potential applications in quantum computing and quantum cryptography have also been demonstrated again with performance limited by photon collection inefficiencies. None of these quiet-light devices are currently used outside the laboratory.

7.2. Future Developments

Given that the basic physics is understood, what are the likely future developments in quiet-light technology in the next decade? Recombination photons can be efficiently and quietly generated in Group III/V compound semiconductor junctions. Most (but not all) potential applications of quiet light are limited to low-photon-loss systems in which partition noise is negligible.

These applications include quantum noise-suppressed optical interconnects and photon transistor logic circuits in conventional and quantum computers, and optical metrology, including velocimetry, spectroscopy, and interferometry. It is therefore likely that research efforts will be directed toward improving photon collection efficiencies with the use of microcavity and other techniques to enhance spontaneous emission rates and beam collimation.

One application which is relatively robust against photon losses is quantum key distribution, using single photons and entangled photon pairs. Here the requirement for secure key exchange is for heralded single-photon, entangled photon-pair, and photon-multiplet emission in which the probability of selected multiplet photon emission is strongly enhanced relative to the Poissonian probabilities. Coupled semiconductor quantum dot and quantum well/microcavity tunneling junction emitters appear to be attractive prospects for these applications.

ACKNOWLEDGMENTS

I acknowledge with thanks the financial support of the Australian Research Council under its Large Grants Program and the encouragement of Yoshi Yamamoto, Robert Hanbury Brown and Rolf Landauer.

REFERENCES

Recombination Noise in Semiconductor Junctions

The University of Canberra – Telstra Tower Free-Space Quantum Key Distribution Test-Bed

by P Lynam from the University of New South Wales' School of Physics at the Australian Defence Force Academy, A Blake from Snowy Technology Pty Ltd, and C Cochran and P J Edwards from the Advanced Telecommunications and Quantum Electronics Research Centre at the University of Canberra

The University of Canberra – Telstra Tower Free-Space Quantum Cryptographic Key Distribution (QKD) Test-Bed, the first Australian facility of its kind, was officially opened in Canberra on Marconi Centenary Day, 12 December 2001. A joint venture between the University of New South Wales (UNSW) at the Australian Defence Force Academy (ADFA), the Canberra Institute of Technology and the University of Canberra, it is part of a multinational project to investigate and develop broadband technologies for the global distribution of quantum crypto-keys using low earth orbit satellites involving US, UK and Australian participants. The 4.2km link between Black Mountain Telstra Tower and the University of Canberra campus is the longest single-photon quantum-key test range in the world.

Introduction

Quantum cryptography, quantum computing and quantum teleportation have become feasible as a result of recent advances in quantum device engineering and quantum information theory. Of these three, quantum cryptography is closest to practical realisation. One form of quantum cryptography uses secret keys exchanged in the form of streams of single polarisation-encoded photons transmitted over publicly accessible optical communication channels. Security is provided by the laws of quantum physics rather than by computational complexity as in the well-known RSA crypto-system. These laws ensure that any attempt to intercept the key and measure the unknown quantum state (for example, the polarisation) of individual photons inevitably introduces noise into the QKD channel and alerts the users to the presence of the eavesdropper and the potential loss of security.

Quantum cryptography principles

Quantum cryptography uses a "one-time pad" – the so-called Vernam cipher – which is exchanged, in the form of a "quantum key", between the parties to a secure communication. Quantum technology provides the means for generating, encoding, transmitting and detecting the single light quanta that constitute the key. Quantum physics, through the operation of the Heisenberg uncertainty principle and related laws, ensures that the quantum states of the encoded photons (and therefore the quantum key bits) cannot be covertly copied without leaving signs of the "break-in". Quantum information theory places rigorous upper bounds on the leakage of Shannon entropy to potential eavesdroppers in the real-world situation of nonideal transmitters, noisy channels and inefficient detectors.

Encryption using one-time pads is known to be absolutely secure providing the key itself is secure. However, the keypad must contain at least as many bits as the message to be encrypted and can only be used once. High key transfer rates are therefore required. A second restriction is that true "single-photon" transmitters are not yet practical. Standard practice is to use very weak pulsed lasers instead. With these, there is a nonzero probability of a signal pulse containing more than one photon. This exposes the QKD system to certain types of sophisticated attack and sets an upper limit to the link attenuation, and therefore to the distance over which a quantum key can be shared between parties with provably high security. Current technology limits this distance to several tens of kilometres of passive low-loss optical fibre.

Figure 1 shows, in block form, a generic polarisation-encoded single-photon QKD transmitter and receiver. Single-photon generators LD₁ through to LDₙ generate an al-
alphabet of $n$ symbols distinguished by their polarisation states. Only one transmitter is activated at any time. Beam combiners $BC_1$ through to $BC_n$ direct a sequence of polarised photons, randomly chosen by the laser diode switching system, to a suitable photon launching platform together with bright synchronising pulses from the BLD. At the receiver, the bright pulses and single-photon signals are demultiplexed on the basis of their wavelength and polarisation state, and detected by single photon counting modules based on avalanche photodiode detectors.

In the B92 scheme devised by Bennett in 1992, a nonorthogonal plane-polarised binary alphabet is used to encode the key, so that $n = 2$ in figure 1. The angle between the two linear polarisation states is chosen to be 45°. This introduces the vital element of uncertainty into the key transfer process. It is essential that the angle not be 90° (the orthogonal polarisation), because this would create completely insecure classical key bits. Eve (the eavesdropper) could then (in principle) make perfect measurements of the binary polarisation states of the transmitted photons by using a polarisation-sensitive beam splitter to select each polarisation state with certainty. She could then cover her tracks by retransmitting exact copies to Bob, the legitimate recipient of the secret key, who would be none the wiser to the resulting complete loss of security. As each photon arrives, Bob, the legitimate recipient of Alice’s transmitted key, randomly sets his receiver to detect either a 1-bit or a 0-bit. He uses a pair of polarisers orthogonal to those of Alice so that a transmitted 1-bit has a 50% probability of registering correctly in his “not 0-bit” detector and a transmitted 0-bit has a 50% probability of being correctly detected in his “not 1-bit” detector.

