Proceedings

High Energy Astrophysics

Supernovae, Remnants, Active Galaxies, Cosmology

Edited by G. Börner

Springer-Verlag

High Energy
Astrophysics

High Energy Astrophysics

Supernovae, Remnants, Active Galaxies, Cosmology

Proceedings of the Second Workshop of the
Max-Planck-Gesellschaft and the Academia Sinica,
Held at Schloß Ringberg, Tegernsee,
July 12–17, 1987

Editor: G. Börner

With 106 Figures

Springer-Verlag Berlin Heidelberg New York
London Paris Tokyo

Professor Dr. Gerhard Börner
Max-Planck-Institut für Physik und Astrophysik,
Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-8046 Garching, Fed. Rep. of Germany

ISBN-13: 978-3-642-73562-2 e-ISBN-13: 978-3-642-73560-8
DOI: 10.1007/ 978-3-642-73560-8

2156/3150-543210

Preface

The second workshop on High Energy Astrophysics of the Max-Planck-Society and the Academia Sinica was held at Ringberg Castle near Tegernsee during the week July 12–July 17, 1987. It is the purpose of these workshops to encourage the exchange of ideas between German and Chinese astrophysicists (scientists from other countries are also welcome, of course), and the pleasant atmosphere and surroundings of Ringberg Castle certainly helped the development of relaxed and stimulating discussions. In addition, we seemed to have picked out the one week of perfect weather during a rainy summer. Thanks are due to Mr. Hörmann and his staff for creating a perfect infrastructure.

The proceedings have required a considerable amount of editing, and I hope that all the papers are comprehensible. Thanks are due to Miss Petra Berkemeyer and Miss Hannelore Müller, who patiently retyped several manuscripts and helped with other editing jobs. I did not feel that a sharp refereeing procedure should be done, although I do not agree with all the papers in this volume.

The interested reader will find that this book gives a fair account of the present status of Chinese activities in the field of high energy astrophysics, i.e. in supernova physics, accretion onto compact objects, active galaxies, and cosmology.

It was our good fortune that three leading theorists (Hillebrandt, Nomoto, Truran) could be persuaded to present their newest ideas, stimulated by SN1987A, on supernova physics. In addition, I feel that one of the highlights of this volume is the article by Li Qibin on historical supernova from Chinese records, a work that here appears in English for the first time. The review article by N. Straumann on "Cosmic Strings" gives an exposition that is nicely self-contained – I have not seen another account of this topic as thorough and readable as this one.

The papers on radiation transport and accretion problems give an up-to-date discussion of several aspects of this field. Several papers deal with the intricate task of extracting information about the distribution of quasars from absorption lines, and several others address the question whether the universe has a fractal structure (I am especially happy that a contribution by Fang Lizhi, who unfortunately was unable to attend the workshop, but who had helped to organize it, could be included here).

Thanks are due to Prof. R. Kippenhahn, the Director of the Max-Planck-Institut für Astrophysik, for his support of the workshop.

München
December 1987

G. Börner

Contents

Part I

Supernovae and Their Remnants

A Recent Study on the Historical Novae and Supernovae

Li Qibin

Beijing Astronomical Observatory, Beijing, People's Republic of China

1. Introduction

The astronomers are being excited by the event of the SN 1987A which is the most brilliant supernova exploded during the last 3 centuries and, is showing some unusual characteristics. In this case people would ask why no supernova has exploded in such a long time in the Galaxy and whether any supernova has shown similar behavior. Whenever astronomers consider these problems they always pay attention to the ancient records, especially, to those in Chinese chronicles which include the richest records of historical astronomical phenomena over the world. Right now my colleagues, Ms Zhuang Weifeng and her co-authors [1] have completed a general compilation of Chinese ancient records of the astronomical phenomena. To compile the book, about 200 scholars have spent 13 years and checked 150 thousands volumes of ancient documents. For this comprehensive book they have collected more than 10 thousands of the relevant records from the ancient documents whichever they could find in China. It includes 10 types of astronomical records, which are novae & supernovae, comets, meteors, showers, meteorites, sunspots, aurorae, solar eclipses, lunar eclipse and occultations. As the editor of the chapter of historical novae and supernovae, I would like to introduce the main results here and discuss the confirmation on historical supernovae according to these materials.

2. A New Catalogue of Historical Novae and Supernovae

2.1 A Brief History

The earliest compilation of the records of "Unusual Stars" in chinese references was done by the chinese scholar Ma Duanlin [2] in the 13th century which listed relevant events recorded from the Han dynasty to the 13th century, which had been translated to the western language by Biot [3,4]. Lundmark [5] noticed that the space distribution of those guest stars was consistent with that

of novae observed in his time and first confirmed that the ancient records on guest stars are real events of novae instead of the forgey of astrologist. Xi Zezong [6] published a catalogue of ancient records of historical novae and supernovae and suggested some identification of historical events with SN remnants in 1955, which was renewed by himself and Bo [7] in which some records from Japanese and Korean documents were supplied in 1965. In 1962 Ho Peng Yok [8] published a list of comets and novae. Clark & Stephenson [9] wrote "The Historical Supernovae" in 1977 in which some records were discussed in some detail. Up to now, the explanation about the ancient records on supernovae has been discussed by many authors. Unfortunately, it is found that some of the records are neglected or wrongly quoted and explained e.g. many comets are usually listed as novae or supernovae. The problems usually result from 1) incomplete material and 2) ambiguity in understanding the semantics in ancient documents. That is why we must investigate again and we must emphasis the completeness, the evaluation on the authentical, and the text research on ancient semantics, especially on the ancient astronomical terminology.

2.2 Text Research

To evaluate the authenticity of the ancient documents, we should discuss the documents separately for three periods. Before the Han dynasty, the astronomical records were scattered in the Tortoise-Bone inscription and some classics such as the Spring & Autumn Annals, the Book of Songs and etc.. We must pick out the astronomical records from all the documents we have ever found. Since the Han dynasty, the main sources of astronomical records come from the official histories which were systematically edited and kept up to now. Fortunately, there are some chapters on astronomy in each book of the official history, which were quoted from the records of Royal Astronomers. We considered, in general, these chapters as the most authoritarian sources.

Sometimes we can find the Real Records which the official histories are based on. Of course they provide a more reliable source of materials. The local histories are also considered as a source of astronomical records which might record some events which have not been observed by the royal astronomers during cloudy days or during the wars. But we have the local histories available only since the Tang dynasty, so that we do not in

general consider the records occurred in the local histories before the Tang dynasty valuable.

By carefully checking many more ancient documents we find some valuable records which have never been listed in the reference on the novae or supernovae. The most remarkable examples are as follows:

1) The Zhoubo star reappeared in AD1016, which might mean that the supernova 1006 had a second explosion ten years later.

2) It is found that the date (the day and the month except the year) and the contents of the records of the events of AD 1404 in the History of the Ming Dynasty and that of AD 1408 in the Real Record of the Ming dynasty are the same word by word. They should be the same event. As the author [10] has mentioned the record of AD 1404 might have been wrongly written about the year because the History of the Ming Dynasty was quoted from the Real Record of Ming Dynasty. The records about the event in AD 1408 in the other documents have confirmed the identification.

2.3 Terminological consideration about the ancient documents

So as to the terminology in the ancient documents, we have to face an important but complicate work. We must study carefully the terms on the names of stars, the words on the shape, brightness of the stars, as well the verbs on the behavior of the stars.

1. "The guest star" seems to be a term with extensive connotation. In most cases the guest star should be regarded as a nova or supernova. However, the guest star sometimes might be used to record some comets, especially, to those without an obvious tail. In a few cases, even meteors might have been recorded as guest stars. We have carefully distinguished the novae and supernovae from the comets termed as guest stars in the ancient records. For most cases, it is not difficult to do so according to the statement on the shape or the movement. Sometimes a celestial event was recorded as a guest star without any further description in a certain chapter or book, while some detailed descriptions on the shape of the tail and the movement for the same event could be found in the other chapters of books. Even the event was clearly recorded as a comet somewhere. To exclude the comets on our list, every record termed as guest star we have ever found has been compared with those that occured at the same time in different chapters or books.

2. "Bei star" was certainly defined as a subclass of the comets. It was written in the History of Jing Dynasty: Bei star is

a subclass of the comets shedding the rays of light in all direction, while the comets with its rays extending forward in one preferential direction are called Hai star". Obviously the Bei stars should be excluded from the list of the novae and supernovae. Unfortunately, a lot of Bei star are usually regarded as the novae or supernovae in the references. About one fourth in Xi and Bo's list [7] as well about one eigth in Clark and Stephenson's list [9] are such cases. Some of the records of Bei stars, for instance, the events in AD 1175 and 1230 were regarded as novae or supernovae, which were originally classified as comets and recorded in the paragraph on comets in the official histories.

3. "Zhoubo star" was defined as "a noble star which is very brilliant and yellow in colour. The country will be thriving and prospering very much when this star appeared", in the History of Jing Dynasty. Therefore, the Zhoubo star could be considered as a supernova. In fact, SN 1572 and SN 604 both are also called the Zhoubo star in the ancient records.

4. Comets were classified into more than thirty classes according to shape, colour, length of the tail and its direction in the sky in the ancient documents. They are easy to distinguish from the others.

5. "Jing star" was obscurely defined and is hard to understand. It was described in the History of Jing Dynasty: "Jing star is like a half moon, helping the moon to shine, and is also defined as a big hollow, or combined of three stars with blue gas and red gas where a yellow star is within. It is a noble star. We believe that most of this kind of events are not novae or supernovae. For several doubtable cases we should leave the explanation of the records for consideration.

To summarize our terminological investigation, we can show our approach in confirming the novae and supernovae in the ancient terminological research as representatives in set algebra:

$$N(\text{novae}) \cup SN(\text{supernovae}) = G(\text{guest stars})/C(\text{comets})$$

and

$$B(\text{bei stars}) \subset C(\text{comets}),$$

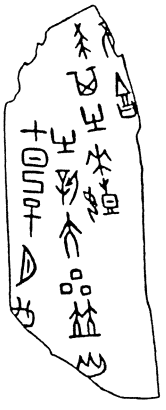
which is in principle different from the following way:

$$N(\text{novae}) \cup SN(\text{supernovae}) = G(\text{guest stars}) \cup B'(\text{beistars})$$

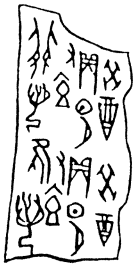
and

$$B'(\text{some beistars}) \subset B(\text{bei stars})$$

taken in some references.



On the Jisi day, the 7th day of the month, a big new star appeared in the company of the Ho star.



On the Xinwei day the new star dwindled.

We have collected 53 records for nova and supernova events of which six records have not been listed in other works. A supplement is given in the general compilation [1] in which the uncertain records are listed.

Table 1. The records of the novae and supernovae in chinese ancient documents

No.	Date in Christien era	Dynasty Emperor Epoch	Year Month Day	Statement	Ref.
1	BC 14 Century			On the Jisi day, the 7th day of the month, a big new star appeared in the company of the Huo star (Antares). On the Xinwei day the new star dwindled.	[11]
2	BC 532 Spring	Xizhou Jingwang	13 Spring	A guest star emerged at the Wunu($\varepsilon\ \mu\ 4\ 3$ Aqr).	[12]
3	BC 134 June 22 -- July 21	Han Wudi Yuanguang	1 6	A guest Star was seen in Fang ($\beta\ \delta\ \pi\ \rho$ Sco).	[13]
4	BC 77 Oct. 17 -- Nov. 15	Han Zhaodi Yuanfeng	4 9	A guest star was between the Doushu and Ji (GC17440 and $\gamma\ \beta\ 5\ 4$ UMi) within the Zigong .	[14]
5	BC 48 May	Han Yuandi Chuyuan	1 4	A guest star, as big as melon, with a bluish-white colour, appeared about 4ft to the east of the second star of Nandou (λ Sgr).	[15]
6	64 May 3	Han Mingdi Yongping	7 3 Gengxu	History Notes recorded: a guest star was seen at the south of Zhuozhifa (η Vir) outside the Maingate of Taiwei Enclosure for 75 days. Its rays shined 2ft away.	[16]
7	70 Dec. 22 -- 71 Jan. 19	Han Mingdi Yongping	13 11	History Notes recorded: a guest had emerged at the Xuanyan (SE of Lyn and the west of Leo) for 48 days.	[17]
8	101 Dec. 30	Han Hedi Yongyuan	13 11 Yichou	There was a small guest star near the forth star of Xuanyuan (Lyn). It was bluish-Yellow in clolur.	[18]

Table 1. (continued)

No.	Date in Christien era	Dynasty Emperor Epoch	Year Month Day	Statement	Ref.
9	107 Sept. 13	Han Andi Yongchu	1 8 Wushen	A guest star appeared at the SW of Hu star (in the Pup) at the Dongjing.	[19]
10	125 Dec. 31 -- 126 Jan. 11	Han Yanguang	4 11	A guest star was seen in Tiansi (Oph, Ser and the South of Her).	[20]
11	126 March 23	Han Shundi Yongjian	1 2 Jiawu	A guest star entered the Taiwei (NE of Vir, the south of Com and the East of Leo).	[21]
12	126	Han Shundi Yongjian	1	A guest star kept guard the Yulin (around 29 Aqr).	[22]
13	185 Dec. 7	Han Lingdi Zhongping	2 10 Guihai	A guest star appeared in the middle of Nanmen (β-σCen). Its size was as large as half a seat-mat, flucfuating multiple colours. It got smaller gradually disappeared during the sixth month of the following year.	[23]
14	222 Nov. 4	Wei Wendi Huangchu	3 9 Jiachen	A guest star was seen within the left gate of the Taiwei Enclosune (to the north of γ_η Vir).	[24]
15	290 April 27 -- May 25	Jin Wudi Taixu	1 4	A guest star was seen in the Zigong (at the UMi).	[25]

Table 1. (continued)

No.	Date in Christien era	Dynasty Emperor Epoch	Year Month Day	Statement	Ref.
16	304 Jun. 19 -- Jul. 18	Jin Huidi Yongxin	1 5	A guest star kept guard the Bi ($\delta\ \varepsilon\ \theta\ \gamma^1\ \lambda\ \alpha$ Tau).	[26]
17	369 March 24 -- April 22	Jin Heixi Taihe	4 2	A guest star was seen at the western enclosure of the Zigong (along the $\alpha\ \kappa\ \lambda$ Dra, 24 UMa and 43 α Cam) and disappeared in the 7th month.	[27]
18	386 April 15 -- May 13	Jin Xiaowu Taiyuan	11 3	A guest star appeared at the Nandou ($\mu\ \lambda\ \varphi\ \sigma\ \tau\ \zeta$ Sgr) and disappeared in the 6th month.	[28]
19	393 Feb. 27 -- March 28	Jin Xiaowu Taiyuan	18 2	A guest star appeared at the centre of the Wei ($\lambda\ \varepsilon\ \zeta\ \theta$ Sco) and disappeared in the 9th month.	[29]
20	396 Feb-May	Wei Taizu Huangshi	1	A comet appeared in the 6th month, before that, a big yellow star appeared between the Bi (around 37 Tau) and the Mao (Pleiades) for more than 50 days. In the 11th month the yellow star reappeared.	[30]
21	421 Jan. 20 -- Feb. 17	Wei Taizong Taichang	5 12	A guest star was seen at the Yi (Crt).	[31]
22	437 March 26	Wei Taiwudi Taiyan	3 1 renwu	A star was seen at the Jing (SW of Gem) in the northeast sky in daylight about 15 -17 . It was redish-yellow in colour and as large as an orange.	[32]

Table 1. (continued)

No.	Date in Christien era	Dynasty Emperor Epoch	Year Month Day	Statement	Ref.
23	541 Feb. 11 -- March 12	Wei Xiaojingdi Yuanxiang	4 1	A guest star appeared an the Zigong (UMi).	[33]
24	561 Sept. 26	Zhou Wudi Baoding	1 9 Yisi	A guest star appeared at the Yi (Crt).	[34] [35]
25	829 Nov.	Tang Wenzong Taihe	3 10	A guest was seen at the Shuiwei (NE of Cnc).	[36]
26	837 April 29	Tang Wenzhong Kaichen	2 3 Jiashen	A guest star appeared below the Dongjing (north of Mon, in the 4th month the Dongjing guest star disappeared.	[36]
27	837 May 3	Tang Wenzong Kaicheng	2 3 Wuzi	Another guest appeared within the Main Gate (of Taiwei Enclosure), near the Ping star ($\xi\ \nu\ \pi\ \rho$ Vir). On the Guiyou day of the 5th month, the Main Gate guest star disappeared.	[36]
28	900 Feb. 4 -- March 3	Tang Zhaozong Guanghua	3 1	A guest star appeared by the side of Huanzhe star within the Ziwei Enclosure, it was as large as a peach, while its rays shined the Huanzhe star (60 Her) so that the latter seemed to be concealed from view.	[37]
29	911 May 31 -- June 28	Liang Taizu Qianhua	1 5	A guest star trespassed against the Dizuo (α Her).	[38]

Table 1. (continued)

No.	Date in Christien era	Dynasty Emperor Epoch	Year Month Day	Statement	Ref.
30	926 April 20	Tang Zhuangzong Tongguang	4 3 Renxu	A guest star trespassed against the Tianku (NE of Cen).	[39]
31	1006 April 3	Song Zhenzong Jingde	3 3 Yisi	A guest star appeared in the southeast sky.	[40]
	1006 May 1	Song Zhenzong Jingde	3 4 2	On the lst day of the 5th month the Royal Astronomer said that he saw a big yellow star appeared to the east of Kulou (east of Cen) and the west of Qiguan (south of Lup), in the begining of night or the 2nd day of the 4th month. It became more bright gradually. The longitude was 3 degree from the Di(mension) in the longitude.	[41]
	1006 May 6	Song Zhenzong Jingde	3 4 Wuyin	A Zhoubo star appeared 1 degree west the Qiguan and East to Kulou. It was as a half moon, shining its rays, which could light up the objects on the earth. In the 9th month it was concealed by the twilight with the spining of the heaven-wheel. In the 11th month it was seen again in Di, since than was concealed. It was often seen in the morning from 11th month and then in the twilight in southwest sky in the 8th month.	[42]
	1006	Song Zhenzong Jingde	3	A big star was seen in the west of Di and brighten as a gold disc. No one had known it. Zhou Keming, the Royal Astronomer, said: this star should be named as Zhoubo according to the Jinzhou Astrology, in which it was writeen that it was sparkling and golden-yellow in colour for this kind of star. Besides, this star was regarded as a noble star which appeared unexpectedly according to the Yuan Ming Bao.	[43]

Table 1. (continued)

No.	Date in Christien era	Dynasty Emperor Epoch	Year Month Day	Statement	Ref.
	1006 June 11	Song Zhenzong Jinde	3 5 Renyin	The Royal Astronomer reported: a Zhoubo star was seen. The record in the Essential Records that Zhoubo star appeared in the 2nd of the 4th month was also reported by the of Royal Astronomer. This record followed the Real Record. The event of Gengchen day of the 4th month of the 9th year of Xiangfu epoch should be refered.	[44]
	1016 May 16	Song Zhengzong Xiangfu	9 4 Gengchen	The Royal Astronomer said the Zhoubo star was seen again. It should be refered to the event in the 1st day of 5th month in Jinde epoch.	[44]
32	1011 Feb. 8	Song Zhengzong Xiangfu	4 1 Dingchou	A guest star was seen in front of the Kui stars of the Nandou ($\varphi\_\sigma\_\tau\_\zeta$ Sgr).	[45]
33	1054 July 4	Song Renzong Zhihe	1 5 Jichou	A guest star appeared at several inches to the SE of the Tianguan. After more than a year it gradually faded away.	[46]
	1054 June 9 -- July 7	Song Renzong Zhihe	1 5	A guest star appeared at the Mao (Pleiades).	[47]
	1054 Aug. 27	Song Renzong Zhihe	1 7 22	Yang Weide, the Royol Astronomer said: I saw a guest star appeared which was yellow with some colour	[48]
	1054 Sept. 5 -- Oct. 4	Liao Xinzong Chongxi	23 8	A guest star appeared at the Mao (Pleiades).	[49]

Table 1. (continued)

No.	Date in Christien era	Dynasty Emperor Epoch	Year Month Day	Statement	Ref.
	1056 March 19 -- April 17	Song Renzong Jiayou	1 3	The Royal Astronomer said, the guest star disappeared, that mean visitors world leave, which had emerged in the east sky and kaept guard Tianguan(η Tau) in the morning in the 5th month of the 1st year of Zhihe epoch, like as Venus in daylight with (redish-white) rays shining around. It lasted for 23 days.	[50]
34	1065 Sept. 11	Liao Daozong Xianyong	1 8 Bingsen	A guest star trespassed against the Tianmiao (Ant).	[51]
35	1069 July 12	Song Shenzong Xiling	2 6 Bingchen	A guest star appeared at the longitnde of the Ji ($\gamma\_\delta\_\varepsilon\_\eta$ Sgr). It trespassed against the Ji and then disappeared on the Dinmao day of the 7th month.	[52]
36	1070 Dec. 25	Song Shenzong Xiling	3 11 Dingwei	A guest star appeared at the Tianqun (NE of Cet). A guest star appeared at the Lou (west of Ari).	[52] [53]
37	1087 July 3 -- Aug. 1	Liao Daozong Xianyong	23 6	A star, like a melon, appeared at the Wenchang (ν_φ_o_15_18 UMa).	[54]
38	1138 June 9 -- July 8	Song Gaozong Shaoxin	8 5	A guest star kept guard the Lou (α_β_γ Ari)	[46]
39	1139 March 23	Song Gaozong Shaoxin	9 2 Rensen	A guest star kept guard the Kang ($\varphi\_\tau\_\kappa\_\lambda$ Vir).	[46]

Table 1. (continued)

No.	Date in Christien era	Dynasty Emperor Epoch	Year Month Day	Statement	Ref.
40	1166 May 1	Song Xiaozong Qiandao	2 3 Guiyou	A guest star, small and white-bluish, appeared at the west of the large star of the Wudizhuo (β Leo) within Taiwe enclosurec. It looked small and it was bluish-white in colour.	[46]
41	1181 Aug. 6	Song Xiaozong Cunxi	8 6 Jisi	A guest star appeared at the Kui mansion, trespassed against the Cuanshe and disappeared on the Guiyou day of the first month of the following year. It lasted for 185 days.	[46]
42	1203 July 28	Song Ningzong Jiatai	3 6 Yimao	A guest star appeared between the Wei stars ($\varepsilon\ \mu\ \zeta\ \eta\ \theta\ \iota\ \lambda$ Sco) in the SE sky it was bluish-white in colour and it was of the same size as Saturn. It kept guard the Wei on the Jiazi day.	[46]
43	1224 July 11	Song Ningzong Jiading	17 6 Jichou	A guest star kept and trespassed against the Wei mansion.	[46]
44	1240 Aug. 17	Song Lizong Jiaxi	4 7 Gengyin	A guest star appeared at the Wei mansion.	[46]
45	1244 May 14	Song Lizhong Cunyou	4 4 Bingzi	A guest star appeared at the Wei mension. It was as large as the Venus.	[55]
46	1248	Song Lizong Cunyou	8	A star, large as Venus, appeared by the Hegu ($\alpha\ \beta\ \gamma$ Aqr).	[56]
47	1388 March 27	Ming Taizu Hongwu	21 2 Bingyin	A star, yellow-redish, appeared in the Dongbi (α And γ _ P	[57]

Table 1. (continued)

No.	Date in Christien era	Dynasty Emperor Epoch	Year Month Day	Statement	Ref.
48	1404 Nov. 14	Ming Taizong Yongle	2 10 Gengchen	A star, yellow in colour and smooth in luster emerged to the SN of Niandao (R_η_θ_4_17 Lyr) brightening as a lamp. It did not move.	[57]
	1408 Sept. 10	Ming Taizong Yongle	6 8 Bingsen	A star, as big as a lamp, bluish in colour, appeared in the west sky at night.	[58]
	1408 Oct. 24	Ming Taizong Yongle	6 10 Gengchen	A big star appeared at SE of the Niandao. It was yellow in colour and smooth in lu ster. It did not move. It was the Zhoubo, a noble star.	[59]
	1408 Oct. 24	Ming Taizong Yongle	6 10 Gengchen	A star, yellow in colour and smooth in luster emerged to the SN of the Niandao brightening as a lamp. It did not move and it was the Zhoubo, a noble star.	[60]
49	1430 Sept. 3	Ming Xuanzong Xuande	5 8 Jiashen	A guest star, dark bluish violet in colour, was seen 1ft off the northeast of Nanhe (α_β_γ CMi).	[61]
	Sept. 9	Ming Xuanzong Xuande	5 8 Gengyin	A star, bluish yellow in colour, was seen about 1ft from the Nanhe.	[62]
				A star, as large as a pellet, dark bluish in colour had appeared by the Nanhe for 26 days and then disappeared. It was seen again in Renwn day of the 3th month of the following year(April 29, 1431).	[57]

Table 1. (continued)

No.	Date in Christien era	Dynasty Emperor Epoch	Year Month Day	Statement	Ref.
50	1431 Jan. 4	Ming Xuanzong Xuande	5 12 Dinghai	A star, as a pellet, yellowish white in colour, smooth in luster had been seen by the Jiuyou (fareast of Eri) for 15 days and then disappeared. It reappeared on the Renwu day of the 3th month of the 6th year of Xuande epoch of Emperor Xuanzong (1431. April 29).	[63]
				Hanyu star, as a pellet, smooth in luster, was seen in the evening. It was yellowish-white in colour and smooth in Iuster.	[64]
	1431 Jan. 15	Ming Xuanzong Xuande	5 12(Run) Wuxu	The Royal Astronomer reported that a Hanyu star, as large as a pellet, had been seen for 10 nights. It get more bright and yellowish-white in colour.	[65]
51	1497 Sept. 20	Ming Xaiozong Hongzhi	10 8 Guisi	In Nanjing, a guest star was seen near tha star Tianji stars (θ-σ-ρ And) in the evening.	[66]
52	1572 Nov. 8	Ming Muzong Longqing	6 10 3	A guest star was seen at the NE sky. It like a pellet, and appeared near the Gedao (τ-ε-δ-φ-θ-ν-oCas) at the longitude of the Bi mension. It had extended rays and became more bright gradually during 19 days. At night of the Renshen day, this star was as big as a lamp with redlish-yellow rays shining around. On the Jiaxu day of the 12th month, the Ministry of Rites reported: the guest star was seen in the daylight with unasual brightness.	[67]
	Nov. 6 -- Dec. 4	Ming Muzong Longqing	6 10	An evil star was seen in the NE sky after midnight every day. It was as bright as the moon.	[68]
	Winter	Ming Muzong Longqing	6 Winter	An evil star was seen.	[69]

Table 1. (continued)

No.	Date in Christien era	Dynasty Emperor Epoch	Year Month Day	Statement	Ref.
53	1604 Oct. 10	Ming Shenzong Wanli	32 9 Yichou	A star, reddish-yellow in colour, like a pellet, was seen at the longitude of the Wei in the heast sky and was concealed in the 10th month. On the Xinyou day it turned to the southeast, and still stay at the longitude of the Wei. It got dim gradually in the and month and went out on the Dingmao day of the 8th month of the following years (Oct. 7, 1605).	[70]

3. Historical Supernovae

3.1 Fuzzy classification

The ancient records are always very sketchy and sometimes ambiguous. Especially the records on the brightness of the star are always not quantative. To distinguish the supernova from the events in the list, we attempt to give a quantative criterion in some extent by means of the fuzzy mathematics. Only the information on the brightness and time when the star appeared are available from the ancient observations.

First, we sieve the events by the statement. A supernovae must be brilliant and seen for a long time so that we take the records which have the description on the brightness and visible duration as the candidates. By sieving in this way we then have 28 candidates for supernovae.

Second, we evaluate the membership degree of the candidates for supernovea by means of fuzzy mathematics consideration.

In fact, the descriptions on the brightness are quite fuzzy in the records. Besides, we can not give a definite minimum about both the brightness and visible duration for a supernova. We can only give a measure about the events to be subordinate to the set of supernovae.

To give the membership degree from the brightness we consider the description of events as three types.

The descriptions, such as "Zhoubo star" or "as venus seen in daylight", "like the moon" or "like a lamp", are the most reliable information because this kind of statement had been used to describe the SN 1572 and 1604 which are undoubtedly recognized as supernovae.

In some of records, the brightness is described by comparison with planets such as Venus, Jupiter, and Saturn. It isn't difficult to decide the brightness.

However, many events are described as large as like something, such as seat mat, melon, grain measurer, fruit and etc.. We attempted to evaluate the brightness in magnitude and give the membership degree in Table 2.

To give the membership degree according to the visible duration of the star, it is useful to look at the distribution of the number of events about the visible duration. Two peaks appeared in the distribution (Fig.1) which is likely to represent a combined distribution of novae and supernovae.

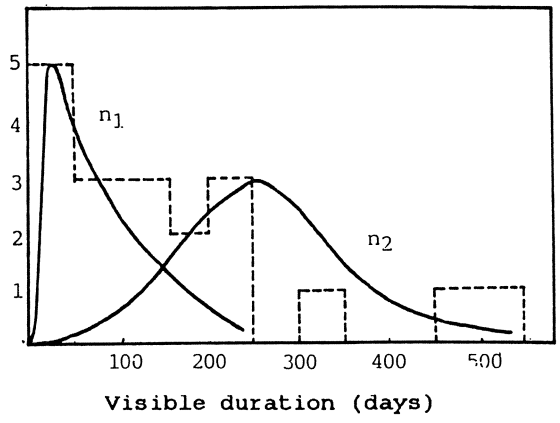


Fig.1 The distribution of the visible duration of supernovae.

Table 2. Brightness Calibration And Membership Degree For SN

descriptions on the brightness	mag.	membership degree
like the Moon	-6	1.0
like half Moon	-6	1.0
like a lamp	-5	1.0
like Venus seen the daylight	-4 - -5	1.0
like Venus	-3 - -4	0.9
large as Jupiter	-1.5 - -2.5	0.6
large as Saturn	-0.5	0.3
large as half a seat mat	-3 - -6	0.8
large as corn measurer	-3.5 - -5	0.7
large as melon	-3 - -4	0.6
large as peach, orange	-2 - -3	0.5
large as pellet	-1 - -3	0.4
a big star	-1 - -4	0.3

We can fit the distribution by two gamma distributions. To simplify, we can take the membership degree as follows

$$f_2(\Delta t) = \begin{cases} 0 & \Delta t \leqslant 50 \\ (\Delta t - 50)/150, & \text{if } 50 < \Delta t < 150 \text{ (days)} \\ 1 & \Delta t \geqslant 150. \end{cases}$$

Finally we take the average of these two membership degrees for supernovae. (Fig.1).

Table 3. The membership degree of the candidates for SN

events	visible	description on brightness	membership degree			
			f_1	f_2	f_1+f_2	$f_1 V f_2$
BC 48		large as melon		0.6	0.6	0.6
AD 64	75		0.17		0.17	0.17
70	48	large as half seat mat	0		0	
185	140		1.0	0.8	1.8	1.0
369	150		0.5		0.5	0.5
386	90		0.27		0.27	0.27
393	210		1.0		1.0	1.0
396	50	big yellow star	0	0.2	0.2	0.2
437		large as orange		0.4	0.4	0.4
483		large as corn measurer		0.8	0.8	0.8
837a	22		0			
837b	44		0			
900		large as peach		0.6	0.6	0.6
1006	240	like half Moon, "Zhoubo"	1.0	0.1	2.0	1.0
1054	540	as Venus seen in daylight	1.0	1.0	2.0	1.0
1069	12		0			
1087		large as corn measurer		0.8	0.8	0.8
1181	185		0.9		0.9	0.9
1203		large as Saturn		0.2	0.2	0.2
1244		large as Venus		1.0	1.0	1.0
1248		large as Venus		1.0	1.0	1.0
1408	102	as lamp, "Zhoubo" star	0.35	1.0	1.35	1.0
1430	26	large as pellet	0	0.4	0.4	0.4
1431	104	large as pellet	0.36	0.4	0.76	0.4
1572	480	Bright as Meen seen in daylight	1.0	1.0	2.0	1.0
1604	330	as Venus seen in daylight, Zhoubo	1.0	1.0	2.0	1.0

Then a sequence results of the membership degree for supernovae among the candidates. From the distribution of the number of events over the membership degree we can see that there are three groups (see Fig.2):

The first group includes 185, 1006, 1054, 1572 and 1604, which always take the highest membership degree no matter which way to obtain the final membership degree is used.

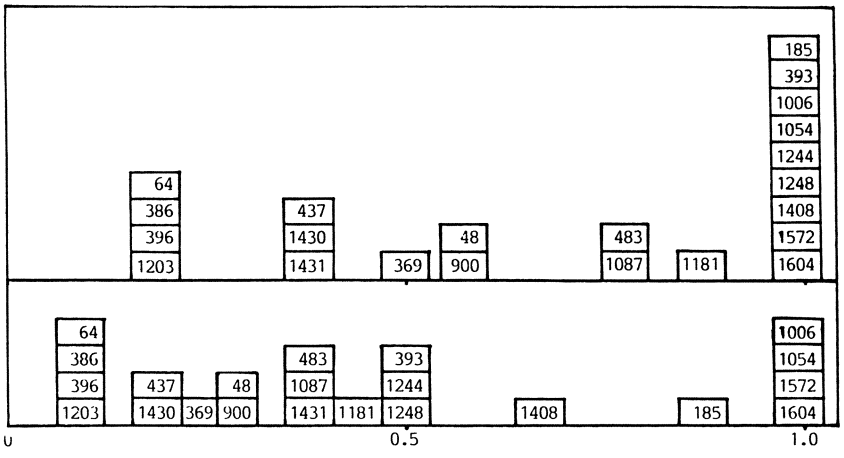


Fig.2 Membership degree of the supernovae

The second group includes the events of 1408, 393, 483, 1087, 1181, 1244, 124 and 1431 which always take the middle membership degree but the order of them depends on the way used with. These are the probable supernovae.

The third group has a smaller degree which can be considered as supernovae only additional evidences have been found.

3.2 Discussion

1. It is shown in the Fig.2 that the explosion of supernovae occurred frequently during two periods, i.e. BC 134-AD 483 and AD 900-1604. Each of these periods lasts about 600-700 years. After those an intermittant period follows, i.e. 6th-10th centuries and 17th-20th centuries, in which no supernova exploded. One probably should indicate a thousand-year periodicity for the explosion of supernovae in the Galaxy.

It is noticed that there is not any obvious drop of the number of the records of comets per century in those intermittant period for supernova, the absence of the supernova in those period is not due to the absence of astronomical records caused by some effect of the artificial facts.

However it raises the question of whether there is any physical reality of the periodicity and the question how the periodicity is explained.

2. The guest star in 1006 has been generally accepted as a historical Supernova.

The ancient records show some remarkable characteristics from which some consequences can be pointed out.

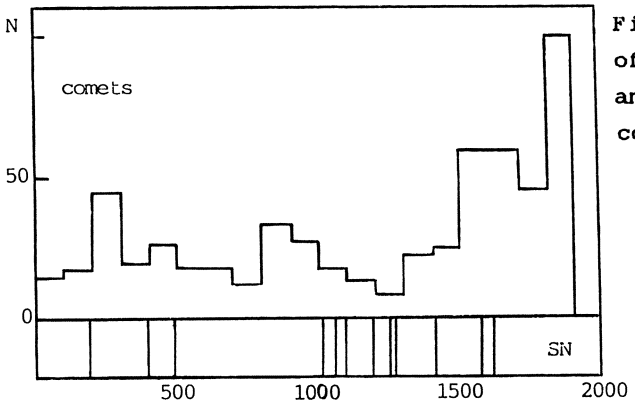


Fig.3 The time sequence of historical supernovae and the frequency of the comets

(1) The guest star in 1006 was also refered to as Zhoubo star. Ten years later, a Zhoubo star reappeared. Moreover, it has been noticed that the record of the event of 1016 should be refered to that in 1006, which means that the Royal Astronomer had identified the Zhoubo star in 1016 to that in 1006. Therefore these two events should be considered as a double explosion of the SN1006. Wang Jianmin had pointed out this fact [7].

(2) It was recorded in the Essential Records in Song dynasty that the guest star became more bright after the explosion. This is a remarkable characteristics of the variation of the brightness of supernovae which appears similar to the SN 1987A but quite different from the other supernovae.

(3) It was recorded for the guest star in 1006 that the star was often in the east sky in the morning from the 11th month and it was concealed in the twilight. The description indicated that the supernova was visible till the 8th month of the following year so that the guest star lasted for more than 450 days instead of 240 days that had been generally considered.

3. The Chinese record on the guest star of 1408 had never been listed as a nova or supernova until the evidence has been given by the author and supported by several authors. The conclusion is proved in the section 2, in which a japanese record (July 14, 1408) has been taken in account for giving the visible duration. I once guessed that Cyg X-1 originated in the explosion of SN1408. However Storm [72], Wang and Seward [73] argued that the guest star should be identified with the supernova remnant CTB80. Examining more carefully the coincidence of the position of the CTB80 with the guest star it seems that their identification is reasonable. Nevertheless I should mention that a pulsar might be at the centre of the CTB80 since the remnant of the SN 1408 is so young.

4. The guest stars in 393, 1203, 1224, 1240 and 1244 were seen in the Wei mansion --- the southeast of the Scorpius. It seems to be not accidental that the four novae or supernovae happened to appear even during the 42 years from 1203 to 1244. Could we consider them as the same supernova in multiple explosion?

5. The events in 396 is likely another example of a double explosion of a SN.

References

1. Zhuang Weifeng et al.: The General Compilation of the Chinese Ancient Records of Astronomical Phenomena (Science and Tech., Jiangshu 1987).
2. Ma Duanlin: Historical Investigation of Public Affairs (Wenxian Tongkao), (1254).
3. M.E. Biot: Catalogue des Etoiles Extraordinaires Observees en Chine depuis les Temps anciens jusqu'a l'an 1203 de Notre Ere; Connais. Temps, Additions, P.60, (1846).
4. M.E. Biot: Catalogue des Etoiles Extraordinaires Observees en Chine depuis l'an 1230 jusqu'a l'an 1640 de Notre Ere, Connais Temps, Additions, P.44 (1846).
5. K.Lundmark: Suspected New Star recorded in Old Chronicles and among Recent Meridian Observations, PASP, Vol.33, P.225 (1921).
6. Xi Zezong: A new Catalogue of Ancient Novae, Acta Astronomica Sinica, 3. P.183 (1955).
7. Xi Zezong and Bo Shuren: Acta Astronomica sinica, Vol.13, P.1, (1965).
8. Ho Peng Yoke: Ancient and Mediaeval Observations of Comets and Novae in Chinese Sources, Vistas in Astronomy, Vol.5, P.127 (1962).
9. D.H. Clark and F.R. Stephenson: The Historical Supernovae (1977).
10. Li Qibin: Cygnus X-1 -- The Remnant of the Historical Supernovae of 1408. Vol.3, P.315 (1979).
11. Chen Mengjia: A Review on the Oracle Inscriptions of Yin Dynast Ruins (1956).
12. Zhushu Jinian: The Bamboo Annals, Part 2, P.25.
13. History of Han Dynasty (Hanshu), Astronomical Chapter, Vol.26, P.1305.
14. Ibid, P.1307.
15. Ibid, P.1309.
16. History of Later Han Dynasty (Houhanshu), Astronomical Chapter, Vol.11, P.3230.

17. Ibid, P.3231.
18. Ibid, P.3237.
19. Ibid, P.3238.
20. Ibid, P.3242.
21. History of Later Han Dynasty (Houhanshu), Biographies, Vol.58, P.1871.
22. History of Later Han Dynasty (Houhanshu), Astronomical Chapter, Vol.11, p.3243.
23. Ibid, Vol.12, P.3260.
24. History of Liu-Song Dynasty (Songshu). Astronomical Chapter, Vol.23, P.680.
25. Ibid, Vol.23, P.695.
26. Ibid, Vol.24, P.703.
27. Ibid, Vol.24, P.719.
28. Ibid, Vol.24, P.724.
29. Ibid, Vol.25, P.726.
30. History of Wei Dynasty (Weishu), Astronomical Chapter, Vol.105, P.2389.
31. Ibid, P.2399.
32. Ibid, P.2404.
33. Ibid, P.2447.
34. History of Zhou Dynasty (Zhoushu), Emperor Wudi, Vol.5, P.65.
35. History of Sui Dynasty (Suishu), Astronomical Chapter, Vol.21, P.604.
36. New History of Tan dynasty (Xingtangshu), Astronomical Chapter 2, Vol.32, P.839.
37. Ibid, P.841.
38. New History of the Five Dynasties (Jiu Wudaishi), Astronomical Chapter, Vol.59, P.76.
39. Old History of the Five Dynasties (Jiu Wudaishi) Biography of Emperor Zhuangzong, Vol.34, P.475.
40. Ma Duanlin: Historical Investigation of Public Affairs Vol.29, P.2327.
41. Comitation of Essential Records of Song Dynasty (Song Huiyao Jigao) Vol.52, P.2080.
42. History of Song dynasty (Songshi), Astronomical Chapter 9, Vol.56, P.1226.
43. Tao Zongyi: Talk in Saloon (Yutang Qinghua), Vol.8, P.312.
44. Supplement to the Mirror of Universal History (Xu Zhizhi Tongjian Changpian).
45. History of Song Dynasty (Songshi), Astronomical Chapter 9, Vol.56, P.1227.

46. Ibid, P.1229.
47. Outline of History (Shishi Zengang), Vol.24, P.21.
48. Compilation of Essential Records in Song Dynasty (Song Huiyao Jigao), Vol.52, P.2080.
49. History of Qidan Country (Qidan Guozhi), Vol.8, P.6.
50. Compilation of Essential Records in Song Dynasty (Song Huiyao Jigao) Vol.52, P.2086.
51. Supplement to the General History (Xu Tongzhi), Vol.171, P.4299.
52. History of Song Dynasty (Songshi), Astronomical Chapter 9, Vol.56, P.1230.
53. Supplement to the Mirror of Universal History (Xu Zhizhi Tongjian Changpian), Vol.217, P.8.
54. History of Qidan Country (Qidan Guozhi), Vol.9, P.4.
55. History of Xuanfu Town (Xuanfu Zhenzhi), Vol.5, P.90.
56. History of Shaoxingfu (Shaoxing Fuzhi), Vol.13, P.26.
57. History of Ming Dynasty (Mingshi), Astronomical Chapter 3, Vol.27, P.404.
58. History of Sichuan Province (Sichuan Tongzhi), Vol.38, P.8.
59. Tan Qian: Guo Que, Vol.14. (Zhonghua 1958).
60. Essential Records of Taizong Emperor in Ming dynasty (Ming Taizong Shilu), Vol.59, P.7.
61. Tan Qian: Guo Que, Vol.21, P.1397 (Zhonghua, 1959).
62. Essential Records of Xuanzong in Ming Dynasty (Ming Xuanzong Shilu). Vol.68, P.4.
63. History of Ming Dynasty (Mingshi), Astronomical Chapter 3, Vol.27, P.405.
64. Tan Qian: Guo Que, Vol.21, P.1406 (Zhonghua, 1958).
65. Essential of Xuanzong Emperor of Ming Dynasty, (Ming Xuanzong Shilu), Vol.74, P.1.
66. Essential of Xiazong Emperor of Ming Dynasty (Ming Xiaozong Shilu).
67. Essential of Shenzong Emperor of Ming Dynasty (Ming Shenzong Shilu). Vol.6, P.1.
68. History of Yifeng County, Vol.4, P.11.
69. History of Ming Dynasty (Ming Shi), Biographies, Vol.215, P.5686.
70. Ibid, Astronomical Chapter 3, Vol.27, P.406.
71. Wang Jianming: Publications of Beijing Observatory, No.6, P.68(1978).
72. R.G. Strom, P.E. Angerhofer and T. Velusamy: Nature, Vol.284, P.38(1980).
73. Z.R. Wang and F.D. Seward: Ap.J., Vol.285, P.607(1984).

Models of Type II Supernova and SN 1987A: Observations Versus Theoretical Predictions

W. Hillebrandt

Max-Planck-Institut für Physik und Astrophysik, Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-8046 Garching, Fed. Rep. of Germany

The supernova 1987A in the Large Magellanic Cloud is unique with regard to its distance, the fact that we know the properties of the progenitor star, and the detection of neutrinos emitted in the event. Moreover, since SN 1987A is classified as type II supernova and because with, maybe, the exception of SN 1054, none of the historical galactic supernovae belonged to that class, it presents a unique opportunity to confront theoretical models with direct observational facts. We will, therefore, discuss core collapse models of massive stars as well as possible interpretations of the neutrino events. It will be demonstrated that, although our general ideas are now well confirmed by observations, certain facts are difficult to understand within the framework of standard scenarios.

1. Introduction

On February 24, 1987, a supernova (SN) was discovered in the Large Magellanic Cloud (Shelton, 1987). It is by now well established that the position of the supernova SN 1987A coincided with that of a blue supergiant Sanduleak (SK) −69 202 (West et al., 1987), which actually consisted of three components. From UV-observations it was confirmed that the brightest component, a B3 Ia supergiant, has indeed disappeared (Panagia et al., 1987). From its bolometric luminosity of about 5×10^{38} erg s^{-1} it can be inferred that the progenitor of SN 1987A had a mass of around 20 $M_\odot$ on the main sequence (Arnett, 1987; Hillebrandt et al., 1987a; Woosley et al., 1987). The surprising fact that the progenitor was a blue rather than a red supergiant may be attributed to lower than solar metallicity (Brunish and Truran, 1982; Hillebrandt et al., 1987; Arnett, 1987) or significant mass loss (Chiosi and Maeder, 1986; Woosley, Pinto and Ensman, 1987) or a combination of both (Truran and Weiss, 1987; Maeder, 1987).