At the conclusion of the key transmission, Bob communicates with Alice on an open channel and advises which of her randomly chosen sequence of key bits registered in either one of his two detectors. Since she already knows the state of each of her transmitted bits, this establishes a shared raw key bit sequence. Bob does not publicly reveal which of his two detectors is activated by any of the quantum bit (qubit) transmissions, as this would give the game away. Their shared raw key must be error-corrected in a subsequent dialogue at the cost of some loss of information. The maximum permissible bit error rate for the B92 protocol is close to 4%. Error rates above this limit result in a corrected key of zero length.

Previous Australian experiments in quantum cryptography

There is currently strong international interest in developing the technology to transmit quantum keys over global distances at high rates to meet the increasing demand for high-security broadband international communications. There is also strong interest in developing secure “last-mile” extensions of existing broadband networks in the CBD and in military situations. Free-space ground to satellite, satellite interorbit and terrestrial point to point links using polarised photons have been widely discussed in these contexts.

University of Canberra laboratory demonstration

In 1999, the Centre for Advanced Telecommunications and Quantum Electronics (CATQER) at the University of Canberra succeeded in implementing the B92 protocol in the laboratory using green gallium phosphide LEDs as qubit generators and photomultipliers as single photon detectors. Figure 2 shows a block diagram of the experimental set up in which the LED photon generators are randomly addressed by the 100kHz switching system to create a random qubit sequence and the receiver configuration is randomly chosen to test for either a “one” or “zero” at any time. This latter operation is accomplished by inserting a passive beam splitter at the input of the receiving system. Single photons are thus randomly directed with 50% probability to either one of the two detectors. A raw key bit rate of 50b/s was achieved in the laboratory with an error rate of less than 3%.

Mt Stromlo satellite tracking station experiments

In 2000, CATQER collaborated with the Australian Surveying and Land Information Group (AUSLIG), Electro Optic Systems Pty Ltd and the UNSW at ADFA in a demonstration of free-space qubit transfer at the AUSLIG Mt Stromlo Satellite Tracking Observatory. This demonstration utilised the 750mm Coude focus observatory telescope as a light collector for a dual polarisation single-photon counting receiver with 10ns range gate resolution, 100 microradian (μrad) field of view and spectral filter width of 0.15nm. The demonstration was successful in that a raw qubit rate of several hundred bits per second was achieved in daylight over a distance of 50m with an error rate of less than 1%. However an attempt to extend the transfer to a distance of 6km was unsuccessful, mainly owing to excessive image motion outside the restricted field of view of the AUSLIG telescope induced by atmospheric turbulence. It was therefore decided to seek funds to set up a fixed free-space test range which could be properly characterised and for which the transmitter and receiver parameters could be optimised.

The University of Canberra – Telstra Tower QKD Test-Bed

Following our laboratory QKD demonstration in 1999, and subsequent single-photon daylight transmissions at Mt Stromlo Satellite Tracking Observatory in 2000, we received support for an experimental free-space link between the University of Canberra campus and the Telstra Tower on Black Mountain. The facility was officially declared open on Marconi Centenary (Radio Foundation) Day, 12 December 2001, with the reception on the university campus of the
Figure 2. Block diagram of the elements of the University of Canberra 892 QKD laboratory demonstration (Pol — polariser, BC — beam combiner, NDF — neutral density filter, BS — beam splitter, PMT — photomultiplier tube, CLK — clock, IF — interference filter).

Figure 3. Global quantum key distribution by satellite (S) using single-photon key distribution channels (OC) for key distribution and public microwave reconciliation channels (MRC) for the post key transmission dialogue.

Objectives
The system will be used to simulate the use of low earth orbit optical satellites as global quantum key couriers (as in figure 3) using state of the art single-photon generators, modulators and receivers. We believe that the 4km path is actually somewhat longer than needed to simulate the image motion and scintillation due to turbulence-induced wave-front distortion expected in a typical ground to low earth orbit satellite link.

Design considerations
The rate at which the raw key can be transmitted, that is the quantum key transfer rate \( R \), together with the associated mean error probability per bit, the quantum key bit error rate (QKBER), are two important operating parameters. The raw key transfer rate is given by:

\[
R = E \eta_i R_i <n> \tag{1}
\]

where \( E \) is the coding efficiency (0.25 bits/photon for the nonorthogonal binary B92 protocol), \( \eta_i \) is the overall photon transmission factor (typically \( 10^{-3} \) or less for a LEOS link), \( R_i \) is the transmitter pulse repetition (key) rate, and \( <n> \) is the mean (Poisson distributed) photon number transmitted per pulse. For \( <n> = 0.01 \) photons per pulse, this gives a raw key transfer rate of only 25kb/s at \( R_i = 10Gb/s \). The importance of using ultra broadband modulators is evident.

The choice of the field of view (FOV) of the receiving telescopes is a compromise between maximising availability in the presence of atmospheric turbulence and minimising the (error-generating) background photon counting rate. The FOV is determined by the ratio of the focal plane light-collecting aperture to the focal length. This aperture is most conveniently set by the dimensions of the active core of the optical fibre which links the telescope to the photon detector. Modelling the 4km atmospheric path suggested that suitable FOVs could be achieved with telescope focal lengths of less than several metres and fibre diameters between 200\( \mu \)m and 1mm without the use of additional optics. Direct coupling to avalanche photo-diode detectors requires fibre core diameters of less than 200\( \mu \)m.