Stars with main sequence masses of about 20 $M_\odot$ form helium cores of 6 $M_\odot$ after core-hydrogen burning and iron cores of about 1.5 $M_\odot$ at the end of their hydrostatic evolution (Arnett, 1977; Weaver, Woosley and Fuller, 1985; Wilson et al., 1986; Hashimoto and Nomoto, 1987). These iron cores will collapse to neutron star densities, thereby radiating away a significant fraction of the gravitational binding energy of a neutron star in form of neutrinos. The facts that neutrinos from SN 1987A indeed have been observed (Hirata et al., 1987; Bionta et al., 1987; Aglietta et al., 1987) and that the energy in neutrinos was around a few 10^{53} erg (Sato and

Suzuki, 1987; De Rújula, 1987; Hillebrandt, 1987) clearly prove the validity of this general picture.

However, attempts to fit all observations by simple theoretical models face difficulties. The very rapid rise of the early visual light curve of SN 1987A requires an explosion energy of (2 ± 1) x 10^{51} erg (Arnett, 1987; Woosley, Pinto and Ensman, 1987; Shigeyama et al., 1987) which was not predicted for stars as massive as SK -69 202. An analysis of the expansion of the photospheric radius during the first few days seems to indicate rather strange changes of the velocity (Schmidt-Kaler, Hanuschik and Dachs, 1987) which cannot be explained by simple explosion models. Finally, if both neutrino bursts associated with SN 1987A were indeed real, rather complicated scenarios have to be invented. All these questions will be addressed in some detail in the following sections.

2. Core Collapse and Supernova Explosions

Type II supernova are thought to be the outcome of the collapse of the cores of massive stars, leaving behind a compact remnant, presumably a neutron star or eventually a black hole. The energy observed in the explosion must then ultimately come from the gain in gravitational binding of the core. Two mechanisms that are potentially able to transform a small fraction of this energy into outward momentum of the stellar envelope have been discussed in the past, neither of which, however, so far gives a satisfactory explanation.

In one class of models the stellar envelope is ejected by a hydrodynamical shock wave generated by the rebounding inner core near nuclear matter density, but this mechanism seems to work only for a very limited range of precollapse stellar models. In fact, one can show quite generally that the crucial quantity which determines whether the shock wave can propagate successfully out to the stellar envelope or will turn into a standing accretion shock is the difference between the mass of the unshocked inner part of the core and the mass of the original iron core (Hillebrandt, 1982, 1984). If this difference exceeds approximately 0.5 $M_\odot$, damping of the shock by nuclear photodissociations becomes so severe that no explosion results.

Given the fact that the main sequence mass of SK -69 202 was around 20 $M_\odot$, stellar evolution models predict that it should have formed an iron core of (1.6 ± 0.2) $M_\odot$ (Wilson et al., 1986; Arnett, 1987; Hashimoto and Nomoto, 1987). For central entropies at the onset of collapse of about (1.0 ± 0.2) k_B/ nucleon, predicted from stellar evolution models, the average lepton fraction of the inner core at bounce will be around 0.35 and, consequently, the mass of the unshocked inner core will be around (0.7 ± 0.1) $M_\odot$ (Bruenn, 1985; Hillebrandt, 1986; Wilson et al., 1986). Most hydrodynamical simulations predict that initially the shock energy is around (5 ± 3) x 10^{51} erg (Hillebrandt, 1982, 1986; Wilson et al., 1986). Photodissociations cost approximately 8 x 10^{18} erg per gramme, or 8 x 10^{51} erg per

0.5 $M_\odot$. From the numbers given above it is therefore not surprising that in none of the numerical simulations performed so far a prompt explosion was found for a 20 $M_\odot$ stellar model.

An alternative explosion mechanism for massive stars has been investigated by Wilson (1985) (see also Bethe and Wilson (1985), Wilson et al. (1986), and Lattimer and Burrows (1984)). It was found that a few hundred milliseconds after core-bounce energy transport by neutrinos may revive a stalling shock. At that time the edge of the unshocked inner core is at a radius of about 20 km and the density and temperature there are about 10^{12} gcm^{-3} and 2 to 3 MeV, respectively. The shock has changed to an almost standing accretion shock at 1.4 $M_\odot$ and a radius of about 500 km, the temperature and the density just behind the front being 0.8 MeV and 10^7 gcm^{-3}, respectively. The shocked matter, composed of free neutrons and protons, is irradiated by a neutrino flux originating from the neutrino sphere at 30 to 70 km. Typical neutrino energies range from 9 to 15 MeV and the neutrino luminosity is around (2-4) x 10^{53} erg s^{-1}. Some of the neutrinos and anti-neutrinos will be captured by free neutrons and protons, respectively, and will be reemitted with lower (thermal) energies, thus heating the matter behind the accretion shock.

Wilson (1985) and Wilson et al., (1986) have computed the hydrodynamic evolution of several stellar models for approximately 1s after core-bounce and found explosions in all cases. Mostly, however, the explosion energy ((2-3) x 10^{50} erg) was too low to explain typical type II supernova light curves, and there is little hope that the delayed esplosion mechanism can explain the very energetic outburst of SN 1987A. This can also be seen from a simple argument. Neutrino heating time scales are typically of the order of several tenths of a second and thus much longer than the hydrodynamical time scale of the matter under consideration. Consequently, the shells being heated will start to expand once a sufficiently high overpressure has been built up to overcome gravity. From then on no significant further heating will occur. Therefore, the energy available to drive the explosion will be of the order of gravitational binding energy prior to the expansion, namely of the order of a few times 10^{50} erg, in agreement with the results of numerical simulations. Our argument would fail only if the neutrino heating could occur on hydrodynamic time scales, which, however is very unlikely.

So in conclusion, it seems that none of the proposed explosion mechanisms can fully explain SN 1987A. The high energy of the outburst seems to indicate a prompt explosion, whereas the high mass of the progenitor makes a delayed explosion more likely, but there may be ways to circumvent these problems.

Firstly, rotation may alter the general pictures outlined so far considerably, and at present it cannot be excluded that SK -69 202 possessed a rapidly spinning core. Recent equations of state predict that the adiabatic index below nuclear matter density is only slightly smaller than the critical value of 4/3 of a non-rotating star. Therefore, already moderate rotation may cause the stellar core to bounce before

nuclear matter density is reached (Müller, Rozyczka and Hillebrandt, 1980; Symbalisty, 1984). The further evolution may be significantly different from that of a non-rotating star. Non-spherical velocity fields will also lead to non-spherical shocks which may introduce large-scale circulations and thus a more efficient neutrino transport on hydrodynamical rather than on diffusion time scales (Müller and Hillebrandt, 1981). Rapidly rotating cores may even be unstable to triaxial deformations. A significant fraction of the rotation energy could then be dissipated into heat at the boundary between the core and the axi-symmetric stellar mantle.

Secondly, the equation of state is not well known above nuclear matter density. Somewhat softer equations of state than currently used in supernova models and somewhat smaller iron cores will favour prompt explosions, as was demonstrated by Baron, Cooperstein and Kahana (1985). Whether or not prompt explosions can be obtained with reasonable changes of the equation of state for 20 $M_\odot$ stellar models has not yet been demonstrated.

3. Neutrinos from SN 1987A and their Interpretation

Events in neutrino detectors have been seen prior to the optical outburst of SN 1987A at Feb. 23.12 (UT) (Aglietta et al.,1987) and at Feb. 23.32 (UT) (Hirata et al., 1987; Bionta et al., 1987), respectively. Two photographs taken at about Feb. 23.44 (UT) (McNaught, 1987) show the supernova at a visual magnitude of about $6^m.4$ only 2.8×10^4 sec and 1.1×10^4 sec, respectively, after the neutrino events were discovered. This close correlation in time indicates that the neutrinos were indeed emitted from the exploding star.

The pulse seen in the Mont Blanc experiment (Aglietta et al., 1987) consisted of 5 events spread over $\Delta t \simeq 7$ sec with measured positron energies between 6 and 8 MeV (Galeotti, 1987). From the second pulse, KAMIOKA (Hirata et al., 1987) detected 12 neutrinos spread over $\Delta t \simeq 13$ sec, whereas the IMB detector (Bionta et al., 1987) saw 8 neutrinos with $\Delta t \simeq 6$ sec. The measured electron (positron) energies in the second pulse where significantly higher and ranged from 6 to 35 MeV (KAMIOKA) to 20 to 40 MeV (IMB). The angular distribution of events in those latter experiments was not quite isotropic, as one would expect from the reaction $\bar{\nu}_e + p \rightarrow e^+ + n$. Finally the Baksan group reported 6 events 30 sec after the IMB detection (Alexeyeva, 1987).

The crucial question now is whether or not the two (or three?) neutrino bursts were real. In order to answer this question one has to check the statistical significance of each detection, the consistency of the various experiments, and the energetics of the neutrino bursts. Although the significance of the KAMIOKA detection is certainly orders of magnitude higher than the Mont Blanc detection, and the absence of a clear signal in the KAMIOKA-detector during the time of the Mont Blanc events may cause a consistency problem, from statistical arguments alone one cannot rule out the possibility that also the first neutrino pulse was indeed real (De Rujula,

1987; Hillebrandt, 1987). However, in order to avoid an inconsistency between the Mont Blanc detection and the fact that no clear signal was seen at that time by the KAMIOKA experiment, one has to postulate that, in the first pulse, the average neutrino energy was fairly low, namely of the order of 6 MeV. Based on this assumption one can estimate the total energy in $\bar{\nu}_e$'s in the first pulse and finds roughly $(3 \pm 2) \times 10^{53}$ erg. Because about the same amount of energy will be radiated in ν_e's, and also ν_μ's and ν_τ's may have contributed to the energy loss of the supernova it has been argued that the forming neutron star cannot supply the required energy and, therefore, the Mont Blanc events must be due to some noise in the detector. One has to keep in mind, however, that these energy arguments are based on the poor statistics of a few events. Because of these rather large uncertainties the Mont Blanc data may still be consistent with the assumption that a neutron star was formed in SN 1987A.

Assuming again that also the KAMIOKA and IMB events were $\bar{\nu}_e$'s, Sato and Suzuki (1987), and many others, have analysed the second burst and have found average neutrino energies between 9 and 15 MeV and a total energy in all three neutrino flavours of about $(1 - 3) \times 10^{53}$ erg. But again these conclusions are rather uncertain, because we do not know for sure which (if any) of the detections were due to (ν, e)-scattering. As was mentioned before, several of the events detected by KAMIOKA were strongly forward directed indicating scattering rather than $(\bar{\nu}_e, p)$-reactions. Because of the much smaller scattering cross-section such events would encrease the integrated neutrino luminosity considerably.

Most attempts to interpret the neutrino data in the framework of astrophysical models start from the assumption that the first pulse was not real but rather was due to noise in the Mont Blanc detector. The interpretation of the second pulse is then straightforward and in good agreement with theoretical predictions (Arnett and Rosner, 1987; Burrows and Lattimer, 1987; Bahcall and Glashow, 1987), and, in particular, with those made by the delayed explosion models. Mayle (1985) (see also Mayle, Wilson and Schramm (1987)) has computed neutrino spectra and luminosities up to 1.4 sec after core bounce, and his results fit the KAMIOKA data quite well, although the average neutrino energy and the luminosity of SN 1987A seem to be somewhat higher than was predicted by the theoretical models. Only the rather long duration of the second neutrino pulse and the time distribution of the events may cause a problem. Burrows and Lattimer (1986) have shown that about 80% of all thermal neutrinos above 7 MeV are emitted from a newly born neutron star during the first 2 sec and 95% are emitted during the first 4 sec of its life. It is, therefore, difficult to explain why after more than 10 sec still high energy neutrinos have been seen. One possibility would be that the last three events seen by KAMIOKA and the last two events observed by IMB were noise, in which case the observed pulse width of about 2 to 3 sec would coincide with theoretical predictions. In contrast, one could argue that neutrinos have a finite rest mass of a few eV/c^2 and, therefore, the pulse has

been smeared over about 10 sec (Hillebrandt et al., 1987b, 1987c). In any case, if the interpretation is correct that the second pulse signaled the formation of a neutron star, it should become visible either as a thermal x-ray source, a pulsar, or a synchrotron nebula after the supernova envelope has become transparent.

A more complicated scenario has to be invented if both neutrino pulses were real. It has to explain why the second pulse was delayed by about five hours and why the energy in both pulses was equal to within factors of about two. A possible explanation would be that the first pulse was emitted when the core bounced at neutron star densities, whereas the second pulse came from the further collapse to a black hole (Hillebrandt et al., 1987c; De Rújula, 1987) or a strange matter configuration (Haubold et al., 1987; De Rújula, 1987; Schmidt-Kaler, Hanuschik and Dachs, 1987). The time delay could be due to neutrino cooling (Hillebrandt et al., 1987c; Haubold et al., 1987) or electromagnetic breaking, provided the proto neutron star was stabilized by rotation and possessed a magnetic field of several 10^{13} Gauss (Hillebrandt, 1987). In particular, the phase transition to a superdense state might supply the right amount of neutrinos on time scales of the order of the neutrino diffusion time. But also the formation of a black hole is not out of range, because hydrodynamic simulations predict that for rather massive iron cores the mass of the newly born neutron star may exceed the critical mass of a cold non-rotating object.

The question, therefore, arises whether there exists additional information which favours this second scenario. There is evidence that the ejecta of SN 1987A were enriched in CNO-processed material (Casatella, 1987), namely in N relative to C. If the progenitor star had possessed significant rotation meridional mixing would provide a natural explanation (Ramadurai and Wilta, 1987). Asymmetries in the P-Cyg profiles of H_α have been interpreted as being due to a non-spherical expansion of the supernova (Cristiani and Gouiffes, 1987; Dachs and Hanuschik, 1987) and observed polarizations of the optical emission during the first days of the outburst point into the same direction (Barrett, 1987). The Bochum group (Schmidt-Kaler, Hanuschik and Dachs, 1987) has determind the expansion velocity of the photospheric radius of SN 1987A from spectral information. They find that their data are very well fitted by a constant velocity of about 11 000 km s^{-1} for the first 4 days of the outburst. However, extrapolating back to the radius of the progenitor star they find that the expansion should have started about 8 hours after the first optical sighting of the supernova as a 6.4 magnitude object by McNaught (1987). This is clearly inconsistent and indicates the supernova started with significantly lower expansion velocities than was observed from day 2 on. A natural explanation of this strange behaviour would be that a first weak shock caused the photosphere to expand slowly. Several hours later a second, strong shock lead to the observed high velocities. Such a two step process would also explain why Wampler et al. (1987) find it easier to fit the early light curve of SN 1987A if the sequence of events leading to the

outburst started several hours earlier than is indicated by the second neutrino pulse, and why hydrodynamic models based on one single strong shock (Arnett, 1987; Woosley et al., 1987; Shigeyama et al., 1987) do not fit the visual luminosities observed during the first two days of the outburst quite well. Unfortunately, spectra have not been taken during the first night and, therefore, it is not possible to determine in an unambigous way what really has happened.

3. Conclusions

SN 1987A is the first direct proof that rather massive stars, $M \simeq 20\ M_{\odot}$, will undergo core-collapse and can eject their envelopes in type II supernova explosions. Such a behaviour was predicted by certain theoretical models, but up to now was still subject to considerable controversies. However, the rather high explosion energy as well as certain properties of the observed neutrino bursts seem to indicate that the explosion mechanism itself is not yet fully understood. Stars in the mass range comparable to SK −69 202 have always been thought to be the main sources of heavy element production. It is a lucky coincidence that the first nearby supernova since SN 1604 belonged to that class. Detailed synthetic spectra will show whether or not theoretical predictions made for heavy element abundances are correct. Finally we may have been witnesses of the birth of a neutron star or a black hole. Interpretations of the observed neutrinos allow for both possibilities, and future observations will reveal which of them is correct.

References

Aglietta et al. 1987, Europhys. Lett. 3, 1315
Alexeyeva, L. 1987, ESO Workshop on SN 1987A, I.J: Danziger, ed., in press
Arnett, W.D. 1977, Ap.J. Suppl. 35, 145
Arnett, W.D. 1987, Ap.J. 319, 136
Arnett, W.D., and Rosner, J. 1987, Phys. Rev. Lett 58, 1906
Bahcall, J.N., and Glashow, S.L. 1987, Nature 326, 476
Baron, E. Cooperstein, J., and Kahana, S. 1985, Nucl. Phys. A440, 744
Barrett, P. 1987, IAU Circular No 4337
Bethe, H.A., and Wilson, J.R. 1985, Ap.J. 295, 14
Bionta, R.M. et al. 1987, Phys. Rev. Lett. 58, 1494
Bruenn, S.W. 1985, Ap.J. Suppl. 58, 771
Brunish, W.M., and Truran, J.W. 1982, Ap.J. 256, 247
Burrows, A., and Lattimer, J.M. 1986, Ap.J. 307, 178
Burrows, A., and Lattimer, J.M. 1987, Ap.J. 318, L 63
Casatella, A. 1987, ESO Workshop on SN 1987A, I.J. Danziger, ed., in press
Chiosi, C., and Maeder, A. 1986, Ann. Rev. Astron. Asrophys. 24, 329
Cristiani, S., and Gouiffes, C. 1987, IAU Circular No. 4350
Dachs, J., and Hanuschik, R. 1987, IAU Circular No. 4367
De Rujula, A. 1987, Phys. Lett. B193, 514
Galeotti, P. 1987, ESO Workshop on SN 1987A, I.J. Danziger, ed. in press
Hashimoto, M. and Nomoto, K. 1987, Univ. Tokyo preprint
Haubold, H.J., Kaempfer, B., Senatorov, A.V. and Voskresenski, D.N. 1987, preprint, Astron. Astrophys., in press
Hillebrandt, W. 1982, in "Supernova: A Survey of Current Research", M.J. Rees and R.J. Stoneham, eds., NATO-ASI c90, Rudel, p. 123
Hillebrandt, W. 1984, Ann. N.Y. Acad. Sci. 422, 197
Hillebrandt, W. 1986, in "High Energy Phenomena Around Collapsed Stars", F. Pacini, ed., NATO-ASI C195, Rudel, p. 73

Hillebrandt, W. 1987, in "Nuclear Astrophysics", W. Hillebrandt et al., eds., Lecture Notes in Physics 287, Springer, p. 335
Hillebrandt, W., Höflich, P., Truran, J.W., and Weiss, A. 1987 a, Nature 327, 597
Hillebrandt, W., Höflich, P., Kafka, P., Müller, E., Schmidt, H.U., and Truran, J.W. 1987 b, Astron. Astrophys. 177, L 41
Hillebrandt, W., Höflich, P., Kafka, P., Müller, E., Schmidt, H.U., and Truran, J.W. 1987 c, Astron. Astrophys. 180, L 20
Hirata, K., et al. 1987, Phys. Rev. Lett. 58, 1490
Lattimer, J.M., and Burrows, A. 1984, in "Problems of Collapse and Numerical Relativity", D. Bancel and M. Signore, eds., NATO-ASI C134, Rudel, p. 147
Maeder, A. 1987, ESO Workshop on SN 1987A, I.J. Danziger, ed., in press
Mayle, R.W. 1985, Ph.D. Thesis, Lawrence Livermore National Laboratory
Mayle, R.W., Wilson, J.R., and Schramm, D.N. 1987, Ap.J. 318, 288
McNaught, R.H. 1987, IAU-Circular No. 4389
Müller, E., and Hillebrandt, W. 1981, Astron. Astrophys. 103, 288
Müller, E., Rozyczka, M., and Hillebrandt, W. 1980, Astron. Astrophys. 81, 288
Panagia, N., Gilmozzi, R., Clavel, J., Barylak, M., Gonzales Giesta, R., Lloyd, C., Sanz Fernandez de Cordoba L., and Wamsteker, W. 1987, Astron. Astrophys. 177, L 25
Ramadurai, S. and Wilta, P.J. 1987, preprint
Sato, K., and Suzuki, H. 1987, Phys. Rev. Lett. 58, 2722
Schmidt-Kaler, T. Hanuschik, R., and Dachs, J. 1987, ESO Workshop on SN 1987A, I.J. Danziger, ed., in press
Shelton I. 1987, IAU-Circular No. 4317
Shigeyama, T., Nomoto, K., Hashimoto, M. and Sugimoto, D. 1987, preprint
Symbalisty, E. 1984, Ap. J. 285, 729
Truran, J.W., and Weiss, A. 1987, in "Nuclear Astrophysics", W. Hillebrandt et al., eds., Lecture Notes in Physics 287, Springer, p. 293
Wampler, E.J., Truran, J.W., Lucy, L.B., Höflich, P., and Hillebrandt, W. 1987, Astron. Astrophys. 182, L 51
Weaver, T.A., Woosley, S.E., and Fuller, G.M. 1985, in "Numerical Astrophysics", J. Centrella et al., eds., Jones and Bartlett, p. 374
West, R.M., Laubeerts, A., Jorgensen, H.E., and Schuster, H.-E. 1987, Astron. Astrophys. 177, L 1
Wilson, J.R. 1985, in "Numerical Astrophysics", J. Centrella et al., Jones and Bartlett, p. 422
Wilson, J.R., Mayle, R.W., Woosley, S.E., and Weaver, T.A. 1986, Ann. N.Y. Acad. Sci. 470, 267
Woosley, S.E., Pinto, P.A., and Ensman, L. 1987, Ap. J., in press
Woosley, S.E., Pinto, P.A., Martin, P. and Weaver, T.A. 1987, Ap. J. 318, 664

Theoretical Model of SN 1987A and Presupernova Evolution of Massive Stars

K. Nomoto[1], *M. Hashimoto*[2], *and T. Shigeyama*[1]

[1]Department of Earth Science and Astronomy,
University of Tokyo, Meguro-ku, Tokyo 153, Japan
[2]Max-Planck-Institut für Physik und Astrophysik, Institut für Astrophysik,
Karl-Schwarzschild-Str. 1, D-8046 Garching, Fed. Rep. of Germany

Hydrodynamical models for SN 1987A are calculated starting from the new presupernova model of the 6 $M_\odot$ helium core with the envelope of mass M_{env}. The photospheric velocity v_{ph} and the light curve in the theoretical models depend on the explosion energy E, M_{env}, and the mass of ^{56}Ni. Models and observations are in reasonable agreement for $M_{env} > 3\ M_\odot$, $E \sim 1.0 - 1.5 \times 10^{51}$ erg, and $M_{Ni} \sim 0.07\ M_\odot$. The best fit among the calculated models is obtained for $M_{env} = 6.7\ M_\odot$ and $E = 1.0 \times 10^{51}$ erg. Mixing of ^{56}Ni in the core and moderate helium enrichment in the envelope are suggested from the light curve and v_{ph}.

1. Introduction

The supernova 1987A in the Large Magellanic Cloud is providing us an excellent opportunity to test the theory of massive star evolution, nucleosynthesis, and supernova explosion. We will report the results of 1) our new set of calculations of the pre-collapse evolution of massive stars based on improved input physics, and 2) hydrodynamical calculation of the supernova explosion and the light curve. Since the light curve of SN 1987A is quite unique, the comparison between the observation and theoretical models makes it possible to determine some previously unknown quantities, i.e., explosion energy E, mass and distribution of ^{56}Ni, envelope mass M_{env}, and the progenitor's radius. Such information on the chemical and dynamical structure of the supernovae would lead us to deeper understanding of the mass loss processes before explosion and the mechanism that transforms collapse into explosion mechanism. These are the major uncertainties in the current theory of massive star evolution.

In §2, our new calculation of the presupernova evolution of massive stars and improved input physics (equation of state, electron capture, and silicon burning network) are described. In §3 - 5, the hydrodynamical models of the explosion of a massive star with a helium core of $M = 6\ M_\odot$ are presented.

2. Presupernova Models

Stars more massive than 11 $M_\odot$ evolve to form an iron core and collapse due to photodisintegration of iron nuclei and electron capture. The outcome of collapse and nucleosynthesis depends sensitively on the iron core mass and a density

structure in the heavy element mantle. However, there exist considerable uncertainties in determining the presupernova models because they depend on the carbon abundance after helium burning which in turn depends on the ^{12}C (α, γ) ^{16}O rate [8] and overshooting at the edge of a convective helium burning core [4]. A larger reaction rate and overshooting would yield a smaller carbon abundance but these factors, especially overshooting, involve large uncertainties.

Since no reliable theory of convection, especially non-local theory, exists at the moment, it is necessary to explore the various possible evolutionary paths within a uncertain range.

Moreover, the mass of the evolved core is close to the Chandrasekhar mass so that the core structure is sensitive to equation of state of strongly degenerate matter and electron mole number Y_e. Therefore inclusion of improved physical input is expected to bring a large effect on the presupernova structure.

We have recently carried out the calculation of massive star evolution. The important improvements compared with the model by Woosley and Weaver [40] (hereafter WW model) are: 1) Coulomb interaction is included in the equation of state, 2) electron capture is included from the beginning of oxygen burning, and 3) the nuclear reaction rate and electron capture rate for silicon burning are obtained from the extensive network calculation for 250 species.

The improvement 1) leads to the decrease in the Chandrasekhar mass and, hence, to the decrease in the core mass at each stage. Due to the improvement 2), even before the oxygen exhaustion, the decrease in electron mole number, Y_e, is significant. Concerning the improvement 3), the behavior of silicon shell burning is considerably different from the WW models, which is crucial to determine the final presupernova structure.

We adopt the Schwarzshild criterion for convective stability and neglect the overshooting in order to explore the cases of possibly largest carbon abundances after helium burning.

The present set of evolutionary calculations has been performed for helium stars of mass M_α = 2.8, 3.0, 3.3, 4.0, 6.0, and 8.0 $M_\odot$. These helium star masses correspond to the main-sequence mass of M_{ms} = 11, 12, 13, 15, 20, and 25 $M_\odot$, respectively. The calculations started from the gravitational contraction of the helium star. In the following, the evolution for M_α = 6.0 $M_\odot$ is described because it is relevant for the progenitor of SN 1987A.

2.1. Helium, Carbon, Neon, and Oxygen Burning

The evolution of these massive stars is rather straightforward. They undergo alternating stages of gravitational contraction and central nuclear burning of hydrogen, helium, carbon, neon, oxygen, and silicon as seen in Figs. 1 and 2 for M_α = 8.0 and 3.3 $M_\odot$, respectively.

The helium burning stage lasts 6.0 x 10^5 yr and a C-O core of 3.80 $M_\odot$ forms. The mass fractions of carbon and oxygen are $X(^{12}C)$ = 0.22 and $X(^{16}O)$ = 0.76,

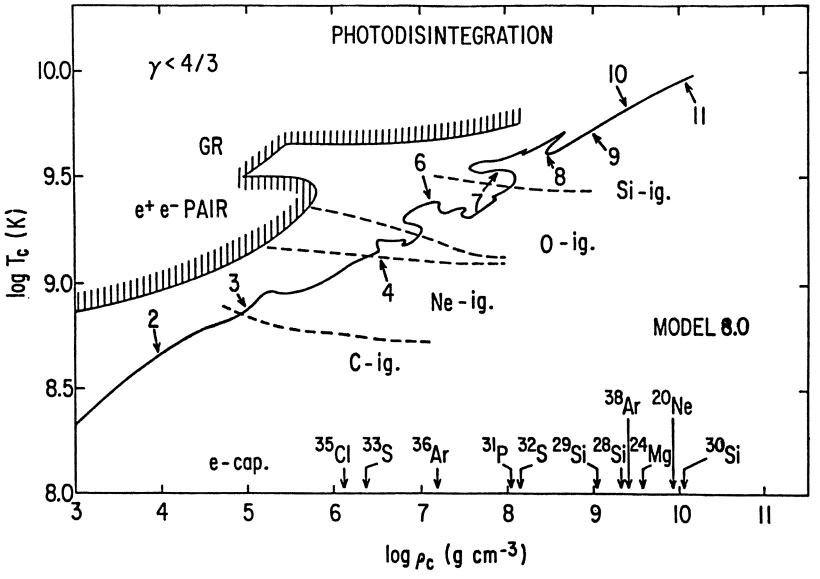


Figure 1 Evolutionary changes in the central density and temperature through the onset of iron core collapse for the 8 $M_\odot$ helium star. Shown are ignition lines for nuclear burning and the threshold densities of electron capture for oxygen and early silicon burning stages.

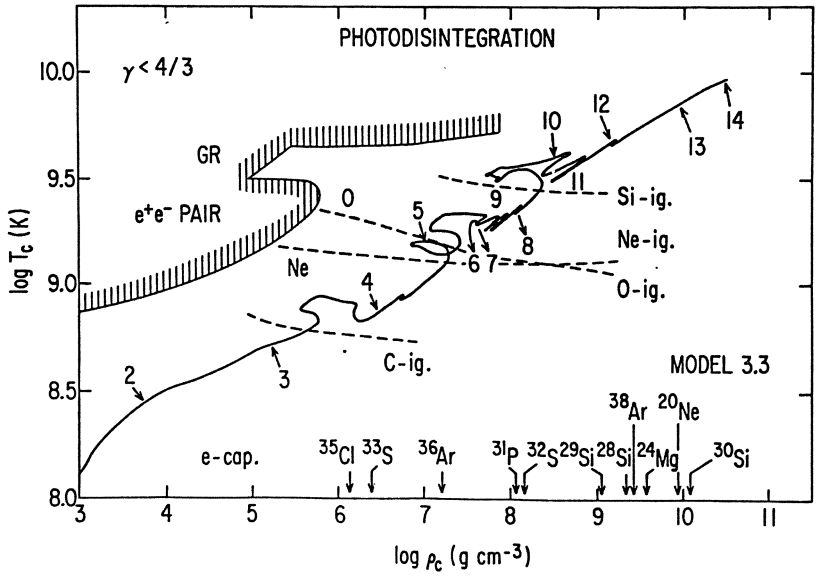


Figure 2 Same as Figure 1 but for the 3.3 $M_\odot$ helium star.

respectively. This amount of carbon is large enough for active (i.e., convective) carbon burning to take place in the center and, later, in the outer shell.

Carbon burning starts in the center (1.4×10^4 yr after exhaustion of helium) and lasts only 450 yr. When carbon is almost depleted in the core, the resultant composition is $X(^{16}O) = 0.66$ and $X(^{20}Ne) = 0.26$. The carbon burning layer shifts to the outer shell and a convective shell develops at $M_r = 1.67\ M_\odot$.

Neon burning is ignited in the center (40 yr after exhaustion of carbon) and lasts only 0.4 yr. After exhaustion of neon in the center, an oxygen-rich core of $X(^{16}O) = 0.79$ and $X(^{24}Mg) = 0.10$ forms.

Oxygen burning synthesizes silicon- and sulfur-rich elements and the product becomes somewhat neutron-rich. The main weak processes involved are positron decays and electron captures. When $X(^{16}O) = 0.01$, Y_e drops to 0.493 (neutron excess $\eta = 1 - 2Y_e = 0.014$) in the convective core where $X(^{28}Si) = 0.54$, $X(^{32}S) = 0.17$, $X(^{34}S) = 0.12$, and $X(^{38}Ar) = 0.13$.

2.2. Silicon Burning and Iron Core Collapse

During the contraction of the Si-S core, electron capture on the products of oxygen burning decreases Y_e down to 0.46 when silicon is ignited. After exhaustion of silicon, a convective silicon-burning shell appears successively. Silicon shell burning is so intense that a convective zone extends even into a part of the oxygen-rich layer. The outer edge of the convective zone is limited by entropy barrier at the oxygen burning shell at $M_r = 1.6\ M_\odot$, where steep density gradient exists. The location of the silicon burning shell, or the iron core mass, depends on the time scales of convective mixing and nuclear reaction. If the mixing is much faster than the reaction, silicon is depleted in the convective shell and the resulting iron core mass is $\sim 1.6\ M_\odot$. On the other hand, if the time scale of mixing is longer than the nuclear burning time scale, the burning shell moves outward as silicon is deplete there. Then the iron core mass ranges $1.4 - 1.6\ M_\odot$. Unfortunately, the mixing time scale is highly uncertain because of the poor convection theory. The distributions of abundance and density for the case of $M_{Fe} = 1.4\ M_\odot$ are shown in Figures 3 and 4, respectively.

At the presupernova stage of $\rho_c = 2 \times 10^{10}$ g cm^{-3}, the masses contained interior to the shell of maximum energy generation for helium (M_{Heb}), carbon (M_{Cb}), neon (M_{Neb}), oxygen (M_{Ob}), and silicon (M_{Sib}) are $M_{Heb} = 3.8\ M_\odot$, $M_{Cb} = 1.8\ M_\odot$, $M_{Neb} = M_{Ob} = 1.6\ M_\odot$, $M_{Sib} = 1.4\ M_\odot$, respectively. Quantitatively similar structure is obtained for $M = 8.0\ M_\odot$.

The gradient of entropy and density is relatively small at the iron core edge. The largest jump in these quantities is seen at the oxygen burning shell. Accordingly, the mass cut that divides the neutron star and the ejecta could be $\sim 1.6\ M_\odot$ for $20 - 25\ M_\odot$ stars.

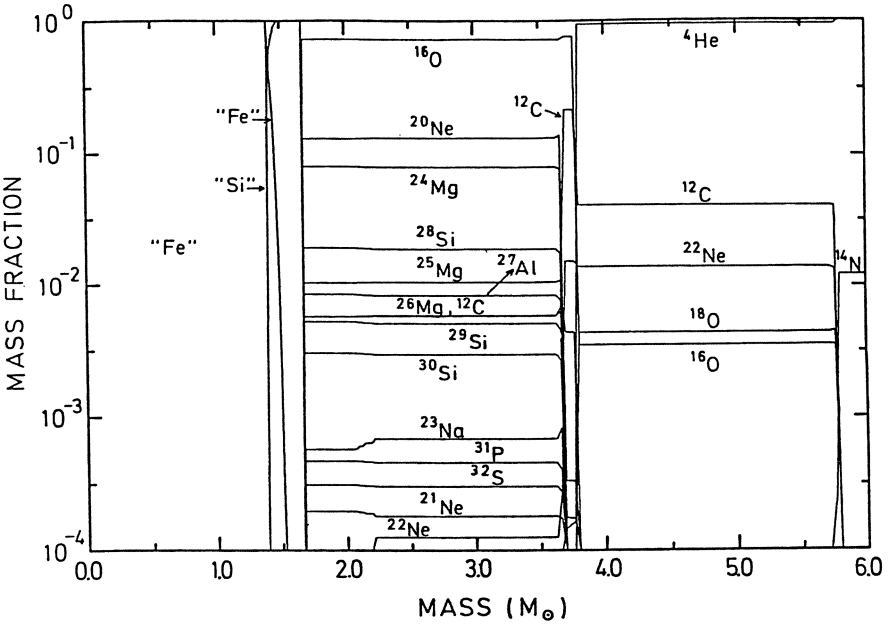


Figure 3: Composition structure of the 6 $M_\odot$ helium star at the iron core collapse.

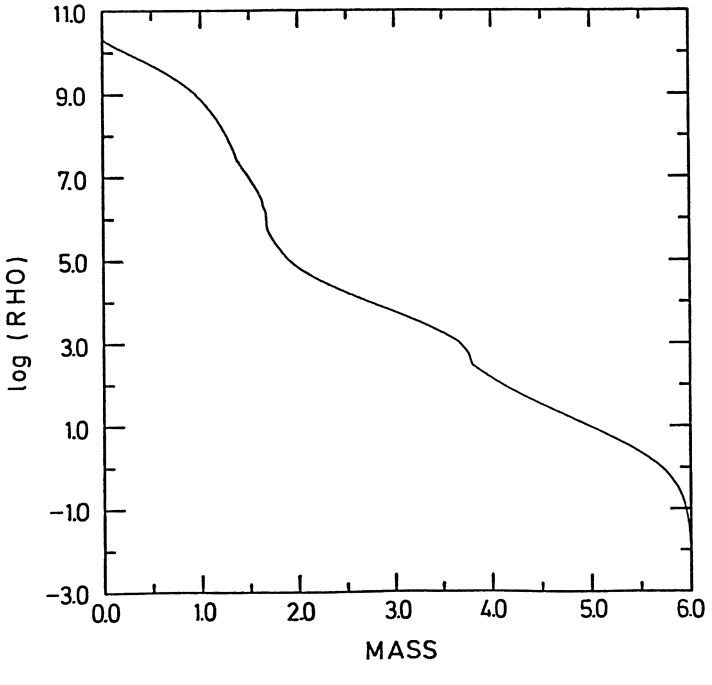


Figure 4: Density distribution of the 6 $M_\odot$ helium star at the iron core collapse

2.3. Stellar Structure: Comparison with the Previous Models

The density distribution of the 8 $M_\odot$ helium star [11] which has 18 % carbon is different from that of the 8 $M_\odot$ helium core of the 25 $M_\odot$ star which has 9 % carbon after helium burning [40] as seen in Fig. 5. (Chemical structure is seen in Fig. 7.) It is noticeable that the density of the oxygen-rich layer in the former model is significantly lower than in the latter because of the steep density gradient at the carbon burning shell.

In the WW25 model, no convective core appears for carbon and neon burning phases, while in our model both carbon and neon burning dominate the neutrino energy losses and thus, a small convective core and then an extensive convective shell form. In particular, convective carbon burning shell extending from M_r = 1.8 $M_\odot$ to 5.4 $M_\odot$ is quite active during neon and oxygen burning. These difference are easily seen in the density distribution.

2.4. Evolution of Stars around 13 $M_\odot$

For the helium star of M_α = 3.3 $M_\odot$ (which correspond to $M_{ms} \sim 13\ M_\odot$), the iron core mass is as small as 1.2 $M_\odot$. The density distribution of the iron core is significantly steeper than that of the WW's 15 $M_\odot$ model as seen in Figure 6. (Chemical structure is seen in Fig. 8.) Electron capture during oxygen burning reduces Y_e down to 0.48 at the stage of X_O = 0.01 where electron capture begins to be included in the WW models. Inclusion of Coulomb interaction in equation of state reduces the pressure by a few percent which is significant because the core mass is close to the Chandrasekhar mass. For such a small iron core, a propmt explosion with the energy of 2.2 x 10^{51} erg results [2].

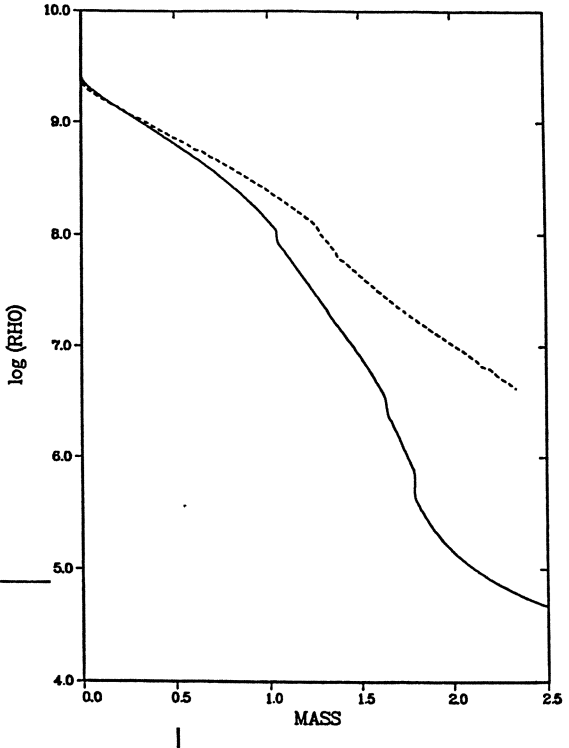
The formation of such a small iron core has been confirmed recently by Woosley [37] who included Coulomb interaction term and obtained the Fe core mass of 1.28 $M_\odot$ for the 15 $M_\odot$ star.

3. Progenitor of SN 1987A and Presupernova Mass Loss

The progenitor of SN 1987A is very likely Sk-69 202 [18, 9]. Its luminosity is about 1 x 10^5 $L_\odot$ [38, 12]. The presupernova luminosities (or at the very advanced stages) of our helium star models of mass M_α are: 1.9 x 10^5, 1.3 x 10^5, 6.1 x 10^4, and 4.0 x 10^4 $L_\odot$ for M_α = 8, 6, 4, and 3.3 $M_\odot$, respectively [25-27, 11]. From this, the progenitor mass is inferred to be $M_\alpha \sim 6\ M_\odot$ (main-sequence mass of $M_{ms} \sim 20\ M_\odot$). The 6 $M_\odot$ helium star was evolved up to the iron core collapse as described in §2 [11]. (See Figs. 3 and 4 for the composition structure and the density distribution at the collapse.)

We constructed the initial model of explosion from the above 6 $M_\odot$ helium star by replacing the inner 1.4 $M_\odot$ with a point mass neutron star and attaching a hydrostatic hydrogen-rich envelope with mass of M_{env}. The composition structure of the ejecta is: the heavy element layer of 2.4 $M_\odot$, helium-rich layer of 2.2 $M_\odot$,

Figure 5 (right).
Density distribution of the 8 $M_\odot$ helium star near the onset of iron core collapse (solid: Hashimoto and Nomoto 1987) compared with that of the 8 $M_\odot$ helium core of the 25 $M_\odot$ star (dashed: Woosley and Weaver 1986).



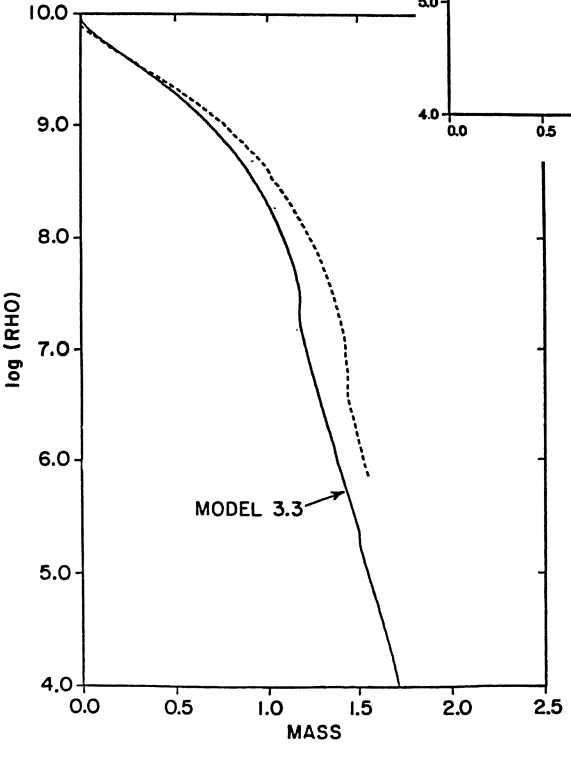


Figure 6 (left). Density distribution of the 3.3 $M_\odot$ helium star at the onset of iron core collapse (solid: Nomoto and Hashimoto 1987) compared with that of the 4 $M_\odot$ helium core of the 15 $M_\odot$ star (dashed: Woosley and Weaver 1986).

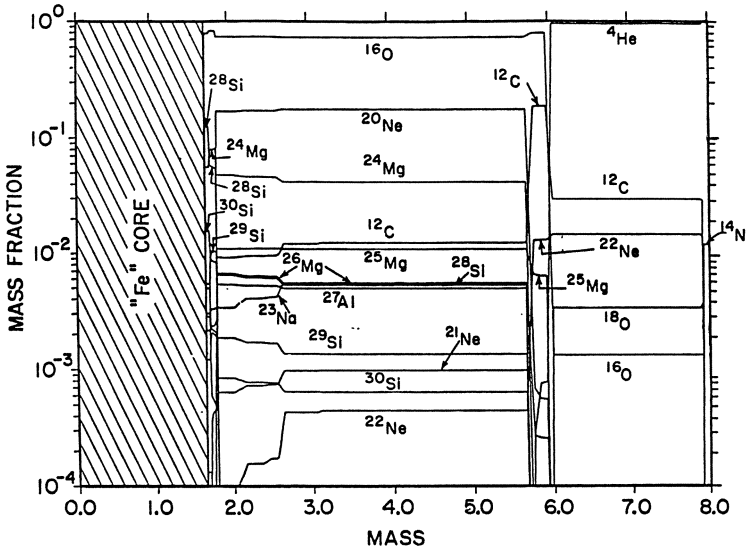


Figure 7 Composition of the 8 $M_\odot$ helium star at the beginning of an iron core collapse. A thin silicon-rich shell is sandwiched between the iron core and the oxygen-rich layer (Hashimoto and Nomoto 1987).

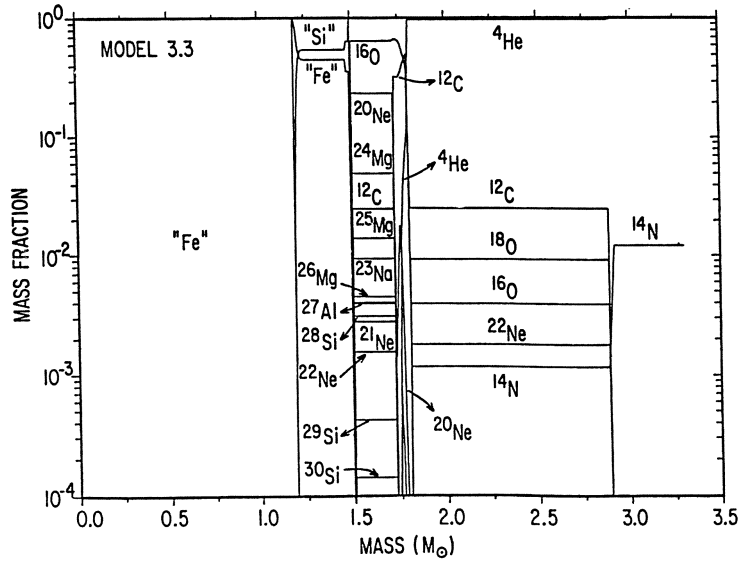


Figure 8 Same as Figure 7 but for the 3.3 $M_\odot$ helium star (Nomoto and Hashimoto 1987).

and the hydrogen-rich envelope of M_{env}. Thus the total mass of the ejecta is $M = 4.6\ M_\odot + M_{env}$.