The effective light collecting aperture and the laser beam width determine the mean link attenuation. The aperture size itself acts as a spatial filter to reduce intensity fluctuations due to scintillation which can be expected to have a variance exceeding the square of the mean intensity. Light-collecting apertures of the order of 0.1m and laser beam-widths of 1mrad are suitable. Their optimum choice also has a strong bearing on the successful implementation of short-range broadband free-space optical wireless bridge technologies for terrestrial use.

Current configuration
As shown in figure 4, the optical transmitting array consists of infrared and red semiconductor laser diodes, and a green frequency-doubled diode-pumped Nd-YAG solid state laser mounted on the collimation plate of a X30 theodolite sight on Black Mountain Telstra Tower, 4km from the university campus where the optical receivers are located. A Xenon flash tube serves as a wide-angle beacon to aid alignment. The IR
S-Band/432MHz Photon-Counting Receiver

Figure 4. Sketch of the physical communication links which constitute the University of Canberra — Telstra Tower Quantum Key Distribution Test-Bed. System communications, control and analysis is handled by computer PC1 located on the university campus. The satellite earth station is available for global extension of the 4km terrestrial point-to-point system.

laser is linearly polarised and pulsed at a rate of 1MHz with a duty cycle of 5%. Neutral density filters placed in front of the laser attenuate the output beam. The transmitted pulse amplitude is controlled remotely down to an average level of less than 0.1 photons per pulse. The red laser is pulsed at a lower rate to provide bright synchronisation signals at the receiver. The flash lamp and green laser are used as beacons to aid in the alignment of the receiving telescopes. They are also monitored to provide information on the optical characteristics of the atmospheric path.

Two of the three optical telescopes planned for the optical receiving station are currently installed. The optical receiving system consists of 200mm and 300mm aperture f/10 and f/6.3 telescopes with 0.5mrad (red) and 0.15mrad (infrared) fields of view defined by multimode optical fibres. The two telescopes are each fibre-coupled to a central photon detection station where a photo-multiplier responds to bright frame sync pulses transmitted by a red (630nm) diode laser, and two silicon APD based Single Photon Counting Modules register the polarised 835nm single-photon pulses from an IR laser.

The received analog and digital signals are logged on to a laptop computer as well as being displayed and analysed in real-time. The bright-pulse amplitude and photon count are monitored in the time and amplitude domains to characterise scintillation and path attenuation. Real-time bit error statistics can also be obtained.

A 111Mb/s half-duplex S-band wireless LAN bridge connected to the University of Canberra ethernet links the two terminals. A WaveRider NCL 1170 wireless bridge extends the university campus LAN to Black Mountain Tower with over-the-air data rates of 111Mb/s and up to 8Mb/s of data throughput using direct sequence spread spectrum modulation. The bridge is connected to the campus net through a 10/100baseT interface physically located in the CATQER satellite dish control room directly below the antenna deck on which the bridge antenna and a 4.5m satellite dish are colocated.

All networking to and from the tower is TCP/IP based. The tower communications node is a 800MHz Pentium computer with 256MB RAM and 20GB hard disk connected to the ethernet port of the wireless bridge terminal. It uses a Linux (Mandrake 8.1) operating system and runs Telnet, ftp and http servers. This facilitates the tested function of the system by enabling: (i) Telnet control of the laser transmitters via the Forth-programmed printer port; (ii) ftp-upload and download of software and data for real-time link characterisation, bit-error diagnostics and privacy amplification of test key transmissions, and; (iii) eventual http-based web access to all functions. The tower node is assigned a fixed IP address and can be accessed via the university LAN and by computers remote from the university campus. Figure 5 summarises the communications connectivity between the various elements of the test-bed.

Future plans

The test bed will also be used to trial the novel single-photon generators required to maintain security and efficiency in high loss QKD systems such as those using low earth orbit satellites. It can be shown that the secure channel efficiency from equation $R_S/R = \eta_n <n>$, cannot exceed $\eta_n$ for conventional weak laser sources. This forces the use of exceedingly weak pulses or else greatly improved single-photon generators in which the
probability of multiple photon pulses is strongly suppressed\textsuperscript{\textcircled{1}}. We plan to test the single-photon generators and broadband modulators that our collaborators in the USA, Japan and Australia are currently developing\textsuperscript{3,\textcircled{11},\textcircled{16}}.

One of the telescopes is capable of tracking orbiting satellites and acquiring infrared telemetry from the recently launched AMSAT communications satellite AO-40. This will be a direct test of many of the basic assumptions embodied in the test-bed concept. It will be possible for example to compare turbulence-induced scintillation and wavefront distortion on the real satellite downlink path with that measured on the test-bed. The received satellite signal, however, will typically be of the order of 1000 photons per bit, compared with only one photon per received qubit.

Our eventual aim is to join with our collaborators in the UK and USA to exchange quantum keys globally using dedicated low earth orbit satellite as shown in figure 3.

Conclusion

It has been argued\textsuperscript{\textcircled{10}} that the physical security associated with line of sight systems such as terrestrial point-to-point and LEOS satellite based systems allows the restrictions referred to above to be relaxed. Others have argued that the security of quantum key channels should be guaranteed absolutely by the laws of quantum physics, rather than by propitious physical circumstances.

Whichever of these views is adopted, free-space line of sight satellite-based quantum key distribution appears to be the only presently known way in which to build global QKD networks and the University of Canberra –Telstra Tower test-bed has a useful role to play in this respect.

Acknowledgements

We acknowledge with thanks the financial assistance of the Australian Research Council and the contributions of J Gleeson, G Ganeshkumar, P J Zelman and W N Cheung.