The mass, M_{env}, and the composition of the hydrogen-rich envelope depends on the mass loss and convective mixing during the presupernova evolution. The evolutionary route to reach the blue supergiant at the explosion is not yet clear. Recent UV observations have shown that the UV emission lines originated from circumstellar material which is nitrogen-rich [19]. This suggests that the progenitor had evolved once to a red supergiant stage, lost a significant fraction of its hydrogen-rich envelope, and then contracted to the size of the blue supergiant. This scenario is consistent with the fact that in the Large Magellanic Cloud there exist many bright red-supergiants corresponding to stars up to $M_{ms} \sim 50\ M_\odot$ [15].

The surface composition changes during the mass loss as seen from the example of the star of $M_{ms} = 25\ M_\odot$ (Fig. 9 taken from [20]). When the remaining stellar mass is 15 - 10 $M_\odot$, the surface abundances are different from the original ones. Helium and nitrogen are overabundant at almost the same stages and the abundance ratios are as large as He/H = 1 - 1.2 and N/C $\sim$40 in mass fraction. During these stages, the star is a F, A, or B supergiant. In the above models, the abundance change occurs during core helium burning and the final state is either a red supergiant or a Wolf-Rayet star [20]. However, the mass loss rate in stars of the Large Magellanic Cloud must be smaller than in the Galaxy because of lower metallicity [21]. Hence starting from $M_{ms} = 20$ - 25 $M_\odot$, the presupernova star could still have its mass of 15 - 10 $M_\odot$ with the large surface He/H and N/C ratios.

Generally, M_{env} and helium abundance of the star of $M_{ms} = 20\ M_\odot$ could range from 14 $M_\odot$ to $\sim$ 0.1 $M_\odot$ and from 0.25 to $\sim$ 0.9, respectively. However, the large N/C ratio in the circumstellar matter suggests that the progenitor had lost more than 40 % of its mass and its envelope is somewhat helium-rich at the explosion. Interestingly the early light curve and the photospheric velocity somewhat depend on the surface abundance and thus provides information on the abundance.

In the following hydrodynamical models, we describe Models 11E1Y4, 11E1Y6, 11E1.5, 11E2, 7E1Y4, where 11 and 7 denote the ejecta mass of $M = 11.3\ M_\odot$ ($M_{env} = 6.7\ M_\odot$) and $7.0\ M_\odot$ ($M_{env} = 2.4\ M_\odot$), respectively; E1, E1.5, and E2 denote explosion energy of $E = 1$, 1.5, and 2×10^{51} erg, respectively; Y4 denotes the surface composition of X = 0.59, Y = 0.40, and Z = 0.01 while Y6 for Y = 0.6. For all models the initial radius of the progenitor is assumed to be $R_0 = 3 \times 10^{12}$ cm.

4. Hydrodynamical Models for SN 1987A

In the hydrodynamical calculation of supernova explosion, the energy E was deposited instantaneously in the central region of the core to generate a strong shock wave. The subsequent propagation of the shock wave, expansion of the star, and the optical light curve were calculated using a hydrodynamic code [41] with a modification to include flux limited diffusion [29].

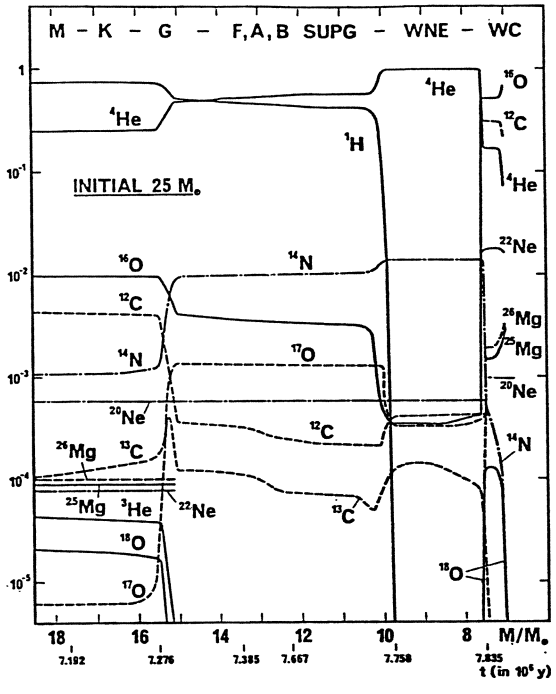


Figure 9: Change in the surface composition for the mass losing star of M_{ms} = 25 $M_{\odot}$ (Maeder 1987a).

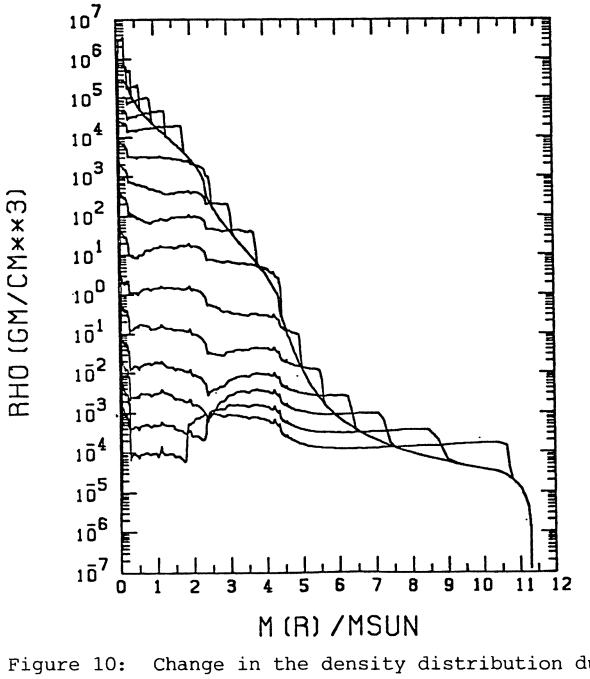


Figure 10: Change in the density distribution during the propagation of the shock wave.

For the heavy element layer recombination of oxygen is taken into account for opacity and energetics. For the helium layer, the helium line opacity of 0.009 $cm^2\ g^{-1}$ sets the lower bound value of the opacity [14]. These line opacities are highly uncertain because of the variation of Doppler shift for a large velocity gradient. The radioactive decay and the deposition of γ-rays are treated by applying the approximate deposition function [33].

The shock wave propagates through the low-density hydrogen-rich envelope producing a reverse shock as seen from the change in the density distribution in Fig.10. This leads to a formation of density inversion as seen in Fig.11 for Models 11E1 and 11E2. Because of Rayleigh-Taylor instability, the core material will be mixed during early stages [40]. Since the time scale and extent of mixing is uncertain, we made several models 11E1Y4 with mixing up to M_r = 0.07 (no mixing), 2.4, 3.4, and 8.0 $M_\odot$. For the mixed layer, we assume uniform composition.

The shock wave arrives at the original surface of the star at t_{prop}, which is approximated for different values of the initial radius R_0 and the ejected mass M, and for different energies of explosion E as [31]

$$t_{prop} \sim 2 \text{ hr } (R_0/3 \times 10^{12} \text{ cm}) [(M/10\ M_\odot)/(E/1 \times 10^{51} \text{ erg})]^{1/2}. \quad (1)$$

The condition of $t_{prop} < 3$ hr (Kamiokande time will be used hereafter) rules out the models with too large R_0 and too low E/M [31]. The progenitor radius larger than 4.5×10^{12} cm is ruled out for Model 11E1 simply from this condition.

After the shock wave reaches the surface, the star starts to expand and soon the expansion becomes homologous as $v \propto r$. In Figure 12, the velocity distribution for the homologous expansion is shown. The velocity gradient against the enclosed mass, M_r, is very steep near the surface, while it is almost flat in the helium layer (1500 - 1700 km s^{-1} for 11E1) and the heavy element core (1200 - 1500 km s^{-1} for 11E1). This is because the core material is decelerated and forms a dense shell due to the reverse shock when the expanding core hits the hydrogen-rich envelope and The expansion velocities of the helium and heavy element layers are so low that the kinetic energy of these layers is only 10 % of the total kinetic energy for 11E1.

As the star expands, the photosphere moves inward in M_r. Because of the steep velocity gradient near the surface, the velocity of the material at the photosphere decreases as seen in Figure 13 for Models 11E1Y4 and 11E1Y6. This may be compared with $v_{ph} = r_{ph}/t$ where r_{ph} is the radius of the black body surface obtained from the photometric observations [24, 6, 10]. The photospheric velocity of Model 11E1Y4 is in good agreement with SAAO data but slightly larger than CTIO values. For E larger than 1.5×10^{51} erg, v_{ph} may be too large for $M \sim 10\ M_\odot$ while E could be larger for larger M. Although the determination of r_{ph} is subject to uncertainties as seen from the difference between SAAO and CTIO data, this comparison gives a useful constraint on E/M of the model.

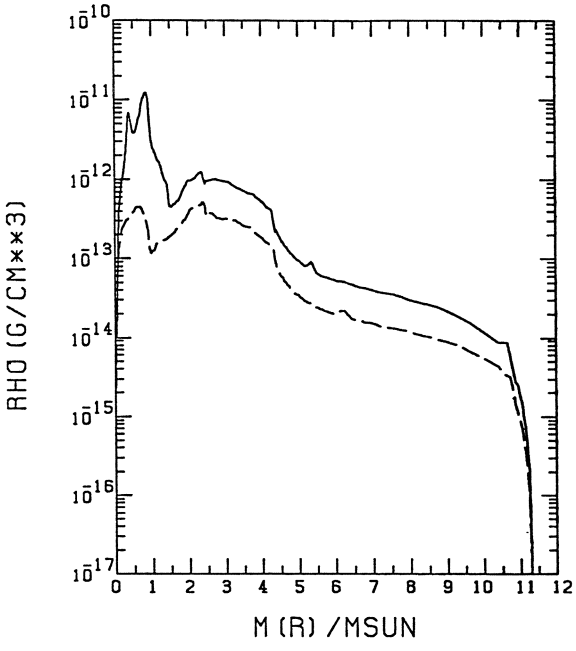


Figure 11: Density distribution of the exploding star at t = 116 d. Shown are Model 11E1 (solid): M = 11.3 M_o, M_{env} = 6.7 M_o, E = 1 x 10^{51} erg and Model 11E2 (dashed): E = 2 x 10^{51} erg.

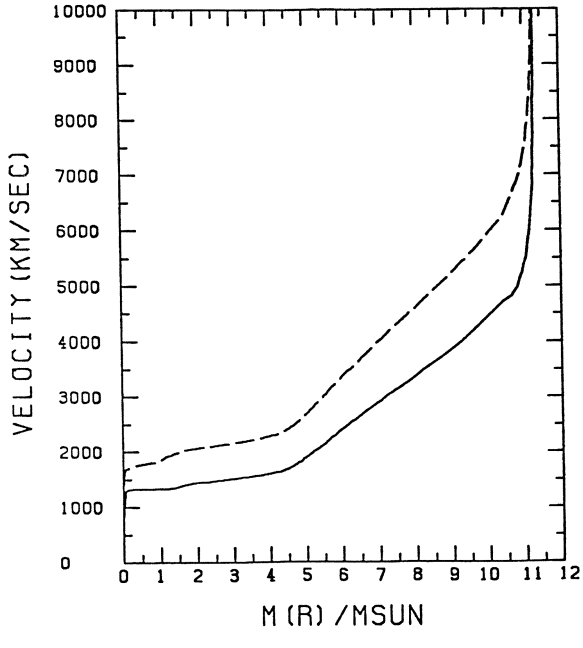


Figure 12: Velocity distribution for Models 11E1 (solid) and 11E2 (dashed) at t = 116 d.

It is interesting to note that v_{ph} depends on Y in the envelope because the helium-line opacity makes r_{ph} larger (Fig. 13). This might indicate that envelope is not extremely helium-rich, although uncertainties are involved in the opacity and r_{ph}. More careful analysis could provide constraints on the composition of the envelope.

In Model 11E1Y4, the photosphere entered the helium layer and then the heavy element layer around t = 90 d (70 d) and 110 d (80 d), respectively, for SAAO (CTIO) data as seen from v_{ph} and the velocity profile in Fig. 12.

5. Light Curve Models

5.1. Shock Heating

The light curve of SN 1987A is quite unique. 1) It took only 3 hours for the visual magnitude to reach 6.4 magnitude after the neutrino burst. 2) At the subsequent plateau, the optical light was much dimmer than typical Type II supernovae. 3) Afterwards optical luminosity increased until about 84 d forming a broad peak. 4) The supernova then declined first relatively rapidly and then slowly at 0.01 mag/day. The slow decline portion of the light curve is clearly the radioactive tail of the ^{56}Co decay.

These light curve features give some constraints on the hydrodynamical model of explosion. First of all, the steep rise in luminosity in 3 hr requires relatively large E/M and small R_0 [31]. Certainly the condition of $t_{prop} < 3$ hr (Eq. 1) should be satisfied. Moreover, when the shock wave arrives at the photosphere, the temperature behind the shock is as high as 10^6 K so that the most of the radiation is emitted in EUV through soft X-ray band. In order for the optical flare-up of the supernova to be seen at t = 3 hr, the ejected gas and the radiation field should have expanded rapidly so that the temperature becomes lower and the radius of the photosphere becomes larger. Therefore, the expansion velocity and thus E/M should be larger than a certain value.

Figures 14 - 16 show the changes in the calculated V magnitude, the bolometric luminosity, and the effective temperature for 11E1Y4 and 11E1Y6. For observed points, the distance modulus of 18.5 and the extinction of 0.6 are adopted [34]. In order for the visual magnitude to reach 6.4 mag in time, $E = 1 \times 10^{51}$ erg is required for $M = 11.3\ M_\odot$. For the envelope with smaller Y, luminosity is lower and hence larger E is required because of larger scattering opacity. Moreover, the model with Y = 0.6 is in better agreement with the overall shape of the observed light curve than Y = 0.4. Such a dependence on Y deserves further careful study because it would provide information on presupernova mass loss.

Although the theoretical models are generally in good agreement with observations at early phase, there remain uncertainties in theoretical model. The supernova atmosphere is scattering dominant so that the color temperature may be significantly higher than the effective temperature [31, 13]. The bolometric

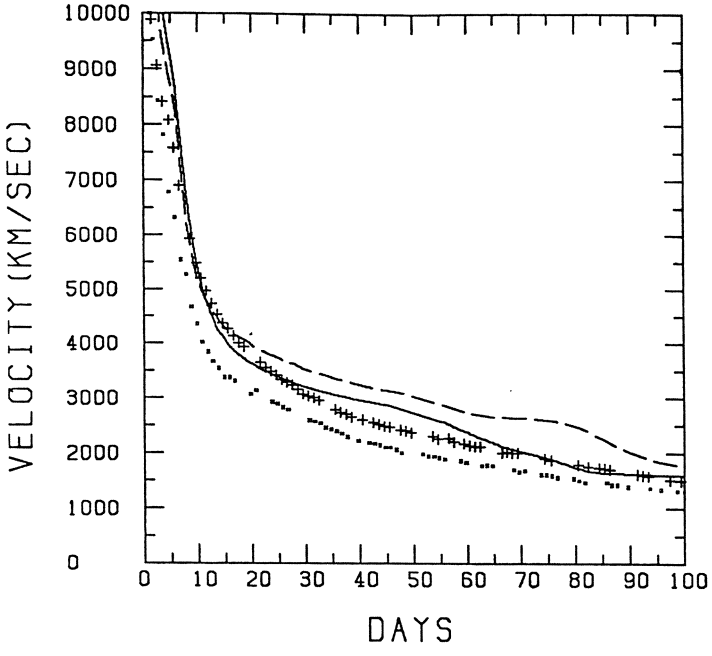


Figure 13: Changes in the expansion velocity of the material at the photosphere for Models 11E1Y4 (solid: E = 1 x 10^{51} erg, M = 11.3 $M_{\odot}$, R_0 = 3 x 10^{12} cm, helium abundance in the envelope Y = 0.4) and 11E1Y6 (dashed: Y = 0.6). Observed values are r_{ph}/t obtained at SAAO (+ mark; Catchpole et al. 1987) and CTIO (small x mark; Hamuy et al. 1987).

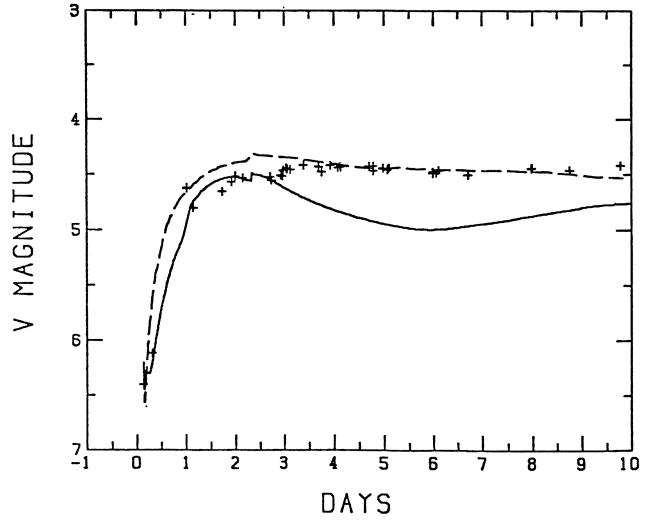


Figure 14: The calculated visual magnitude for Models 11E1Y4 (solid) and 11E1Y6 (dashed). Observed data are taken from ESO (Cristiani et al. 1987), CTIO (Blanco et al. 1987), and SAAO (Menzies et al. 1987) except for the early two points measured on films (McNaught 1987; Zoltowski 1987).

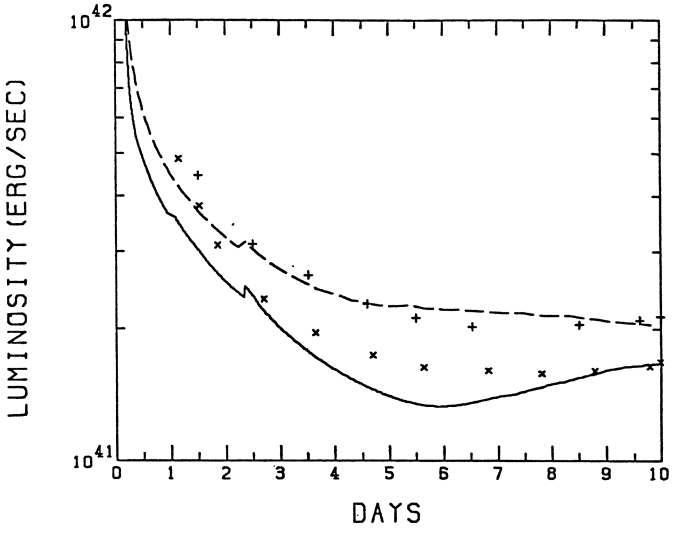


Figure 15: Same as Fig. 14 but for the bolometric luminosity. The solid and dashed lines are Models 11E1Y4 and 11E1Y6, respectively. Observations at SAAO (Menzies et al. 1987; Catchpole et al. 1987) and CTIO (Hamuy et al. 1987) are indicated by + and small x, respectively.

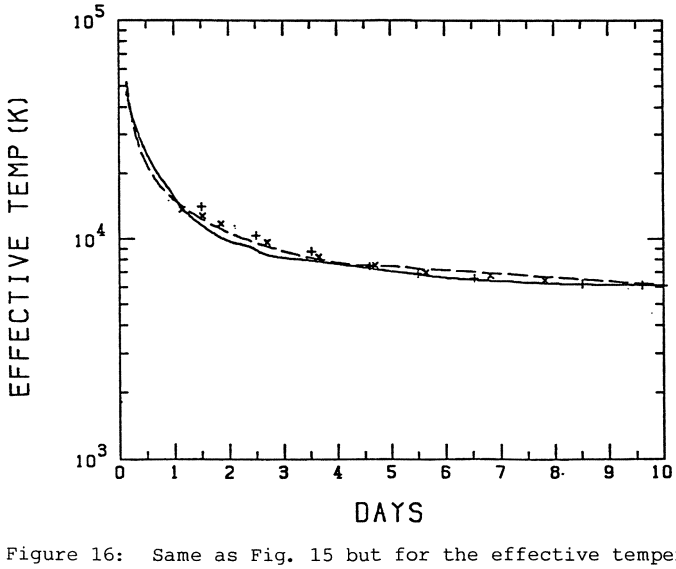


Figure 16: Same as Fig. 15 but for the effective temperature.

correction is sensitive to the color temperature because it is as high as 4 - 6 x 10^4 K. If we apply the bolometric correction based on the color temperature, theoretical visual luminosity is lower than in Fig. 14 so that larger E/M is required as derived in Shigeyama et al. [31]. To examine this problem, we need multi-frequency transfer calculations.

In short, the early light curve up to t $\sim$ 20 d can be accounted for by diffusive release of energy which was deposited in the envelope during the passage of the shock wave. (For Model 11E1, the initial radiation field with energy of 4.4×10^{50} erg is established.)

5.2. Light Curve Powered by Radioactive Decays

After these early stages, the observations show the increase in the bolometric luminosity. This indicates the existence of an energy source that continually heats up the expanding star. The light curve tail after 120 d clearly shows that this energy source is the decay of ^{56}Co.

The theoretical light curve with radioactive decays of ^{56}Ni and ^{56}Co was calculated assuming the production of 0.071 $M_\odot$ ^{56}Ni. Figures 17 - 20 compare the calculated bolometric light curve with observations [6, 10] for several cases of E, M_{env}, and Ni distribution. As the star expands the theoretical optical light at this phase changes in the following way:

The photosphere becomes deeper as the recombination front proceeds through the hydrogen-rich envelope deeper in mass. At the same time heat wave is propagating from the interior. As a result, the theoretical bolometric luminosity starts to increase at a certain stage and reaches a peak value of $L_{pk} \sim 1 \times 10^{42}$. After the peak, the theoretical luminosity decreases rather steeply compared with the broad peak of SN 1987A. Finally the light curve enters the tail.

It should be noted that the observed peak luminosity is larger than the energy generation rate due to Co-decay

$$L = 1.43 \times 10^{43} \, M_{Co}/M_\odot \, e^{-t/113.7d} \text{ erg s}^{-1}, \qquad (2)$$

which gives $L = 5.4 \times 10^{41}$ erg s^{-1} at t = 70 d for $M_{Co} = 0.071 \, M_\odot$. This implies that previously deposited energy from Co-decay is also radiated away during the broad peak of the light curve [39]. How this additional energy is radiated away is rather sensitive to the hydrodynamics, opacity, etc. The difficulty to reproduce the broad peak may be related to the fact that total photon energy emitted from the supernova is smaller than 10^{49} erg, less than 1 % of the kinetic energy of expansion [6, 10]; the optical flux may easily affected by slight change in the hydrodynamics. Also the flux-limited diffusion approximation used in the calculation may need to be improved. Contribution of oxygen recombination to the light curve [30] was found to be negligible in our calculation.

Despite this difficulty, both the peak luminosity and the tail can be consistently accounted for by the radioactive decay of ^{56}Co. The amount of ^{56}Ni initially produced is 0.071 $M_\odot$ and 0.066 $M_\odot$ if we fit the light curve tail to SAAO data and CTIO data, respectively. The γ-ray optical depth at t = 170 d is still as large as 12 for Model 11E1Y4 (Fig. 21). Therefore the light curve tail is almost identical to the Co-decay tail given by Eq. (2).

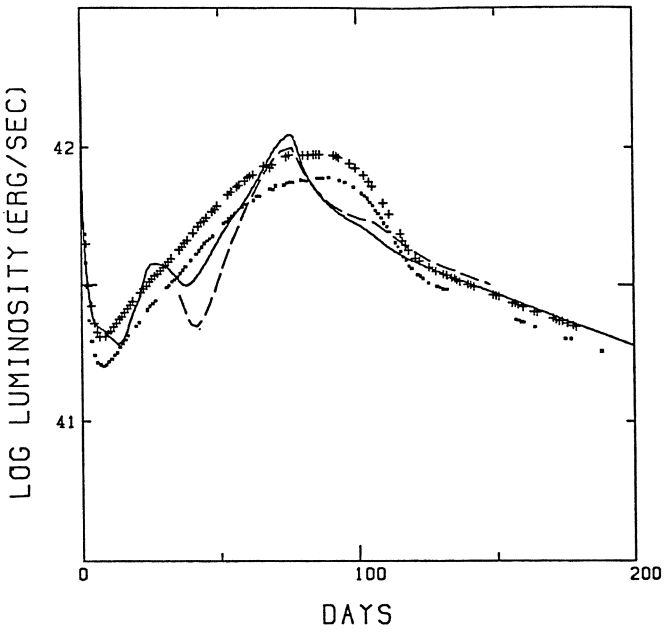


Figure 17: Bolometric light curve for several cases of Model 11E1Y4. Production of 0.071 M_o ^{56}Ni is assumed. For the dashed curve ^{56}Ni is assumed to be confined in the innermost layer while the solid curve assumes that ^{56}Ni is mixed uniformly in the heavy element core (M_r < 2.4 M_o). Observed points are taken from SAAO (+; Catchpole et al. 1987) and CTIO (x; Hamuy et al. 1987).

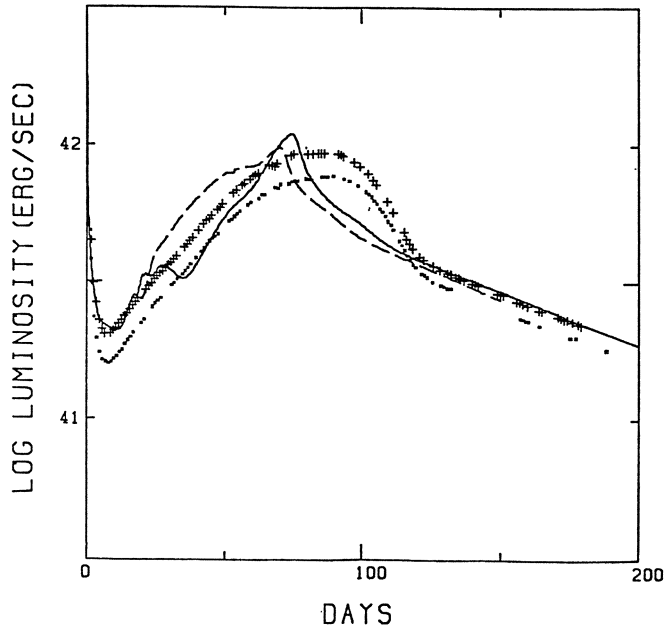


Figure 18: Same as Fig. 17 but for the cases where ^{56}Ni is mixed uniformly at M_r < 3.4 M_o (solid) and at M_r < 8 M_o (dashed).

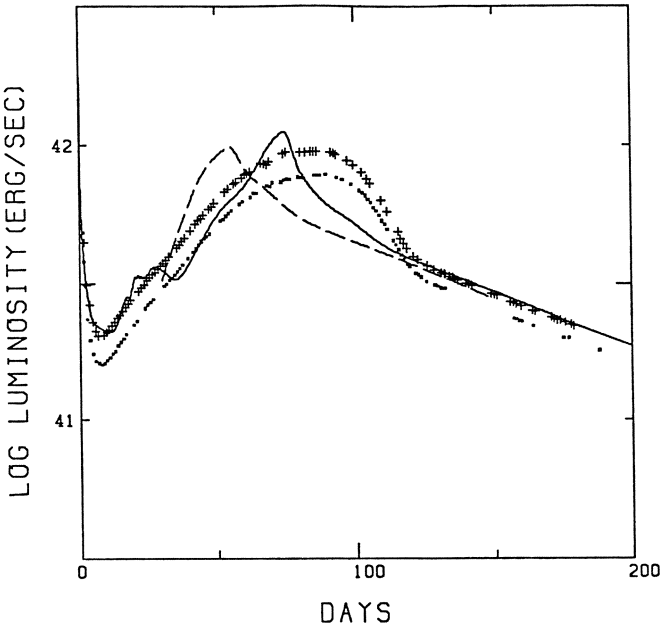


Figure 19: Effect of M_{env} on the light curve. The solid line is Model 11E1Y4 (M = 11.3 M_o, M_{env} = 6.7 M_o; mixing of ^{56}Ni at M_r < 3.4 M_o) and the dashed is Model 7E1Y4 (M = 7 M_o, M_{env} = 2.4 M_o, E = 1 x 10^{51} erg).

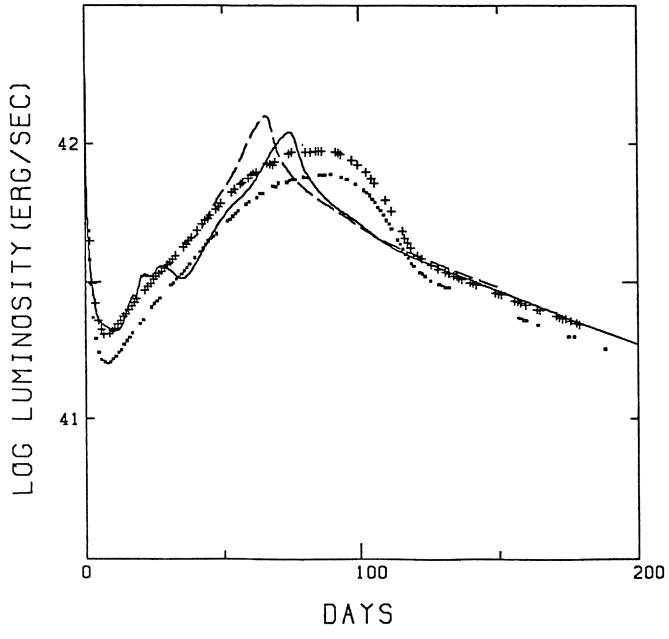


Figure 20: Effect of explosion energy E on the light curve. The solid line is the same as in Fig. 15 and the dashed is for Model 11E1.5Y4 (E = 1.5 x 10^{51} erg).

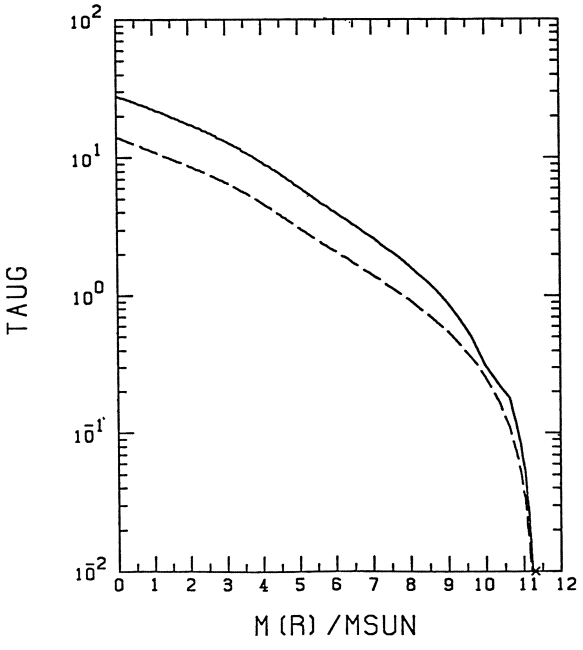


Figure 21: Thomson scattering optical depth as a function of M_r at t = 116 d for Model 11E1Y4 (solid) and 11E2Y4 (dashed).

The light curve depends on the time scale of the heat transport in the expanding matter and thus on the opacity. We may consider the case where the lower bound of the opacity in the helium and heavy element layer is larger. This might be possible, if the effects of velocity gradient on the line opacity is large. The result for such enhanced opacity is not so satisfactory because L_{pk} is about L given by Eq. (2) which is too low to be compatible with the observed L_{pk}. On the other hand, if we set a smaller lower bound for the opacity of helium layer, the increase in light curve to the peak is too steep.

5.3. Light Curve and Hydrodynamics

As discussed above, the light curve shape is sensitive to the hydrodynamics and thus a useful tool to infer 1) distribution of the heat source ^{56}Ni, 2) explosion energy E, and 3) mass of the hydrogen-rich envelope M_{env}.

1) Ni - Co distribution: As a result of Rayleigh-Taylor instability (§4), it is likely that ^{56}Ni is mixed into outer layers to some extent. Figures 17 - 18 show how the light curve depends on the distribution of ^{56}Ni. If we assume that ^{56}Ni is confined in the innermost layer of the ejecta, the increase in luminosity is delayed, a dip appears in the curve, and the light curve shape of the rising part is too steep as compared with the observations (dashed line in Fig. 17). On the other hand, if ^{56}Ni is mixed into outer layers, heat is transported to the envelope earlier. As a result the optical light increases earlier and the light

curve shape is less steep, being in better agreement with the observation (Fig. 18). This may suggest that heat source might actually be mixed into outer layers. Non-spherical effect could play an important role.

2) Envelope mass: For Model 7E1 with the hydrogen-rich envelope of 2.6 $M_\odot$, a peak luminosity of $\sim 10^{42}$ erg s^{-1} is reached too early at about t = 50 d because the expansion velocities of the helium and heavy element layers are larger than in 11E1 and the photosphere approaches the helium layer earlier. Therefore M_{env} should be larger than $\sim$ 3 $M_\odot$. Probably we need M_{env} = 5 - 8 $M_\odot$ for good agreement with the observations (also [37]). This indicates the importance of the deceleration of the expanding core by the hydrogen-rich envelope [39]. On the other hand, M_{env} should be lower than a certain value because the nitrogen-rich layer had been lost from the star before the explosion. The important question is if there exists an allowable range of M_{env} to satisfy both conditions.

3) Explosion energy: For Model 11E1.5Y4 (E = 1.5 x 10^{51} erg), the luminosity peak of 1.3 x 10^{42} erg s^{-1} is reached at t = 63 d which is a little too early compared with the observation. The cases with E = 1.0 - 1.3 x 10^{51} erg show a better fit to the observations (Fig. 20). This constraint is consistent with the condition obtained from the photospheric velocity.

To summarize, the light curve of late stages suggests relatively slow expansion velocity of the helium and heavy element layers. This is also suggested from the observed expansion velocity at the photosphere v_{ph}. This condition imposes the upper bound of E/M and the lower bound of M_{env}. On the other hand, the initial steep rise and plateau luminosity imposes the lower bound of E/M. These conditions can be satisfied as follows:

1) The late light curve suggests the upper bound of explosion energy to be about 1.5 x 10^{51} erg for M $\sim$ 10 $M_\odot$. On the other hand, the early light curve requires E > 1.0 x 10^{51} erg. Thus E = 1.0 - 1.5 x 10^{51} erg for M = 11.3 $M_\odot$ could satisfy both conditions. For larger M and M_{env}, E would be larger.

2) The late light curve may require M_{env} > 3 $M_\odot$. However, this M_{env} would be close to the upper bound of M_{env} in order for the nitrogen-rich layer to be exposed by mass loss. At this moment, it is not clear whether we can find solution to satisfy these conditions of M_{env}. Certainly M_{env} is an important quantity to determine.

6. Concluding Remarks

The hydrodynamical models and the resulting light curve models impose several interesting constraints on M_{env} and E, i.e., on the presupernova mass loss history and the explosion mechanism.

In particular, the requirement of M_{env} > 3 - 5 $M_\odot$ suggests that the progenitor returned from the red-supergiants to the blue before losing most of its hydrogen-rich envelope. This might be inconsistent with the evolutionary models [21, 36]

where M_{env} is smaller than 1 $M_\odot$. The radius of the progenitor is sensitive to many factors, especially convection, more careful evolutionary calculation with mass loss is required. If significant mass loss had occurred in the progenitor, Figure 9 predicts large ratios of He/H, N/C, N/O, and $^{13}C/^{12}C$ in the surface region of the progenitor. This can be checked observationally and theoretically.

The explosion energy, M_{env}, and the mixing of ^{56}Ni can be inferred from the observations of X-rays and γ-rays originated from radioactive decays [22]. Very recently, the X-ray astronomy satellite Ginga discovered an X-ray source in the region of SN 1987A [43]. Because of its hard spectrum, the source is very likely SN 1987A. The X-rays from this source have been gradually increasing since t = 130 d and seems to reach the peak around t = 190 d [44]. If the hard component of the X-rays is due to Compton degradation of γ-rays emitted from the ^{56}Co decay, the detection is much earlier than predicted [45,46]. Such an early emergence of X-rays and a relatively slow increase in intensity can be qualitatively accounted for by the radioactive decay model if we assume that ^{56}Co has been mixed into outer layers [17,47].

The evolutionary scenario with significant mass loss during the red supergiant phase predicts the existence of a low density inner circumstellar matter and a denser outer matter around SN 1987A. The soft component of the X-rays observed with Ginga [44] might be due to the collision between the supernova ejecta and the circumstellar material [16,48]. If this is the case, the X-ray luminosity and timing [43] indicate that the mass loss occured several hundreds years before the explosion and the mass loss rate over wind velocity is 3×10^{-6} $M_\odot$ y^{-1} / 10 km s^{-1} [48].

The suggested explosion energy of 1.0 - 1.5 x 10^{51} erg could be obtained by both prompt [3] and delayed explosion mechanism [35]. From the presupernova core structure (Figs. 3-4), the iron core is in the range of 1.4 - 1.6 $M_\odot$, while the very steep density jump appears at $M_r \sim 1.6$ $M_\odot$. Therefore mass cut that divides the neutron star and the ejecta may be about 1.6 $M_\odot$ irrespective of the explosion mechanism.

We would like to thank Drs. R. Chatchpole and R. Williams for sending data taken at SAAO and CTIO, respectively, prior to publication. We are grateful to Prof. Y. Tanaka and Dr. H. Inoue for Ginga information. We are indebted to Dr. A. Yahil for providing us the hydrodynamics code and Professors G.E. Brown and D. Sugimoto for stimulating discussion. M.H. thanks Drs. W. Hillebrandt and E. Muller for the hospitality at the Max-Planck-Institut. This work has been supported in part by the Grant-in-Aid for Scientific Research (62540183) of the Ministry of Education, Science, and Culture in Japan and by the Space Data Analysis Center, Institute of Space and Astronautical Sciences.

References

1. Arnett, W.D. 1987, Astrophys. J., **319**, 136.
2. Baron, E., Bethe, H., Brown, G.E., Cooperstein, J., and Kahana, S. 1987, preprint.
3. Baron, E., Cooperstein, J., and Kahana, S. 1986, Phys. Rev. Lett.
4. Bertelli, G., Bressan, A., and Chiosi, C. 1985, Astr. Astrophys.
5. Blanco V.M. et al. 1987, Astrophys. J., in press.
6. Catchpole, R. et al. 1987, M.N.R.A.S., submitted.
7. Cristiani, S. et al. 1987, Astr. Astrophys., **177**,L5.
8. Fowler, W.A. 1984, Rev. Mod. Phys., **56**, 149.
9. Gilmozzi, R. et al. 1987, Nature, **328**, 318.
10. Hamuy, M., Suntzeff, N.B., Gonzalez, R., and Martin, G. 1987, Astrophys. J., submitted.
11. Hashimoto, M., and Nomoto K. 1987, in preparation.
12. Hillebrandt, W., Hoflich, P., Truran, J.W., and Weiss, A. 1987, Nature, **327**, 597.
13. Hoflich, P. 1987, in Proc. of the 4th workshop on Nuclear Astrophysics.
14. Huebner, W., Merts, A., Magee, M., Jr., and Argo, M. 1977, Los Alamos Sci. Lab. Rept. No. LA6760M; Opacity table for Nomoto mixture.
15. Humphreys, R.M., and Davidson, K. 1978, Astrophys. J., **232**, 409.
16. Itoh, H., Hayakawa, S., Masai, K., and Nomoto, K. 1987, Publ. Astron. Soc. Japan, **39**, 529.
17. Itoh, M., Kumagai, S., Shigeyama, T., and Nomoto, K. 1987, in IAU Colloquium 108, Atmospheric Diagnostics of Stellar Evolution, ed. K. Nomoto (Springer).
18. Kirshner, R., Sonneborn, G., Crenshaw, D.M., and Nassiopoulos, G.E. 1987, Astrophys. J. Lett., in press.
19. Kirshner, R. et al. 1987, IAU Circ. No. 4435.
20. Maeder, A. 1987a, Astr. Astrophys. **173**, 247.
21. Maeder, A. 1987b, in SN 1987A, ed. J. Danziger (ESO, Garching).
22. McCray, R., Shull, J.M., and Sutherland, P. 1987, Astrophys. J. Lett., **317**, L73.
23. McNaught, R.H. 1987, IAU Circ. No. 4389.
24. Menzies, J.M. et al. 1987, M.N.R.A.S., in press.
25. Nomoto, K., and Hashimoto, M. 1986, Prog. Part. Nucl. Phys., **17**, 267.
26. Nomoto, K., and Hashimoto, 1987a, in Chemical Evolution of Galaxies with Active Star Formation, Sci. Rep. Tohoku Univ. **7**, 259.
27. Nomoto, K., and Hashimoto, 1987b, in Proc. Bethe Conference on Supernovae, Physics Report, in press.
28. Nomoto, K., Itoh, H., Hayakawa, S., Masai, K., and Shigeyama, T. 1987, in SN 1987A, ed. J. Danziger (ESO, Garching).
29. Pomraning, G.C. 1983, Astrophys. J., **266**, 841.

30. Schaeffer, R., Casse, M., Mochkovitch, R., and Cahen, S. 1987, preprint.
31. Shigeyama, T., Nomoto, K., Hashimoto, M., and Sugimoto, D. 1987, Nature, **328**, 320.
32. Wampler, E.J., Truran, J.W., Lucy, L.B., Hoflich P., and Hillebrandt, W. 1987, Astr. Astrophys., **182**, L51.
33. Weaver, T.A., Axelrod, T.A., and Woosley, S.E. 1980, in Type I Supernovae, ed. J.C. Wheeler (University of Texas; Austin), p. 113.
34. West, R.M., Lauberts, A., Jorgensen, H.E., and Schuster, H.E. 1987, Astr. Astrophys., **177**, L1.
35. Wilson J.R., and Mayle, R. 1987, preprint.
36. Wood, P.R., and Faulkner, D.J. 1987, Proc. Astr. Soc. Australia.
37. Woosley, S.E. 1987, preprint.
38. Woosley, S.E., Pinto, P.A., Martin, P.G., and Weaver, T.A. 1987a, Astrophys. J., **318**, 664.
39. Woosley, S.E., Pinto, and Ensman, L. 1987b, Astrophys. J., submitted.
40. Woosley, S.E., and Weaver, T.A. 1986, Ann. Rev. Astr. Ap., **24**, 205.
41. Yahil, A., Johnston, M.D., and Burrows, A., 1987, J. Comp. Phys. in press.
42. Zoltowski, F. 1987, IAU Circ. No. 4389.
43. Makino, F. 1987, IAU Cir. 4447.
44. Dotani, T. and the Ginga team 1987, Nature, submitted.
45. Xu, Y., Sutherland, P., McCray, R., Ross, R.P. 1987, Astrophys. J. Lett. submitted.
46. Ebisuzaki, T., Shibazaki, N. 1987, Astrophys. J. Lett. submitted.
47. Ebisuzaki, T. 1987, private communication.
48. Nomoto, K., Shigeyama, T., Hayakawa, S., Itoh, H., Masai, K. 1987, in Proc. 4th George Mason Workshop on Supernova 1987A in the LMC.

Stellar Evolution and Nucleosynthesis Implications of Supernova 1987A

J.W. Truran

Max-Planck-Institut für Physik und Astrophysik, Institut für Astrophysik,
Karl-Schwarzschild-Str. 1, D-8046 Garching, Fed. Rep. of Germany and
Department of Astronomy, University of Illinois,
Urbana, IL 61801, USA

The implications of Supernova 1987A in the Large Magellanic Cloud are examined with respect to theoretical models of the evolution of massive stars and of nucleosynthesis in massive stars and associated Type II supernovae. The outburst is identified with the explosion of a supergiant star of approximately 20 $M_\odot$. We address the question as to how the presence of a blue supergiant progenitor for Supernova 1987A can best be understood in the context of stellar evolution theory. We also review the nucleosynthesis expectations for Type II supernovae including, specifically, the production of nuclei of mass A = 56 as ^{56}Ni and its implications for both the light curve of Supernova 1987A and gamma ray astronomy.

1. Introduction

The outburst of Supernova 1987A in the Large Magellanic Cloud is providing an unprecedented opportunity for theoretical astrophysicists to test their models of massive star evolution, gravitational collapse, neutron star formation, the supernova explosion mechanism itself, and explosive nucleosynthesis. Observations of the photometric and spectral development over a broad range of wavelengths are providing information which is crucial to efforts to understand the structure and evolution of the stellar progenitor, the hydrodynamics of the core explosion and the ensuing passage of the shock outward through the stellar envelope, and nucleosynthesis in massive stars and supernovae. The detection of neutrinos from Supernova 1987A signaled the birth of extra-solar-system neutrino astronomy and exciting developments in X-ray and γ-ray astronomy may be forthcoming.