References

Quantum key distribution using quantum-correlated photon sources

P.J. Edwards*, G.H. Pollard, and W.N. Cheung
Advanced Telecommunications and Quantum Electronics Research Centre, University of Canberra, Canberra ACT 2601, Australia

Received 14 July and Received in final form 20 November 2001

Abstract. Quantum key exchanges using weak coherent (Poissonian) single-photon sources are open to attack by a variety of eavesdropping techniques. Quantum-correlated photon sources provide a means of flagging potentially insecure multiple-photon emissions and thus extending the secure quantum key channel capacity and the secure key distribution range. We present indicative photon-counting statistics for a fully correlated Poissonian multibeam photon source in which the transmitted beam is conditioned by photon number measurements on the remaining beams with non-ideal multiphoton counters. We show that significant rejection of insecure photon pulses from a twin-beam source cannot be obtained with a detector having a realistic quantum efficiency. However quantum-correlated (quadruplet or octuplet) multiphoton sources conditioned by high efficiency multiphoton counters could provide large improvements in the secure channel capacity and the secure distribution range of high loss systems such as those using the low earth orbit satellite links proposed for global quantum key distribution.

PACS. 03.67.Dd Quantum cryptography – 02.50.Fz Stochastic analysis – 42.50.Lc Quantum fluctuations, quantum noise, and quantum jumps

1 Introduction

Current quantum key distribution systems commonly employ weak Poissonian laser and light-emitting diodes as quasi single-photon sources. However it is well known that such systems are susceptible to “intercept/resend”, “beam-splitter” and “photon number splitting” attacks [1-3]. If the transmitted key bits contain two or more photons, a hypothetical Eve, constrained only by the laws of physics and not by the realities of contemporary technology could gain significant Shannon information without disturbing the quantum channel and so compromise the security of the key.

In lossy systems such as those employing low earth orbit satellites as key couriers, it is generally recognised that the key exchange will be totally insecure if the number of transmitted multiple photon signals exceeds the number of received single photon signals [3]. For the weak coherent (Poissonian) photon sources currently used this leakage of Shannon entropy increases with the mean photon number \( \langle n \rangle \), which must therefore be kept small in order to maintain secure key exchange over a lossy channel. In consequence the probability of single photon emission, and therefore the transmission efficiency, will be correspondingly small, most pulses being empty of photons. For example a multiple-photon “leakage” probability

\[ L = P(n > 1) \approx \frac{(n)^2}{2} < 0.005, \]

the single photon probability

\[ S = P(n = 1) \approx \langle n \rangle < 0.1, \]

and the “no photon” probability

\[ P(n = 0) \approx (1 - \langle n \rangle) > 0.9, \]

unavoidably low, and the “no photon” probability

\[ P(n = 0) \approx (1 - \langle n \rangle) > 0.9, \]

unavoidably high. This enforced trade between potential entropy leakage (multiple-photon probability) and channel efficiency (single-photon probability) becomes even more unfavourable at low single-photon efficiencies when the bit error rate due to background photons and dark counts further restricts the secure key bit transfer rate.

Improved methods of single-photon generation for which the single photon probability \( S \) and the ratio \( S/L \) are both higher than for Poissonian sources are currently under active investigation [4-8].

We recently proposed [9] a novel scheme based on an extension of the correlated twin beam concept [10-12] in which one of two quantum-correlated beams formed by parametric down conversion (PDC) is used to condition the other beam [3]. Our conditional single-photon generating scheme requires strong photon number-correlations between a multiphoton beam with separated spatially separated beams each of which is monitored by a high efficiency multiphoton counter capable of differentiating between single-photon and multiphoton pulses [12-14].

Multiple-photon events are detected and either deleted during the quantum key transmission or else subsequently discarded in the error correction and privacy amplification dialogue following the transmission. Although not yet demonstrated in the laboratory, a possible
realisation might be a tandem array of two or more quantum wells, microcavities, or quantum dot light-emitters [15]. This would represent an extension to the mesoscopic scale of the strong quantum correlation observed between the bright beams emitted from arrays of macroscopic semiconductor junction light-emitting diodes and diode lasers when these are electrically coupled together [15-20].

We first briefly review the limitations of an unconditioned Poissonian photon source. We then obtain indicative statistics for a fully correlated multiplet photon source conditioned by one or more high quantum efficiency noiseless photon counters.

2 Single Poissonian photon beam

Consider an ideal semiconductor light-emitting diode or ideal laser diode driven by a periodic train of identical current pulses. These generate a train of weak light pulses each of which contains a Poisson-distributed number of photons, with the same mean photon number per pulse $\langle n \rangle$. This is a typical realisation of a quasi-single-photon QKD source.

For later comparison with the conditioned multiple beam statistics we note the Poisson probability that a pulse will contain exactly $n$ photons is

$$P(n) = \frac{\langle n \rangle^n}{n!} \exp(-\langle n \rangle),$$

so that:

$$N_1 = P(0) = \exp(-\langle n \rangle) \approx (1 - \langle n \rangle);$$

$$S_1 = P(1) = \langle n \rangle \exp(-\langle n \rangle) \approx \langle n \rangle;$$

$$L_1 = P(> 1) = 1 - (1 + \langle n \rangle) \exp(-\langle n \rangle) \approx \langle n \rangle^2/2;$$

where the approximations refer to small mean numbers, $\langle n \rangle < 1$.

Figure 1 shows the variation of $N$, $S$, $L$ and $L/S$ with $\langle n \rangle$. Note that while the maximum value of the single photon probability $S = 1/e = 0.37$ for $\langle n \rangle = 1$, such a high value of $\langle n \rangle$ is unusable because of the correspondingly high leakage probability of multiple photons, $L = 0.26$. Each multiple photon pulse carries a potentially insecure bit which can in principle be intercepted and copied clandestinely. The ratio $S/L$ is therefore one convenient parameter to characterise the quality of a single-photon generator.

A conservative and necessary (but not sufficient) condition for the secure exchange of a key is that the number of multiple bits available for covert copying by an eavesdropper at the transmitter must not exceed the number of single photon reaching the receiver [3].

For single-photon channel transmission probability $\eta_T$, this means that $\eta_T P(\langle n \rangle = 1) > P(\langle n > 1)$, or $\eta_T S/L > 1$. A conservative lower limit for the parameter $S/L$ required for secure key exchange over a lossy channel is then the channel attenuation ($1/\eta_T$). A convenient figure of merit ($Q_s$) for a single-photon quantum key source is the product of the maximum permissible secure channel loss ($1/\eta_T = S/L$) and the single-photon transmitter efficiency, $S$:

$$Q_s = S^2/L.$$

Evidently $S/L$ must certainly exceed 1, corresponding to a lossless channel with ideal photon detection. Reference to equations (1-4) and Figure 1 shows that this limits $\langle n \rangle$ to values less than 1.33 for a secure lossless, ideal noiseless Poisson channel. Practical lossy channels necessitate much higher values of $S/L$. These can only be achieved by greatly reducing the value of $\langle n \rangle$ and so sacrificing transmitter efficiency. In the limit of small $\langle n \rangle \ll 1$, $S = \langle n \rangle$, $S/L = 2/\langle n \rangle$ and $Q_s = 2$. A channel loss ($1/\eta_T$) of greater than 100 (20 decibels) therefore restricts $S$ to less than 0.02. Secure key transfer is thus achieved at the cost of low channel efficiency. The minimum tolerable channel transmission factor is,

$$\eta_T(\text{min}) = L_1/S_1 = P(\langle n > 1)/P(\langle n \rangle = 1) = [\exp(\langle n \rangle) - (1 + \langle n \rangle)]/\langle n \rangle. \quad (6)$$

From equation (6) above, since $L_1/S_1 \approx \langle n \rangle/2$ for small $\langle n \rangle$, the corresponding secure BB84 [21] Poisson channel efficiency is then,

$$\varepsilon(\text{max}) \approx \eta_T(\text{min})\langle n \rangle/2 = \langle n \rangle^2/4 = \eta_T^2(\text{min}). \quad (7)$$

A single-mode fibre QKD channel with an attenuation of, say, 0.5 dB/km (a conservative estimate of the loss encountered in public switched optical networks operating at long wavelengths) can therefore be no longer than

$$R = 20 \log_{10}(1/\eta_T(\text{min})) = 10 \log_{10}(1/\varepsilon), \quad (8)$$
with $R$ in km and $\varepsilon = (n)^2/4$, the corresponding secure channel efficiency for noise-free key exchange using the BB84 protocol.

For secure QKD systems operating with $(n) = 0.1$, the maximum permissible channel loss is then only 13 dB, corresponding to a noiseless BB84 Poisson channel efficiency, $\varepsilon$ of less than 0.0025 bits per transmitted symbol.

It is apparent that this severely limits the efficiency and therefore the (bit rate) capacity of a secure lossy channel. In practice the efficiency and capacity will be significantly lower because of the need to discard erroneous bits and to institute privacy amplification in order to maintain security in the presence of thermal and background photon counting noise.

Earth satellite - based QKD systems have been proposed [23] in which the channel attenuation is typically greater than 30 dB ($n_T < 10^{-5}$). The corresponding secure channel efficiency ($\varepsilon$) will then (from Eq. (7)), be less than 10$^{-8}$, even assuming ideal detection.

It has been argued that the physical security associated with “line of sight” systems of this kind allow the condition expressed in equation (6) to be relaxed [24]. Other authors [3] argue that the security of quantum key systems should be guaranteed by the laws of physics, rather than by physical circumstances and current technology. If this latter view is accepted, the capacity of Poissonian channels will be limited so severely by equations (6–8), that they are unlikely to be useful for any purpose other than low loss, low capacity, short range quantum key distribution. Single photon sources employing single quantum dots, and diamond colour centres are therefore being developed [4–8] with reported $S/L$ ratios approaching 10$^2$.

3 Conditional single-photon sources

Considerations such as those above indicate the desirability of realising photon transmitters for which the ratio $L/S = P(n > 1)/P(n = 1)$ is much lower and the figure of merit $Q_s = S^2/L$ is much higher than that for Poissonian sources. Ideally such sources should emit only single photons on demand.

In what follows, we shall examine the potential gains in secure channel capacity and range to be had from the use of conditioned single photon sources in a quantum key transmitter. For simplicity we shall assume fully correlated Poissonian photon sources, each with the same mean emission rate. As we shall show, conditioned twin-beam sources suffer from the disadvantage that an unrealistically high quantum detector efficiency is required. However, one can take advantage of the repeated measurement of photon number provided by a larger array to improve the identification (and subsequent rejection) of multiple-photon signals.

In the following sections we discuss one method of achieving this based on quantum- correlated multiple beams. The relative probability ($L/S$) of insecure multiple-photon signals is strongly suppressed and the figure of merit $Q_s$ is raised above the Poissonian values by a conditioning process, allowing secure quantum keys to be exchanged over longer distances at higher rates.

The process is shown schematically in Figure 2. Quantum-correlated photon emitters $B_1, ..., B_m$ are shown electrically connected together and driven by repetitive pulse generator $G$. The transmitter beam, $B_1$, is coupled to a modulating device before being launched into an optical fibre or free-space. If each transmitted pulse (from $B_1$) contains $n_1$ photons then for fully correlated beams the same photon number $n_1$ will be replicated at $B_2, ..., B_m$ and will be separately counted by the $(m-1)$ noiseless multiphoton counters $C_2$ through $C_m$, each with the same quantum efficiency (single-photon counting efficiency), $\eta$. For simplicity we assume that $n_1$ is a Poisson variable. It then follows that photon counts $n_2, n_3, ..., n_m$ are also Poissonian with the same mean $(n_j) = \eta(n_1)$ for $j = 2, ..., m$.