Our concern in this paper is with the implications of Supernova 1987A for stellar evolution and nucleosynthesis theories. One significant outgrowth of observational studies of Supernova 1987A has been the unambiguous identification of the stellar progenitor. This alone has provided some surprises for stellar evolution theorists. In particular, the identification [1,2] of a blue supergiant progenitor for Supernova 1987A – the B3 Ia supergiant Sanduleak -69 202 – has raised questions concerning our understanding of the details of the later stages of evolution of massive stars. The general opinion of stellar evolution theorists was that stars of masses > 10 $M_\odot$ compatible with the progenitors of Type II supernovae should evolve to red supergiants prior to the ignition of carbon in the central regions and thus should subsequently explode as red supergiants. We note,

alternatively, that the calculations of LAMB et. al. [3], for stars of solar metallicity evolving in the absence of mass loss, indicated that significant redward evolution need not be experienced for stars in the appropriate mass range. The inclusion of mass loss effects encourages evolution to the red, unless the mass loss rates are so high that the hydrogen envelope of the star is almost completely removed by the end of helium burning. Such high mass loss rates are observed only in the case of Wolf-Rayet stars, which typically have much higher initial masses than SK -69 202.

We first require an estimate of the allowed range of luminosity and associated stellar mass for the supernova progenitor. The fact that Supernova 1987A is situated in the Large Magellanic Cloud, at a well determined distance, allows an accurate estimate to be obtained. The identification of Sanduleak -69 202 as of spectral type B3 Ia implies a surface effective temperature of approximately 15,000 K. With the choices of an apparent visual magnitude m_v = 12.24 mag [4], a distance modulus m-M = 18.5 for the LMC, and a reddening correction A_v = 0.6 mag, a photometric analysis of the several components of the Sanduleak -69 202 system by WEST et. al. [1] yielded an absolute visual magnitude M_v = -6.8 mag for the B3 Ia star. The further assumption of a bolometric correction of 1.3 yields an absolute bolometric magnitude for the supernova progenitor M_{bol} = -8.1 mag and a corresponding luminosity of 5×10^{38} erg s^{-1} or 1.3×10^5 $L_\odot$.

An estimate of the mass of the progenitor now follows from a survey of published models of massive star evolution. As we will learn from our subsequent discussion, it is necessary to give careful attention to results obtained for various assumptions concerning initial metallicity, the rate of mass loss, and the treatment of convection. The models of BRUNISH and TRURAN [5,6] indicate that a 15 $M_\odot$ star will reach the onset of carbon burning at a luminosity $\approx 3.1 \times 10^{38}$ erg s^{-1} ($\approx 8 \times 10^4$ $L_\odot$) quite independent of initial metallicity and the presence of a moderate rate of stellar mass loss. WEAVER et. al. [7] found a 15 $M_\odot$ star of solar metallicity to have a luminosity of 3.7×10^{38} erg s^{-1}, while the 20 $M_\odot$ model discussed by WILSON et. al. [8] had a luminosity of 5.7×10^{38} erg s^{-1}. Similar results are reflected in the evolutionary tracks for massive stars undergoing mass loss, as described by CHIOSI and MAEDER [9]. We will therefore proceed on the assumption that the progenitor of Supernova 1987A had a mass of $\approx$ 20 $M_\odot$.

2. Evolution of the Stellar Progenitor

A question of particular interest to stellar evolution theorists concerns why the progenitor was a blue supergiant. In this section, we briefly review three possible explanations for such an occurrence: (i) it is a consequence of the lower metallicity characteristic of the Large Magellanic Cloud; (ii) it is a consequence of a high rate of mass loss, which enables the progenitor to evolve to the red and then return to the blue; and (iii) it is attributable to the effects of convection

on the evolution of supergiants. Some possible observational constraints on the supergiant phase evolution of the progenitor star are also noted.

Metallicity effects provide a promising possible explanation for the occurrence of a blue supergiant progenitor. Guided by the results of evolutionary calculations by BRUNISH and TRURAN [5,6] which predicted that stars of mass ~ 20 $M_\odot$ of low metallicity can reach the end of their evolution as blue supergiants, HILLEBRANDT et. al. [10] suggested that the occurrence of a blue supergiant progenitor in a low metallicity population such as that which characterizes the Large Magellanic Cloud could be understood in this manner. This has been confirmed by subsequent calculations by TRURAN and WEISS [11]. The results of calculations [10] performed without mass loss for models of 15 and 20 $M_\odot$ and different metal contents are shown in Fig. 1. For comparison, the metal abundances in H II regions [12] in the Large Magellanic Cloud indicate a metallicity $Z \approx Z_\odot/4$. Clearly, in both cases, a moderately low metal abundance compared to the sun results in a blue supernova progenitor. The models confirm earlier results by BRUNISH and TRURAN [5,6] and STOTHERS and CHIN [13], which indicated that a low metallicity favors a "bluish" evolution. Indeed, these models fail to achieve any red supergiant evolution whatsoever. This fact appears to be in contradiction with the occurrence of cool stars in the LMC. It therefore would be very interesting to determine whether SK -69 202 actually has been a red supergiant at earlier phases.

BRUNISH and TRURAN [5,6], furthermore, have investigated the influence of moderate mass loss on the evolution of stars of both populations. Although their final models at the end of the carbon burning phase, after suffering the loss of approximately 10% of their initial mass, have a tendency to be slightly cooler than the models without mass loss, the basic results are unchanged. MAEDER [14]

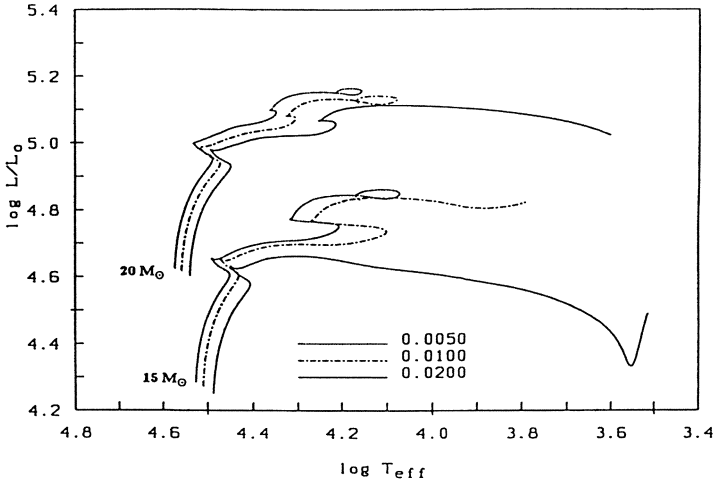


Fig. 1 Evolution for 15 $M_\odot$ and 20 $M_\odot$ stellar models to the onset of carbon burning for different metallicities

finds the same weak influence on Population I models becoming red supernova progenitors when utilizing moderate mass loss. However, when he used a higher mass loss rate, models losing 50% or more of their initial mass were found to reach the end of their evolution as blue supergiants. MAEDER's [15] recent result for the evolution of the progenitor of SN 1987A, shown in Fig. 2, confirms that heavy mass loss yields a blue supergiant. However, the total amount of hydrogen left at the time of core collapse is only ~ 3 $M_\odot$ (the star has lost mass from 20 $M_\odot$ to 9 $M_\odot$). This value may be too small to reconcile with models of the light curve and spectral evolution of SN 1987A.

An alternative approach which yields a blue progenitor that previously has been a red supergiant is based upon a treatment of semiconvection which differs from the standard (Schwarzschild) treatment. Using the Ledoux criterion and some particular assumption concerning the mixing timescale, WOOSLEY et. al. [16] succeeded in obtaining the desired result. BARAFFE and EL EID [17] presently are also investigating evolution with the Ledoux criterion for convection. Equivalent calculations by TRURAN and WEISS [18] for a 20 $M_\odot$ star are shown in Fig. 3. They apply the same criterion and also assume that mixing takes place only in the region where $\nabla_{rad} > \nabla_{Ledoux} = \nabla_{ad} + \beta/4\text{-}3\beta \cdot d\ln\mu/d\ln P$. This corresponds to the extreme case of assuming that mixing in semiconvective regions, i.e. where $\nabla_{ad} < \nabla_{rad} < \nabla_{Ledoux}$, is infinitely slow. Note that, for these assumptions, the star is found to evolve to the red and then to return to the blue, even prior to the termination of core helium burning (Fig. 3). The other extreme, instantaneous mixing as in the fully convective regions, corresponds to the application of Schwarzschild's criterion, i.e. $\nabla_{rad} > \nabla_{ad}$. In fact, LANGER et. al. [19], using a diffusion approximation for semiconvection, mentioned that in most cases the mixing is very fast, so that the Schwarzschild case is recovered. Evolution with the Schwarzschild criterion is shown for comparison in Fig. 4; here, the star

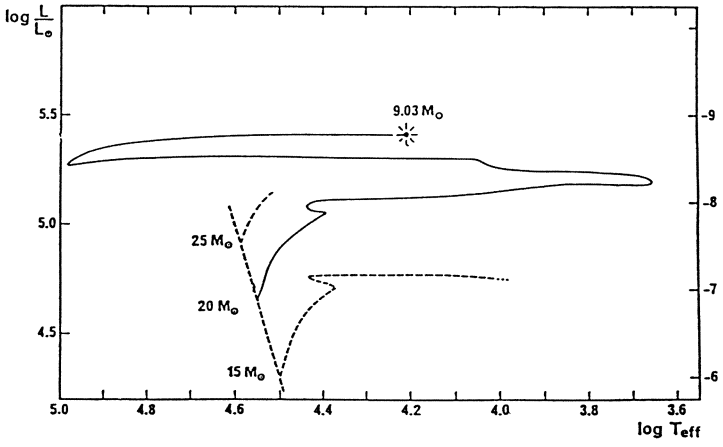


Fig. 2 Evolution of a possible progenitor for SN 1987A with high mass loss (adapted from Maeder 1987)

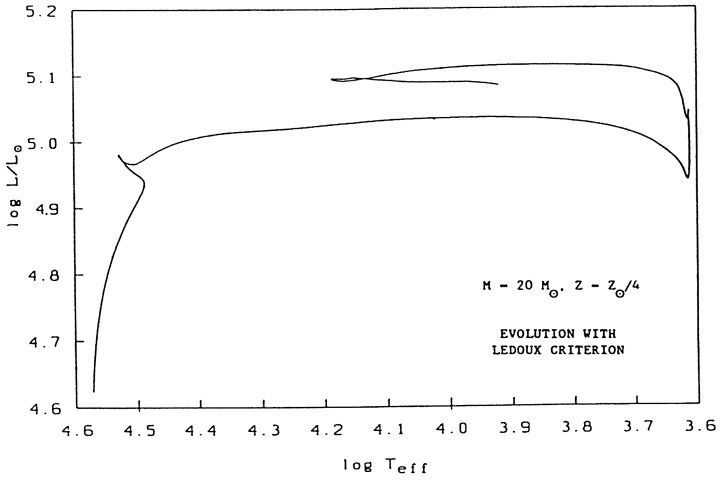


Fig. 3 Evolution through helium burning with the choice of the Ledoux criterion

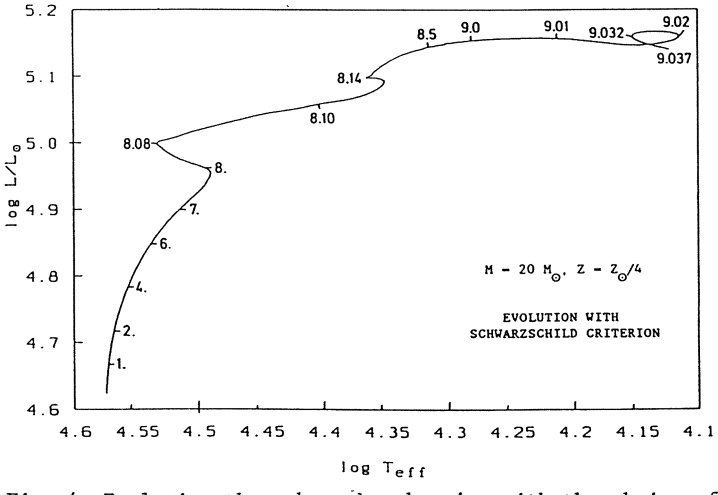


Fig. 4 Evolution through carbon burning with the choice of the Schwarzschild criterion (numbers give ages in 10^6 years)

never evolves to the red and is to be found in the blue supergiant region at the end of core carbon burning.

The supergiant phase of evolution of the progenitor of Supernova 1987A is thus seen to provide an interesting test of our understanding of stellar evolution theory. A critical question here concerns the extent of evolution experienced by the supernova progenitor as a red or blue supergiant. Was SK -69 202 ever a red supergiant? While it was apparently a blue supergiant immediately prior to the outburst, its earlier evolution in the supergiant region of the Hertzsprung-Russell diagram is a matter of discussion and debate. There are some general constraints on the character of this evolution which we believe may help guide us to an understanding of the nature and evolutionary history of Sanduleak -69 202.

The question quite naturally arises as to whether the stellar evolution calculations reviewed here for metal deficient stars imply a significant distortion of the Hertzsprung-Russell diagrams for low metallicity populations in general and for the Large Magellanic Cloud in particular. HUMPHREYS and DAVIDSON [20] have argued that massive star evolution has resulted in very similar supergiant populations in the Galaxy and the LMC, in spite of the small differences in the heavy element abundance levels. Of particular concern might be the fact that the observed M_{bol} verses log T_e H-R diagrams for both the LMC [20] and the SMC [21] reveal the presence of a significant population of red supergiants at luminosities consistent with stars in the mass range ~ 15-20 $M_\odot$. For the case of the LMC, we believe that this observed behavior might be simply a consequence of a spread in metallicity of a factor of a few, since our numerical results reveal that there exists a strong dependence of the extent of redward evolution on metallicity Z for values in the range $0.1\ Z_\odot \leq Z \leq Z_\odot$. There also exists some uncertainty as to whether the H II region abundances [12] for oxygen and a few other elements accurately reflect the level of abundances of heavier metals like iron. A somewhat broader spread in metallicity might be necessary to allow an understanding of the H-R diagrams for the SMC, which is known to have a lower Z than the LMC. This could be a problem for our models. However, we believe that the observations of both the LMC and the SMC suggest a relatively larger population of supergiants in the temperature range T_e ~ 10,000-20,000 K, compatible with the results of our calculations.

The alternative interpretation of this data is generally based on calculations of massive star evolution at high mass loss rates. Evolutionary tracks for massive stars of solar composition undergoing mass loss, reviewed by CHIOSI and MAEDER [9], reveal a more rapid evolution to the red subsequently followed by quite significant blueward excursions. For such models, the breadth of the blue supergiant region (and its relative population) is generally attributed to the presence of these blue loops. The occurrence of an immediate blue supergiant progenitor of Supernova 1987A is not easy to understand in this context, since mass loss effects are not expected to be so significant for stars of mass ~ 20 $M_\odot$. However, MAEDER [15] has shown that, for the choice of an appropriate high rate of supergiant phase mass loss, the evolutionary models can reproduce both the observed distributions of red and blue supergiants and the occurrence of a blue supernova progenitor.

What evidence might exist for the presence of high rates of mass loss for stars in the Magellanic Clouds? The predicted rates of mass loss for main sequence stars of masses ~ 20 $M_\odot$ are not high enough to effect any significant change in the envelope mass. Stars which remain in the blue supergiant regime of the H-R diagram and never evolve to red supergiants, as in the calculations of BRUNISH and TRURAN [5,6], may therefore not experience significant levels of mass loss. There are, however, several indications that stars in the LMC and SMC do suffer

significant mass loss. BESSEL [22] has noted that the pulsation periods for red supergiants in the SMC imply masses which are less than those one would normally infer from the luminosities, suggesting that mass loss has indeed occurred. KUDRITZKI [23] has determined high helium to hydrogen ratios for two BO stars in the LMC, again indicating that substantial envelope depletion is realized at least for stars of somewhat larger masses ~ 30 $M_{\odot}$.

These observations suggest that significant mass loss may accompany phases of red supergiant evolution. Other features of such evolution are also of interest. Models that evolve to red supergiants prior to carbon ignition develop convective envelopes which act to bring processed matter from the CNO-hydrogen-burning shell to the surface. Abundance signatures of such mixing will include an increased He/H ratio and high N/C and N/O ratios, such as have been reported for other blue supergiants in the LMC by KUDRITZKI [23]. High mass loss rates can further emphasize these signatures. The detection of such abundance trends in the spectrum of Supernova 1987A during the phase when the spectrum is dominated by hydrogen would serve to confirm the convective character of the envelope at an earlier epoch and thus support the argument that Sanduleak -69 202 was once a red supergiant. Alternatively, the facts that no such peculiar abundance ratios have been detected to date and that the spectrum remains hydrogen rich appear to support the view that significant mass loss did not occur for the supernova progenitor.

3. Condensed Remnant Expectations

Another interesting and still outstanding question with respect to Supernova 1987A concerns the character of the remnant. The massive star progenitors of supernovae of Type II are generally predicted to give rise to condensed remnants: neutron stars or black holes. This is a simple consequence of the fact that thermonuclear burning stages in such stars proceed through the formation of an iron core in which, lacking further nuclear fuel, collapse ensues. Model predictions for the helium core masses, iron core masses, and remnant neutron star masses, for stars of main sequence masses in the range 12-25 $M_{\odot}$, are provided in the review by WOOSLEY and WEAVER [24].

Of particular interest here is the question whether the remnant core mass can exceed the limiting mass for a stable neutron star ~ 2 $M_{\odot}$ [25] and thus imply black hole formation rather than neutron star formation. The fact that WOOSLEY and WEAVER [24] quote a remnant mass of 1.96 $M_{\odot}$, for the case of a 25 $M_{\odot}$ star, is suggestive of this possibility. We note that predictions of the remnant masses are a sensitive function of the core entropy structure, which in turn depends upon whether the core is convective or radiative. A concomitant of convective burning is efficient energy transport and cooling. This generally favors the transition to a partially degenerate gas such that the critical mass for contraction is the Chandrasekhar mass, and relatively small burning cores are realized. In contrast,

for radiative burning cores the Chandrasekhar mass does not play such a dominant role. Core carbon burning in the presence of low ^{12}C concentrations, such as might be expected using the currently accepted value for the rate of the $^{12}C(\alpha,\gamma)^{16}O$ reaction, yields radiative burning cores and thus larger iron core masses. This effect ultimately yielded the 1.96 $M_\odot$ neutron star mass quoted for a 25 $M_\odot$ star. WOOSLEY and WEAVER [24] argue that this effect becomes more important for stars of mass $M \gtrsim 20\ M_\odot$ than for lighter ones, but there still seems to remain the possibility of black hole formation in models of mass compatible with that inferred for Supernova 1987A.

What is the best guess as to the remnant of Supernova 1987A? A straightforward interpretation of the Kamiokande and IMB neutrino events follows from the assumption of neutron star formation and its implied intensity and timescale for the emitted neutrino pulse. The remnant expectations in this instance are clear: on the timescale of one or two years required for the expanding supernova envelope to become transparent, the underlying neutron star may become detectible as a consequence of pulsar beaming, thermal X-ray emission, or the heating and excitation of the surrounding nebular remnant.

4. Nucleosynthesis Expectations

The hydrogen features in the spectrum of Supernova 1987A confirm that it is a Type II supernova. Such supernovae are the expected products of the evolution of massive stars $M \gtrsim 10\ M_\odot$. The higher temperatures prevailing in the cores of massive stars are sufficient to ensure that carbon, oxygen, neon, and silicon thermonuclear burning phases will proceed through the formation of a core composed of iron-peak nuclei, surrounded by a mantle of matter enriched in intermediate mass elements. The predicted mass [24] of heavy elements ejected in the supernova phase for a 15 $M_\odot$ model is 1.24 $M_\odot$, while that for a star of 20 $M_\odot$ is about 2.5 $M_\odot$. Such matter should presumably become viewable when the expanding hydrogen-rich envelope becomes optically thin.

Detailed nucleosynthesis calculations (see, for example, the reviews by TRURAN [26] and WOOSLEY and WEAVER [24]) indicate that oxygen and nuclei from neon to calcium (the expected products of explosive carbon and oxygen burning) are produced in approximately solar proportions in massive stars and associated Type II supernovae, while both carbon and the iron peak nuclei are somewhat underproduced. ^{12}C underproduction is a consequence of the rate of the $^{12}C(\alpha,\gamma)^{16}O$ reaction and the implied $^{12}C/^{16}O$ ratio resulting from core helium burning; this also directly influences the predicted abundances of the nuclei neon, sodium, magnesium, and aluminum - the products of carbon burning. While there is general qualitative agreement regarding these nucleosynthesis expectations, published numerical calculations differ significantly in detail. This is evident from a comparison of the presupernova nucleosynthesis predictions from the studies by ARNETT [27] and WOOSLEY and WEAVER [28] which are shown in

Table 1
PRESUPERNOVA NUCLEOSYNTHESIS
ABUNDANCE

MASS ($M_\odot$)	^{12}C	^{16}O	Ne	Mg	Si	S	Ar	Ca	Ref.
15	0.57	1	0.43	0.35	5.1	8.7	11.	13.	a
	2.6	1	1.5	6.0	3.9				b
20	0.29	1	0.23	0.21	2.4	2.6	2.4	1.6	a
	0.83	1	3.5	5.0	3.9				b
25	0.19	1	0.38	0.52	1.4	1.5	1.6	1.5	a
	0.76	1	3.0	2.5	2.4				b

(a) WOOSLEY and WEAVER (1986b); (b) ARNETT (1978)

Table 1. Here, the production of nuclei from ^{12}C to ^{40}Ca, normalized to oxygen in the sun, is presented as a function of stellar mass. Note that, while the two studies used a similar $^{12}C(\alpha,\gamma)^{16}O$ rate, the abundance predictions differ considerably. The level of uncertainty associated with stellar model calculations of nucleosynthesis is presumably reflected in this comparison. Nevertheless, we can expect that the spectrum of Supernova 1987A should ultimately reveal the presence of significant enrichments of nuclei from oxygen to calcium, when matter from the outer regions of the hydrogen-exhausted core reaches the photosphere.

The mass fraction in the form of iron peak nuclei predicted to be ejected in a Type II supernova event is a sensitive function of the prevailing temperature and density conditions in the vicinity of the mass cut, at the boundary of the collapsing core. The shock strength in these regions will serve to determine both the extent to which silicon and intermediate mass nuclei are processed to iron peak nuclei and the velocity imparted to the matter. Strong shocks will thus favor increased yields of iron peak nuclei. Observations of heavy element abundance patterns in extremely metal deficient stars [29,30] suggest that iron peak nuclei are typically underproduced in massive stars and associated Type II supernovae by perhaps a factor of three.

The extent of iron production specifically in Supernova 1987A is relevant to two issues: the possibility of detection of gamma rays from the decay of ^{56}Co and the possible role of decay heating in defining the late stages of the decline of the light curve. We note here the significant prediction of explosive nucleosynthesis calculations that iron peak nuclei are formed in a proton-rich environment in such a manner that many of the stable isotopes in solar system matter are products of the decay of proton-rich unstable nuclei at their respective mass numbers [31,26,32]. In particular, nucleosynthesis calculations predict that nuclei of mass A = 56 are formed in situ as ^{56}Ni ($\tau_{1/2}$ = 6.10 days; average decay energy 1.72 MeV) which, following supernova ejection, decays through ^{56}Co ($\tau_{1/2}$ = 78.76 days, average decay energy 3.58 MeV) to ^{56}Fe. The longer

lifetime of ^{56}Co allows the possibility of detection of gamma rays from ^{56}Co a year or more after the initial explosion, by which time the expansion will hopefully have diluted the ejecta sufficiently to permit these high energy photons to escape. If we assume 0.1 $M_\odot$ of ^{56}Ni to have been ejected (2.1 x 10^{54} atoms), the expected flux of gamma rays from ^{56}Co decay for a supernova at a distance of 50 kpc is given as a function of time by

$$0.73\exp(-t/113\text{days})\,\text{cm}^{-2}\text{s}^{-1}\left(\frac{M_{56}}{0.1M_\odot}\right)\left(\frac{R}{50\text{kpc}}\right)^{-2}$$

Note that the predicted flux after two years, ~ 1.1 x 10^{-3} cm^{-2} s^{-1}, is sufficient to be detected with available facilities. This would provide an extremely important test of the predictions of models of Type II supernovae. Fits to the exponential tail of the light curve of Supernova 1987A [33,34] are consistent with the production of 0.075 $M_\odot$ of ^{56}Ni, close to the value assumed above. The exciting possibility thus exists that gamma rays from ^{56}Co decay can indeed be detected.

Type II supernovae also provide the likely site of the formation of the r-process heavy nuclei by neutron capture processes [26,35,36], and thus of the interesting nuclear chronometers ^{232}Th, ^{235}U, ^{238}U, and ^{244}Pu. Unfortunately, the total mass ejected in this form in a single supernova event is expected to be $\lesssim$ 10^{-5} $M_\odot$ [37] and thus should not be detectible.

We note finally that WILLIAMS [38] has argued from the spectrum for the presence of a significant barium abundance in the ejected envelope of Supernova 1987A. If this should turn out to be a correct interpretation of the spectrum, it will certainly provide a challenge to nucleosynthesis theorists. The production of s-process nuclei like Ba is predicted to occur in the helium-burning cores of massive stars during the late stages of helium burning. However, it would be hard to understand how this could give rise to an enrichment of Ba in the hydrogen-rich envelope matter that has dominated the early spectra.

5. CONCLUDING REMARKS

We can conclude from the results of numerical studies that the evolution of a star of mass ~ 20 $M_\odot$ and initial metal composition Z = 0.25-0.5 $Z_\odot$, compatible with the metallicity of the LMC, is consistent with the occurrence of a blue supergiant progenitor for Supernova 1987A. Indeed, the predicted properties of such stars are quite in agreement with those observed for the presumed stellar progenitor Sanduleak -69 202, and their typical envelope structures, characterized by steeper density gradients and smaller photospheric radii, have been found to be consistent with the spectral and light curve development of Supernova 1987A. In the context of these models, a spread in metallicity of a factor of two for LMC supergiants is sufficient to permit an understanding of the fact that there exist red supergiants in the LMC at the same luminosity as that of the blue supergiant progenitor of SN 1987A.

Models calculated with different treatments of convective mixing or with the assumption of a high rate of mass loss can in principle provide blue progenitors which have passed through the red, but in both instances this requires a rather careful tuning of the stellar parameters. The answer to the question as to whether the progenitor was ever a red supergiant can perhaps best be provided by observations of Supernova 1987A itself. If Sanduleak -69 202 was indeed a red supergiant at some earlier stage of its history, significant mass loss and envelope mixing by convection should have occurred, and both the mass of envelope (hydrogen rich) matter and particularly its composition should provide evidence for these behaviors.

Supernova 1987A is also expected to teach us a great deal about nucleosynthesis. Analyses of the spectrum should ultimately reveal the presence of elements of intermediate mass, oxygen through calcium and beyond, as well as of iron peak nuclei. The accurate fit of the exponential tail of the light curve to the lifetime of ^{56}Co is strongly suggestive, and the ultimate detection of gamma rays from ^{56}Co decay would absolutely confirm the unique prediction of nucleosynthesis theory [31] that mass A = 56 is formed in stars as ^{56}Ni. An accurate mass estimate of 0.075 $M_{\odot}$ of ^{56}Fe formed in Supernova 1987A would then follow.

Acknowlegements

This research was supported in part by the United States National Science Foundation under grant AST 86-11500 at the University of Illinois. The author wishes to express his thanks to the Alexander von Humboldt Foundation for support by a U.S. Senior Scientist Award and to Professor R. Kippenhahn for the hospitality of the Max-Planck-Institut für Astrophysik, Garching bei München.

References

1. R. M. West, A. Lauberts, H. E. Jørgensen, H.-E. Schuster: Astron. Astrophys. 177, L1 (1987)
2. N. Panagia, R. Gilmozzi, J. Clavel, M. Barylak, R. Gonzalez-Riesta, C. Lloyd, L. San Fernandez de Cordoba, W. Wamsteker: Astron. Astrophys. 177, L25 (1987)
3. S. A. Lamb, I. Iben, Jr., W. M. Howard: Astrophys. J. 207, 209 (1976)
4. J. Isserstedt: Astron. Astrophys. Suppl. 19, 259 (1975)
5. W. M. Brunish, J. W. Truran: Astrophys. J. 256, 247 (1982)
6. W. M. Brunish, J. W. Truran: Astrophys. J. Suppl. 49, 447 (1982)
7. T. A. Weaver, G. B. Zimmermann, S. E. Woosley: Astrophys. J. 225, 1021 (1978)
8. J. R. Wilson, R. Mayle, S. E. Woosley, T. A. Weaver: Ann. N.Y. Acad. Sci. (1986)
9. C. Chiosi, A. Maeder: Ann. Rev. Astron. Astrophys. 24, 32 (1986)
10. W. Hillebrandt, P. Höflich, J. W. Truran, A. Weiss: Nature 327, 597 (1987)

11. J. W. Truran, A. Weiss: In Nuclear Astrophysics, ed. by W. Hillebrandt, E. Müller, R. Kuhfuss, J.W. Truran, (Springer, Berlin 1987) p. 81
12. R. Stothers, Ch.-W. Chin: Astrophys. J. 198, 407 (1975)
13. R. J. Dufour: In Structure and Evolution of the Magellanic Clouds, ed. by S. van den Bergh, K. S. de Boer, (Reidel, Dordrecht 1984) p. 353
14. A. Maeder: Astron. Astrophys. 102, 401 (1981)
15. A. Maeder: Preprint (1987)
16. S. E. Woosley, P. A. Pinto, L. Ensman: Astrophys. J. (1987)
17. I. Baraffe, M. El Eid: Private communication (1987)
18. J. W. Truran, A. Weiss: Preprint (1987)
19. N. Langer, M. El Eid, K. J. Fricke: Astron. Astrophys. 45, 179 (1985)
20. R. M. Humphreys, K. Davidson: Astrophys. J. 232, 409 (1979)
21. R. M. Humphreys: Astrophys. J. 265, 176 (1983)
22. M. Bessel: Private communication (1987)
23. R. Kudritzki: Private communication (1987)
24. S. E. Woosley, T. A. Weaver: Ann. Rev. Astron. Astrophys. 24, 205 (1986)
25. G. Baym, C. Pethick: Ann. Rev. Astron. Astrophys. 17, 415 (1979)
26. J. W. Truran: Ann. Rev. Nucl. Part. Sci. 34, 53 (1984)
27. W. D. Arnett: Astrophys. J. 219, 1008 (1978)
28. S. E. Woosley, T. A. Weaver: In Radiation Hydrodynamics in Stars and Compact Objects, ed. by D. Mihalas, K.-H. Winkler (Reidel, Dordrecht 1986)
29. M. Spite, F. Spite: Ann. Rev. Astr. Astrophys. 23, 225 (1985)
30. J. W. Truran, F.-K. Thielemann: In Stellar Populations, ed. by C. A. Norman, A. Renzini, M. Tosi (Cambridge University Press, Cambridge 1987) p. 149
31. J. W. Truran, W. D. Arnett, A. G. W. Cameron: Can. J. Phys. 45, 2315 (1967)
32. S. E. Woosley: Saas Fee Lecture Notes, Preprint (1987)
33. W. D. Arnett: Private communication (1987)
34. S. E. Woosley: Private communication (1987)
35. W. Hillebrandt: Space Sci. Rev. 21, 639 (1978)
36. G. J. Mathews, R. A. Ward: Rep. Prog. Phys. 48, 1371 (1985)
37. J. W. Truran, A. G. W. Cameron: Astrophys. Space Sci. 14, 179 (1971)
38. R. E. Williams: Astrophys. J. 320, L117 (1987)

X-Ray Observations of Supernova Remnants – The State of the Art –

W. Brinkmann

Max-Planck-Institut für Physik und Astrophysik, Institut für Extraterrestrische Physik, D-8046 Garching, Fed. Rep. of Germany

X-ray observations of young supernova remnants (SNR) provide the most direct tool to study their evolution, their chemical composition, and their interaction with the interstellar medium. We will show that the remnants of type I and type II supernova explosions differ considerably in their physical properties and raise different questions. Using two well studied typical remnants, Tycho's SNR for the type I and the the Crab nebula for the type II remnants, we will discuss the current observational and theoretical state of the art.

1. Introduction

If events like the recent supernova 1987A would appear more frequently our interest in the study of supernova remnants (SNR) would be perhaps considerably lower. But so far, observations of historical SNR of known age seem to be the only way to obtain direct conclusions about the explosion itself and the structure of both the star and the surrounding interstellar medium. Although much progress has been made in recent years, particularly by observations with the Einstein, EXOSAT and Tenma satellites, there are still rather large discrepancies in the interpretation of the data. It will be shown that this is due to insufficient theoretical modelling of the SNR as well as to deficiencies in current observational techniques.

In a recent compilation by van den Bergh [1] 135 objects have been listed which are considered to be galactic SNR according to their appearance and their non-thermal radio spectra. Whether all of them are actually SNR has to be proven by future observations. From 40 of these objects optical emission has been observed and 33 are X-ray emitters. Weiler [2] has estimated that about 80 % of the radio SNR are shell like, 5 % are centrally filled (or Crab - like) and the rest are a combination of both.

In the shell type SNR the emitted radiation is associated with the shock wave driven by the initial explosion into the interstellar medium. The temperatures reached in the shocks give rise to thermal emission in the X-ray range. As these remnants originate from type I supernova explosions where the whole star is disrupted in the explosion, no stellar remnant is expected in their centres. The emitted spectrum thus reflects the physical conditions and the chemical composition of the progenitor star and the ambient interstellar medium.
In the next chapter we will discuss the state of the art of deconvolving the observed spectra in order to understand the physical and chemical conditions of the system at the time of explosion.

The Crab - like or plerionic remnants are fuelled by the interaction with the central rotating, highly magnetized neutron star. The structure, dynamics and the emitted spectrum are predominantly determined by the energy input from the pulsar. As a matter of fact, the detection of the pulsar in the Crab nebula in 1969 solved two long lasting problems: on one hand

the question of the origin of the continous energy supply of $\geq 10^{38}$ erg/sec, required to power the nebular expansion and radiation. On the other hand it confirmed the idea that pulsars are rapidly rotating neutron stars: just this amount of energy will be lost by electromagnetic radiation from a "typical" neutron star with the measured angular velocity of the Crab pulsar. However, how this energy is actually transferred to the radiating particles is still an unsolved problem and it will be discussed in chapter 3.

2. Tycho's Remnant

The remnant of the type I explosion observed by Tycho de Brahe in 1572, represents one of the most beautiful examples of a SNR of this class. Its nearly perfect spherical shape shows that the explosion went off in a largely homogeneous medium so that, in turn, a spherical (one dimensional) treatment of the remnant evolution seems to be appropriate. A detailed analysis of the Einstein data by Gorenstein et al [3] confirms roughly the "standard" picture, although the mass they derive for the progenitor star is larger than the $\simeq 1.4 M_{\odot}$ expected from the deflagration of a white dwarf (Nomoto [4]).

The analysis of various experiments with higher spectral resolution further showed (Becker et al [5], Pravdo et al [6], Hamilton et al [7], and Tsunemi et al [8]) that the chemical abundance cannot be solar, but must be enriched with iron peak elements. Using the data from Nomoto's W7 model (Nomoto et al [9]), and mixing the innermost zones of that model, Itoh [10] was able to reproduce the spectral behaviour, observed by Tenma (Tsunemi et al [8]) by following numerically the temporal evolution of the hydrodynamical and non - equilibrium ionisation structure up to Tycho's present age.

However, several problems of theoretical as well as observational nature remain open, leaving the correctness of the obtained estimates uncertain. Let's first start to discuss some of the theoretical uncertainties.

The first relates directly to the underlying model. Type I supernovae are thought to be caused by exploding white dwarfs where a deflagration front, starting at the centre of the white dwarf disrupts the whole star (Nomoto [4]). The main uncertainty in these models is, however, the velocity of the deflagration front which, at present, is not calculated from first physical principles but is simulated by a time dependent mixing length theory of convection with a free paramater α, defined as the ratio of the mixing length to the pressure scale height (in particular, the W7 model is characterized by $\alpha = 0.7$).

Secondly, the required mixing of the elemental abundances of the inner zones of the model in order to reproduce the observed X-ray spectrum (as well as a similar procedure to mix the outer layers to fit the observed optical spectra of SN 1981 b (Branch [11]), indicate that some kind of mixing of matter during the stellar disruption has occured. This effect is expected because of the decaying nature of the deflagration wave and the hydrodynamics has to be simulated by at least a 2-dimensional hydrocode.

It is well known that the ionisation state of the plasma in the outgoing shock front is not at its equilibrium value at the corresponding temperature, because the time scale to ionize the heavy elements is of the order of $10^4/n_e$ years (Canizares [12]), larger than the age of the remnant.

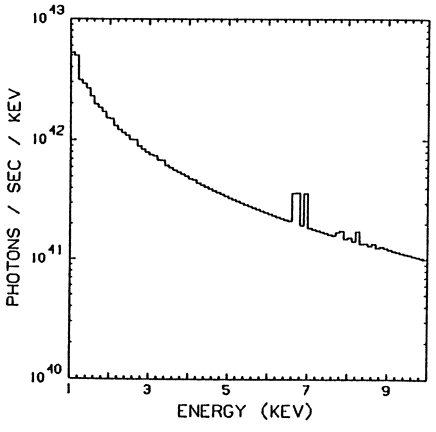


Figure 1a: Integrated photon spectrum, Equilibrium Ionisation

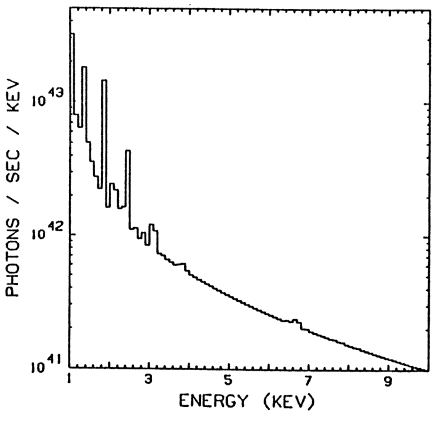


Figure 1b: Integrated photon spectrum, Non - equilibrium Ionisation

Figure 1 demonstrates this effect for a calculation done with the parameters typical for Tycho and assuming solar abundances for the circumstellar matter as well as for the ejecta (Brinkmann and Fink [13]). Figure 1a shows the emergend photon spectrum, integrated over the whole remnant for an assumed ionization equilibrium, figure 1b the corresponding spectrum for the same model but with the ionization profile calculated simultaneously with the hydrodynamical evolution (Note the different abszissas !). Both the slope of the continuum and the equivalent widths of the lines differ drastically.

Non equilibrium ionization calculations require exact reaction rates for all elements, which in many cases exist only emperically (see Mewe [14] and references therein). The biggest uncertainty, however, seems to be the state of the electron gas in the shock front itself: The timescale to equilibrate ion and electron temperatures is of the order of $m_i/m_e\ T_i^{3/2}/n_o$, that is 3×10^5 years for $T_i = 10^8$ and a particle density of $n_o = 10\ cm^{-3}$. McKee [15] suggested that plasma instabilities would cause a much more rapid equilibration, but no specific processes have been worked out. It should be noted that, even if these instabilities are sufficiently fast, they will very likely not produce Maxwellian distribution functions - as always assumed in the determination of the ionization rates.

The shock fronts entering the calculations are always assumed to be hydrodynamical, i.e. the flow variables ρ and v vary discontinously. The jumps are determined from the usual Rankine - Hugeniot conditions, relating the flow variables on either side of the discontinuity, possibly augmented by some numerical viscosity introduced by the hydro codes. It is, however, not clear at all how the structure of a high Mach number ($M \gtrsim 100$) nearly collisionless, ionizing shock is. Particle simulations are at present not able to reach these high Mach numbers at all.

The flat, power law type radio spectrum of most SNR demonstrates the additional presence of relativistic electrons and magnetic fields in the shock regions. Finally, inhomogenuities in the ejecta and in the ambient medium, as well as hydrodynamical instabilities in the outflow would require in many cases two - or three dimensional numerical modelling.

These difficulties indicate that, even given a perfectly well measured X-ray spectrum, the modelling of the supernova remnant expansion incorporates various theoretical uncertainties. It seems as if the gross features of the supernova evolution can be represented adequately, but details, requiring an accurate parameter determination which, in turn, can be used to discriminate between different models and possibilities are so far hidden in the 'noise' of imperfect theoretical knowledge.

The second class of problems is related to our temporal observational possibilities, where we are faced with the trade off between resolution and counting statistics. Roughly spoken, current detectors have either high spectral resolution but very low spatial resolving power (like the GSPC on EXOSAT or Tenma) or vice versa (like the Einstein HRI). Both types of systems give only limited information about the physical parameters of the SNR and instruments with higher spectral and spatial resolution are required to get unique answers. This is demonstrated in Figure 2 where two line of sight spectra of the above mentioned non equilibrium ionization modell of Tycho are shown.

Plotted are two line-of-sight photon spectra as seen by an instrument with a spatial resolution of 10×10 arc secs at two different positions on the remnant: one (Fig. 2a) near the centre, the other at the position of the shock. Note, that although Figure 2b shows the spectrum at the shock, the hottest region of the remnant, no iron line is seen at $\sim$ 6.7 keV, due to the insufficient time to ionize iron to the required stage.

It is evident from Figures 1 and 2 that all 'observed' spectra differ considerably, although they are taken from the same model. Trying to fit these spectra, as it would be done during the reduction of observational data, one would conclude that more component thermal bremsstrahlungs fits plus additional lines or ionisational equilibrium models are required with temperatures in the between 3 keV and $\lesssim$ 15 keV. Although these temperatures are actually in the range encountered in the hydrodynamical calculations, none of them is in any sense representative just for this model.

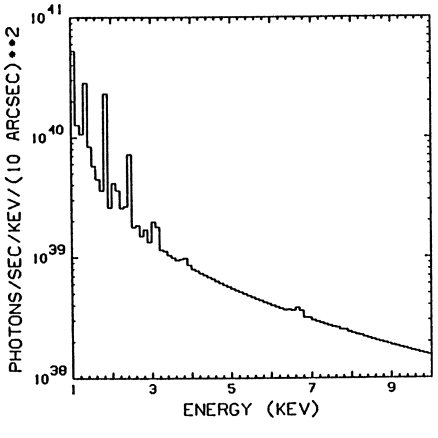


Figure 2a: Line-of-sight photon spectrum, centre of remnant

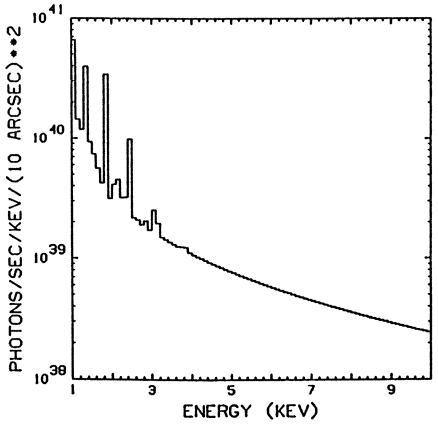


Figure 2b: Line-of-sight photon spectrum, shock region

As the equivalent widths of individual lines are dependent on the underlying continuum, the deduced chemical abundance values are clearly model dependent as well. This means, the determination of line strengths without the knowledge of a well defined continuum is highly uncertain. To pin down the physical parameters of the remnant high spectral as well as spatial resolution is necessary. As the radial variations of the equivalent widths of the lines are equivalent to their temporal evolution, i.e. their ionisation time $n_e t$, their measurement together with a well defined continuum flux give strong model constraints on the temperature - density - structure of the remnant. A spectral mapping of Tycho with the high spatial resolution assumed in Figure 2 across the remnant would allow a determination of the ionisation history of the matter and, vice versa, give the physical information required to model the object reliably.

3. The Crab Nebula

In contrast to the shell - type remnants, emitting thermal radiation from shock heated gas, Crab like or centrally filled remnants show non thermal radiation from the radio to the X-ray range with a centrally - peaked spatial distribution. The generally accepted explanation is that we are seeing Synchrotron radiation from relativistic electrons, injected by the central pulsar into the surrounding nebular field. The shrinking size of the Crab nebula with increasing photon energy is easily explained as caused by the decreasing lifetime of the higher energetic radiating electrons. This lifetime argument, on the other hand, demands continous injection of relativistic particles into the nebula.

The lifetime of plerionic remnants is therefore determined by the slowing down time scale $\tau = P \,/\, 2\dot{P}$ of the pulsar. The object we see now with its low expansion velocity of $\simeq$ 1000 km/sec is very likely a pulsar created secondary phenomenon. So far, all attempts to see traces of the original blast wave of the explosion have failed.

As mentioned above, the enormeous energy requirement to balance the losses of the emitted radiation and driving the expansion of the nebula of a few times 10^{38} erg/sec is just matched by the rotational energy loss of the rapidly spinning neutron star. Whereas this energy can be supplied, in principle, by the strong low frequency (30 Hz) electromagnetic wave, generated by a rotating magnetic vacuum dipole, the Synchrotron nature of the radiation requires the continous injection of relativistic electrons as well, and it is not clear at all how the energy is shared between these two components.

The energy is certainly not injected isotropically into the nebula as the central region of the Crab nebula shows distinct features. First of all we have an emission "hole" around the pulsar, with reduced emission at all frequencies. Secondly, on both sides of the pulsar along the major axis of the nebular we see the so called whisps, a time varying phenomenon which indicates a magneto - hydrodynamic interaction of the pulsar with its surroundings. The question, whether the temporal changes observed (Scargle [16]) represent real material flow motion or only a varying exitation of radiation is so far unsolved.