Figure 2 shows the logic of this conditioning process in which gate $G_2$ is closed if $n_2 > 1$, gate $G_3$ is closed if $n_3 > 1$, ..., and gate $G_m$ is closed if $n_m > 1$. If any one of the $(m-1)$ independent counts $n_m$ through $n_m$ exceeds 1, the corresponding pulse is removed from the sequence of photon pulses from $B_1$, leaving the conditioned sequence $\{B_1(B_2, ..., B_m)\}$ available for the key exchange. A photon pulse from $B_2$ will pass through all gates and be included in the key transfer only if the counters register either one photon per pulse or none. The logic gates do not necessarily represent physical devices or operations although these are certainly not excluded. Insecure pulses can therefore be identified and removed prior to transmission or subsequently discarded in the post key transmission dialogue.

The multiple beams generated by tandem arrays of macroscopic semiconductor junctions are generally sub-Poissonian [16] while PDC twin beams have excess number variance. In order to simplify the analysis we assume Poissonian source statistics so that the analysis is “worst case” in this respect.

We also assume noiseless detection. This latter assumption is realistic in that a state-of-the-art multiphoton detector such as the visible light photon counter [14] has nanosecond resolution and a dark count of order 10$^4$ s$^{-1}$. The noise count in any one detector will then be of order 10$^4$ s$^{-1}$ for a pulse repetition rate of 10$^8$ s$^{-1}$, typically less than 0.01% of the single photon counting rate. The effects
of noise on both Alice's and Bob's detectors can thus be minimised, as in other single-photon systems, by precise time-gating.

Our third simplifying assumption will be to assume multiphoton counters characterised by a single parameter, the single-photon counting "quantum" efficiency, $\eta$. With this assumption the two-photon counting efficiency is $\eta^2$, the $N$-photon counting efficiency is $\eta^N$ and the photon counts remain Poissonian. Coupling efficiencies will generally be less than 100%. In our calculations we have therefore used external single-photon detection efficiencies as low as 50%. In later examples (shown in Figs. 5 and 6), we have taken $\eta = 0.875$ [13], corresponding to an ideal two-photon counting efficiency of 0.77. This latter situation is regarded as an upper limit. Note that the correlated photons emitted from different junction in Alice's semiconductor multiplet source may have different wavelengths. Thus, the wavelength of the transmitted photon could be tuned (by choice of band gap) to 1550 nm to minimise fibre transmission loss, while the wavelength of the conditioning photons could be tuned to 700 nm for maximum detection efficiency using current multiphoton counters.

4 Quantum-correlated photon twin beam

Consider two fully number-correlated pulsed photon beams $B_1, B_2$ ($m = 2$ in Fig. 2). These might be realised by the use of parametric down conversion [11] or, as suggested, by pulsed electrically coupled mesoscopic light-emitters [15].

For the twin beam case the conditional probabilities $N_2 = P(n_1 = 0 | n_2 = 0)$ and $S_2 = P(n_1 = 1 | n_2 = 0)$ are then easily found from Bayes' theorem to be:

$$N_2 = [(1/(1 + \eta(n_1))) \exp[-(n_1)(1 - \eta)]]$$

$$S_2 = [(n_1)/(1 + \eta(n_1))] \exp[-(n_1)(1 - \eta)].$$

So that

$$L_2 = P(n_1 > 1 | n_2 = 0) = 1 - N_2 - S_2 = 1 - [(1 + (n_1))/(1 + \eta(n_1))] \exp[-(n_1)(1 - \eta)].$$

The leakage ratio:

$$L_2/S_2 = (D_2 \exp[(n_1)(1 - \eta)] - [(1 + (n_1))]/(n_1)$$

in which the function $D_2(\langle n_1 \rangle, \eta) = (1 + \eta(n_1))$ is plotted in Figure 3 for a range of single-photon counting efficiencies $\eta$ and mean photon numbers, $\langle n \rangle$. The uppermost curves (for $\eta = 0$), correspond to the unconditioned single beam. Figure 4 shows the conditioned single photon emission probability, $S_2 = P(n_1 = 1)$. It shows the increase in single-photon probability in the higher photon-number, regimes made possible by an efficiently conditioned Poissonian source.

It is evident that high quantum efficiency is needed in a conditioned twin beam source to effect a major improvement in either channel efficiency or security against
a beam-splitter attack. From equation (12), for \( \langle n \rangle < 1 \), 
\[
L_2/S_2 \approx (\langle n \rangle)(1-\eta^2)/2 \approx (1-\eta^2)^{-1} \text{ less than for the unconditioned single beam.}
\]
For \( \eta \) close to 1, \( L_2/S_2 \approx (\langle n \rangle)(1-\eta) \), giving a reduction by a factor of \( 1/(2(1-\eta)) = 50 \) for \( \eta = 0.99 \). This is a very significant improvement since for the same \( \langle n \rangle \) it would permit a fifty fold (17 dB) increase in channel attenuation (corresponding to a 34 km increase in fibre length, or a seven fold increase in free-space range) without any loss of security, albeit with the inevitable reduction in channel capacity. Alternatively, for the same channel loss it would permit a fifty fold increase in channel capacity through increased \( \langle n \rangle \). Unfortunately, such a high value of counting efficiency is well beyond the capabilities of current technology and we must look to other means of beam conditioning which make more relaxed demands on detector photon counting efficiency.

### 5 Quantum - correlated multiplet beam

It has been shown [15-20] that arrays of semiconductor light emitting junctions provide photon “multiplet” beams in which, the number of correlated beams is not restricted to 2 as for PDC twin beams. We shall show that repeated \((m-1)\) photon number measurements made with such a photon multiplet beam source allows effective conditioning with much lower (and more practical) detector efficiencies. It should, be pointed out however that although macroscopic quantum correlations have been demonstrated with bright multiplet beams of this type, successful extension to the single-photon number domain remains to be shown. However, Sumitomo et alia [22] have recently confirmed by Monte Carlo simulation the concept of efficient heralded twin-photon production by a pair of series-coupled array of mesoscopic light-emitting diodes.