A rough geometrical picture of the pulsar - nebula interaction proposed by Aschenbach and Brinkmann [17] seems to be confirmed by recent X-ray observations (Brinkmann et al [18]). The observation showed that even at high energies ($E_\nu \gtrsim 50$ keV) the X-ray Crab has a finite extension of $\gtrsim 60"$, in contrast to the simple model of a point like injector (Makishima et al [19]).

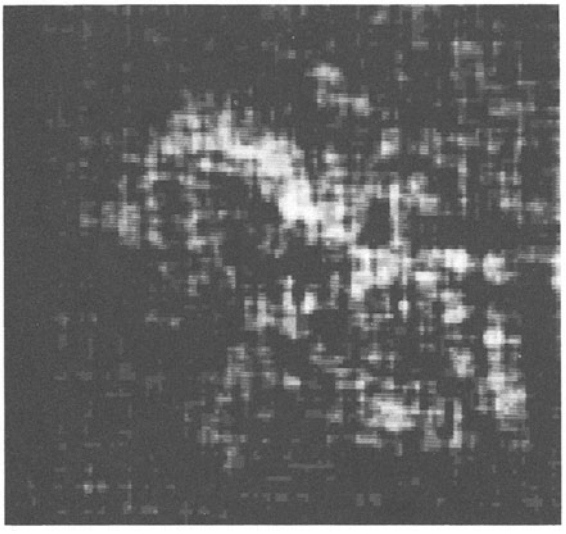


Figure 3: The X-ray emission of the Crab nebula from processed off - pulse Einstein HRI data. For details see [18].

Figure 3 shows the picture obtained by processing the Einstein HRI off-pulse Crab data (Brinkmann et al [18]). A ring like structure, including the central hole is clearly visible. The dark parts of the picture contain signals below 10% of the maximum intensity. The overall configuration as well as the ratios of the axes agree closely with the predictions of the above mentioned model, in which electrons are accelerated and then transported outwards, radiationlessly, in the strong electromagnetic wave of the pulsar. When the ram pressure of this relativistic wind equals the thermal pressure of the surrounding nebula a shock forms, producing an amplification of the magnetic field and a randomization of the directed plasma motion (Rees and Gunn [20]). Taking for the pulsar the field configuration of an oblique rotator, the outflow of the electrons will be concentrated near the rotational equatorial plane, producing the toroid-like structure of the emission region. The height of this toroid is determined by comparing the lifetime of the radiating electrons with their effective radial propagation time.

The brightness asymmetry in the optical picture near the pulsar and the X-ray offset is explained by an enhancement of the magnetic field strength in the nothwestern parts of the nebula, caused by a compressional effect of the pulsar running into the surrounding medium.

Although in the original model the geometrical configuration of the emitting region is at least qualitatively correct, a detailed comparison of the data with numerical simulations showed that the underlying particle injection and transportation mechanism cannot be as simple as anticipated. It is more probable that a more complex mechanism including particle re-acceleration, as proposed by Kennel and Coroniti [21], is involved or that the energy injection mechanism is quite different (Michel [22]).

The study of the central regions of the Crab nebula shows that the interaction between a pulsar and its surrounding medium is physically and geometrically very complex. None of the

theoretical models seem to be able to make definite predictions about the size and shape of other Synchrotron nebulae. Both, the pure vacuum model as well as any simple wind model have various conceptional difficulties, especially as a complete understanding of the most basic ingredient, the pulsar magnetosphere, is still missing.

As the plasma injection process can be approached neither by a purely electromagnetic vacuum description nor by a fully relativistic MHD treatment, future relativistic particle simulations seem to provide the most promising method to clarify these open questions. From an observational point of view, X-ray pictures with high spectral and spatial resolution are required, but the chances to achieve this goal in the near future appear, unfortunately, quite low.

4. Conclusions

We have shown that X-ray spectra with high spectral and spatial resolution are the best way to obtain better information about both types of supernova remnants.

For SNR of type I explosions it was demonstrated that X-ray spectra integrated over the whole remnant do not reveal the physical parameters of a supernova remnant with the required accuracy. Radially resolved spectra, together with a detailed hydrodynamic simulation and a simultaneous calculation of the ionisation structure seems to be the only way to obtain "physically reliable" parameters for these objects.

Synchrotron nebulae whose best and most extensively studied prototype is the Crab nebula present still a major theoretical challenge, closely linked to our general understanding of the pulsar phenomenon itself.

References

1. S. van den Bergh, 1983, In: Supernova remnants and their X-Ray Emission, IAU symposium no. **101**, eds. J. Danziger and P. Gorenstein, D. Reidel, p. 597
2. K. Weiler, 1983, In: Supernova remnants and their X-Ray Emission, IAU symposium no. **101**, eds. J. Danziger and P. Gorenstein, D. Reidel, p. 299
3. P. Gorenstein, F. Seward, and W. Tucker, 1983, In: Supernova remnants and their X-Ray Emission, IAU symposium no. **101**, eds. J. Danziger and P. Gorenstein, D. Reidel, p. 1
4. K. Nomoto, 1980, In: Type I Supernovae, ed. by J.C. Wheeler, Austin, Univ. Texas, p. 164
5. R.H. Becker, S.S. Holt, B.W. Smith, N.E. White, E.A. Boldt, R.F. Mushotzky, and P.J. Serlemitsos, 1980, Ap.J. 240, L5
6. S.H. Pravdo, B.W. Smith, P.A. Charles, and I.R. Tuohy, 1980, Ap.J. 235, L9
7. A.J.S. Hamilton, C.L. Sarazin, and A.E. Szymkowiak, 1986, Ap.J. 300, p. 713
8. H. Tsunemi, K. Yamashita, K. Masai, S. Hayakawa, and K. Koyama, 1986, Ap.J. 306, p. 248
9. K. Nomoto, F.K. Thielemann, and K. Yokoi, 1984, Ap.J. 286, p. 644
10. H. Itoh, In: Proc. IAU Colloquium no. **101**, in print
11. D. Branch, 1984, In: Workshop on Challenges and New Developments in Nucleosynthesis, Chicaco, Univ. of Chicago Press, in press

12. C.R. Canizares, 1984, In:X-ray Astronomy '84 , ed. by Y. Tanaka, Bologna

13. W. Brinkmann and H.H. Fink, In: Proc. IAU Colloquium no. **101**, in print

14. R. Mewe, 1987, Contributions presented at the 33rd Scottish Universities Summer School in Physics at St. Andrews, SRON preprint

15. C.F. McKee, 1974,Ap.J. 188, p. 335

16. J.D. Scargle, 1969, Ap.J. 156, p. 401

17. B. Aschenbach and W. Brinkmann, 1975, Astron. Astr. 41, p. 147

18. W. Brinkmann, B. Aschenbach and A. Langmeier, 1985, Nature 313, p. 662

19. K. Makishima, Y. Ogawara, M. Matsuoka, M. Oda, S. Miyamoto, R.M. Pelling, L.E. Peterson, and W.S. Penciesas, 1981, Space Sci. Rev. 30, p. 259

20. M.J. Rees and J.E. Gunn, 1974, Mon. Not. R. astr. Soc. 167, p. 1

21. C.F. Kennel and F.V. Coroniti, 1984, Ap.J. 283, p. 694

22. F.C. Michel, 1985, In: The Crab Nebula, ed. M. Kafatos, Cambridge Univ. Press, p. 55

Supernova Explosions Inside Pre-existing Wind-Driven Cavities: Spherically Symmetric Models

G. Tenorio-Tagle, P. Bodenheimer, and J. Franco†*

Max-Planck-Institut für Physik und Astrophysik,
Institut für Astrophysik und Institut für Extraterrestrische Physik,
D-8046 Garching, Fed. Rep. of Germany

This paper presents some preliminary results on the evolution of supernova explosions propagating inside wind-driven bubbles. The hydrodynamic simulations are performed with a Lagrangian one dimensional code and include the evolution of a strong stellar wind from the Type II supernova progenitor. This wind, with $\dot{M} = 10^{-5}$ $M_\odot$ yr^{-1} and $V = 10^3$ kms^{-1}, creates a large and expanding interstellar bubble with a massive outer shell. The SN ejecta evolve in the complex density distribution of the bubble and generate a series of reflected shock waves as the blast wave encounters different density stratifications.

I. INTRODUCTION

The mechanical and radiative energy injection from massive stars is apparently regulating the ionization, temperature structure, and velocity dispersions in the general interstellar medium (see reviews by McCray and Snow 1979, and Tenorio-Tagle and Bodenheimer 1988). In particular, supernovae ejecta generate strong shock waves and large expanding shells which can lead to the multi-phase structure observed in the gaseous disk (Cox and Smith 1974; McKee and Ostriker 1977). These strong shocks have been suspected of inducing star formation (Herbst and Assousa 1977) and provide an appealing scenario to accelerate cosmic rays (Blandford and Ostriker 1978). Also, depending on the supernova rate, the expanding shells can break out of the disk and may create "fountains" or winds at galactic scales (Chevalier and Oegerle 1979; Bregman 1980).

These important features are not exclusively created by expanding SN shells, however. Stellar winds also induce strong shocks with their corresponding effects (Weaver et. al. 1977; Abbott 1982) and, given that massive stars tend to be clustered in OB associations, the combined action of stellar winds and SN explosions results in larger and more energetic interstellar structures (Bruhweiler et. al. 1980). Moreover, this collective energy injection provides an appropriate

---

*Permanent address: Lick Observatory, University of California, Santa Cruz, CA 95064, USA

†Permanent address: Instituto de Astronomia, UNAM, Apdo. Postal 70-264, 04510 Mexico D.F., Mexico

feedback control to the bulk properties of the gaseous disk, including the star forming activity and the mass exchange among the different gas phases (Franco and Shore 1984; Tenorio-Tagle and Palous 1987). Thus, aside from its importance in the study of individual SN shells, the evolution of supernova remnants inside wind-driven shells appears as a fundamental tool in our understanding of the Milky Way and gaseous galaxies.

There exists a large number of analytical and numerical studies with the details of blast waves propagating in different media. These include constant density gas with typical disk conditions (see review by Chevalier 1977), dense and homogeneous molecular clouds (Shull 1980; Wheeler, et. al. 1980), remnants breaking out of molecular clouds (Tenorio-Tagle et. al. 1985) a cloudy interstellar medium (McKee and Cowie 1975, Cowie and McKee 1977), power-law ambient density distributions (Cox and Franco 1981; Edgar and Cox 1984), power-law distributions for both the atmosphere of the progenitor and the ambient medium (Chevalier 1982), and a combination of a shallow power-law medium with a small interstellar shell (Dickel and Jones 1985). However, the evolution inside the wind-driven cavity left by the supernova progenitor has not been considered in detail. Here we present preliminary one-dimensional hydrodynamic simulations of this event. Section II presents a brief review of the evolution of supernova remnants (SNR) and wind-driven shells. A simulation is shown in section III together with a brief discussion of the results.

II. EVOLUTION OF SNRs AND STELLAR WINDS.

The evolution of a SNR and of a wind-driven cavity can always be divided into three phases: a free-expansion phase, an adiabatic phase, and a momentum conserving phase. The time dependence of radius and velocity in the latter two phases depends on the density stratification. The following subsections give some examples of the evolution in different types of ambient media.

a) SN shells in a constant density medium

i) Free-expansion phase: The ejecta are expelled with a large speed, V_s, and develop two strong shocks moving in opposite directions: the leading shock which moves outwards and collects ambient gas and a reverse shock which penetrates and thermalizes the ejecta. The leading shock moves with a constant speed until the swept-up mass becomes dynamically important (i.e. comparable to the mass of the ejecta, M_e) and decelerates afterwards. A reference timescale for this phase is usually selected as the time when the swept-up interstellar matter is equal to M_e

$$t_{fe} \simeq V_s^{-1} \left[\frac{3M_e}{4\pi\rho_o}\right]^{1/3} \simeq 200\ V_9^{-1} M_1^{1/3} n_o^{-1/3}\ \text{yr}, \qquad (1)$$

where $\rho_o = \mu n_o$ is the ambient mass density, $V_9 = V_s/10^9 \text{cms}^{-1}$, and M_1 is the mass of the ejecta in solar masses.

ii) Adiabatic or Sedov phase: After the kinetic energy of the ejecta has been thermalized by the reverse shock, the leading shock remains strong but now it is driven by the high thermal pressure of the hot shocked gas. As long as radiative losses can be neglected, the leading shock follows the standard adiabatic Sedov (1959) solution

$$R_s = \left[\frac{2.02\ E_o}{\rho_o}\right]^{1/5} t^{2/5} = 13\ E_{51}^{1/5}\ n_o^{-1/5}\ t_4^{2/5}\ \text{pc}, \tag{2}$$

$$V_s = \frac{2}{5}\frac{R_s}{t} = 508\ E_{51}^{1/5}\ n_o^{-1/5}\ t_4^{-3/5}\ \text{kms}^{-1}, \tag{3}$$

where E_{51} is the energy of the explosion in units of 10^{51}erg, and t_4 is the evolutionary timescale in units of 10^4yr. A large fraction of the total energy, about 2/3, is in the form of thermal energy in the hot cavity whereas the rest, $\sim$ 1/3, is the kinetic energy of the expanding external shell. This phase comes to an end when cooling becomes important at the approximate time

$$t_{cool} \simeq \frac{3kT}{8n_o\Lambda} \simeq \frac{9\mu V_s^2}{128 n_o \Lambda} \simeq 5.3 \times 10^4\ V_7^2\ n_o^{-1}\ \Lambda_{23}^{-1}\ \text{yr}, \tag{4}$$

where $V_7 = V_s/10^7\text{cms}^{-1}$, and Λ_{23} is the cooling function in units of 10^{-23}erg cm^3s^{-1}.

iii) Momentum conserving phase: As the internal energy is radiated away by the expanding SNR, the outer shell cools down and continues its expansion due to momentum conservation. The momentum gathered during the previous adiabatic phase is now locked in the dense and cold shell, $M_s(t_{cool})V_s(t_{cool})$ = constant, which decelerates rapidly until it approaches the random velocity of interstellar clouds. The radius and velocity during this phase are

$$R_s = \left[R_s^4(t_{cool}) + \frac{3M_s(t_{cool})V_s(t_{cool})}{\pi\rho_o}\,t\right]^{1/4} \propto t^{1/4}, \tag{5}$$

$$V_s = \frac{1}{4}\frac{R_s}{t} \propto t^{-3/4}\ . \tag{6}$$

where time is measured from t_{cool}.

The relative importance of these phases in a given SNR depends on the assumed ambient density. For instance, for densities above $\sim 10^4\text{cm}^{-3}$, such as those expected in molecular clouds, radiative losses are already important by the end of the free-expansion phase and the adiabatic phase may be nonexistent (Wheeler et. al. 1980). For low density ambient media, on the other hand, there may be relatively long transition stages between the three main phases. The transition between the adiabatic and the momentum conserving phases is usually called "pressure-driven snowplow" (Oort 1951), and evolves as $R_s \propto t^{2/7}$.

b) Evolution in a cloudy medium

The properties of a cloudy medium have been approximated by a homogeneous medium filled with small spherical clouds (McKee and Cowie 1975; McKee and Ostriker 1977). The main assumption of this model is that the clouds engulfed by the expanding remnant do not affect the remnant's dynamics, but cloud evaporation results in a density increase of the hot interior. These evaporation-dominated remnants evolve as $R_s \alpha t^{3/5}$ in the adiabatic stage, instead of the $R_s \alpha t^{2/5}$ prescribed by the Sedov solution (i.e. equation 2). Such a time dependence is similar to the one predicted for adiabatic wind-driven shells (see §IIe) and it is also identical with that found for adiabatic blast waves evolving in a power-law medium with $\rho \alpha r^{-5/3}$ (see §IIc). In the late stages (in the transition from adiabatic to momentum conserving) the remnant evolves as $R_s \alpha t^{2/7}$, in agreement with the pressure-driven snowplow (see §IIa).

c) Power-law ambient density

A density gradient continuously modifies the shock conditions but self-similar solutions can also be constructed for power-law stratifications (Sedov 1959). The corresponding adiabatic solutions for $\rho_o = Ar^{-w}$ (A is a constant and r is the distance from the explosion site) have a time dependence (c.f. Cox and Franco 1981).

$$R_s = \left[\left[\frac{3-w}{3}\right] \left[\frac{5-w}{5}\right]^2 \frac{2.03\ E_o t^2}{(1-0.23w)A} \right]^{1/(5-w)} \alpha\ t^{2/(5-w)} . \qquad (7)$$

This solution reduces to eq.(2) for w=0, as expected, and resembles the evaporation-dominated time dependence for w=5/3.

The evolution for w<0 corresponds to radially increasing density gradients, and could be used as a guideline to the study of SNRs evolving in large pre-existing cavities. The cases with positive values of w, on the other hand, can be used to study the effects of the extended atmosphere from the progenitor and the evolution inside centrally condensed clouds (Chevalier 1982). The solutions for the shock evolution in two other geometries, cylindrical and plane parallel, are given for power-law density distributions by Cavaliere and Messina (1976).

d) Effects of an interstellar shell

Recently, Dickel and Jones (1985) have performed numerical simulations including the effects of a small interstellar shell surrounding the SN progenitor. Their model, which is specifically designed to reproduce the radio and x-ray emission of Tycho's supernova remnant, shows one important new feature. As the leading shock encounters the interstellar shell a fraction of its energy is transmitted into the shell while the rest bounces back into the hot interior in a reflected shock

wave. The reflected wave modifies the internal pressure structure and generates secondary reflections.

e) Stellar winds.

The evolution of a wind-driven bubble follows the same phases described above although, instead of the impulsive energy injection, now there is a continous energy and momentum injection. The density structure of the bubble is rather complex but there are three well distinguished zones: i) a free-expanding wind zone, ii) a shocked-wind region, which is the hottest part and contains most of the mass ejected by the star, and iii) a dense outer shell which contains shocked interstellar gas. The shocked-interstellar and the shocked-wind regions are separated by a contact discontinuity. The detailed evolution of all these regions in a homogeneous medium has been described by Weaver et. al. (1977). In the case where the interior evolves adiabatically but cooling is included in the dense shell, the solution for the outer interstellar shock, which corresponds to the leading shock in the SN case, is

$$R_s \simeq 27\ L_{36}^{1/5}\ n_o^{-1/5}\ t_6^{3/5}\ \text{pc}, \qquad (8)$$

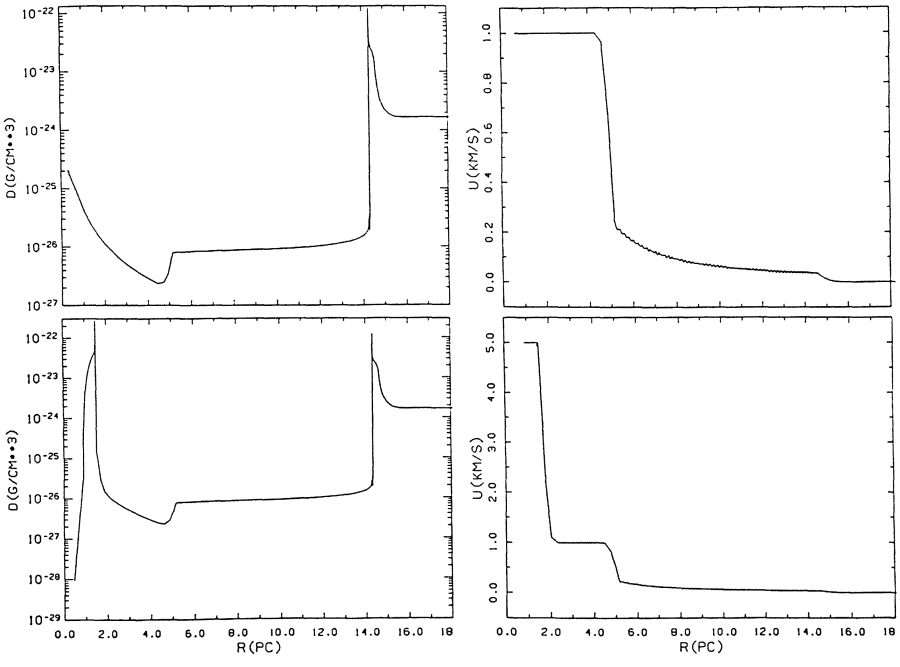


Fig. 1. The initial conditions. Density (left) and velocity (right; in units of 10^3kms^{-1}) are plotted as a function of radius. Upper frames: the wind solution after a time of 2.2×10^5yr. Lower frames: the solution shortly after the injection of the supernova energy.

where $L_{36} = \frac{1}{2}\dot{M}_\omega V_\omega^2$ is the wind luminosity in units of 10^{36}erg s^{-1}, and t_6 is the evolutionary timescale in units of 10^6yr. The density and velocity structures of one of these wind-driven bubbles, with $\dot{M}_\omega = 10^{-5}\ M_\odot yr^{-1}$ and $V_\omega = 10^3 kms^{-1}$ at $t = 2.2 \times 10^5$yr, is presented in Figure 1 (upper frames).

There the structure of the wind cavity, the inner and outer shocks, and the contact discontinuity can easily be recognized. In the free wind region the density drops as R^{-2} and has a constant velocity ($10^3 kms^{-1}$). The inner shock then decelerates the flow while thermalizing the wind, leading to higher densities and lower velocities. The shocked wind region extends up to the contact discontinuity where the densest part of the thin, cool shell of interstellar swept up material lies. The outer shock provides the swept up matter with a velocity of the order of 40 kms^{-1}.

III. THE TIME EVOLUTION

a) The initial conditions

The structure caused by the strong stellar wind (Fig. 1) was used as the background medium to be encountered by the ejecta from a supernova explosion. The model assumes that the supernova progenitor had a short wind phase (2.2×10^5yr) immediately after which it exploded as a supernova. One can show that the photoionization caused by the star will not affect the density distribution caused by the wind; thus during the evolution of the wind phase photoionization was not taken into account. The explosion itself was simulated by the insertion of 4 $M_\odot$ with a velocity of $5 \times 10^3 kms^{-1}$ into a small volume at the center of the grid (see Fig. 1, lower frames). The energy of the explosion ($E_{SN} = 10^{51}$erg) is thus initially assumed to be in the form of kinetic energy. The wind and supernova phases were simulated with a spherically symmetric Lagrangian hydrodynamic calculation, including interstellar cooling. The numerically calculated positions of the inner and outer wind shocks as a function of time matched well with the analytic solution in a constant density medium (Weaver et al. 1977; see eq. 8). Other tests showed that the code agreed with the analytic solutions for a supernova explosion in a constant density medium, both in the Sedov phase and in the cooling phase.

b) A supernova evolving in a wind-driven cavity

The expansion of the ejecta leads to the formation of a strong shock which quickly propagates through the entire free-wind region and, about 10^3yr after the explosion, interacts with the inner shock produced by the former wind. After the interaction a transmitted shock moves rapidly into the shocked wind region and a reverse shock travels with and decelerates the ejecta. The sequence of plots in Figure 2 shows the steady progression of the ejecta toward the shell of swept-up

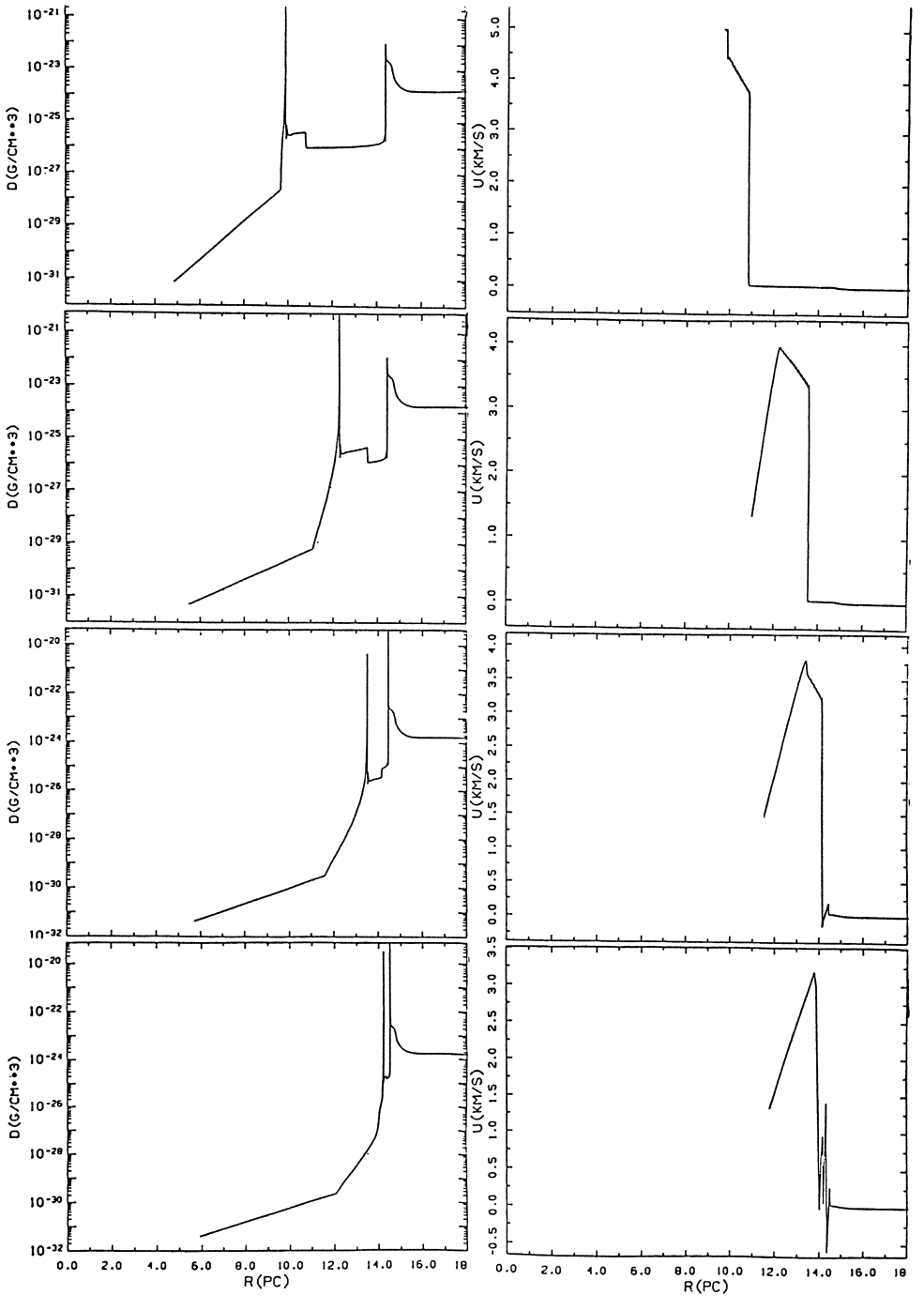


Fig. 2. Density and velocity (in units of $10^3 kms^{-1}$) as a function of radius at the following times during the evolution of the supernova remnant: (1.9, 2.4, 2.7, and 2.9) x 10^3 yrs after the explosion.

interstellar matter. When the transmitted shock reaches this shell it generates a reflected shock and a secondary transmitted shock, which moves into the dense shell. The latter is noticeable at this stage only in the velocity plot (3^{rd} panel), while the reflected shock affects both the density and velocity profiles as it moves through the shocked wind gas towards the ejecta.

Eventually the reflected shock does meet the ejecta, and there it is again reflected and moves toward the outer shell. The ejecta are further decelerated by a backward-facing shock that strengthens the first reverse shock. This process repeats; as the ejecta approach the slowly moving outer shell the original transmitted shock is reflected several times between them. An indication of these many reflections is provided by the complex velocity field immediately behind the dense outer shell (Fig. 2, last panel). The reflections lead to the thermalization of the ejecta and thereby to a large pressure gradient that generates large negative velocities and a rapid flow of both ejecta and wind matter back toward the explosion site. Figure 3 shows several views of this inward motion. Note that the dense outer shell continues to move outwards and that eventually it presents a double density maximum. This effect is caused by the propagation of the original transmitted shock through the dense shell toward the leading outer shock.

The total amount of kinetic energy involved in the backward-moving flow ($\simeq 0.6$ E_{SN}) is rather large. It is eventually thermalized across a shock front that forms as the flow reflects at the grid origin. Once more, thermalization leads to a large pressure gradient and this time to a net outward motion (Figure 4).

The flow now resembles that near the beginning of the evolution in the sense that many shocks and shock reflections appear as the outward kinetic energy becomes thermalized upon encountering the complex density distribution. The conversion of thermal energy into kinetic energy and vice versa, leading to inward and outward motions, repeats itself several times. The total energy gradually drops as some energy is radiated away. After every bounce the outer shell is further compressed, and it acquires a maximum speed of less than $200 kms^{-1}$ after the original transmitted shock catches up with the original outer wind shock.

The preliminary results just presented will be extended to cover the parameter space. In this respect one of the most important parameters is the initial location of the shell of swept up matter, or the amount of mass swept up by the wind compared to the mass of the supernova ejecta. Massive shells, as shown above, force a rapid transition into momentum conservation while a significant amount of energy is radiated away after every bounce. Such an evolution may explain the filamentary appearance of supernova remnants. A complete report, together with detailed astrophysical implications of the model, is now in preparation.

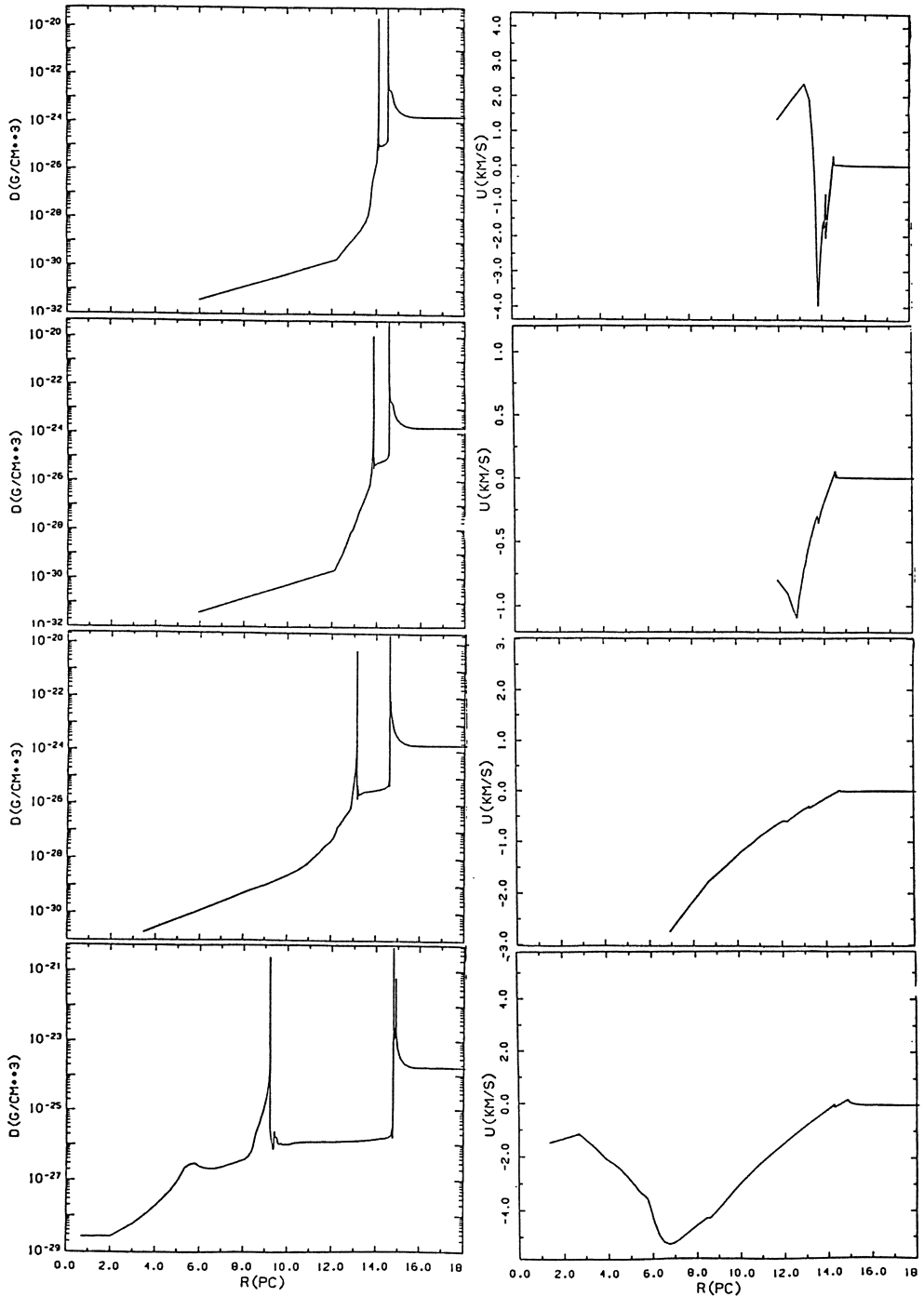


Fig. 3. Density and velocity (in units of $10^3 kms^{-1}$) as a function of radius at the following times during the evolution of the supernova remnant: (3.1, 3.2, 3.5, and 4.5) x 10^3 yr after the explosion.

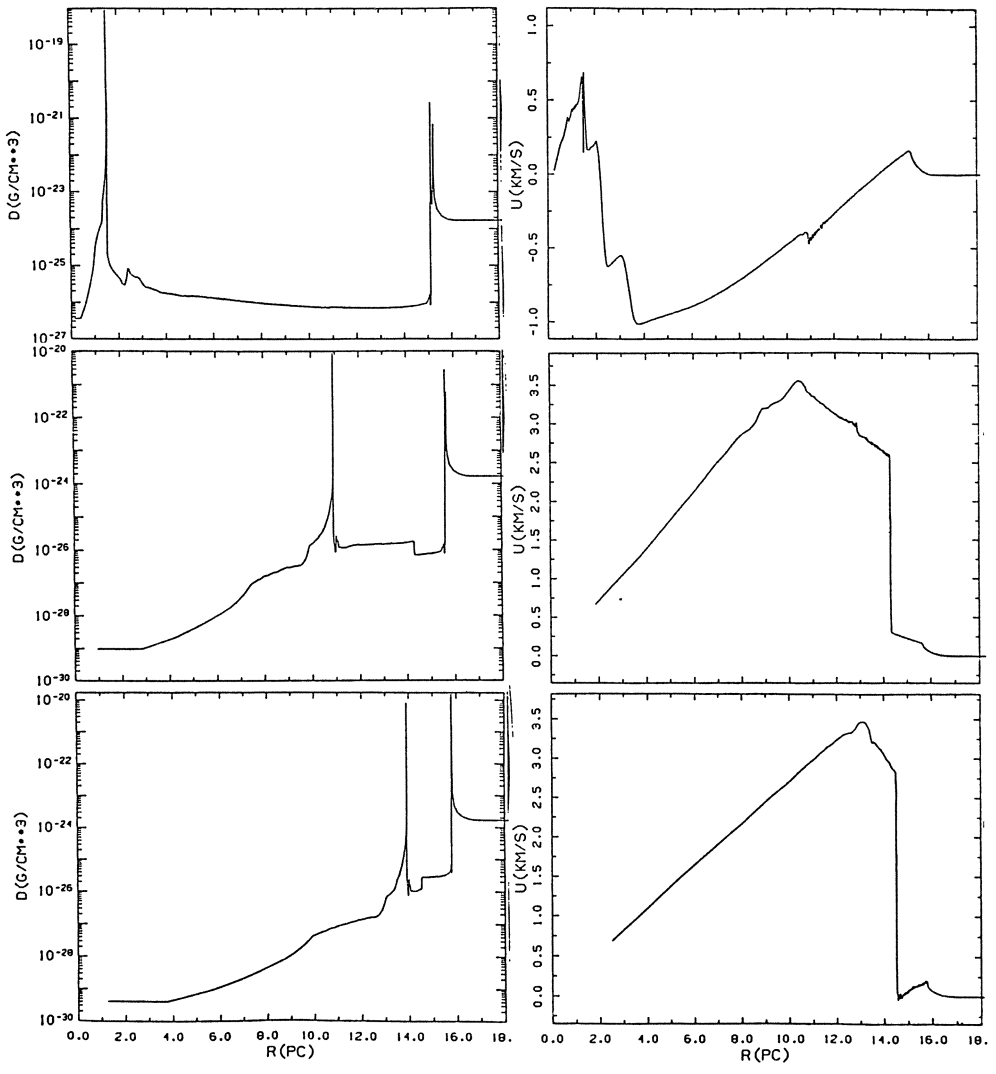


Fig. 4. Density and velocity (in units of $10^3 kms^{-1}$) as a function of radius at the following times during the evolution of the supernova remnant: (6.8, 9.5, and 10.4) x 10^3yr. after the explosion.

Acknowledgements

J.F. thanks the Max-Planck-Institut für Astrophysik, Garching and CONACYT-Mexico for financial support. G.T.-T. is supported by the Deutsche Forschungsgemeinschaft grant M0416. P.B. thanks the Alexander von Humboldt Foundation for a U.S. Senior Scientist Award and the Max-Planck-Institut für Physik und Astrophysik, Garching, for their hospitality.

REFERENCES

Abbott, D. 1982, Ap.J., 263, 723.

Blandford, R.D., and Ostriker, J.P. 1978, Ap.J. (Letters), 221, L29.

Bregman, J. 1980, Ap.J., 236, 577.

Bruhweiler, F.C., Gull, T.R., Kafatos, M., and Sofia, S. 1980, Ap.J. (Letters), 238, L27.

Cavaliere, A., and Messina, A. 1976, Ap.J., 209, 424.

Chevalier, R.A. 1977, Ann. Rev. Astron. Ap., 15, 175.

_____1982, Ap.J., 258, 790.

Chevalier, R.A., and Oegerle, W. 1979, Ap.J., 227, 398.

Cowie, L.L., and McKee, C.F., 1977, Ap.J., 211, 135.

Cox, D.P., and Franco, J. 1981, Ap.J., 251, 687.

Cox, D.P., and Smith, B.W. 1974, Ap.J. (Letters), 189, L105.

Dickel, J.R., and Jones, E.M. 1985, Ap.J., 288, 707.

Edgar, R.J., and Cox, D.P. 1984. Ap.J., 283, 833.

Franco, J., and Shore, S.N. 1984, Ap.J., 285, 813.

Herbst, W., and Assousa, G.E. 1977, Ap.J., 217, 473.

McCray, R., and Snow, T.P. Jr. 1979, Ann. Rev. Astron. Ap., 17, 213.

McKee, C.F., and Cowie, L.L. 1975, Ap.J., 195, 715.

McKee, C.F., and Ostriker, J.P. 1977, Ap.J., 218, 148.

Oort, J.H. 1951, Problems of Cosmical Aerodynamics, Dayton, Ohio: Central Documents Office.

Sedov, L.I. 1959, Similarity and Dimensional Methods in Mechanics, London: Academic Press

Shull, J.M. 1980, Ap.J., 237, 769.

Tenorio-Tagle, G., and Bodenheimer, P. 1988, Ann. Rev. Astron. Ap., in press.

Tenorio-Tagle, G., Bodenheimer, P., and Yorke, H.W. 1985, Astron. Ap., 145, 70.

Tenorio-Tagle, G., and Palous, J. 1987, Astron. Ap., 186, 287.

Weaver, R., McCray, R., Castor, J., Shapiro, P., and Moore, R. 1977, Ap.J., 218, 377.

Wheeler, J.C., Mazurek, T.J., and Sivaramakrishnan, A. 1980, Ap.J., 237, 781.

Inverse Compton Scattering (ICS) Plays an Important Role in Pulsar Emission

G.J. Qiao

Department of Geophysics, Peking University, People's Republic of China

Summary: Theoretical and observational evidence shows that the ICS plays a very important role in pulsar emission. Here we present some relevant points.

(1)*The Energy loss of $e^{\pm}$ by ICS in a thermal radiation field limits the value of the Lorentz factor γ of $e^{\pm}$.

(2)*ICS is an efficient mechanism to produce γ-rays and the observed micro wave radiation.

(3) **Absorption features of radio pulse profiles indicate that ICS plays an important role in the radio emission of pulsars.

(4)*ICS gives a reasonable explanation for core emission.

(5)The optical spectra of PSR 0531+21 can be fitted by ICS naturally.

(*Strong evidence. **Strongest evidence)

1. Introduction

Inverse Compton Scattering (ICS) has been known as an efficient mechanism for the generation of high energy photons in the scattering of low-energy photons by relativistic electrons (or positrons). In the theory of pulsar emission, CHENG & RUDERMAN [1] present an ICS model for the optical emission of the Crab pulsar. But up to now, it has not been recognized as an important mechanism in pulsar emission. In fact, there is some evidence that ICS plays a very important role in pulsar emission. Some of this evidence is presented in this paper.

2. Energy Loss Limits the Lorentz Factor γ of $e^{\pm}$

It is well known in pulsar emission theory that $e^{\pm}$ have high γ values. For example, in the RS Model (RUDERMAN and SUTHERLAND, [2]) $\gamma \sim 10^6$ to 10^7.

But this is questionable. If the energy gain is not greater than the energy loss in the gap, $e^{\pm}$ can not be accelerated to higher γ values.

2.1 The Energy of $e^{\pm}$ Obtained in the Gap

According to the RS model, in the limit of very small h (the thickness of the gap) i.e. $h \ll r_p$, the potential $\Delta\phi$ along a field line traversing the gap is

$$\Delta\phi = 2\pi \rho_e h = \Omega_* B_s h^2 C^{-1} = (1.6\times10^{12}\ \text{Volts})\ B_{12}^{1/2}\ P^{-1/2}\ S_6^{4/7} \qquad (1)$$

Here ρ_e is the positive charge density, which was in the gap region before the gap was formed and is now held by the surface. $B_{12}=B_0/10^{12}$, B_0 is the magnetic field strengths, in Gauss. P is the period of the pulsar, S is the field line curvature, in cm, and $S_6=S/10^6$. Ω_* is the angular velocity of the neutron star.

$$\gamma_p << h < R, \tag{2}$$

where R is the radius of the neutron star, γ_p is the radius of the polar cap. The potential difference between bottom and top of the gaps is (see RS)

$$\Delta\phi = \frac{1}{2C}\Omega_* B_0 (\gamma_p^2 - \gamma_\perp^2) \tag{3}$$

Here $r_\perp = r\sin\theta$ (see Fig. 2).

In fact, if the angular velocity of the isolated magnetosphere polar-column region Ω is a constant, we have

$$\Delta\phi = \frac{1}{2C}\Delta\Omega B_0 \gamma_p^2 \left(1 - \frac{\gamma_\perp^2}{\gamma_p^2}\right) \tag{4}$$

Where $\Delta\Omega=\Omega_*-\Omega$, $(\Omega \leq \Omega_*)$. We find that $\Delta\Phi$ reaches its maximum value $\Delta\Phi_{max}$ at the center of the gap, $\Delta\phi_{max}$, $\Delta\phi_{max} \sim \frac{1}{2C}\Delta\Omega B \gamma_p \sim 10^{12}$ Volts (as $\Delta\Omega \sim \Omega_*$). From the center to the edge of the gap, the value of $\Delta\Phi$ falls down rapidly. The power of $e^\pm$ obtained in the gap is

$$P_{ob} = P_{ob} = e\Delta\phi/\tau \approx \frac{e}{h/c} \tag{5}$$

$$P_{ob} \leq 10^6 \left(1 - \frac{\gamma_\perp^2}{\gamma_p}\right) \text{ erg/s, when } h=10^4\text{cm}, \ \Delta\Phi_{cen}=10^{12} \text{ Volts}$$

$$P_{ob} \leq 10^4 \left(1 - \frac{\gamma_\perp^2}{\gamma_p}\right) \text{erg/s, as } h=10^4\text{ cm}, \ \Delta\Phi=10^{10} \text{ Volts.}$$

2.2 The Energy Loss of $e^\pm$ by ICS in the Thermal Radiation Field in the Gap

The electrons or positrons lose their energy by ICS with the thermal photons in the gap. This energy loss of $e^\pm$ was calculated by XIA, QIAO et al. [3], the result is given in Table 1. Comparing the values in Table 1 and the values given in Eq. (5) we can see that if the temperature in the gap is not too low, the energy loss of $e^\pm$ must limit further acceleration.

Table 1. $B=0.1B_q$. The power loss of an electron in an isotropic thermal radiation field (I) and in the thermal radiation field for $\varphi=60^\circ$ (II) and $\varphi=30^\circ$ (III) respectively. Here $B_q=4.414\times10^{13}$ Gauss. φ is the angle of the radiation cone.