Equations (9-12) for the conditioned twin beam source can be extended (Appendix A) to multiplet \((m\text{-beam})\) sources. For the triplet case \((m = 3)\), where we identify and reject multiple photon signals if either one of two detectors register more than one count, we can rewrite equation (12) as

\[
L_3/S_3 = \left( D_3 \exp([\langle n \rangle](1-\eta^2)^2 - [(1 + \langle n \rangle)])/\langle n \rangle \right)
\]

with

\[
D_3 = [1 + 2\langle n \rangle \eta(1-\eta) + \langle n \rangle \eta^2 (1 + \langle n \rangle)(1-\eta^2)^2].
\]

Similarly from Appendix A, for the quad source \((m = 4),\)

\[
L_4/S_4 = \left( D_4 \exp([\langle n \rangle](1-\eta^2)^2 - [(1 + \langle n \rangle)]/\langle n \rangle \right)
\]

with

\[
D_4 = [1 + 3\langle n \rangle \eta(1-\eta)^2 + 3\langle n \rangle \eta^2 (1-\eta) \times (1 + \langle n \rangle)(1-\eta^2) + \langle n \rangle \eta^2 (1-\eta^6)].
\]

Table 1. Single-photon quantum key transmitter efficiency \( S \), multiphoton leakage ratio \( L_m/S_m \), and figure of merit \( Q_s = S_m^2/L_m \), in the small \( \langle n \rangle \) approximation for multiplet \((m\text{-beam})\) sources conditioned by \((m-1)\) noiseless multiphoton counters with quantum efficiencies \( \eta = 0.875; 0.500 \).

<table>
<thead>
<tr>
<th>( M )</th>
<th>( S )</th>
<th>( L_m/S_m )</th>
<th>( Q_s(0.875) )</th>
<th>( Q_s(0.500) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 &lt;( \langle n \rangle &lt; 1 )</td>
<td>( 2 &lt;\eta &lt; 2 )</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2 &lt;( \langle n \rangle &lt; 1 &lt;\eta &lt; 2 )</td>
<td>( 2 &lt;\eta &lt; 2 &lt;\eta &lt; 2 )</td>
<td>36.4</td>
<td>3.56</td>
<td></td>
</tr>
<tr>
<td>3 &lt;( \langle n \rangle &lt; 1 &lt;\eta &lt; 2 )</td>
<td>( 2 &lt;\eta &lt; 2 &lt;\eta &lt; 2 )</td>
<td>155</td>
<td>4.74</td>
<td></td>
</tr>
<tr>
<td>8 &lt;( \langle n \rangle &lt; 1 &lt;\eta &lt; 2 )</td>
<td>( 2 &lt;\eta &lt; 2 &lt;\eta &lt; 2 )</td>
<td>5.14 \times 10^6</td>
<td>15.0</td>
<td></td>
</tr>
</tbody>
</table>

In the small \( \langle n \rangle \) approximation, \( L_3/S_3 \approx 2\langle n \rangle(1-\eta^2), \) so that, for example, if two identical detectors were used to monitor two fully correlated beams of the triplet a more realistic quantum efficiency of only 0.93 would be required to obtain the 50 fold suppression of multiple photon signals referred to above. Proceeding further, for the quad beam, \( L_4/S_4 \approx 4\langle n \rangle(1-\eta^2), \) allowing the quantum efficiency to be relaxed even further, to \( \eta = 0.85, \) while for an octuplet beam with seven independent beam counters, the required efficiency is only \( \eta = 0.65. \)

In general, for the \( m \)-fold beam the leakage ratio (to first order in \( \langle n \rangle \)),

\[
L_m/S_m = \langle n \rangle (1-\eta^2)^{m-1}/2
\]

and the figure of merit of the conditioned source,

\[
Q_s = 2(1-\eta^2)^{1-m}.
\]

It follows that the leakage continues to drop by the factor \((1-\eta^2)^2 \approx 2(1-\eta)\) for every additional high efficiency conditioning counter used. If we take \( \eta = 0.875, \) the maximum single-photon counting detector efficiency currently reported [13], the leakage drops by a maximum factor of 4.27 with each additional conditioning beam. Table 1 summarises these results for the case of small \( \langle n \rangle \), with \( \eta = 0.875 \) and \( \eta = 0.5 \) as examples. The advantages to be had from using a high order multiplet source conditioned by high efficiency multiphoton counters in a high loss quantum channel are evident.

Figure 5 shows the leakage \( L_m = P(>1) \) plotted for \( m = 0, 1, 2, 3, 4, \) and Figure 6 shows the ratio \( L_m/S_m \) for \( \eta = 0.875. \) It is evident that the leakage suppression is maintained for mean photon numbers close to unity. This permits high values of the figure of merit \( Q_s = S_m^2/L_m \) to be achieved. It is also evident from equations (17, 18) that we can profitably trade beam number for detector quantum efficiency to maintain a given degree of suppression. Returning to the previous discussion, a fifty fold reduction in leakage requires either a single 99% (twin beam) detector efficiency, two 93% efficient detectors for a triplet beam, four 79% efficiency detectors for a quintuplet, seven 65% efficiency detectors or nine 59% efficient detectors.
6 Conclusions

The secure QKD outreach can be extended by reducing the occurrence of multiple-photon signals in the key bit sequence. One possible technique, discussed here, is the use of a photon beam conditioning technique using correlated multibeam (multiplet) photon sources monitored by high quantum efficiency multiphoton detectors. In practice, for practical quantum counting efficiencies of less than 90%, multiplet (quadruplet, or even octuplet) beams must be employed. These could provide source parameters $S/L = P(1)/P(>1)$ of $10^4$ or more as required for secure quantum key distribution by earth satellite, for reasonable values of transmitter efficiency, $S = P(1) > 0.1$, corresponding to a single-photon source figure of merit $Q_s > 1000$. Entangled photon pair systems have been shown to be immune to beam-splitting attacks [25], however it is difficult to see how global key distribution could be implemented by earth satellites using entangled photon twins.