	γ	10	50	100	200
	T(K)	P(erg/s)			
I	$5\,10^4$	0.20e−9	0.30e−8	0.86e−7	0.79e−6
	10^5	0.12e−8	0.34e−6	0.30e−5	0.25e−2
	$5\,10^5$	0.12e−4	0.97e0	0.21e3	0.26e4
	10^6	0.38e−3	0.85e3	0.10e5	0.33e5
	$5\,10^6$	0.21e5	0.10e7	0.15e7	0.17e7
	10^7	0.10e7	0.60e7	0.66e7	0.72e7
II	$5\,10^4$	0.14e−9	0.32e−10	0.53e−9	0.78e−8
	10^5	0.32e−9	0.37e−8	0.31e−7	0.89e−6
	$5\,10^5$	0.14e−5	0.45e−4	0.94e−3	0.27e2
	10^6	0.30e−4	0.38e−2	0.10e3	0.10e5
	$5\,10^6$	0.12e0	0.51e6	0.15e7	0.17e7
	10^7	0.11e5	0.60e7	0.64e7	0.47e7
III	$5\,10^4$	0.96e−10	0.10e−11	0.16e−12	0.18e−10
	10^5	0.22e−9	0.24e−11	0.13e−9	0.17e−8
	$5\,10^5$	0.15e−8	0.37e−6	0.26e−5	0.31e−4
	10^6	0.97e−6	0.14e−4	0.13e−3	0.14e−1
	$5\,10^6$	0.41e−2	0.18e2	0.31e5	0.68e6
	10^7	0.11e0	0.12e6	0.26e7	0.69e7

3. ICS is An Efficient Radiation Mechanism

3.1 ICS is Very Efficient in Producing γ-rays

The energy loss of $e^{\pm}$ discussed above would associate with the production of γ-rays. The energy of the outgoing photons in ICS is (QIAO et al. [4]):

$$\hbar\omega'^{L}=\frac{1}{\mathrm{Sin}^2\theta'^{L}}\left[\gamma mC^2(1-\beta\mathrm{Cos}\,\theta'^{L})+\hbar\omega^{L}(1-\mathrm{Cos}\,\theta^{L}\,\mathrm{Con}\,\theta'^{L})-\left\{\left[\gamma mc^2(1-\beta\mathrm{Cos}\,\theta'^{L})+\hbar\omega^{L}(1-\mathrm{Cos}\,\theta^{L}\mathrm{Con}\,\theta'^{L})\right]-\mathrm{Sin}^2\theta'^{L}\left[\hbar\omega^{L2}\mathrm{Sin}^2\theta^{L}+2\gamma mc^2\cdot\hbar\omega^{L}(1-\beta\mathrm{Cos}\,\theta^{L})\right]\right\}^{1/2}\right] \quad \text{when } \mathrm{Sin}\,\theta'^{L}=0 \qquad (6,a)$$

$$\hbar\omega'^{L}=\frac{1+\beta}{2}\gamma\frac{\hbar^2\omega^{L2}\mathrm{Sin}^2\theta^{L}+2\gamma mc^2\hbar\omega^{L}(1-\beta\mathrm{Cos}\,\theta^{L})}{mc^2+\gamma\hbar\omega^{L}(1+\beta)(1-\mathrm{Cos}\,\theta^{L})} \quad \text{when } \theta'^{L}=0 \qquad (6,b)$$

$$\hbar\omega'^{L}=\frac{\hbar^2\omega^{L2}\mathrm{Sin}^2\theta+2\gamma mc^2\hbar\omega^{L}(1-\beta\mathrm{Cos}\,\theta^{L})}{2\left[\gamma mc^2(1+\beta)+\hbar\omega^{L}(1+\mathrm{Cos}\,\theta^{L})\right]} \quad \text{when } \theta'^{L}=\pi \qquad (6,c)$$

Where "L" represents the Lab. frame. θ (or θ') is the angle between the incoming (or outgoing) photons and the direction of the magnetic field. $\hbar\omega$ ($\hbar\omega'$) is the energy of incoming (outgoing) photons $\beta=V/C$. V is the velocity of $e^{\pm}$, m is the rest mass of $e^{\pm}$.

In our case $\gamma>>1$, and $\theta^{\perp}\backsim\pi$, $\theta^{\parallel}\backsim 0$, we have $\hbar\omega'^{\perp}\sim 4\gamma'^{2}\hbar\omega^{\perp}$.

If the termal radiation is coming from a black body of temperature T, so $\hbar\omega'^{\perp}\sim 4\gamma'^{2}\hbar\times 3.7\times 10^{//}$ T. When $T=10^{6}K$, $\gamma=10^{2}$, $\omega'^{\perp}\sim 1.6\times 10^{22}(s^{-1})$; When $T=10^{6}K$, $\gamma=2\times 10^{2}$, $\omega'^{\perp}\sim 6.4\times 10^{22}(s^{-1})$, and $\hbar\omega'^{\perp}>2mc^{2}$. It means that these γ-rays can produce $e^{\pm}$ pairs in strong magnetic field. To produce such γ-rays through ICS the Lorentz factor γ of $e^{\pm}$ does not have to reach $\gamma\backsim 10^{6}$, just $\gamma\gtrsim 10^{2}$ is enough.

3.2 ICS is an Efficient Mechanism to Produce the Observed Microwave Radiation

A. Low Frequency Wave

According to the RS model the sparking must take place in the inner gap. It would associate with a low frequency wave, $\omega\backsim 6\times 10^{5}s^{-1}$. The amplitude is very large ($\backsim\Delta\phi/e$).

B. The Production of the Observed Radio Wave by ICS

The radio luminosity produced in ICS by $e^{\pm}$ with the low frequency wave is (QIAO, [5]):

$$L_{radio}\cong\frac{\alpha_1\alpha_2}{2(4\pi)}\frac{\eta\omega'}{\gamma^{3}mc\omega}\sigma_{th}\frac{(\Delta\phi)^{4}}{h}\approx\alpha_1\alpha_2\eta\cdot 10^{37}\ erg/s \qquad (7)$$

Where $\eta\geq 1$, $\alpha_1<1$, $\alpha_2<1$, $\omega'=\gamma^{2}\omega$, $\Delta\phi\backsim 10^{12}V$, $\omega\backsim 6\times 10^{5}S^{-1}$, $\gamma\backsim 10^{2}$. For most pulsars, $L_{radio}<10^{30}erg/s$. So this is a very efficient mechanism to produce the observed radio wave.

3.3 Curvature Radiation (CR) is an Inefficient Mechanism Which Does Not Play Such an Important Role As It Does in the RS Model

A. CR Is an Inefficient Mechanism to Produce γ-Ray Photons in Inner Gap

The energy loss by CR in the inner gap is $P_{cur}\cdot h/c$, and $f_{cr}=(P_{cur}h/c)/e\Delta\phi=2\gamma^{4}e^{2}h/3s^{2}e\Delta\phi$. Where s is the field line curvature. As $\gamma\backsim 10^{6}$, $h\backsim 10^{4}cm$, $s\backsim 10^{6}cm$, $\Delta\phi\backsim 10^{12}$ Volts, we have $f_{cr}\backsim 10^{-3}$.

Compare as we discussed in section 2 and section 3 for ICS $f_{ICS}=P_{ICS}\cdot h/e\Delta\phi c\backsim 1$.

B. CR Is an Inefficient Mechanism to Produce the Observed Micro Wave

Owing to the relativistic beaming effect, an observer will receive only a pulse of radiation from a given particle corresponding to its passage along an arc of angular extent $1/\gamma$, whose tangent is toward the observer. The observed time scale is $dt\backsim s/\gamma^{3}c$, and the ratio f_{CR}, $\Gamma=P_{cur}dt/\gamma mc^{2}=2e^{2}/smc^{2}\backsim 10^{-19}$. Here we take $s\backsim 10^{6}cm$. Compare with this the power of ICS is $P_{ICS}=\sigma_T(1)Cn_{\omega}\hbar\omega'$, and

$$f_{ICS,T}=\frac{P_{ICS}\,dt}{\gamma mc^{2}}=\alpha_1(\Delta\phi)^{2}\frac{\sigma_{th}}{4\pi h^{2}\gamma^{4}mc^{2}}\sim 10^{-5}\alpha_1 \qquad (8)$$

$$\frac{f_{CR,T}}{f_{ICS,T}} \sim 10^{-14} \cdot \frac{1}{\alpha_1} \qquad (9)$$

Here $\alpha_1<1$, but of order 1, σ_{th} is the Thomson cross section. Compared with ICS, CR is a very inefficient mechanism to produce the observed radio emission of pulsars and γ-rays in the inner gap.

4. Absorption Features of Radio Pulse Profiles Show That ICS Plays an Important Role in the Production of the Observed Microwave Radiation

4.1 Absorption Features of Radio Profiles

The narrowing of pulse widths becomes slower at high frequencies and even reversed for some pulsars (see Fig. 1). This phenomenon, called "absorption feature", was reported by some authors [7-10] and recognized by Rankin [11] to be a significantly new and truly general property of pulsar emission profiles.

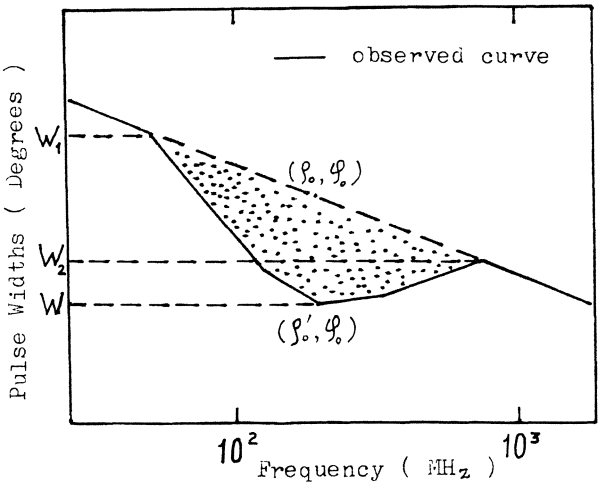


Fig. 1. The "absorption feature" of pulsar emission. $W_1=2\Theta_1$, $W_2=2\Theta_2$

4.2 An ICS Explanation for "Absorption Features"

The cross section of ICS is as follows (QIAO et al. [4]) (For $B/B_g<<1$, $\hbar\omega/mc^2<<1$):

$$\sigma_T^L(1)=(1-\beta\cos\theta^L)\,\sigma_{th}\left\{\sin^2\theta+\frac{1}{2}\cos^2\theta\left[\left(\frac{\omega}{\omega+\omega_B}\right)^2+\left(\frac{\omega}{\omega-\omega_B}\right)^2\right]\right\}^R \qquad (10)$$

$$\sigma_T^L(2)=\frac{1}{2}(1-\beta\cos\theta^L)\,\sigma_{th}\left[\left(\frac{\omega}{\omega+\omega_B}\right)+\left(\frac{\omega}{\omega-\omega_B}\right)\right]^R \qquad (11)$$

$$\sigma_T^L=\frac{1}{2}(1-\beta\cos\theta^L)\,\sigma_{th}\left\{\sin^2\theta+\frac{1}{2}(1+\cos^2\theta)\left[\left(\frac{\omega}{\omega+\omega_B}\right)^2+\left(\frac{\omega}{\omega-\omega_B}\right)^2\right]\right\}^R \qquad (12)$$

Where "R" and "L" represent the rest frame of electron and the Lab frame respectively, here we indicate by "1" the linear polarization of the photon parallel to

the plane defined by the magnetic field and the incoming photon, and by "2" the linear polarization orthogonal to this plane. $B_q=4.414\times10^{13}$ Gauss. ω_B is magnetic gyrofrequency. For θ equal to zero, the cross section is equal to zero too, so for each magnetic field line there must be a point at which ICS does not take place, giving rise to an "absorption feature". (See Fig. 2.) $\omega^R=\gamma\omega^L(1-\beta\cos\theta^L)<<\omega_B$, and $\sin^2\theta^R=\sin^2(\theta^R\pm\pi)=\sin^2\theta^L/\gamma^2(1\pm\beta\cos\theta^L)^2$. From (16) we have $\sigma_1(2)=0$, and (in the Lab frame):

$$\sigma_T(1)=\frac{\sin^2\theta}{\gamma^2(1\pm\beta\cos\theta)^2}=\begin{cases}0 & \text{For } \theta=0\\ \sigma_{th}/\gamma^2 & \text{For } \theta=\pi/2\end{cases} \qquad (13)$$

For a dipole field line, $\rho_e=r_e\cos^2\phi$. At a point $p(\rho_0,\ \phi_0)$ the tangential Eq. is

$$\rho=\frac{\rho_0^2}{\rho_0\cos(\varphi-\varphi_0)-\rho'\sin(\varphi-\varphi_0)}=\frac{r_e\cos^3\varphi}{\cos\varphi_0\cos(\varphi-\varphi_0)+2\sin\varphi_0\sin(\varphi-\varphi_0)} \qquad (14)$$

Here $\rho'=\partial\rho/\partial\phi$. If R=AB (in Fig. 2) we have:

$$\frac{R'}{r_e}=\frac{\cos^3\varphi_0}{\cos\varphi_0\cos(\pi/2-\varphi_0)+2\sin\varphi_0\sin(\pi/2-\varphi_0)}=\frac{\cos^2\varphi_0}{3\sin\varphi_0} \qquad (15)$$

$$\operatorname{tg}\varphi_0=\frac{1}{4}\left[3\operatorname{ctg}\theta_0+\sqrt{9\operatorname{ctg}^2\theta_0+8}\right] \qquad (16)$$

From observational data we can find the value of $\theta_1=\frac{1}{2}W_1$, $\theta_2=\frac{1}{2}W_2$, and estimate the positions of points $p_0(\rho_0,\phi_0)$, and (ρ',ϕ') (Fig. 2). At the inner field lines, $\theta=\frac{1}{2}W$, using (16) we get the value of ϕ.

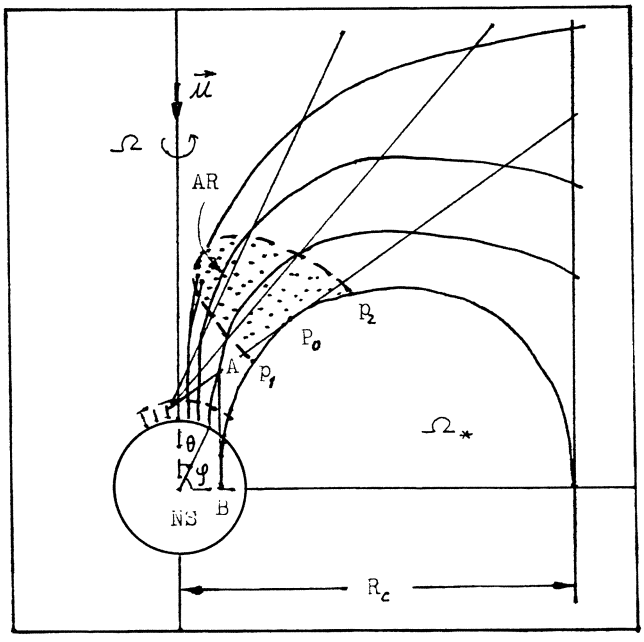


Fig. 2. The "absorption region" AR is the absorption region

From the calculation we fined that: A. ICS can give a reasonable explanation for an "absorption feature". The "absorption region" can be calculated quantitatively. B. The location of the radiation region for most pulsars are very close: $\backsim 10^7$vm to 10^8cm (to the cente of the neutron Star).

5. ICS Can Give a Reasonable Explanation for Core Emission

There exist two distinct physical types of pulsar emission, the cone emission and the core emission. The core emission has been recognized by RANKIN [12,13] as an essential feature of pulsar emission, but the current theories of pulsars cannot give any understanding for it. QIAO [5] presents a model of ICS which can give a reasonable explanation for both core and cone emission, especially for : (1) geometry; (2) subpulse modulation properties; (3) Spectra; (4) absorption features; (5) polarization.

5.1 Core Emission and Cone Emission

The core emission is an essential feature of pulsar emission. All evidences show it is in some sense prior to that of the cone. The characters are as follows: A. Position. the cone emission region is near the center and the hollow cone surrounds the core. The core emission comes from a much lower region near the surface of the star. B. Spectrum. The conal emission exhibits a somewhat flatter spectrum than that of the core emission. C. Polarization. The core emission exhibits circular polarization. The emission from the hollow cone accounts for the characteristics S-shaped polarization angle swing. D. Subpulses. There is no evidence of periodic subpulse modulation in the core emission, but the hollow-cone beam exhibits drifting subpulses.

5.2 The Production of Core and Conal Emission by ICS

A. The Position and the Drifting of Sparks The discharges take place in the region near the edge of the inner gap. the occurrence of sparks depends on two things: the energy of the γ-rays must exceed $2mc^2$ and B must be large enough.

The high energy pairs $e^{\pm}$ can produce such γ-rays in two ways: CR of ICS.

For CR, $Eph \backsim 1.5\gamma^3$ C/S, as $s \backsim 10^6$cm, for $Eph \geqslant 2mc^2$, we need $\gamma \geqslant 1.5\times10^6$. For ICS, $Eph \backsim 4\gamma^2 h\omega$, we need $\gamma \geqslant 0.4\times10^2$. Owing to: (1). From the analysis of "absorption features" we find that the conal emission come from the region very close to the outest open field lines. (2). The value of B is not large enough in the center of the gap, which is also shown that the spark would not take place near the center of the gap. (3). The potential difference in the gap fall down from the center to the edge, near the edge $\Delta\phi$ is too small to give $\gamma \geqslant 1.6\times10^6$. (4). As we pointed out in section 2, ICS would limit the value of γ. We conclude that ICS would produce such γ-rays.

The mean free path ℓ of γ-ray photons can be estimated as (SUTHERLAND, [6]):

$$\bar{\ell} = \frac{4.4}{(e/\hbar c)}\left(\frac{\hbar}{mc}\right)\left(\frac{B_\ell}{B_\perp}\right)\exp\left(\frac{4}{3x}\right) \qquad \text{for } x<<1$$

$$\bar{\ell} = \frac{3.4}{(e^2/\hbar c)}\left(\frac{\hbar}{mc}\right)\left(\frac{B_\ell}{B_\perp}\right)x^{1/3} \qquad \text{for } x>>1 \qquad (17)$$

$$X = \frac{h\nu \sin\theta \cdot B}{2mc^2 \quad B_\ell} = \frac{h\nu_\perp \quad B}{2mc^2 \quad B} = \frac{h\nu \quad B_\perp}{2mc^2 \quad B_\ell} \qquad (18)$$

For $h\nu_\perp = 2mc^2$, we have $X \backsim 1/15$.

From (17) and (18), we can see that ℓ is decreasing when B increases. For dipole magnetic field lines, $B_\perp \propto \frac{1}{s} \propto \sin\varphi$, this means that from the center to the edge of the gap B is increasing, at the center B $\backsim 0$ and $\ell \to \infty$.

Thus the position of sparking would preferably be very close to the edge of the gap.

The subpulse drifting would take place in the conal emission. We have (QIAO, [5])

$$\frac{P_3}{P} = \frac{2\, r_p\, hB}{C\Delta\phi(r_\perp)} \cdot \frac{1}{P} \qquad (19)$$

Here P_3 is the subpulse drifting period. For the typical value of $r_p \backsim 5\times10^4$cm, $h \backsim 5\times10^3$cm, $B \backsim 10^{12}$ Gauss, $\Delta\phi(r_p) \backsim 10^{10}$ Volts, and $P \backsim 1$sec. we get: $\frac{P_3}{P} \backsim 5$. This is a typical value of P_3 from observations.

B. The Production of Core Emission

Near the center of the inner gap there are some high energy positrons with a Lorentz factor of about $\backsim 10^2$ (see Section 2). In ICS between these positrons and the low frequency radio wave produced in the sparking, the photons with $\omega' \backsim \gamma^2\omega \backsim 6\times10^9 s^{-1}$ would be produced. These outgoing photons are the core emission. In the center, the magnetic field lines are parallel to the line of sight, and in the cold plasma approximation, this outgoing radio wave has circular polarization.

C. Conal Emission

The outgoing positrons produced in the sparking move out along the magnetic field lines. ICS would take place between these positrons and the low frequency wave produced in the sparking. This would give the hollow cone emission. In the cold plasma approximation, the radiation is linearly polarized, and presents an S-shaped polarization angle swing.

Owing to the sparking in this region, the moving out $e^\pm$ is pulsating and drifting, so drifting subpulse will be observed in the radiation.

6. Optical Spectra of the Crab Pulsar Can Be Fitted by ICS Naturally

CHENG and RUDERMAN [1] suggested that the optical radiation of the Crab Pulsar comes from ICS of radio emission scattered by the secondary $e^\pm$ near the outer gap. Because the radio emission of the Crab pulsar is very unstable, the outgoing

optical photons produced in ICS would be unsteady. but the observed optical radiation is very stable. QIAO et al. [4] suggested that in the model of CHENG and RUDERMAN the low frequency wave produced in the sparking (Section 3) instead of the observed radio radiation, is scattered by secondary $e^{\pm}$ in ICS near the outer gap. This model can give the main characteristics of the optical radiation: Luminosity, polarization, stability and spectra (ZHU and QIAO [14])

7. Conclusion

The evidence presented here shows that the ICS must play a very important role in pulsar emission. The outline of the RS Model is correct, but the production of γ-rays in the sparking and observed microwave may come from ICS process instead of CR.

Stimulating conversations with many participants of the Workshop on High Energy Astrophysics of Max-Planck-Society--Academia Sinica, July 1987, München. I am grateful to Prof. G. Börner. I thank Max-Planck-Academia Ainica for the support. I thank National Science Foundation of China for the support through a grant No. 1860604.

8. Literature

1. A.F. Cheng, M.A. Ruderman: Astrophys. J. 216, 865 (1986)
2. M.A. Ruderman, P.G. Sutherland: Astrophys. J. 196, 51 (1975)
3. X.Y. Xia, G.J. Qiao, X.J. Wu, Y.Q. Hou: Astron. Astrophys. 152, 93 (1985)
4. G.J. Qiao, X.J. Wu, W. Bao, X.Y. Xia: Scientia Sinics, 7, 732 (1986)
5. G.J. Qiao: Proceedings of 4th Asian-Pacific Regional Meeting of IAU (accepted)
6. P.G. Sutherland: Fundamentals of Cosmic Physics, 4, 95 (1979)
7. A.D. Kuzmin, V.A. Izikove, Yu.P. Shitor, A.G. Lybe, J.G. Davies: IAU Symposium 95, Pulsars, ed. by W. Sieber, R. Wielebinski, Reidel (Dordrecht 1980) p.455
8. N. Bartel, et al.: Astron. Astrophys. 93, 85 (1981)
9. N. Bartel: Astron. Astrophys. 97, 384 (1981)
10. A.D. Kuzmin, V.M. Malofeev, V.A. Izekova, W. Sieber, R. Wielebinski: Astron. Astrophys. 161, 183 (1986)
11. J.M. Rankin: Astrophys. J. 274, 359 (1983)
12. J.M. Rankin: Astrophys. J. 274, 333 (1983)
13. J.M. Rankin: Astrophys. J. 301, 901 (1986)
14. Y.D. Zhu, G.J. Qiao: Submitted to Acta Astronomica Sinica.

Line Emission from Magnetized Neutron Stars and Electron Cyclotron Maser Instability

Wang De Yu[1] *and Mao Din Yi*[2]

[1]Purple Mountain Observatory, Nanjing, People's Republic of China
[2]Hohai University, Nanjing, People's Republic of China

ABSTRACT The Electron cyclotron maser instability might occur above the magnetized accretion column of neutron stars, due to a few electrons with a non-thermal velocity distribution. The coherent emission with frequency close to the electron cyclotron frequency and its second harmonic frequency might be amplified by this instability. In the case of line emission from Her X-1, the growth rates of the instability have been calculated using loss-cone and hollow beam velocity distributions. The growth rates increase proportionally with the electron density in the accretion column, and vary rapidly with emission angle. The observations about the intensity and the energy of emission line correlated with Her X-1 pulse phase have also been explained.

1. Introduction

The hard X-ray spectrum of Her X-1 has been extensively investigated since a rather sharp spectral feature at 58 keV and possibly a second one at ~110 keV were discovered by MPI in 1976 [1]. Further observationand analysis were performed by MIT [2], MPI [3], Saclay [4] etc.. Some new interesting observational results on the spectral features of the Her X-1 source have been reported: (1) The spectral feature depends on the pulse phase of Her X-1, (the phase of 1.24s pulsation) and is particularly pronounced only in a narrow range of the pulse phase [2],[3]. (2) The emission line centroid varies with pulse phase, the measured variation can be represented by a sinusoidal function [3]. These results seem to indicate that the line emission is radiated in a narrow range of angle with respect to the neutron star s magnetic axis. (3) The energy variation of the emission line is correlated with the intensity of the line L_e [4], as well as the continuum luminosity L_c from Her X-1 [4]. (4) The ratio of L_e/L_c varies also proportional with the increase of L_c [4]. It will provide some serious constraints for theoretical models of cyclotron line emission from magnetized neutron stars, if these results are confirmed by further observation.

The existing theoretical models of the line feature from Her X-1 fall into three classes. In the first class the X-ray at 58 keV is interpreted as the electron quantum synchrotron radiation in a magnetic field of 5×10^{12}G [5]. The second one attempts to account for the hard X-ray spectrum in terms of electron cyclotron resonance absorption [6],[7]. In recent years, we have suggested that a population of nonthermal electrons from the accreting column trapped above the magnetic cap co-exists with a denser background plasma. The electron cyclotron maser instability can be excited by these nonthermal electrons, and the radiation

with frequency near the electron cyclotron frequency and its second harmonic frequency might be amplified by this instability. It is responsible for the line emission from magnetized neutron stars.

2. The Electron Cyclotron Maser Instability Theory on the Line Emission from a Magnetized Neutron Star

In fact, the plasma flows from a massive companion through the accretion disk, then falls onto the magnetic polar cap of the X-ray pulsar, forming an accretion column. This accretion column may be divided into an upper and lower part [10]. The influence of the radiation pressure, which is acting on the surface of the neutron star, is negligible in the upper part of the accretion column, the plasma flow there might be considered a "free" fall under the gravitational attraction of the neutron star.

However, a part of the electrons form a hollow beam velocity distribution due to the fact that they corotate with the intense magnetic field of the neutron star; or form a loss-cone velocity distribution due to the magnetic field gradient along the magnetic axis. Therefore, the electron velocity distribution might form a nonthermal distribution in the upper accretion column. On the other hand, the radiation pressure balances the gas pressure of the accreting flow in the lower accretion column, and a shock wave may be produced in this part. In this case, the plasma flow should approach a thermal equilibrium velocity distribution.

The electron cyclotron maser instability may occur in the upper accretion column. It is supposed, that the ions only play the role of a neutral background in the plasma, their kinetic effect may be negligible, and the nonthermal electron density n_s << the electron density of the background n_o. The resonance condition of the electron cyclotron maser instability may be written as

$$\gamma\omega_r - m\Omega_e - k_{\parallel}u_{\parallel} = 0 \qquad (1)$$

where $u_{\parallel}=\gamma v_{\parallel}$, $\gamma=(1 - v^2/c^2)^{-1/2} = (1 + u^2/c^2)^{1/2}$ is the relativistic factor, ω_γ is the resonance frequency, Ω_e is the electron cyclotron frequency and m is the harmonic number. This resonance condition corresponds to the resonance interaction of electron and wave, it may be expressed as a resonance ellipse in the momentum u space:

$$\frac{u_{\perp}^2}{a^2} + \frac{(u_{\parallel} - u_o)^2}{b^2} = 1 \qquad (2)$$

where

$$\frac{u_o}{c} = \frac{m\Omega_e N\cos\Theta}{\omega_\gamma(1-N^2\cos^2\Theta)}$$

$$\frac{a^2}{c^2} = \frac{m^2\Omega_e^2}{w_\gamma^2(1-N^2\cos^2\Theta)} - 1$$

$$\frac{b^2}{c^2} \quad \frac{a^2}{c^2(1-N^2\cos^2\Theta)}$$

θ is the angle between the emissive direction and magnetic field, N is the refractive index of the background plasma. The cold plasma approximation for the background plasma has been adopted in the following discussion. We can find easily from equation (2), that not only the electron velocity with $u_\perp \geqslant u_{\perp C}$ can contribute to the line emission, but also with $u_\perp < u_{\perp C}$ ($u_{\perp C}$ is a critical perpendicular velocity for the single electron emitting cyclotron emission). Therefore, the electron number which could contribute to the resonance line emission will increase vastly compared with single electron process.

A loss-cone velocity distribution function F ($u_{\parallel}$, $u_\perp$) has been chosen as a nonthermal electron distribution

$$F(u,u) = \begin{cases} F_> = A \exp(-u^2/\alpha^2); & u_\perp^2 > (u_\parallel^2 + \Phi)/d \\ F_< = A \exp(-u^2/\alpha^2) \exp[(u_\perp^2 d - u_\parallel^2 + \Phi)/d\beta^2]; & u_\perp^2 \leqslant (u_\parallel^2 + \Phi)/d \end{cases} \qquad (3)$$

where $d = B_M/B-1$, $\Phi = 2(u-u_1)/M$, U_1, B_M are the gravitational potential and the magnetic field at the magnetic mirror point respectively. A is the normalization constant.

In the case where the growth rate of the instability ω_i is much less than the resonance frequency ω_γ, the growth rates of the maser instability might be written as [11]:

$$\left[\frac{\omega_i}{\Omega_e}\frac{n_o}{n_s}\right]_m = \frac{4\pi^2}{G}\left[\frac{c}{\alpha}\right]^2\left[\frac{\omega_{pe}}{\Omega_e}\right]^2\left[\frac{\Omega_e}{\omega_\gamma}\right]^2 \sum_{j=m}^{\infty} \left\{\left[\int_{u_j(-)}^{u_\parallel(j-)} du_\parallel + \int_{u_\parallel(j+)}^{u_j(+)} du_\parallel\right]\right.$$

$$j\left[j\left[\frac{\alpha^2}{\beta^2} - 1\right]\frac{\Omega_e}{k\sin\theta} - \left(1 + \frac{\alpha^2}{d\beta^2}\right)\cot\theta u_\parallel\right] u_\perp(As) F<\Psi j$$

$$\left. - \int_{u_\parallel(j-)}^{u_\parallel(j+)} du_\parallel j\left[\frac{m\Omega_e}{k\sin\theta} + \cot\theta u_\parallel\right) u_\perp(As) F>\Psi j\right]\Bigg\}; \quad \Delta>0$$

$$\left[\frac{\omega_i}{\Omega_e}\frac{n_o}{n_s}\right]_m = \frac{4\pi^2}{G}\left[\frac{c}{\alpha}\right]^2\left[\frac{\omega_{pe}}{\Omega_e}\right]^2\left[\frac{\Omega_e}{\omega_\gamma}\right]^2 \sum_{j=m}^{\infty} \int_{u_\parallel(-)}^{u_j(+)} du_\parallel\, u_\perp(As) F<\Psi_j$$

$$\left[j\left[\frac{\alpha^2}{\beta^2} - 1\right]\frac{\Omega_e}{k\sin\theta} - \left[1 + \frac{\alpha^2}{d\beta^2}\right]\cot\theta\, u_\parallel\right]; \qquad \Delta<0$$

where $\Delta \equiv \left[1 + \frac{1}{d}\right]\left[\frac{j\Omega_e}{\omega_\gamma}\right]^2 - \left[1 + \frac{\Phi}{dc^2}\right]\left[1 + \frac{1}{d} - j^2\cos^2\theta\right]$

the meaning of the other parameters are the same as in reference [11].

3. Transformation from the Fixed Frame to the Observational Frame

All the theories of line emission from magnetized neutron stars are developed in the frame of axes fixed on the neutron star. The transformation between this fixed frame and the frame in which the observation is carried out must be considered because the neutron star rotates with high angular velocity. This transformation may be written in the form

$$\begin{aligned} \cos\phi \sin\theta &= \sin\theta_0 \cos(\phi_0 - X) \\ \sin\phi \sin\theta &= \cos\delta \sin\theta_0 \sin(\phi_0 - X) + \sin\delta \cos\theta_0 \\ \cos\theta &= \cos\delta \cos\theta_0 - \sin\delta \sin\theta_0 \sin(\phi_0 - X) \end{aligned} \qquad (5)$$

where (θ_0, ϕ_0) and (θ, ϕ) denote the polar angle and the azimuthal angle specifying the direction of radiation in the observation coordinate system and the fixed coordinate system respectively, δ is the angle between the magnetic axis and the rotation axis of the neutron star and $X=\omega_0 t$, ω_0 is the angular velocity of the neutron star.

4. Numerical Calculation and Discussion

In order to compare with the cyclotron emission line from Her X-1 an average energy of the nonthermal electron distribution of 150 keV has been taken, it corresponds to $c/\alpha = 1.7$, $c/\beta = 11$; the radius of the neutron star $r_M = 10^6$cm, the mass of the neutron star $M = 1\ M_\odot$, the gravitational potential $\phi/c^2 = 0.55$, the maser instability occurs at $r = 1.2\ r_M$, the magnetic field $B = 4.5 \times 10^{12}$ Gauss and $d = 0.728$ at ths point.

The results show that the growth rates of the FX mode in the maser instability are much larger than the growth rates of the O mode in the above mentioned parameters, and the polarization of the line radiation from the FX mode is large. However, the difference between the hollow beam and the loss-cone nonthermal distribution is small.

The growth rates of the maser instability vary with the electron density and emission angle. This is shown in Fig. 1 and Fig. 2. It is apparent from Fig. 1 that the ratio of the growth rate $\omega_i(m=2)/\omega_i(m=1) = 0.05$ is very small. This means that the second harmonic emission line (110 keV) is quite weak, if the conditions $n_s << n_o$, and $n_o < 10^{26}/cm^3$ hold. The growth rates of the instability increase with the electron density (or accretion rate M) as shown in Fig. 1, in contrast to the transition rates between the Landau levels which are independent of n_o. In general, the luminosity of the emission line can be expressed by $L_e = L_o \exp(\omega_i \tau)$ (τ is the saturation time of the instability). Therefore, the line radiation produced from the maser instability is coherent, and the luminosity of line radiation will increase with the nonthermal electron density in an exponential manner, if the variation of the saturation time with the electron density can be neglected. It is characteristic of a collective excitation, and may explain qualitatively that the ratio L_e/L_c varies with L_c [4].

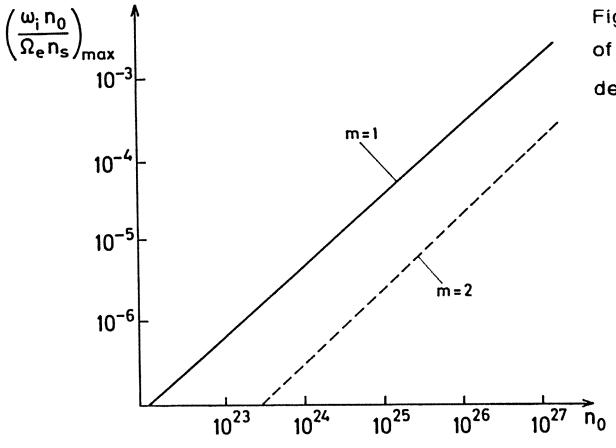


Fig. 1 The maximum growth rates of the FX mode versus the electron density n_o

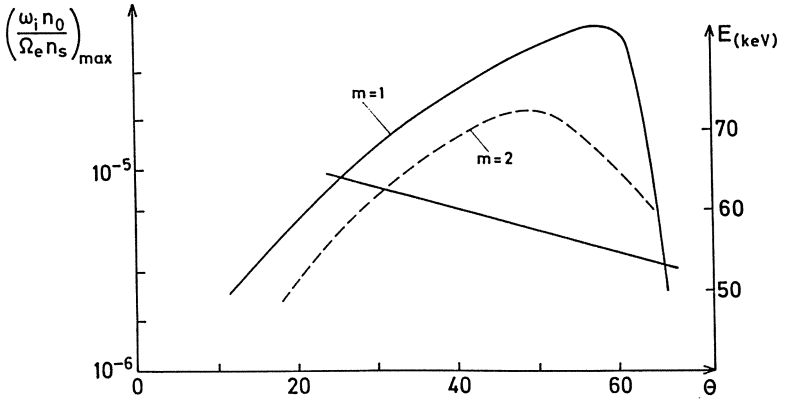


Fig. 2 The maximum growth rates and the energy of the emission line as a function of angle with respect to the magnetic field

The growth rates depend intensely on the angle relative to the magnetic field, as shown in Fig. 2. This means that the emission line from maser instability can be observed only in a limited angle region, or the intensity of the emission line will depend strongly on the pulse phase of the X-ray pulsars. The energy of the emission line depends approximately linearly on emission angle as shown in Fig. 2. Taking the transformation (5), it can be found in Fig. 3 that the emission line can be observed only in a part of the 1.24s pulsations and its energy varies with time in a sinousidal pattern. In contrast, the angular distribution of the emission line from Landau levels transition is rather smooth, and the energy of the emission line is approximately constant [12]. It can be seen by means of the transformation (5), that the emission line would be observed over a wide range of the pulse phase and the energy would be constant. This is not consistent with the observations [2], [3].

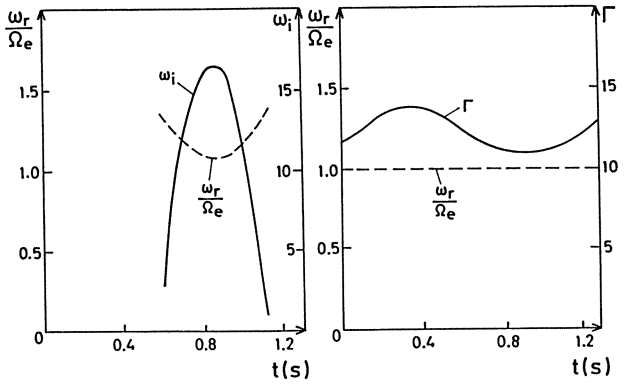


Fig. 3 The energy and the intensity (relative value) of the emission line as a function of pulse phase in the observational coordinate system for the maser instability model (left); for the emission due to Landau level transitions (right)

This is an attractive model for some observational results. The emission line from Her X-1 might be explained by the electron cyclotron maser instability model. However, many physical processes such as quantum effects and radiation transfer effects have not been considered in this model.

References

1. J. Trümper, W. Pietsch, C. Reppin, W. Voges, R. Staubert and E. Kendziorra: Ap.J. 219, L105 (1978)
2. D.E. Gruber, J.L. Matterson, P.L. Nolan, F.K. Knight, W.A. Baity, R.E. Rothschild, L.E. Peterson, J.A. Hoffman, A. Scheepmarker, W.A. Wheaton, F.A. Primini, A.M. Levine and W.H.G. Lewin: Ap.J. 240, L127 (1980)
3. W. Voges, W. Pietsch, C. Reppin, J. Trümper, E. Kendziorra and R. Staubert: Ap.J. 263, 803 (1983)
4. N. Prantzos and P. Durouchoux: Astron. Astrophys. 136, 363 (1984)
5. M. Basko and R. Sunyaev: Astron. Astrophys. 42, 311 (1975)
6. P. Meszaros: Space Sci. Rev. 38, 325 (1983)
7. W. Nagel: in this workshop
8. D.Y. Wang: "The Origin and Evolution of Neutron Stars" IAU Symp. 125, eds by D.J. Helfand and J.H. Huang (1986)
9. D.Y. Wang and D.Y. Mao: Publication of Purple Mountain Observatory 5, 1 (1986)
10. J. Frank, A.R. King and D.J. Raine: "Accretion Power in Astrophysics" Cambridge University Press (1985)
11. D.Y. Mao and D.Y. Wang to be published in Chines Journal of Nuclear Fusion an Plasma Physics (1087)
12. K.R. Lang Astrophysical Formulae 2nd ed. (Springer-Verlag, 1983)

Radiation Processes in Strong Magnetic Fields

W. Nagel

Max-Planck-Institut für Physik und Astrophysik, Institut für Astrophysik,
Karl-Schwarzschild-Str. 1, D-8046 Garching, Fed. Rep. of Germany

Since the discovery of cyclotron lines in the spectrum of the X-ray pulsar Her X-1 [1] much effort has gone into theoretical work on radiation processes and radiation transport in hot ($\sim 10^8$ K) and strongly magnetized ($\sim 10^8\,\mathrm{Vs\,m^{-2}}$) plasma (see MÉSZÁROS [2] for a review). Apart from its astrophysical relevance for pulsars, and possibly γ-ray bursts, this problem also offers a unique opportunity of learning about fundamental physics: the interaction of light and matter. I have been studying this problem for quite a number of years now, but only recently did I find a consistent and promising way to attack it—using quantum field theoretical methods. Let me try to convince you that radiation processes in magnetic fields are a challenging research project—a project which is just at its beginning, and a lot of interesting exercises in quantum electrodynamics are still in store. But before I come to these fundamental aspects let me talk about the astrophysical background of the problem.

Figure 1 shows a few simple model spectra which I computed some years ago [3] with the intention of producing something similar to the Her X-1 spectrum. These spectra were computed for uniform hot plasma layers of various thicknesses ($\tau_{\mathrm{Thomson}} = 10^{-2} \ldots 10^3$). You can see that only for very thin slabs does one get an emission line, which soon becomes self-reversed, and turns into an absorption line for thick layers. My conclusion from these calculations was

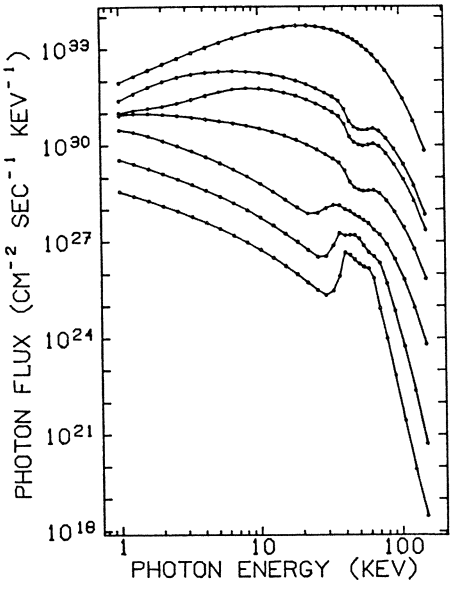


Fig. 1: Computed X-ray spectra from uniform layers of hot magnetized plasma, $\rho = 10^{-2}\,\mathrm{g\,cm^{-3}}$, $T = 10\,\mathrm{keV}$, $\hbar\omega_B = 50\,\mathrm{keV}$, thicknesses 10^{-2}, 10^{-1}, ... , 10^3 in Thomson units. The uppermost curve represents the Wien spectrum.
(from NAGEL 1981, Astrophys. J. **251**, 295)

that the feature(s) in the Her X-1 spectrum should be interpreted as absorption lines. While I still believe that this is true, these model spectra have turned out to be wrong in almost every respect. The line shape, for example, is controlled by a side-effect: the non-orthogonality of the polarization modes at the angle $\theta = 60°$ which was adopted for the numerical calculations. Vacuum polarization and spin-flip transitions, which were ignored here, have a profound effect on the line shape [4]. Moreover, it turned out later that we had left out an important radiation process: two-photon emission, which is in fact the dominant source of continuum photons. I hope it will become clear in the course of my talk how this came about—why an important effect had simply been forgotten.

The fundamental radiation process in a magnetic field is cyclotron radiation. Non-relativistically, an electron cascades down from one Landau level to the next, emitting a single photon at each step. But an electron can return from the first excited Landau level to the ground state also by emitting two photons simultaneously. This process occurs less frequently than normal cyclotron emission, but, as was first pointed out by KIRK *et al.* [5], it is important as a source of continuum photons. The frequency of the emitted photons is variable, because only the sum, $\omega_1 + \omega_2 = \omega_B$, is fixed.

Preliminary estimates [6] showed that under conditions thought to be typical for an accreting neutron star like Her X-1 the two-photon emissivity would significantly exceed the bremsstrahlung rate. More detailed calculations of the two-photon cross sections were carried out by John Kirk and Michelle Storey. Typically, in two-photon emission from the first excited Landau level, one of the photons is quite soft, while the other has nearly the line energy. In order to obtain an 'effective' absorption coefficient and emissivity for the soft photons, they took averages over the hard ('cyclotron') photons, which have short mean free paths and can be assumed to have a thermalized (Bose-Einstein) distribution. This averaging led to another surprise: the resulting cross sections had an angular and polarization dependence remarkably similar to that of bremsstrahlung. It is possible to derive an expression for the two-photon emissivity in close analogy with bremsstrahlung. Since I find this very instructive, let me briefly present this here.

Perhaps I should warn you that this derivation is somewhat unorthodox—I like to call it 'pseudo-classical', because it is just the opposite of the well-known 'semi-classical' approach. Let us treat the electrons classically, but assume that cyclotron radiation consists of little 'bullets' carrying momentum $\hbar\omega_B/c$. Imagine that the electrons are in a 'thermal bath' of cyclotron radiation. Whenever a line photon is absorbed or emitted, the electron suffers a recoil. Thus the electrons are vigorously 'shaken' by the line radiation. From the spectrum of the velocity fluctuations one can deduce the spectrum of the emitted radiation.

Since an electron in a magnetic field is just a harmonic oscillator, the velocity fluctuations are quite easy to work out (see Fig. 2). Introducing $v_\pm = (v_x \pm iv_y)/\sqrt{2}$, one gets very simple equations of motion:

$$\frac{d}{dt}v_\pm = \pm i\omega_B v_\pm$$

and correspondingly the response to a 'kick' at time $t = 0$ is

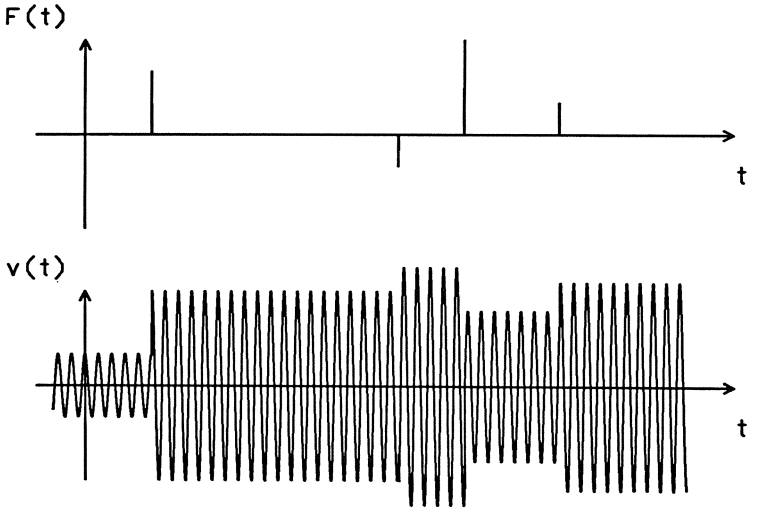


Fig. 2: An example of how a velocity component of an electron in a magnetic field may vary in time. It is assumed that collisions with ions or cyclotron line photons lead to instantaneous changes of the velocity.

$$G_{\pm}(t) = \Theta(t)\, e^{\pm i\omega_B t}, \qquad G_{\pm}(\omega) = \frac{i}{\omega \pm \omega_B + i0}.$$

The white input noise due to the random pulses is filtered by the oscillator, and the power spectrum of the velocity fluctuations is therefore

$$\int dt\; e^{i\omega t}\langle v_+^*(0)v_+(t)\rangle = \lambda\left(\frac{\hbar\omega_B}{mc}\right)^2 \frac{1}{(\omega+\omega_B)^2},$$

$$\int dt\; e^{i\omega t}\langle v_-^*(0)v_-(t)\rangle = \lambda\left(\frac{\hbar\omega_B}{mc}\right)^2 \frac{1}{(\omega-\omega_B)^2},$$

$$\int dt\; e^{i\omega t}\langle v_z^*(0)v_z(t)\rangle = \lambda\left(\frac{\hbar\omega_B}{mc}\right)^2 \frac{1}{\omega^2},$$

where λ is the rate at which the pulses occur. Actually, velocity jumps occur at different rates for directions transverse and along the magnetic field, owing to the anisotropy of the cyclotron cross sections (see KIRK *et al.* [7] for these details). The emissivity is then given by

$$\epsilon = \frac{\alpha\hbar\omega^2}{2\pi c^2}\sum_{\mu,\nu} e_\mu^* e_\nu \int_{-\infty}^{\infty} dt\; e^{i\omega t}\langle v_\nu(0)v_\mu(t)\rangle$$

$$= \frac{2\alpha^2}{5\pi}\hbar\omega_B\left(\frac{\hbar\omega_B}{mc^2}\right)^3 n(\omega_B)\left[\frac{\omega^2}{(\omega+\omega_B)^2}|e_+|^2 + \frac{\omega^2}{(\omega-\omega_B)^2}|e_-|^2 + \frac{4}{3}|e_z|^2\right],$$

where

$$n(\omega_B) = \left[e^{(\hbar\omega_B-\mu)/T} - 1\right]^{-1}$$

is the occupation number of the cyclotron line photons. In this form, the similarity to bremsstrahlung [8,9] becomes quite obvious. At low frequencies ($\omega \ll \omega_B$) the angular and polarization dependence of the emissivity is controlled by the component of the polarization vector along the magnetic field, e_z. Only along the field can the electrons move freely; transverse motion is strongly inhibited. For this reason, all radiation processes (scattering, bremsstrahlung, two-photon emission) have a rather similar angular dependence.

Figure 3 shows the angular dependence of the two-photon emissivity as computed with the simple formula just derived, and the more complicated expressions given by KIRK *et al.* [7]. Although the derivation is so primitive, it gives a pretty good idea of the magnitude and angular dependence of the emissivity. There are some discrepancies at large angles, because the Kubo formula used above for the emissivity is based on the dipole approximation. (The more general formula is given below.)

The most important feature of the two-photon emissivity is the appearance of the density of cyclotron photons. Two-photon emission is caused by collisions with line photons, as opposed to bremsstrahlung which is due to collisions with ions. Under the conditions likely to prevail in the accretion column of Her X-1, line photons are far more numerous than ions. Therefore two-photon emission will produce many more soft photons than bremsstrahlung. On the other hand, the preceding derivation makes it understandable why this process had not been thought of for such a long time. In the case of resonance lines in the optical or ultraviolet, with which astrophysicists are very familiar, the momentum of a line photon is always taken up by a whole atom. Hence the velocity changes are much smaller. And the momentum of an optical photon

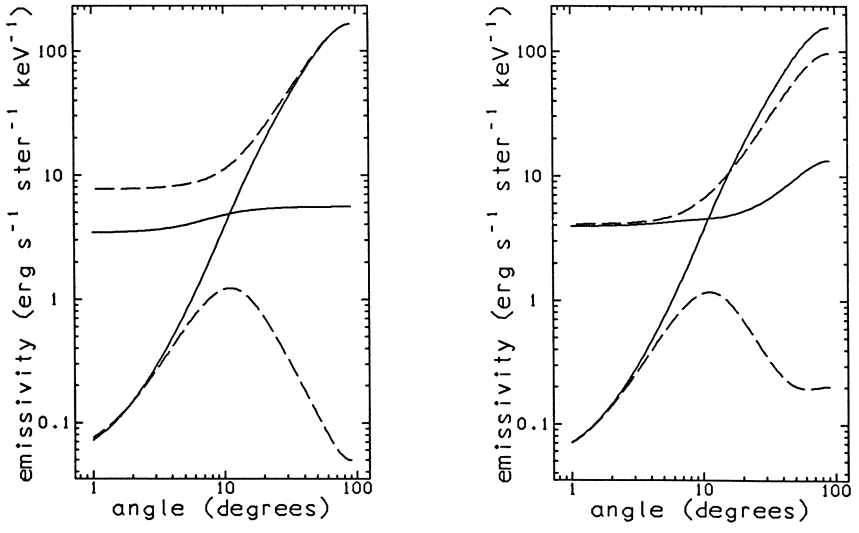


Fig. 3: Angular dependence of the 'effective' two-photon emissivity. Left: as computed from the simple formula given in the text. Right: using the more accurate expressions given by KIRK *et al.* [7]. The occupation number of cyclotron line photons is taken to be $n(\omega_B) = 1$, density $\rho = 10^{-2}\,\mathrm{g\,cm^{-3}}$, temperature $T = 10\,\mathrm{keV}$, and cyclotron frequency $\hbar\omega_B = 50\,\mathrm{keV}$. Emissivities are given at $\hbar\omega = 10\,\mathrm{keV}$ (upper curves) and 1 keV (lower curves) for the ordinary (solid) and the extraordinary polarization mode (dashed lines). Vacuum polarization leads to an exchange of the two modes at 2.4 keV.

is, of course, minute compared to a photon in the hard X-rays. For these reasons two-photon emission is negligible as a photon source in normal stellar atmospheres.

Two-photon emission solves the old problem that with bremsstrahlung as a photon source one had to go to rather high densities ($\gtrsim 1\,\mathrm{g\,cm^{-3}}$) to produce model spectra for Her X-1 with realistic luminosities ($\sim 10^{37}\,\mathrm{erg\,s^{-1}}$). With two-photon emission it is now possible to get reasonable luminosities also from the relatively thin plasma thought to fill the accretion column.

But these processes also create a problem: they make it impossible to compute opacities without already knowing the radiation field. The calculation of opacities becomes coupled to the solution of the transport equation, since e.g. the soft photons can modify the cyclotron line shape. In the model calculations of KIRK *et al.* [7] this difficulty was by-passed by assuming a certain equilibrium density of line photons (which was treated essentially as a free parameter). Ideally, one would like to calculate the densities of line and soft photons consistently—each kind affecting the radiative cross sections of the other. The work required to solve the transfer equation is then almost negligible compared to that of computing the opacities, because multiple integrals over frequencies and directions are required. But there is a problem even more serious than this numerical difficulty: How must one formulate the problem? What are the equations which we should solve?

It seems natural to start from two Boltzmann equations: one for the photons, and another one for the electrons. But as soon as you have written down these equations, you have fallen into a trap. Why? Can anything be wrong with such a basic thing as the Boltzmann equation? This is certainly a strange question to ask for someone who has worked on radiative transfer problems. I only noticed that there was a problem when I started to think about the rate coefficients that go into the kinetic equations. It is quite straight-forward to write down a scattering cross section for electrons with a particular speed. But the total scattering cross section is infinite, if you integrate over the electron velocity distribution. This is an old problem, of course, and already WEISSKOPF and WIGNER [10] showed that it is necessary to include a damping term in order to obtain a reasonable line shape. However, this damping term ruins our book-keeping system provided by the Boltzmann equations, since it is no longer possible to separate the various radiation processes like absorption, scattering, bremsstrahlung, etc. All of them contribute to the damping. If you try to expand in powers of the coupling constant α, you are back at square one, because the cross sections diverge. MELROSE and KIRK [11] wrote down both the divergent, and the physically meaningful expressions (with the damping included). But they emphasized that the damping represents just an approximation, and that it should not be used for accurate work. The question then that I would like to answer is: What is the true cyclotron line shape? How can the opacity be calculated?

In the theory of spectral line shapes it is customary to express the absorption coefficient as a Fourier integral of the dipole moment auto-correlation function. I do not find it very meaningful to talk about a dipole moment of a free electron. Moreover, we need to go beyond the dipole approximation if we want to discuss also the higher harmonics of the cyclotron line. The solution of the problem lies in the dielectric tensor, or in the conductivity tensor, which is

related to it. It is well known (at least to solid state physicists) that the conductivity tensor can be expressed in terms of the current-density fluctuations of the medium, giving rise to a so-called Kubo formula [12]. By taking the hermitian part of the conductivity tensor we can thus obtain a Kubo formula for opacities:

$$\kappa = \frac{\mu_0 c}{2\hbar\omega} \sum_{\mu\nu} e^*_\mu e_\nu \int_{-\infty}^{\infty} dt \int d^3x \, e^{-i(kx-\omega t)} \, \langle j_\mu(x,t) j_\nu(0,0) - j_\nu(0,0) j_\mu(x,t) \rangle .$$

Here it is understood that we take quantum mechanical expectation values as well as statistical averages of the current-density operators j_μ. The order of the operators is obviously significant—otherwise the commutator would vanish.

For the emissivity we can write down an analogous expression,

$$\epsilon = \frac{\mu_0 \omega^2}{8\pi^2 c} \sum_{\mu\nu} e^*_\mu e_\nu \int_{-\infty}^{\infty} dt \int d^3x \, e^{-i(kx-\omega t)} \, \langle j_\nu(0,0) j_\mu(x,t) \rangle .$$

We already used a simplified variant of this formula in the derivation of the two-photon emissivity above. I would like to convince you that these formulas are very general and useful, since they describe *all* radiation processes. It is just necessary to compute the correlation functions in sufficient detail.

As a simple demonstration let us first consider the cyclotron emissivity. If we use the dipole approximation (i.e. let $k \to 0$), then we need only the velocity fluctuations,

$$\int d^3x \, \langle j_\mu(x,t) j_\nu(0,0) \rangle = N e^2 \langle v_\mu(t) v_\nu(0) \rangle ,$$

which are very easy to determine. This is because the operators $v_\pm = (v_x \pm i v_y)/\sqrt{2}$ act as ladder operators

$$v_+ \, |n\rangle = \left(\frac{\hbar\omega_B}{m} \right)^{\frac{1}{2}} \sqrt{n+1} \, |n+1\rangle , \qquad v_- \, |n\rangle = \left(\frac{\hbar\omega_B}{m} \right)^{\frac{1}{2}} \sqrt{n} \, |n-1\rangle ,$$

moving an electron from one Landau level to the next. It is easy to see that in thermal equilibrium (temperature $T = 1/\beta$) one has

$$\langle v_+(0) v_-(t) \rangle = \langle n \rangle \, \frac{\hbar\omega_B}{m} \, e^{-i\omega_B t} ,$$

$$\langle v_-(0) v_+(t) \rangle = \langle n+1 \rangle \, \frac{\hbar\omega_B}{m} \, e^{i\omega_B t} ,$$

$$\langle v_z(0) v_z(t) \rangle = \frac{T}{m} ,$$

where $\langle n \rangle = 1/\left[\exp(\beta\hbar\omega_B) - 1 \right]$. Now we find

$$\epsilon = \frac{\alpha\hbar\omega^2}{2\pi c^2} N \sum_{\mu\nu} e^*_\mu e_\nu \int_{-\infty}^{\infty} dt \, e^{i\omega t} \langle v_\nu(0) v_\mu(t) \rangle$$

$$= \frac{\alpha\hbar^2\omega_B^3}{mc^2} N \langle n\rangle |e_-|^2 \delta(\omega - \omega_B).$$

Note that I retained only the positive frequency terms. At temperature $T = 0$ the emissivity drops to zero, as it must, even though the velocity fluctuations are still non-zero (due to vacuum fluctuations). They contain only negative frequencies at $T = 0$.

What we have obtained is an extremely simplified expression for the cyclotron emissivity. The limit $k \to 0$ is tantamount to the assumption that an electron always stays at the same place—so there is no Doppler broadening of the line. It requires a little more work to write down the full expressions involving the current correlations. This is straight-forward, in principle. What is less straight-forward is the way scattering cross sections can be computed from the Kubo formulas for opacity and emissivity. Only recently did I learn how to do it, and I won't go into all the technical details, which you can find elsewhere [13].

The essential idea is to introduce a time-ordering along a closed time-path [14], so that products of operators will always appear in the correct order. In effect one duplicates the operators: for every operator $A(\underline{t})$ on the 'forward' time branch there is another one on the 'backward' time branch $A(\bar{t})$ which is identical, except that it has the opposite behaviour under time ordering. If we denote a time ordered product by $[\![\cdots]\!]$, we define

$$[\![A(\underline{t})B(\underline{t}')]\!] = \Theta(t-t')A(t)B(t') \pm \Theta(t'-t)B(t')A(t)$$
$$[\![A(\bar{t})B(\underline{t}')]\!] = A(t)B(t'),$$
$$[\![A(\underline{t})B(\bar{t}')]\!] = \pm B(t')A(t),$$
$$[\![A(\bar{t})B(\bar{t}')]\!] = \Theta(t'-t)A(t)B(t') \pm \Theta(t-t')B(t')A(t)$$

where $\Theta(t)$ is the usual Heaviside step function.

With this definition, correlation functions can be expressed in the form

$$\langle v(t)v(0)\rangle = \langle [\![v(\bar{t})v(\underline{0}) \exp -i\tfrac{e}{\hbar} \oint dt' H'(t')]\!]\rangle$$

with $H'(t) = eA(t)v(t)$, and where $\oint dt' = \int_{-\infty}^{\infty} d\underline{t} - \int_{-\infty}^{\infty} d\bar{t}$ indicates integration over a closed time-path.

The third basic idea is the application of Wick's theorem. Expectation values of operator products can be factorized into elementary Green's functions. An expression like

$$\oint dt_1 \oint dt_2 \, \langle [\![v(t)v(0)A(t_1)v(t_1)A(t_2)v(t_2)]\!]\rangle$$

gives rise to two terms of the form

$$\oint dt_1 \oint dt_2 \, \langle [\![v(t)v(t_1)]\!]\rangle \langle [\![A(t_1)A(t_2)]\!]\rangle \langle [\![v(t_2)v(0)]\!]\rangle$$

because there are two possible ways of pairing up the operators. The building blocks of the quantum field theoretical machinery are therefore Green's functions such as

$$\langle [\![v_\mu(t)v_\nu(0)]\!]\rangle = \frac{\hbar\omega_B}{m}\left[\delta_{\mu-}\delta_{\nu+}\Theta_o(t)e^{-i\omega_B t} + \delta_{\mu+}\delta_{\nu-}\Theta_o(-t)e^{i\omega_B t}\right],$$

(for cold plasma, $T = 0$) and

$$\langle [\![A_\mu(t) A_\nu(0)]\!] \rangle = \frac{\mu_0 c}{2\omega^2} F_0 \left[e_\mu e_\nu^* e^{-i\omega t} + e_\mu^* e_\nu e^{i\omega t} \right],$$

if there is a wave with flux F_0 which is completely polarized and has the polarization vector e_μ. Here, $\Theta_o(t-t') = 1$, if t is a 'later' time on the closed time-contour than t', and zero otherwise. I have simplified the notation and given only those parts which are actually needed to calculate the differential scattering cross section; see NAGEL [13] for more details.

At first sight, the whole formalism looks rather complicated, but the basic ideas are simple, and the application of the machinery is quite straight-forward (if tedious). What I found most interesting is that with this method you can in fact calculate the well known scattering cross sections in a magnetic field [15]. Thus line and continuum opacities can be calculated from the same fundamental formula. The beauty of the Kubo formulas is that they contain *all* radiation processes—if we evaluate the correlation functions with sufficient care. Moreover, they are really the only way to arrive at meaningful expressions for the cyclotron line shape. The usual way of adding up squares of Feynman diagrams for different processes is only an approximation which leads into the trouble that I have discussed. There is but one observable opacity—separating parts due to 'absorption' or 'scattering' (or etc.) is meaningless.

This implies that one cannot use the Boltzmann equations to set up the problem, because this book-keeping system requires a careful separation of the various radiation processes appearing on the right hand side. Instead, one should start from the equations of motion for the field and current correlation functions, $\langle [\![A_\mu(x,t) A_\nu(0,0)]\!] \rangle$ and $\langle [\![j_\mu(x,t) j_\nu(0,0)]\!] \rangle$. Current fluctuations give rise to field fluctuations, and vice versa. It should be possible to derive self-consistency relations for the spectra of these fluctuations. Perhaps these equations are not much harder to solve numerically than the usual radiative transfer equation. Then radiation processes of arbitrarily high order could be included (if only in an approximate way), like e.g. in the Hartree-Fock/self-consistent field method. This is an interesting possibility that deserves to be explored further.

References

1. J. Trümper, W. Pietsch, C. Reppin, W. Voges, R. Staubert, E. Kendziorra: *Astrophys. J. (Letters)* **219**, L105 (1978)
2. P. Mészáros: *Space Science Reviews* **38**, 325 (1984)
3. W. Nagel: *Astrophys. J.* **251**, 288 (1981)
4. P. Mészáros, W. Nagel: *Astrophys. J.* **298**, 147 and **299**, 138 (1985)
5. J.G. Kirk, D.B. Melrose, J.G. Peters: *Proc. Astron. Soc. Australia* **5**, 478 (1984)
6. J.G. Kirk: *Astron. Astrophys.* **158**, 305 (1986)
7. J.G. Kirk, W. Nagel, M. Storey: *Astron. Astrophys.* **169**, 259 (1986)
8. W. Nagel: *Astrophys. J.* **236**, 904 (1980)
9. W. Nagel, J. Ventura: *Astron. Astrophys.* **118**, 66 (1983)
10. V. Weisskopf, E. Wigner: *Zeits. f. Physik* **63**, 54 and **65**, 18 (1930)
11. D.B. Melrose, J.G. Kirk: *Astron. Astrophys.* **156**, 268 (1986)

12. J.M. Ziman: *Elements of Advanced Quantum Theory* (Cambridge University Press, Cambridge 1969)
13. W. Nagel: *J. Phys. A* (submitted)
14. K. Chou, Z. Su, B. Hao, L. Yu: *Physics Reports* **118**, 1 (1985)
15. J. Ventura: *Phys. Rev. D* **19**, 1684 (1979)

Radiation Dominated Shocks in a One-Dimensional Plasma Flow

H. Riffert

Astronomy Department, Pennsylvania State University,
525 Davey Laboratory, University Park, PA 16802, USA and
Max-Planck-Institut für Physik und Astrophysik, Institut für Astrophysik,
Karl-Schwarzschild-Str. 1, D-8046 Garching, Fed. Rep. of Germany

The spectrum of a radiation field passing a shock-front in a hot tenuous plasma is influenced drastically by the bulk motion of the gas. This may appear in many astrophysical environments, for example in accretion columns near the surface of neutron stars (BASKO and SUNYAEV [1], in spherical accretion flows around AGNs or black holes (MÉSZÁROS and OSTRIKER [2], BLONDIN [3]), in the terminal shock-waves of supernovae (COLGATE [4]) or in the radiation dominated era of the early universe (ZEL'DOVICH et al. [5]). If the plasma density is sufficiently low then the main coupling mechanism between the radiation and the gas is Compton scattering. In addition, in a converging flow the photons can systematically gain energy by the first order Fermi process. The details of this mechanism have been worked out by BLANDFORD and PAYNE [6], and the repeated energy increase of photons diffusing many times through the shock leads to a high energy power law in the post-shock spectrum. However, the redistribution of energy due to Compton scattering has to be neglected in order to obtain the power law solution. Certain aspects of this effect have been considered by LYUBARSKII and SUNYAEV [7] under the assumption of a constant electron temperature and a flow velocity structure varying linearly with optical depth. The resulting spectrum obtained from the injection of soft photons contains a low energy power law plus a high frequency exponential cut-off, which is a common feature for Compton scattering with thermal electrons (SUNYAEV and TITARCHUK [8]). This model does not, however, represent a complete solution of the shock problem; its validity is mainly restricted by the assumption of a constant temperature. In the case of pure scattering a local Compton equilibrium will lead to a temperature profile in the shock depending on the spectral properties of the radiation field (ZEL'DOVICH and LEVICH [9], ILLARIONOV and SUNYAEV [10])

$$kT = \frac{\hbar}{4} \frac{\int_0^\infty n(1+n)\omega^4 d\omega}{\int_0^\infty n\omega^3 d\omega}, \tag{1}$$

where $n(\omega)$ is the photon occupation number. The temperature distribution will in turn alter the local spectrum: the exponential cut-off due to Comptonization depends on the position in the shock, and the photons that diffuse into a certain volume element have been scattered by electrons with a temperature that is different from the local value.

The simultaneous calculation of the flow structure, the electron temperature and the photon spectrum has to be performed numerically. For simplicity we consider a one-dimensional stationary shock structure where Compton scattering is the only relevant interaction process between the plasma and the radiation. The electron distribution is adjusted to a local Maxwell distribution by the recoil effect on a very short time scale (ZEL'DOVICH and LEVICH [9]). In the

limit of the diffusion approximation the following transport equation has to be solved together with the temperature condition (1) (we will, however, neglect the induced scattering processes, being proportional to n^2)

$$-\frac{1}{3}\frac{\partial^2 n}{\partial \tau^2}+\frac{v}{c}\frac{\partial n}{\partial \tau}-\frac{1}{3c}\frac{dv}{d\tau}\omega\frac{\partial n}{\partial \omega}=\frac{\hbar}{m_e c^2}\cdot\frac{1}{\omega^2}\frac{\partial}{\partial \omega}\left[\omega^4\left(n+\frac{kT}{\hbar}\frac{\partial n}{\partial \omega}\right)\right] \tag{2}$$

where v is the flow velocity and τ is the optical depth with respect to Thomson scattering.

The right hand side describes the photon scattering in the Fokker-Planck limit (KOMPANEETS [11]), and the left hand side contains the diffusion, advection and compression heating of the radiation. The (spatial) diffusion coefficient has been calculated in the Thomson limit and is, therefore, independent of the photon energy. As a consequence, it is possible to obtain an analytic solution of the shock structure which is independent of the particular spectrum (BLANDFORD and PAYNE [6]); the velocity for example reads

$$v(\tau)=v_-\left\{1-\frac{3}{7}\left(1-\frac{1}{M^2}\right)\left[1+\tanh\left(\frac{\tau}{\tau_s}\right)\right]\right\} \tag{3}$$

where $M^2=\frac{3}{4}\rho_- v_-^2/p_-$ is the Mach number, which depends on the pre-shock values of the density ρ_-, the velocity v_-, and the pressure p_-. The typical shock width τ_s is given by

$$\frac{1}{\tau_s}=\frac{3}{2}\frac{v_-}{c}\left(1-\frac{1}{M^2}\right).$$

The asymptotic behaviour of equation (2) far behind the shock leads to a constant temperature T_+ and an equilibrium Wien spectrum

$$n_+=A\exp\left(-\frac{\hbar\omega}{kT_+}\right),$$

and for an initial Wien spectrum with the temperature T_- the post-shock value depends on the Mach number only

$$\frac{T_+}{T_-}=\frac{1}{49M^2}\left(8M^2-1\right)\left(M^2+6\right). \tag{4}$$

The figures 1 and 2 show two numerical solutions of the transport equation (2) for the self-consistent velocity profile (3) with a Mach number of 10 and different initial velocities and temperatures. At high frequencies the spectra decrease exponentially, because for $\hbar\omega \gg kT$ the photons loose their energy very efficiently by the electron recoil. The spectral features and the temperature profiles are quite different for those two solutions. In figure 1 the temperature reaches a maximum value close to the shock centre and decreases to the final value (4) on a length scale much larger than the shock width. The temperature function in the figure (2), however, is monotonically increasing. A qualitative explanation of this behaviour can be obtained by analyzing the post-shock structure in a perturbation approach. For small deviations from the final equilibrium, both the temperature and the spectrum decrease exponentially to this state, and there appear two characteristic length scales one of which is the shock width and the other one can be interpreted as a Comptonization length. The ratio of both length scales is

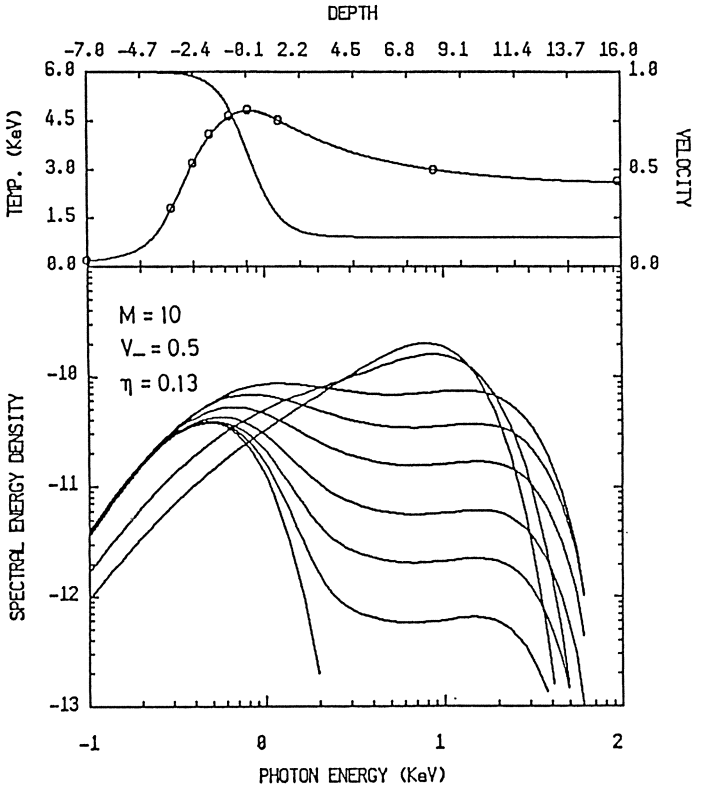


Fig. 1: Spectra, velocity, and temperature in a radiation dominated shock. The upper frame shows the velocity and the electron temperature as a function of the optical depth coordinate. The spectral energy density (in arbitrary units) as a function of energy is shown in the lower frame for various positions in the shock. (The corresponding depth points are indicated by circles on the temperature curve.) The initial spectrum is a Wien spectrum with a temperature of $kT = 0.156\,\mathrm{keV}$; the pre-shock velocity is $v = 0.5\,c$, and the Mach number is $M = 10$.

$$\eta = \frac{M^2+6}{7(M^2-1)}\left[\sqrt{1+\frac{8}{9}\cdot\frac{8M^2-1}{M^2+6}\cdot\beta M^2}-1\right]$$

and β is the ratio of the thermal to the kinetic energy of the electrons in the pre-shock regime. Although η has been derived under the condition that deviations from the equilibrium are small, it is useful to classify the shock structure in general, as numerical results show.

For $\eta \ll 1$ (figure 1) Compton scattering is not effective enough to establish a local equilibrium; this is the case for cold and fast flows

$$\beta \ll \frac{1}{M^2} \ll 1.$$

The spectra for all $\tau < 0$ in figure 1 contain a maximum at about 14 keV which is almost independent of the depth. This corresponds to a thermal spectrum of $kT \sim 4.7$ keV which is approximately the maximum temperature appearing at $\tau \sim 0$. Therefore, the spectrum in front

of the shock is dominated by hot photons diffusing upstream from the shock centre where the temperature is highest. This means in particular that there is an excess of high energy photons in this regime which effectively preheats the plasma electrons so that the temperature shoots over its final adiabatic value T_+.

If the length scale ratio η is larger than one (figure 2) the spectra are locally adjusted to an almost equilibrium distribution at least in the high energy range; the temperature approaches T_+ in a monotonic way.

As a significant result our self-consistently calculated solutions show that one does not obtain power law parts in the spectra of radiation dominated shocks. If the spectra deviate strongly from a local thermal equilibrium the existence of a non-constant temperature profile completely destroys any power law behaviour which appears in the case of no Compton scattering (BLANDFORD and PAYNE [6]).

We have assumed here an idealized configuration of an infinite one-dimensional plasma flow which can only qualitatively be applied to investigate realistic astrophysical situations. In order to observe a spectrum from the central part of the shock, the flow must become optically thin in the post-shock regime, i.e., the plasma must expand to a lower density. However, we expect

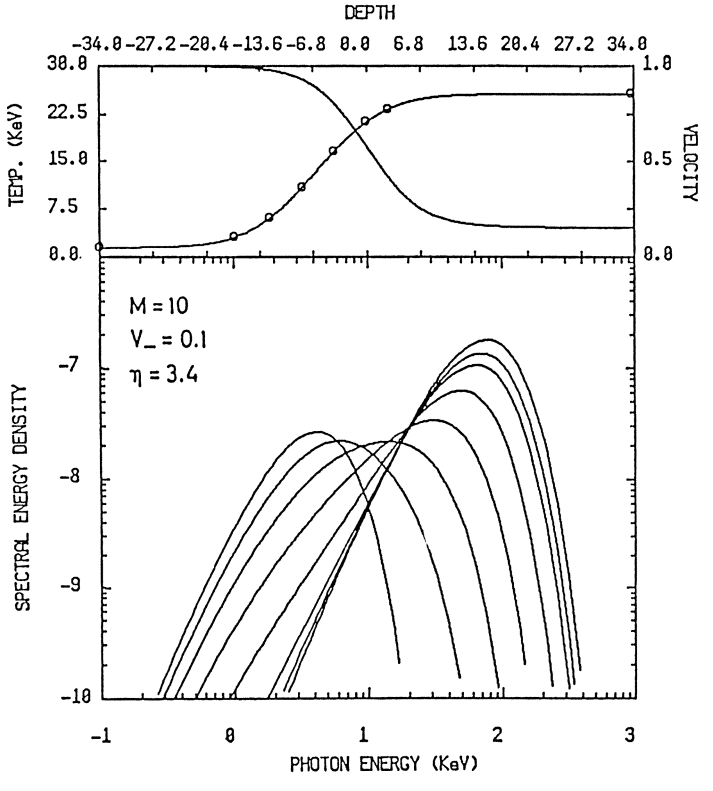


Fig. 2: Spectra, velocity, and temperature in a radiation dominated shock. Presentation is the same as in figure 1. Pre-shock values: temperature $kT = 1.39\,\mathrm{keV}$, velocity $v = 0.1\,c$, Mach number $M = 10$.

that the general shock structure for the radiation dominated case is similar to the solution we have presented here, and the possibility of two principal shock types, depending on the initial conditions, will still remain valid.

REFERENCES

1. Basko, M.M., and Sunyaev, R.A. 1976, *Monthly Notices Roy. Astron. Soc.* **175**, 395.
2. Mészáros, P., and Ostriker, J.P. 1983, *Astrophys. J.* **273**, L59.
3. Blondin, J.M. 1986, *Astrophys. J.* **308**, 755.
4. Colgate, S.A. 1975, *Astrophys. J.* **195**, 493.
5. Zel'dovich, Ya.B., Illarionov, A.F., and Sunyaev, R.A. 1973, *Sov. Phys. JETP* **35**, 643.
6. Blandford, R.D., and Payne, D.G. 1981, *Monthly Notices Roy. Astron. Soc.* **194**, 1041.
7. Lyubarskii, Yu.E., and Sunyaev, R.A. 1982, *Sov. Astron. Lett.* **8** (5), 330.
8. Sunyaev, R.A., and Titarchuk, L.G. 1980, *Astron. Astrophys.* **86**, 121.
9. Zel'dovich, Ya.B., and Levich, E.V. 1970, *JETP Lett.* **11**, 35.
10. Illarionov, A.F., and Sunyaev, R.A. 1975, *Sov. Astron.* **18** (4), 413.
11. Kompaneets, A.S. 1957, *Sov. Phys. JETP* **4**, 730.

Part II

Active Galaxies

Quasar Models

M. Camenzind

Landessternwarte Königstuhl,
D-6900 Heidelberg 1, Fed. Rep. of Germany

1. Introduction

Despite the many attempts to explain the overall structure of quasars, we are still not in the position to give only an approximate view about the innermost parsec in those sources. This is a somewhat discouraging aspect for those people who work laboriously for a longer time in this field in comparison with the situation encountered in many other branches of theoretical astrophysics.

Quasars are jet sources indicating that plasma must be driven away from a central region in a galaxy into the kpc-region. But even extensive observations of about 200 jets in extragalactic objects could not reveal the basic physical properties of jets (see e.g. BRIDLE and PERLEY /1/). In particular, neither the nature of the plasma (ion plasma or electron- positron mixture) nor the speed of this outflow are known. We only know that the jet speed has to be close to the speed of light in sources, which show superluminal motion. Here, the minimal Lorentz factors have to be between 2 and 10 (or even 20 for 3C 345, BIRETTA et al. /2/). It turns out to be extremely difficult to obtain such a high acceleration in a simply radiation pressure driven wind. In addition, with such a mechanism the collimation of the outflow must be achieved with external "pressure walls", such as formed by the funnel of thick accretion disks.

A few years ago, I proposed a much more powerful mechanism for the initial acceleration of such outflows (CAMENZIND /3/). This mechanism is based on the ***magnetic sling effect*** in a rapidly rotating magnetosphere. When plsama can be injected beyond the corotation radius in such a magnetosphere, we can achieve extremely high speeds when relativity is included in the problem formulation. The essential ingredients in the magnetic sling effect are physically quite simple. As in the Newtonian case, the rotating plasma will decouple from the magnetosphere at the Alfven point, but relativity forces this Alfven point always to be located inside the light cylinder. The final outflow velocity is then essentially just the rotation speed obtained at the Alfven point. The location of this Alfven point along a magnetic flux tube is now just a question of the density of the plasma. When the magnetosphere is heavily populated with plasma, the Alfven point is far inside the light cylinder and only moderate outflow velocities are achieved. When the magnetosphere is however only sparsely populated with plasma, the Alfven point is driven towards the light cylinder and outflow speeds very close to the speed of light can occur.

This rotating magnetosphere also provides the collimation force for the outflow. The rotating magnetosphere filled up with plasma produces a toroidal field which then acts as a pinch force against the inertial force of the outflowing plasma. It is however not a simple question to decide on which typical distance this collimation can be achieved.

In the following we describe first two situations in the center of a galaxy, which can generate large scale rotating magnetospheres. The basic equations for a plasma equilibrium are given, and we provide the first numerical solutions of this problem for the extreme relativistic limit. Finally, we outline various consequences of this jet model for the observations of quasars.

2. Basic Requirements for Quasar Models

In the past twenty years, many models have been proposed for the innermost parsec in quasars. It is, however, not easy to construct a realistic model which satisfies all the constraints imposed by observations. These constraints can be summarised in the following way:

(i) coexistence of inflow and outflow for jet sources;

(ii) coexistence of the UV-photosphere with jets;

(iii) the existence of magnetic fields in the nuclear jets.

The detection of the broad UV-bump in low-polarisation quasars, which are at the same time jet-sources, may indicate that accretion must occur with high angular momentum and that the outflow is perpendicular to this accretion disk. MALKAN and SARGENT /4/ and MALKAN /5/ have modelled this UV-excess in terms of optically thick thermal emission from the inner part of the accretion disk. The specific form of the UV-bump is not known, since there are too many uncertainties in the form of the optical synchrotron spectrum and in the contributions of other components from this wavelength region (see e.g. CAMENZIND and COURVOISIER /6/, EDELSON and MALKAN /7/). On the basis of these observations alone, the central UV- source in quasars can still be due to a rapid rotator or a supermassive disk.

The strong UV-radiation field in the center of a galaxy inhibits the effective pair production by gamma rays. From this we expect that the plasma feeding the outflow is a normal stellar plasma consisting of electrons and ions. We assume therefore for the following that jets consist of normal plasma.

The properties of nuclear jets in quasars - and in other radio sources - cannot be explained without the inclusion of magnetic fields. The easiest way to produce magnetic fields is to start with the presence of a strong magnetosphere around the central source. Plasma can then be injected into such a rotating magnetosphere either from the surface of a non-collapsed object or from the surface of an accretion disk. The outflowing plasma will drag along the magnetic field.

3. Two Models

There are essentially only two types of models, which can satisfy the above constraints. In the first instance, we consider the magnetosphere of a non-collapsed rapidly rotating compact object in interaction with an accretion disk. These types of objects are well known on the stellar mass scale. A rapidly rotating magnetized neutron star can eject plasma from the accretion disk which surrounds the object (CAMENZIND /8/). On the supermassive scale, the formation of rapidly rotating compact objects is very probable in the center of a galaxy, when black holes are not formed at the initial stage of galaxy formation. Matter around the center of a galaxy has in general too much angular momentum for the onset of a dramatic collapse (CAMENZIND /3/). The contraction of a central condensed object in a galaxy must occur on a long time scale dictated by angular momentum loss. The specific angular

momentum $a_* = J_*/(GM/c)$ of these objects is still quite high

$$a_* = \sqrt{2}\, k^2 \frac{<\Omega>}{\Omega_K} \sqrt{R_*/R_S} \tag{1}$$

when the mean rotation frequency $<\Omega>$ is very near to the Keplerian angular velocity Ω_K at the surface ($k^2 \simeq 0.1$). The fatal collapse towards a Kerr black hole with $a_H < 1$ will only occur if $a_* \simeq 1$. Mass loss must therefore be an essential ingredient in the evolution of supermassive rotators or supermassive disks (CAMENZIND /8/). These objects will carry rapidly rotating magnetospheres, and plasma will be injected either from the surface of the object or from a surrounding accretion disk. When the injection point is very near or beyond the corotation radius, plasma cannot accrete onto the central object, but will be driven away by the strong centrifugal forces. This *magnetic sling effect* is the most powerful way to produce high energetic plasma flows.

A similar interplay between magnetic fields and plasma flow can also occur when the magnetosphere is formed in an acretion disk around a central black hole (BLANDFORD and PAYNE /9/). The magnetosphere is then formed either by diffusive processes (see e.g. LOVELACE et al. /10/), or by dynamo action in a strongly turbulent and differentially rotating disk. Plasma is now injected directly from the surface of the disk.

In both models, we are forced to use relativistic MHD, since the magnetosphere is rotating so fast. As a consequence, the light cylinder is very near to the plasma injection point. Newtonian calculations are only meaningful when the light cylinder moves to infinity - or when at least all critical radii are far inside the light cylinder. In addition, gravitational effects from the background geometry are negligible in a first approximation for centrifugally driven winds. The background space time is therefore Minkowskian.

4. The Formation of Relativistic Jets

The magnetic field for axisymmetric configurations can be represented in terms of the magnetic flux function Ψ and the toroidal magnetic field B_T

$$\underline{B} = \frac{1}{R}(\nabla\Psi \wedge \underline{e}_T) + B_\phi \underline{e}_T \tag{2}$$

The motion of the plasma follows then from the equations of motion

$$T^{\alpha\beta}_{;\beta} = 0 \tag{3}$$

$$N^{\alpha}_{;\alpha} = 0 \tag{4}$$

for the total energy momentum tensor

$$T^{\alpha\beta} = (\rho + P)\, u^\alpha u^\beta - P\, g^{\alpha\beta} + T^{\alpha\beta}_{em} \tag{5}$$

together with Maxwell's equations. These equations are in general quite complicated, unless we work in the MHD-limit and assume certain symmetries. Since we are interested in equilibrium configurations for axisymmetric plasma flows, the set of the above equations can be considerably reduced. In particular, stationarity and axisymmetry require the existence of 5 constants of motion, which in the MHD-limit are also constant along the magnetic flux tubes (CAMENZIND /11/)

$$N^\beta \Omega^F_{,\beta} = 0 \qquad N^\beta \eta_{,\beta} = 0 \tag{6}$$

$$N^\beta E_{,\beta} = 0 \qquad N^\beta L_{,\beta} = 0 \qquad N^\beta S_{,\beta} = 0 \tag{7}$$

These are the angular velocity Ω^F of the field lines, which determines the light cylinder for each field line, the particle injection rate η, total energy and angular momentum and the entropy per flux-tube. The parameters $\Omega^F(\Psi), \eta(\Psi)$ and $S(\Psi)$ are given by inner boundary conditions, while the total energy and angular momentum follow from the conditions that the outflow of a wind solution passes through the Alfven point and the fast magnetosonic point. Due to the existence of 5 constants of motion, the equation of motion for the poloidal velocity u_p is integrable along a given flux-tube and can be transformed to a polynomial of degree 16 for adiabatic ion flows with $\Gamma = 5/3$ (CAMENZIND /12/)

$$\sum_{n=0}^{16} A_n\left(\frac{R}{R_L}; E, L, p_*; f/\sigma_*\right) u_p^{n/3} = 0 \tag{8}$$

Similar equations can be derived for other adiabatic indices. The flow is then essentially determined by the flux-tube function

$$f = \frac{B_{p*} R_*^2}{B_p R^2} \tag{9}$$

and Michel's magnetisation parameter

$$\sigma_* = \frac{(B_{p*} R_*^2)^2}{4\pi m c J_*(\Psi) R_L^2} \tag{10}$$

where R_* is the injection radius, B_{p*} the poloidal magnetic field and J_* the particle flux at the injection point. σ_* is the essential parameter of the problem. It determines the asymptotic Lorentz factor in a flux-tube with

$$\gamma_\infty(\Psi) \simeq \sqrt{1 + \sigma_*^{2/3}(\Psi)} \tag{11}$$

When $\sigma_* \gg 1$, we xpect highly relativistic outflows, when $\sigma_* \ll 1$ Newtonian outflows.

Fig. 1 shows the solutions of the hot wind equation for moderate σ_*-values.

The physical wind solution starts with low velocity at the injection point, crosses an unphysical branch at the Alfven point inside the light cylinder and crosses this branch a second time at the fast magnetosonic point. When the ion plasma is relatively cold (in the relativistic sense !), the fast magnetosonic point moves to infinity. In the relativistic sense, realistic ion plasmas are always cold (even at a temperature of $10^9\ K$). This means that the fast magnetosonic point is always outside of the computable region. These solutions also show that the light cylinder itself is no critical point of the wind equation. However, for fixed injection points, the Alfven point is driven towards the light cylinder with increasing σ_* -parameter. This means that the decoupling of the plasma from the rotating magnetosphere occurs very near to the light cylinder. The outflow will reach therefore a high speed.

One of the essential points of the computation of self-consistent jet equilibria is the particular form of the toroidal current j_ϕ, which determines via Ampere's equation

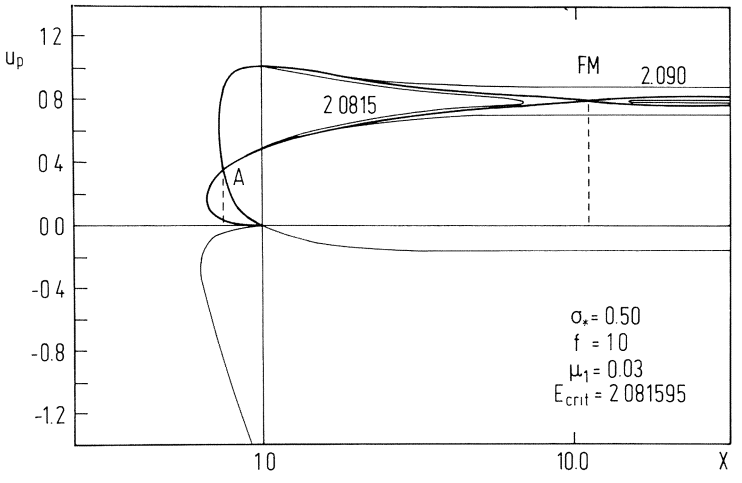


Fig. 1: The solutions of the hot wind equation

$$\nabla \wedge \underline{B}_p = \frac{4\pi}{c}\, \underline{j}_T \qquad , \qquad \underline{j}_T = j_\phi \underline{e}_T \tag{12}$$

the structure of the magnetosphere. With the introduction of the magnetic flux Ψ, this equation can be transformed to a divergence equation

$$R\nabla \cdot \left(\frac{1}{R^2}\nabla\Psi\right) = -\frac{4\pi}{c}\, j_\phi \tag{13}$$

For nonrotating magnetic structures and static equilibria this equation is well known as the *Grad-Schlüter-Shafranov* equation. In this static case, the current is simply given by two functions of Ψ

$$j_\phi = cR\,\partial_\Psi P(\Psi) + cB_\phi(\Psi)\,\partial_\Psi(RB_\phi(\Psi)) \tag{14}$$

Plasma confinement in the laboratory is based on this equation.

When plasma is moving relativistically in the magnetic field, the form of this current is quite complicated. A derivation of j_ϕ for this case has been given in CAMENZIND /12/. Apart from pressure effects, there are essentially three different components which contribute to j_ϕ

(i) inertial currents;

(ii) currents related to the toroidal field;

(iii) Goldreich-Julian currents due to the rotation of the magnetosphere.