Reasonably efficient multiphoton conditioning counters are currently available. It appears therefore that if quantum-correlated multiple Poissonian photon beams can be realised, then the $L/S$ ratio may be lowered by several orders of magnitude with a commensurate increase in channel capacity and/or permissible secure channel loss using existing detector technology.

However weak photon beams correlated at the single-photon level remain to be demonstrated. Electrically coupled arrays of single-electron/photon turnstile devices such as quantum dot emitters or mesoscale light-emitting semiconductor junctions may provide suitable quantum-correlated multiplet beams [9,15,22].

This work was supported by the Australian Research Council under its Large Research Grants Scheme. The authors acknowledge with thanks the assistance of G. Ganesharajah.

Appendix A

We sketch here the derivation of a general expression for the multiphoton "leakage" $L_m$, and the "leakage ratio" $L_m/S_m$, the ratio of the conditional probabilities $P(n_1 > 1)/P(n_1 = 1)$ given that none of the $(m-1)$ noiseless photon counters with single-photon quantum efficiency $\eta$ register more than a single count.

From Section 3, the probability of a null count in every one of the $(m-1)$ counters,

$$P\{n_2 = 0, n_3 = 0, \ldots, n_m = 0\} = \sum_{j=0}^{\infty} P[n_1 = j \text{ and } \{n_2 = 0, n_3 = 0, \ldots, n_m = 0\}] = \sum_{j=0}^{\infty} \left(\frac{\langle n_1 \rangle^j \exp(-\langle n_1 \rangle)}{j!}\right) \left[(1-\eta)^{m-1}\right]^j = \exp(-\langle n_1 \rangle) \left[1 - (1-\eta)^{m-1}\right].$$

Also, the probability of a single count in the $m$th counter and null counts in every one of the remaining $(m-2)$ counters,

$$P\{n_2 = 0, n_3 = 0, \ldots, n_{m-1} = 0, n_m = 1\} = \sum_{j=1}^{\infty} P[n_1 = j \text{ and } \{n_2 = 0, n_3 = 0, \ldots, n_{m-1} = 0, n_m = 1\}] =$$
becomes, after insertion of the Poisson and binomial probability functions, summation and collection of terms:

\[
E_0(\langle n_1 \rangle (1-\eta)^{-m-2}) \langle n_1 \rangle \eta (1-\eta)^{-m-2} \exp(-\langle n_1 \rangle),
\]

where \(E_0(x) = \sum_{j=1}^{\infty} \langle j^2 x^j \rangle (j-1)!\), so that \(E_0(x) = \exp(x); E_1(x) = (1+x) \exp(x); \) etc.

Similarly,

\[
P\{n_2 = 0, n_3 = 0, ..., n_{m-2} = 0, n_{m-1} = 1, n_m = 1\} = \sum_{j=1}^{\infty} P\{n_1 = j \text{ and } n_2 = 0, n_3 = 0, n_{m-2} = 0, n_{m-1} = 1, n_m = 1\}
\]

\[
= \{E_1(\langle n_1 \rangle (1-\eta)^{-m-1}) \langle n_1 \rangle \eta^2(1-\eta)^{-m-3} \exp(-\langle n_1 \rangle)\}.
\]

Further, proceeding in the same way,

\[
P\{n_2 = 0, n_3 = 0, ..., n_{m-3} = 0, n_{m-2} = 1, n_{m-1} = 1, n_m = 1\} = \sum_{j=1}^{\infty} P\{n_1 = j \text{ and } n_2 = 0, n_3 = 0, n_{m-2} = 1, n_{m-1} = 1, n_m = 1\}
\]

\[
= \{E_2(\langle n_1 \rangle (1-\eta)^{-m-1}) \langle n_1 \rangle \eta^3(1-\eta)^{-m-4} \exp(-\langle n_1 \rangle)\}.
\]

Hence,

\[
P\{n_2 = 0 \text{ or } 1, n_3 = 0 \text{ or } 1, ..., n_m = 0 \text{ or } 1\} = \{\exp(\langle n_1 \rangle (1-\eta)^{-m-1}) + \langle n_1 \rangle \sum_{j=1}^{m-1} \binom{m-1}{j} \eta^j(1-\eta)^{m-1-j} \times E_{j-1}(\langle n_1 \rangle (1-\eta)^{-m-1}) \exp(-\langle n_1 \rangle)\} = C_m,
\]

say.

It then follows immediately that the conditional probability of exactly one photon per pulse from \(B_1\),

\[
S_m = P\{n_1 = 1, n_2 = 0 \text{ or } 1, ..., n_m = 0 \text{ or } 1\} \times \{P\{n_2 = 0 \text{ or } 1, ..., n_m = 0 \text{ or } 1\}\}^{-1} = \langle n_1 \rangle \exp(-\langle n_1 \rangle)/C_m,
\]

and

\[
L_m = 1 - P\{n_1 = 0 \text{ or } 1/n_2 = 0 \text{ or } 1, n_3 = 0 \text{ or } 1, ..., n_m = 0 \text{ or } 1\} = 1 - (1 + \langle n_1 \rangle)C_m^{-1} \exp(-\langle n_1 \rangle).
\]

Thus, the leakage ratio for the \(m\) beam case,

\[
\frac{L_m}{S_m} = \frac{[C_m \exp(\langle n_1 \rangle) - (1 + \langle n_1 \rangle)\langle n_1 \rangle]}{\langle n_1 \rangle}
\]

\[
= \frac{D_m \exp(\langle n_1 \rangle (1-\eta)^{-m-1}) - (1 + \langle n_1 \rangle)}{\langle n_1 \rangle},
\]

as in equations (12, 13, 15), where

\[
D_m = C_m \exp(\langle n_1 \rangle [1 - (1-\eta)^{-m-1}]\right).
\]

References