The Goldreich-Julian current and part of the inertial currents can be absorbed by the left hand side of Ampere's equation. In this way, we obtain the relativistic version of the *Grad-Schlüter-Shafranov* equation (see CAMENZIND /12/)

$$R\nabla \cdot \left(\frac{1 - M^2 - (R/R_L)^2}{R^2}\nabla\Psi\right) = -\frac{4\pi}{c}\, j \tag{15}$$

M is here the relativistic Alfven Mach-number. For cold flows and a rigidly rotating magnetosphere the remaining current j is of a simple form

$$j = j^{(n)} + j^{(0)} \tag{16}$$

with

$$j^{(n)} = \gamma n' c R \left(\partial_\Psi E - \Omega \, \partial_\Psi L\right) \tag{17}$$

$$j^{(0)} = c R \, \frac{B_\phi^2 + M^2 B_p^2}{4\pi} \, \partial_\Psi \, ln(\eta) \tag{18}$$

The toroidal magnetic field itself follows from the integration of the equations of motion

$$R \, B_\phi = 4\pi\eta(\Psi) \, E(\Psi) \, \frac{R^2\Omega^F - L/E}{1 - M^2 - (R/R_L)^2} \tag{19}$$

The above current and the toroidal magnetic field are regular at the Alfven surface, where the relativistic Mach-number M satisfies the relation

$$M_A^2 = 1 - (R_A/R_L)^2 \tag{20}$$

The position of the Alfven surface itself follows from the total angular momentum

$$R_A^2 \, \Omega^F = L/E \tag{21}$$

It turns out that the relativistic Grad-Schlüter-Shafranov equation is elliptic in the region between the surface of the slow magnetosonic points and the surface of the fast magnetosonic points. Outside of this latter surface, this equation becomes hyperbolic. Since we consider cold ion flows, which are driven by centrifugal forces, the Grad-Schlüter-Shafranov equation is indeed elliptic on the whole integration domain accessible by numerical techniques. This situation would change when we considered electron-positron plasmas confined by rotating magnetospheres. This equation is now solved numerically with the method of *finite elements* (for more details on this question, see CAMENZIND /12/). This method is particularly suitable for the modelling of boundary surfaces of arbitrary shape. In this respect, it is important to note that the Alfven surface is an inner boundary surface, which divides the entire solution into an inner part and an outer solution.

5. Extreme Relativistic Jets

The GSS equation is now made dimensionless by introducing the total magnetic flux a_D involved in the problem

$$\Psi \rightarrow a_D \, \Psi \quad , \quad 0 \leq \Psi \leq 1 \tag{22}$$

$$X = R/R_L \quad , \quad Z = z/R_L \tag{23}$$

Then we have to solve the following equation

$$X \nabla \cdot \left(\frac{1 - M^2 - X^2}{X^2} \nabla \Psi \right) = - \frac{1}{\sigma^2(\Psi)} \, j \tag{24}$$

together with the hot wind equation (8). j is now a dimensionless current. The parameter $1/\sigma^2(\Psi)$ acts as the effective coupling constant between the magnetic flux and its source current j with

$$\sigma(\Psi) = \frac{a_D\, c}{4\pi\mu\eta(\Psi)\, R_L^2} \tag{25}$$

This parameter is similar to the magnetisation parameter σ_*, which determines the properties of the flow. Both are related by

$$\sigma_* = \sigma(\Psi)\, B_{p*} R_*^2 / a_D \tag{26}$$

In particular, when no particles are injected into the magnetosphere, $\eta \to 0$, this coupling constant vanishes. This is the extreme relativistic limit with no effective coupling.

We studied solutions of the extreme relativistic limit for a constant injection law, $\eta = \eta_0 = const$, and therefore also with $\sigma = \sigma_0 = const$. In this case, the mass flux contained within the flux surface Ψ is then given by

$$\dot{M}_j(\Psi) = \frac{a_D^2}{2c\,\sigma_0\, R_L^2}\,\Psi \tag{27}$$

In this way, we obtain an estimate for the mass flux in the jets of quasars as a function of σ_0

$$\dot{M}_j(\Psi) = 0.26\, M_O\, yr^{-1}\, a_{D,34}^2\, R_{L,16}^{-2}\, \frac{\Psi}{\sigma_0} \tag{28}$$

The total magnetic flux involved in jet formation for strong quasars is typically of the order of $10^{34}\, Gauss\, cm^2$ as determined from the synchrotron emission of the kpc-jets. For $\sigma_0 \gg 1$, the mass flux in the jets is extremely low, while for $\sigma_0 \simeq O(1)$ we would get heavy jets with low final velocities.

The collimation of the outflow is also a question of the σ-parameter. At present, this question can only be studied numerically by investigating self-consistent solutions of the GSS equation and of the wind equation. We found that in the extreme relativistic limit collimation is achieved already within a distance of about $100\, R_L$ for $\sigma_0 = 100$. Since in this limit the Alfven surface coincides practically with the light cylinder, the magnetosphere is heavily squeezed inside the light cylinder so that the field lines cross the light cylinder at right angles. This means that collimation can only start outside of the light cylinder and takes at least a few tens of light cylinder radii for this process. The light cylinder radius is therefore the crucial quantity for the scale of the collimation process. This is in contrast to many models discussed in the literature, where jets are cone-like emanating from a black hole e.g. The above conclusions concerning the *relatively soft jet formation* around supermassive objects are not dependent on the particular nature of this object, which generates the magnetosphere. Jet formation in the magnetosphere of an accretion disk shows similar global properties.

6. Observational Consequences

The overall spectra of strong quasars find a natural interpretation in terms of our models. Besides the central UV-source and the X-rays emitted by Comptonized UV-photons, the nuclear jet will produce a quasi-stationary self-absorbed synchrotron spectrum. In 3C 273 this self-absorption frequency is in the range of 50 GHz and the optically thin synchrotron emission has a spectral index of -0.67. This emission must be related to a kind of a beam dump, where

much of the outflowing energy is already transformed to relativistic particles in shocks around the central parsec. It is important to note that this synchrotron component is not beamed in our model.

Magnetized jets are subject to various instabilities, which can be classified in terms of the azimuthal mode number m. The lowest instability with $m = 0$ is the pinch instability producing knots moving with the plasma along the nuclear jet. They could be the origin of the moving knots seen by VLBI in superluminal sources such as 3C 273 and 3C 345. Emission from those knots is strongly beamed in the forward direction and therefore only visible in jets seen under low inclination angle. There must be many quasars with stationary synchrotron spectra in the mm-region but only small contributions from moving knots in the cm-range. The kink instability with $m = 1$ is of no less importance, since it could deform the nuclear jet from a straight line into a more helical structure without disruption.

The growth of instabilities in relativistic magnetized jets has not yet been analysed. One cannot simply scale Newtonian results into the extreme relativistic regime. I hope that this fact has been demonstrated with the jet formation in rapidly rotating magnetospheres. When rapid rotation is involved in MHD, the asymptotic limit of this relativistic MHD is not simply Newtonian MHD. This is a crucial fact. It means in particular that we are not able to calculate the propagation of these relativistic jets with presently available numerical schemes. The development of time-dependent relativistic MHD codes is a challenge for the future.

7. Literature

1. A.H. Bridle, R.A. Perley: Ann. Rev. Astron. Astrophys. **22**, 319 (1984)
2. J.A. Biretta, R.L. Moore, M.H. Cohen: Astrophys. J. **308**, 93 (1986)
3. M. Camenzind: Astron. Astrophys. **156**, 137 (1986)
4. M. Malkan, W.L.W. Sargent: Astrophys. J. **254**, 22 (1983)
5. M. Malkan: Astrophys. J. **268**, 582 (1985)
6. M. Camenzind, T.J.-L. Courvoisier: Astron. Astrophys. **140**, 341 (1984)
7. R.A. Edelson, M. Malkan: Astrophys. J. **308**, 59 (1986)
8. M. Camenzind: in **Interstellar Magnetic Fields**, ed. by R. Beck and R. Gräve, Springer Proceedings (Springer, Berlin, Heidelberg, 1987) p. 229
9. R.D. Blandford, D.G. Payne: Mon. Not. Roy. Astron. Soc. **199**, 883 (1982)
10. R.V.E. Lovelace, J.C.L. Wang, M.E. Sulkanen: Astrophys. J. **315**, 504 (1987)
11. M. Camenzind: Astron. Astrophys. **162**, 32 (1986)
12. M. Camenzind: Astron. Astrophys. **184**, 341 (1987)

Accretion Disks in the Centers of Galaxies

W.J. Duschl

Max-Planck-Institut für Physik und Astrophysik, Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-8046 Garching, Fed. Rep. of Germany

We discuss the problems of α-accretion disks in soft potential wells of the centers of galaxies. If matter is fed into the disk only from large radii, accretion is very inefficient, as there is no shear that could drive it. If on the other hand matter falls into the disk at all radii, the disk may become so massive that it can change the potential, and make accretion possible. We show that such a process is capable of explaining the observed massive objects in the centers of normal galaxies. Similar processes may also be of relevance for the explanation of active galactic nuclei.

1. Introduction

In α-accretion disks (SHAKURA and SUNYAEV [1]) the mechanism that drives the radial transport of mass and angular momentum is a turbulent viscosity that acts in a shearing medium. A geometrically thin gas disk placed in a soft potential well has a small radial gradient of angular velocity; under such a condition there is practically no shear that could drive an efficient accretion. ICKE [2] has shown that in such an environment one hardly can speak of an accretion disk as there is almost no radial transport of angular momentum and mass. This is true as long as the mass is put into the accretion disk only from its outer edge and/or the mass of the disk is negligible compared to the respective background mass that produces the soft potential well. In the following we show that the situation changes as soon as matter falls into the disk at all radii and thus the disk is capable of becoming very massive. We shall discuss the evolution of such disks in the centers of normal galaxies. After a time scale of 10^{10} years there has formed a central body of the order of 10^6 $M\odot$.

2. Disk Accretion – The Problem

A detailed description of the theoretical treatment of accretion disks in soft potential wells will be published by DUSCHL [3] .

In this Section we describe the problems of disk accretion in such an environment. The basic problem results from the fact that "soft potential" means

$$\lim_{s \to 0} \frac{d\,\Phi\,(s)}{d\,s} = 0\,; \tag{1}$$

where Φ is the potential, and s the radial coordinate in a cylindrical system. Equation (1) is equivalent to

$$\lim_{s \to 0} \Omega\,(s) = 0\,, \tag{2}$$

with the angular velocity Ω.

In the following we regard a azimuthally symmetric, geometrically thin, isothermal α-accretion disk. The evolution of the surface density, Σ, is described by a diffusion-type equation (e.g. PRINGLE [4]):

$$\frac{\partial \Sigma}{\partial t} + \frac{\alpha \cdot v_c}{s} \cdot \frac{\partial}{\partial s}\left[\left(\frac{d\, s^2 \cdot \Omega}{d\, s}\right)^{-1} \cdot \frac{\partial}{\partial s}\left(h \cdot \Sigma \cdot s^3 \cdot \frac{d\, \Omega}{d\, s}\right)\right] = \dot{\Sigma}_i \ . \tag{3}$$

$\dot{\Sigma}_i$ is the mass infall into the disk from outside, v_c the sound velocity, and h the vertical scale height. From (3) one can define an accretion time scale, τ_{accr}, that describes the typical time over which matter moves towards the center due to the accretion process described in (3):

$$\tau_{accr} = \frac{1}{\alpha \cdot h \cdot v_c} \cdot \left| \frac{\left(\frac{d\, s^2 \Omega}{d\, s}\right)}{\left(\frac{d\, \Omega}{d\, s}\right)} \right| . \tag{4}$$

In the case of a non-selfgravitating Keplerian accretion disk this definition leads to the wellknown expression (see e.g. KIPPENHAHN and THOMAS [5] ; they call this time scale $\tau_{visc,s}$)

$$\tau_{accr,K} = \frac{s^2 \cdot \Omega}{3 \cdot \alpha \cdot v_c^2} \ . \tag{5}$$

But if – as in the present case – the radial derivative of the angular velocity vanishes this is equivalent to

$$\lim_{s \to 0} \tau_{accr} = \infty \ , \tag{6}$$

which means that accretion cannot transport matter through the disk towards the center. This is the case as discussed by ICKE [2] .

A different scenario was introduced by BAILEY [6] ; starting from some of the basic ideas that he discussed, in the next Section we introduce a possible solution of this problem.

3. A Possible Way Out

The typical situation that leads to the problem as discussed in Section 2. is the following: An accretion disk is placed in a soft potential well, and is fed from its radially outer edge with matter; due to the very long accretion time scale this matter practically does not move in radial direction, i.e. actually no accretion takes place. To overcome this problem one has to find a possibilty to get – at least – some of the matter into the inner regions of the accretion disk in another way. At the beginning this does not help either as the potential there is also flat. But if one assumes that the mass infall into the accretion disk at different radii is a continuous process one has to take into account another fact, namely that the accretion disk may become so massive that it is no longer negligible compared to the local background mass that defines the soft potential well. Now there is another component that also defines the potential, namely the contribution from the accretion disk's mass, and – in later stages of the evolution – a central mass. Altogether one now gets a potential, Φ, of the following form:

$$\Phi(s) = \Phi_s(s) - \frac{G \cdot (M_c + M_d(s))}{s} \ ; \tag{7}$$

Φ_s is the original soft potential, M_c the central mass, and $M_d(s)$ the disk's mass within the radius s. For simplicity here the disk mass is assumed to be spherically distributed; this approximation

introduces only negligible errors. One finds that there is the possiblity that the second, Keplerian term in the potential is dominant. Then one no longer has a rigidly rotating gas disk but a real accretion disk where shear plays the required role. We shall find that the Keplerian term first dominates in the very center of the potential, later on the size of the accretion zone grows larger and larger and thus involves more and more of the accretion disk. When this situation is reached the matter within the accretion zone is redistributed in the following sense: as in the zone outside there is no accretion yet, only mass from within plus the matter just falling in can be accreted: matter flows towards the center of the accretion disk thus depleting the accretion disk. How large a fraction of the matter within the accretion zone is involved in this process depends on the accretion time scale. One can approximate to a fairly good degree the mass flow rate by the disk's mass divided by the accretion time scale:

$$\dot{M}(s) = \frac{M_d(s)}{\tau_{accr}(s)} \,. \tag{8}$$

In this sense, $\dot{M}(s)$ is a mass flow rate averaged over the accretion zone in the disk. The driving mechanism for the growing of the accretion zone's radius is the further infall of matter into the accretion disk.

While this is a process that would be limited only by the amount of matter stored in the disk, and on the other hand would be continuously intensified by the increase of the radius of the accretion zone, there also exists a counteracting mechanism that makes accretion less efficient, namely selfgravitation. When the mass of the disk is no longer negligible compared to the background mass one also reaches at a similar time the situation that in the vertical direction the selfgravitation of the disk dominates over the gravitation of the background mass. As we shall show later, selfgravity increases the accretion time scale.

The onsets of accretion and selfgravity take place at similar phases in the evolution of the disk but not necessarily at the same time, as the onset of accretion involves the total mass of the disk within the respective radius while for selfgravity a comparison of the local gravitational forces is relevant. In the next Section we shall discuss this difference more quantitatively.

4. Normal Galaxies – Theory

For simplicity we assume the galaxy to be spherically symmetric; its density distribution $\rho(r)$ (r: radial coordinate in a spherical system) is given. Via Poisson's Equation one gets the potential, $\Phi(r)$. DUSCHL [7] has shown that the density distribution has to fulfill the condition

$$\lim_{r \to 0} \frac{d \, \log \rho(r)}{d \, \log r} \geq -1 \tag{9}$$

to give a soft potential, i.e. to fulfill (1). One furthermore gets the equilibrium velocity, v_ϕ, for this potential. We also assume an axially symmetric rotation law for the galaxy, v_*, thus implicitly introducing a velocity dispersion.

The spherical symmetry of the galaxy, together with the axial symmetry of the rotation curve defines a symmetry plane of the galaxy. The accretion disk that we shall regard in the following will build up in this plane. As we shall have geometrically thin accretion disks, and shall use the thin

disk approximations (see e.g. PRINGLE [4]), we shall describe the disk in a cylindrical coordinate system; quantities originally defined in the spherical system but now used in the cylindrical one are meant in the sense that one takes them only at small poloidal angles and replaces r by s.

The mass distribution of the galaxy, ρ, is assumed to loose matter at a relative rate $\dot{\mu}$. This has to be understood in the following way: a) $\dot{\mu}$ is taken with respect to the total density, i.e. to all forms of mass in the galaxy, be it stars, be it gas ...; $\dot{\mu}$ has to be taken such that $\rho \cdot \dot{\mu}$ describes the possible mass loss; b) the time dependence of $\dot{\mu}$ describes the evolution of the stellar component of the galaxy's mass as far as it is of relevance for the problems discussed here. With respect to the actual value of ρ we assume that $\dot{\mu}$ is so small that the mass distribution in the galaxy does not change during the regarded evolution of the disk; and c) $\dot{\mu}$ is not the actual mass loss rate from the star but is taken as an effective rate with respect to the accretion disk, i.e. $\dot{\mu}$ describes only that part of the lost matter that actually contributes to the mass infall into the disk.

For the viscosity we take the usual α-ansatz (SHAKURA and SUNYEAV [1]), i.e. we describe the dynamical viscosity, ν, by

$$\nu = \alpha \cdot v_c \cdot h \,, \tag{10}$$

and assume $\alpha = 1$; this value has proven as a good approximation in modelling accretion disks in dwarf novae (MEYER and MEYER-HOFMEISTER [8, 9]) and symbiotic stars (DUSCHL [10, 11]). We assume the disk to be isothermal at a temperature T; as BAILEY [6] has shown this is an acceptable approximation. The scale height, h, is given by (see e.g. PRINGLE [4] , and PACZYŃSKI [12])

$$h = \begin{cases} \frac{v_c}{\Omega}, & \text{if selfgravity is negligible,} \\ \frac{v_c^2}{2 \cdot \pi \cdot G \cdot \Sigma}, & \text{if selfgravity dominates.} \end{cases} \tag{11}$$

Selfgravity dominates as soon as the surface density becomes larger than some critical value, Σ_{cr}; this critical value is determined by setting the vertical gravitational acceleration due to the galaxy's background mass, the central mass, and the disk's mass, g_M, equal to the acceleration due to the local surface density, g_Σ:

$$g_M(s) = \frac{G \cdot (M_c + M_d(s) + M_G(s)) \cdot h(s)}{s^3} \,, \tag{12a}$$

$$g_\Sigma(s) = 2 \cdot \pi \cdot G \cdot \Sigma(s) \,. \tag{12b}$$

Here $M_G(s)$ is the galaxy's mass within a sphere of radius s. From (12) one gets the value of Σ_{cr}.

We assume that the mass lost by the matter of the galaxy falls into the accretion disk conserving its angular momentum. So one can calculate a rate of mass infall per area into the accretion disk, $\dot{\Sigma}_G$.

In the next Section we shall discuss numerical models for normal galaxies.

5. Normal Galaxies – Models

5.1. The Galaxy

We assume a density distribution, $\rho(r)$, fulfilling condition (9) for soft potentials:

$$\rho(r) = \rho_0 \cdot \left(1 + \left(\frac{r}{R}\right)^2\right)^{-3/2} . \tag{13a}$$

ρ_0, and R, are scaling factors; we take:

$$\rho_0 = 1.9 \cdot 10^{-20} \text{ g/cm}^3, \tag{13b}$$

$$R = 6.2 \cdot 10^{20} \text{ cm} . \tag{13c}$$

For the outer radius of the galaxy we assume

$$r \leq 50 \cdot R . \tag{13d}$$

The rotation law, $v_*(s)$ is defined as follows:

$$v_*(s) = \begin{cases} 1.0 \cdot 10^7 \text{cm/s} \cdot (0.2 \cdot s/R), & \text{if s} \leq 5 \cdot \text{R,} \\ 1.0 \cdot 10^7 \text{cm/s}, & \text{if s} > 5 \cdot \text{R.} \end{cases} \tag{14}$$

Thus we have a galaxy with an outer radius of 10 kpc, a central density of 280 M$\odot$ /pc^3, a total mass of $1.0 \cdot 10^{11}$ M$\odot$, and a rotation velocity that increases linearly from 0 to 100 km/s within the innermost kpc to stay constant for larger radii.

5.2. Mass Loss

We assume a mass loss rate that is constant in time; for its value we take

$$\dot{\mu} = 7.0 \cdot 10^{-20} \text{ sec}^{-1}, \tag{15}$$

which is equivalent to a total mass infall rate into the disk from the entire galaxy of $2.1 \cdot 10^{-1}$ M$\odot$ /year. The radial distribution of the mass infall rate per area for different inner masses, M_i, is shown in Fig. 1.; the "inner mass" is defined by

$$M_i(s) = M_c + M_d(s) . \tag{16a}$$

There not the actual value $\dot{\Sigma}_i$ is given but a dimensionless one, $\dot{\sigma}_i$, which is defined as follows:

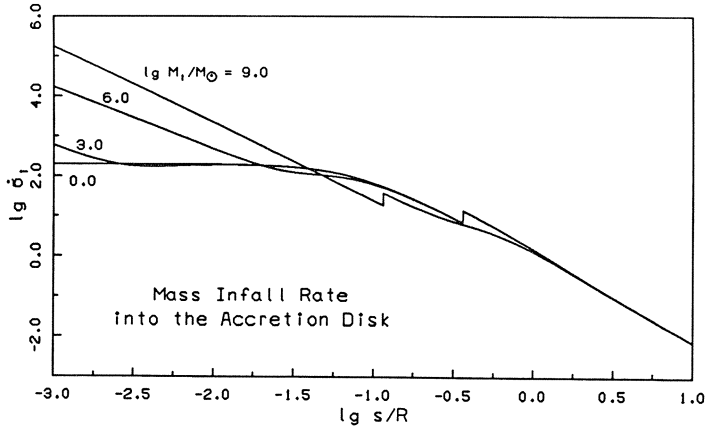


Figure 1: The dimensionless mass accretion rate, $\dot{\sigma}_i$, as a function of the radius, s, for different inner masses, M_i.

$$\dot{\sigma}_i = \frac{\dot{\Sigma}_i}{2 \cdot \dot{\mu} \cdot \rho_0 \cdot R} \cdot \tag{16b}$$

The discontinuity of $\dot{\sigma}_i$ around log $(s/R) \approx$ -1.0 reflects the discontinuity of the radial derivative of the rotation curve (14).

5.3. Accretion Disk

We asssume the temperature of the disk, T, to be 30 K; BAILEY [6] has shown that this is a reasonable value.

5.4. Evolution of the Standard Model

The quantities defined in the previous Subsections describe a "Standard Model" of a normal galaxy. Fig. 2 shows the evolution of the central mass, M_c, and the mass accretion rate, $\dot{M}$, from t = 10^7 to 10^{10} years. In Fig. 3 the evolution of the accretion time scale, τ_{accr}, is given for the same period during the evolution.

The evolution shown in Fig. 2 is characterized by the interaction of three processes: At early times the accretion sets in while selfgravity not yet is capable of dominating it, then the accretion rate – and correspondingly the central mass – is strongly increasing. Later selfgravity dominates. Then $\tau_{accr} \propto \Sigma$ ((4) and (11)); this means that the accretion rate decreases as time goes on as the local surface densities become larger. But as a counteracting process also the disk's accretable mass increases; this increases $\dot{M}$; the latter one is the domain in which the disk is in the later phases of the evolution shown in Fig. 2.

Two approximations may at the beginning seem to be quite arbitrary ones, namely a) to assume that the viscosity parameter $\alpha = 1$, and b) that the disk is isothermal at 30 K.

DUSCHL [3] has shown how the evolution depends on the particular choices; the result is shown in Fig. 4 (Fig. 4, as well as the others of this paper, follow the presentation in DUSCHL [3]). One finds that the evolution depends almost linearly on α, and goes approximately as $\sqrt{T}$; both results can be understood from the definition of the accretion time scale in (4).

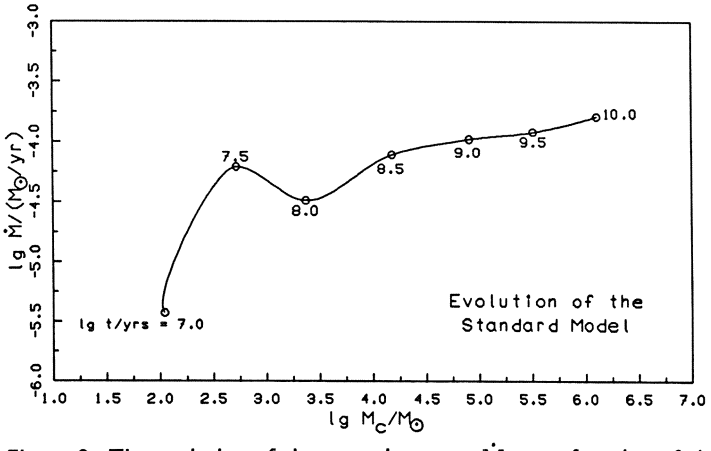


Figure 2: The evolution of the accretion rate, $\dot{M}$, as a function of the central mass, M_c, and of time, t.

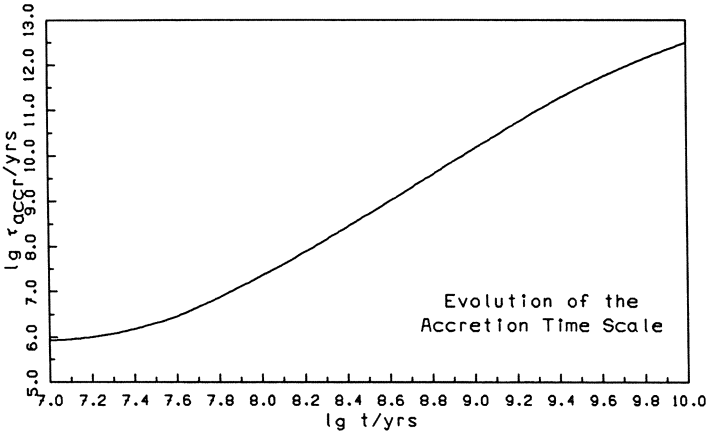


Figure 3: The evolution of the accretion time scale, τ_{accr}, as a function of time, t.

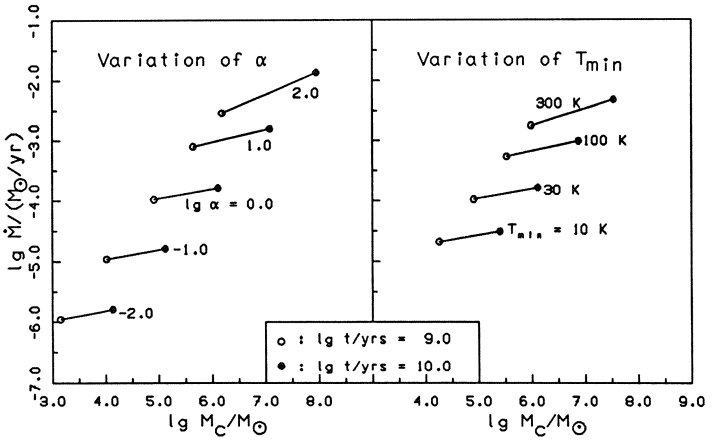


Figure 4: The evolution of the central mass, M_c, and the accretion rate, $\dot{M}$, from $t = 10^9$ to 10^{10} years for different values of the viscosity parameter, α, and the disk's temperature, T.

Although the value of α cannot be determined from a real theoretical description of turbulence as yet, one can obtain information about it from the observed accretion disks in close binaries (see above) and finds that $\alpha = 1$ is a reasonable choice in those systems.

As one sees from (4) and (11) α and T do not influence the timescale independently from each other; the really relevant quantity is a product of α and some power of T – depending on whether selfgravity dominates or not.

6. Summary

The results presented in the previous Section shows that one can typically expect a central mass of the order of 10^6 M⊙ with an accretion rate of 10^{-3} to have evolved in the center of a normal galaxy within a time span of the order of 10^{10} years.

This result is in good agreement with observations of the central regions of different systems, as e.g. our galaxy (GENZEL and TOWNES [13]), the Andromeda galaxy, M31, (KORMENDY [14]),

or even the compact companion galaxy to M31, M32, (TONRY [15]). All these observations point towards a central object of 10^6 M$\odot$; additional observations in our galaxy (LO and CLAUSSEN [16]) show that also the mass inflow rate is of a reasonable order of magnitude.

The model presented here overcomes the problems of the ones that explain the formation of a central object by the collapse of the galaxy's core; That process would need too long a time scale to explain what we observe today. (CHANDRASEKHAR [17] , SPITZER and HART [18] , and KORMENDY [19]).

6. Acknowledgements

I thank Drs. Friedrich Meyer and John Kirk for carefully reading the manuscript; it benefitted much from their suggestions.

References

1. N.I. Shakura, R.A. Sunyeav: AstronAstrophys 24, 377 (1974)
2. V. Icke: AstronAstrophys 78, 21 (1979)
3. W.J. Duschl: AstronAstrophys, in press (1987)
4. J.E. Pringle. AnnRevAstronAstrophys 19,137 (1981)
5. R. Kippenhahn, H.-C. Thomas, AstronAstrophys 114, 77 (1982)
6. M.E. Bailey: Thesis, University of Edinburgh, Scotland, U.K. (1978)
7. W.J. Duschl: in: 10^{th} European Regional Astronomy Meeting, Prague, in press (1988)
8. F. Meyer, E. Meyer-Hofmeister: AstronAstrophys 104, L10 (1981)
9. F. Meyer, E. Meyer-Hofmeister: AstronAstrophys 132, 143 (1984)
10. W.J. Duschl: AstronAstrophys 163, 56
11. W.J. Duschl: AstronAstrophys 163, 61
12. B. Paczyński: ActaAstronomica 28, 91
13. R. Genzel, C.H. Townes: to be published in AnnRevAstronAstrophys 25 (1987)
14. J. Kormendy: in Supermassive Black Holes (Ed.: M. Kafatos), Cambridge University Press, Cambridge, UK. (1987)
15. J.L. Tonry: AstrophysJ 283,L27 (1984)
16. K.Y. Lo, M.J. Claussen: Nature 306, 647 (1987)
17. S. Chandrasekhar: Principles of Stellar Dynamics, Dover Press, New York, NY, USA (1960)
18. L. Spitzer, M.H. Hart: AstrophysJ 164, 399
19. J. Kormendy: in Supermassive Black Holes (Ed.: M. Kafatos), Cambridge University Press, Cambridge, UK. (1987)

Synchrotron Emission from Shock Waves in Active Galactic Nuclei

P.L. Biermann

Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69,
D-5300 Bonn, Fed. Rep. of Germany

The origin of the sharp near infrared cutoff in the continuous energy distribution of many compact non-thermal sources (BL Lacs, OVVs, red quasars and certain jets) is considered under the assumption that particle acceleration takes place in shocks. This model predicts a highest frequency ν^* of electron synchrotron emission which depends principally on the shock velocity and the ratio a of photon to magnetic energy density in the acceleration region. For near relativistic flows and reasonable values of a a spectral cutoff is predicted in the range $3\ 10^{14} < \nu < 2\ 10^{15}$ Hz. The model thus leads to 1) near relativistic flows, 2) a gradual steepening of optical continuum spectra as one follows a jet outwards, 3) a correlation between spectral hardening and luminosity, 4) a correlation between optical spectrum and X-ray emission, 5) a possible synchrotron contribution to the X-ray emission in Quasars from secondary particles, and 6) the production of very high energy particles such as observed in cosmic rays, of up to $\sim 10^{12}$ GeV.

1. Introduction

In a series of papers, Rieke and his collaborators (Rieke et al. 1976, 1979, Rieke and Lebofsky 1980, Bregman et al. 1981, Rieke et al. 1982) and Sitko et al. 1983 have demonstrated that many BL Lacs, OVVs and red quasars show a sharp cutoff in their continuous emission spectrum near $3\ 10^{14}$ Hz. A similar cutoff has been found in the emission of jets and hot spots (M 87: Stocke et al. 1981; Cen A: Brodie et al. 1983; 3C273: Röser and Meisenheimer 1986; 3C33: Simkin 1978, 1986, Meisenheimer and Röser 1986, Crane et al. 1987; PKS0521-36: Keel 1986, Cayatte and Sol 1987). In Pictor A a cutoff is consistent with the data, but not seen (Röser and Meisenheimer 1987). It is generally supposed that the radiation arises from incoherent electron synchrotron emission.

Substantial efforts have been made to account for this cutoff phenomenon. Webb et al. (1984), discussed the competition between acceleration and synchrotron losses within the framework of diffusive shock acceleration theory (Axford 1981, Drury 1983). Further discussions have been given by Pérez-Fournon (1984, 1985), Bregman (1985), Björnsson (1985), Pérez-Fournon and Biermann (1986), Meisenheimer and Heavens (1986), Heavens and Meisenheimer (1987), and, from a different point of view, by Schlickeiser (1984). Consideration of synchrotron losses leads to an upper bound to possible electron energy and hence, potentially, to a spectral

cutoff. A detailed theory for the cutoff has been given by Biermann and Strittmatter (1987 a). The present discussion follows that paper and extends the results.

2. The Proposed Model

The most difficult problem in making detailed predictions from shock acceleration theory is the determination of the particle mean free path. This mean free path determines the rate of gain of energy. The mean free path λ is normally computed in the small angle, resonant scattering approximation (Drury 1983) where the particle deflection is assumed to be dominated by Alfvén waves (i.e. the turbulent magnetic field) with wavelength equal to the gyroradius of the particle concerned. In these circumstances λ is given by $\lambda = r_g \, (B^2/8\pi)/(I(k)k)$, where r_g is the gyration radius of the particle under consideration, B the strength of the magnetic field, and I(k) the magnetic energy density per unit wave number k in the turbulent magnetic field. The problem of determining λ thus reduces to that of specifying the spectrum of turbulence.

2.1 The Plasma Turbulence on Large Scales

We need to determine the turbulence on the scale of the gyroradius of the particles concerned, i.e. of order 100 MeV and higher energies for electrons and of order GeV and higher for protons. This corresponds to $\gtrsim 3\ 10^5\ B^{-1}$ cm (B in Gauß), which is $\gtrsim 7.10^9$ cm in the solar wind near Earth ($B \simeq 50\ \mu G$), $\gtrsim 1.10^{11}$ cm in the interstellar medium ($B \simeq .3\ \mu G$), $\gtrsim 3.10^9$ cm in jets (M87: $B \simeq 10^{-3}$ G, Stocke et al. 1981), and $\gtrsim 3\ 10^5$ cm in active nuclei cores ($B \simeq 1$ G, Angel and Stockman 1980).

A self consistent (though possibly not unique) specification of the turbulence may be given, as follows. For strong shocks, the predicted particle spectrum implies that the particle energy is contained primarily in the most energetic particles (i.e. the differential particle energy spectrum is $N(E) \propto E^{-p}$ with $p \neq 2$). Furthermore, since energy loss mechanisms are more efficient for electrons than for protons (see Section 3) the highest energies will be achieved by the protons.

Since the highest energy protons have the greatest mean free path and hence the largest upstream range, it is reasonable to assume that the upstream perturbation to the flow is caused by these particles and results in the onset of turbulence. The usual turbulent cascade process then results in a turbulent energy spectrum at successively smaller scales. Dimensional arguments and solar wind observations (Matthaeus and Goldstein 1982, Matthaeus et al. 1982, Smith et al. 1983, Goldstein et al. 1984, Montgomery 1986, Montgomery et al. 1986) provide strong constraints for a fully developed turbulent cascade in a plasma where the magnetic field has about the same energy density as (or less than) the thermal energy density. The solar wind in situ measurements show a very low wavenumber range with a k^{-1} spectrum, usually interpreted as an inverse cascade, and an extended range with a $k^{-5/3}$ spectrum. The wavenumbers over which the $k^{-5/3}$

spectrum is observed correspond to wavelengths of about 3 10^9 to 10^{12} cm, just the scale required above. The data are not consistent with a $k^{-3/2}$ spectrum; such a spectrum has been derived by Kraichnan (1965) for the case when the magnetic field strongly dominates over thermal gas pressure. Very long baseline observations of H_2 O-Maser spots at 22 GHz will soon be able to discern between Kraichnan and Kolmogorov turbulence in the interstellar plasma. Comparisons of the data with quantitative model calculations favor the Kolmogorov turbulence spectrum (Gwinn et al. 1987). Interstellar plasma turbulence is already established to be consistent with a powerlaw of Kolmogorov-type over the scales 10^{10} to at least 10^{15} cm (Rickett et al. 1984), again consistent with the above requirement. For near equipartition of energy densities (Montgomery 1986) the Kolmogorov argument can be carried through also in the MHD case and thus suggests that a $k^{-5/3}$ spectrum may be quite common in astrophysical environments.

Hence it seems reasonable to assume that the inertial range spectrum (energy density per unit wave number) of MHD turbulence is of Kolmogorov type just as for hydrodynamical turbulence (Kolmogorov 1941, Heisenberg 1948, Sagdeev 1979). In the following discussion, we therefore assume the turbulent energy spectrum to be of the form $I(k)=I_o/k_o\ (k/k_o)^{-ß}$ with ß=5/3. Here $2\pi/k_o$ ($=r_{g,max}$, the gyration radius of the most energetic protons) corresponds to the outer scale of turbulence.

2.2 The Acceleration Time (parallel shocks)

We define b as the ratio of turbulent to ambient magnetic energy density. For strong shocks the acceleration time is then given (Drury 1983) by

$$\tau_{acc} \simeq E/(dE/dt) \simeq \frac{80}{3\pi}\left(\frac{c}{U^2}\right)\left(\frac{r_g}{b(ß-1)}\right)\left(\frac{r_{g,max}}{r_g}\right)^{ß-1} \qquad (1)$$

where U is the upstream velocity in the flow and where $r_g=\gamma mc^2/eB$ is again the radius of a gyration of the particle in question, m is the particle mass and e the electron charge. (Note that for ß≃5/3, the highest energy particles do indeed have the greatest λ).

3. The Cutoff Frequency

3.1 Acceleration with Synchrotron Losses only

First we consider particle acceleration in the presence of synchrotron losses only. The synchrotron loss time τ_{syn} for protons or electrons respectively is, e.g., given by Rybicki and Lightman (1979).

The upper bound to the proton energy $\gamma_{p,max}$ is then obtained by setting $\tau_{p,syn}=\tau_{acc}$. The analysis may then be repeated setting $\tau_{e,syn}=\tau_{acc}$ for the electrons. This yields an upper limit ν^* to the frequency of electron synchrotron emission (averaging over all possible aspect angles). With ß≃5/3, we obtain $\nu^*=3\ 10^{14}\ (3b\ (U/c)^2)$ Hz.

On physical grounds it appears likely that b≤1. Similarly, Webb (1985 a, b, c, 1987) has recently shown that high energy particle acceleration is only effective for $(U/c)^2 < 1/3$. It therefore appears that 3 10^{14} Hz is a strong upper limit to ν^*.

3.2 Acceleration with Synchrotron and Photon Interaction Losses

In many astrophysically important situations, the ambient photon density is sufficiently high that photon interactions become a significant energy loss mechanism for accelerated particles. For electrons, the inverse Compton process becomes important once the photon energy density approaches that in the magnetic field. For protons, the situation is somewhat more complicated since energy losses in inelastic proton photon collisions dominate at high photon densities. Stecker (1968) has given the loss time $\tau_{p\gamma}$ for proton–photon interactions as

$$\frac{1}{\tau_{p\gamma}} = \int_{\varepsilon_{th}/2\gamma_p}^{\infty} d\varepsilon\ n(\varepsilon)\ \frac{c}{2\gamma_p^2\ \varepsilon^2} \int_{\varepsilon_{th}}^{2\gamma_p\varepsilon} k_p(\varepsilon')\ \ \sigma(\varepsilon')\ \ \varepsilon'\ \ d\varepsilon' \qquad (2)$$

where $n(\varepsilon)$ is the number density of photons per unit energy interval for photons of energy ε, ε_{th} is the threshold energy for inelastic collisions, $k_p(\varepsilon')$ is the inelasticity, and $\sigma(\varepsilon')$ is the cross section in the relativistic proton frame.

For purposes of illustration we adopt a photon spectrum suggested by observations of 3C273 and Cen A (Bezler et al. 1984, Ballmoos et al. 1987) namely of $n(\varepsilon) = N_0/\varepsilon_0\ (\varepsilon/\varepsilon_0)^{-2}$ for $\varepsilon_0 \leq \varepsilon \leq \varepsilon^*$, and $n(\varepsilon) = 0$ otherwise, where ε_0, ε^* correspond to radio and X-ray energies respectively. With the further assumption that $\varepsilon_0 < \varepsilon_{th}/2\gamma_p$ equation (2) can be simplified. The integral in equation (2) can be rewritten as the sum over the different interaction channels (cf. Cambridge Bubble Chamber group 1967, Armstrong et al. 1972) to give an average cross section $\overline{\sigma_{\gamma p}}$. We note that p–p losses as well as pair creation losses should properly be included in the sum. However, we limit ourselves here to the dominant pion producing interaction channels ($\varepsilon_{th} \approx m_\pi\ c^2$). The average cross section $\overline{\sigma_{\gamma p}}$ is about 900 μb with an estimated error of 20%.

We define a as the ratio of photon to magnetic energy density, and

$$A = \frac{\overline{\sigma_{\gamma p}}}{\sigma_T} \frac{(m_p/m_e)^2}{\ln\ (\varepsilon^*/\varepsilon_0)} \approx \frac{\overline{\sigma_{\gamma p}}}{\sigma_T}\ \ 1.6\ 10^5 \approx 200 \qquad (3)$$

We emphasize that the value A is only weakly dependent on the properties of the source. Repeating the analysis gives a maximum electron synchrotron cutoff frequency, for ß = 5/3, of

$$\nu^* = 3\ 10^{14} \left[3\ b\ \left(\frac{U}{c}\right)^2\right]\ \ f(a)\ \text{Hz} \qquad (4)$$

where $f(a) = (1+Aa)^{1/2}/(1+a)^{3/2}$. ν^* depends strongly on the value of ß, changing by about a factor of ten for changing ß to 5/3 ± 0.1. For A = 200, the function

f(a) is close to unity for aA << 1, has a maximum of 2/3 $(A/3)^{1/2} \simeq 5.4$ at a ≃ 0.5, and monotonically decreases for larger values of a. Very large values of a are, however, unlikely to occur in practice because of very large implied luminosity at high photon energies.

4. Consistency Tests

The theory of particle acceleration in shocks thus predicts an upper bound to the frequency of electron synchrotron emission which is in the range of infrared or visible wavelengths. This appears to be consistent with the observations of many compact nonthermal sources such as BL Lac objects. The general model also makes other predictions which provide for the checks for consistency (for details see Biermann and Strittmatter 1987a).

4.1 Length Scales

It is required that the outer scale of turbulence ($\sim r_{g,max}$) should not exceed the scale of the system and that the time scales for loss or acceleration should not exceed the variability timescales observed.

This condition leads to a lower limit on the magnetic field strength compatible with the observations for the M 87 jet as well as active nuclei.

4.2 Time Scales and Variability

Variability has been observed in the M87 jet (see Sulentic et al. 1979 and Warren-Smith et al. 1984) with a time scale of order 10 years. In the active nuclei time scales of 3 10^4secs are suggested by the observations (Angel and Stockman 1980). Again this is consistent with our model.

4.3 Secondary Pair Production from p-p Collisions

We also have to check whether secondary pair-production from pions generated in proton-proton collisions dominates over primary electron acceleration, as claimed by Zdziarski (1986). Losses due to proton-proton collisions occur with the timescale $\tau_{pp} \simeq 1/\sigma_{pp}\, n_T\, c$ where σ_{pp} (~40 mb) is the p-p collision cross-section and n_T the number density of thermal protons. p-p collisons as a loss process are thus negligible if the thermal density obeys

$$n_T \prec 1.4\ 10^7\ cm^{-3}\ [b(\beta-1)\ (1+Aa)]^{1/2}\ U/c\ B^{3/2} \qquad (5)$$

As demonstrated in Biermann and Strittmatter (1987a), secondary electron emission becomes unimportant for low thermal densities with a condition very similar to eq. (5).

We conclude from the above that the overall model is both self consistent and not in obvious conflict with the observational constraints.

5. Conclusions

5.1 Near-Relativistic Flow

If our basic interpretation is correct, we infer from the observed cut-off frequencies that the sources in question must contain near relativistic flows and close to maximal particle energies. This is consistent with the standard interpretation of active nuclei such as BL Lacs and OVVs (Angel and Stockman 1980, Kellermann and Pauliny-Toth 1981, Eckart et al. 1985). It does however imply that the flow in the M87 jet and other similar jets as well as hot spots (Röser and Meisenheimer 1986, Meisenheimer and Röser 1986) is also near relativistic. Conversely, the model predicts that many jets and hot spots may have optical synchrotron emission (Biermann and Strittmatter 1987a,b).

5.2 Spectra

When comparing observed spectra with the emission expected from a modelled shock, it is important to note that we observe the integrated emission from a whole shock region in most cases for lack of spatial resolution. Considering then the region downstream from a shock, the cutoff energy of the energy distribution of electrons decreases with distance from the shock leading to an emission spectrum (integrated over the whole region in analogy to Blandford and Königl 1979) which is the original $\nu^{-\alpha}$ (α =(p-1)/2) spectrum below a loss-bend frequency ν_b, $\nu^{-(\alpha+1/2)}$ between the frequencies ν_b and ν^*, the cutoff frequency, and a sharp turndown at ν^*. Thus, for, e.g., p=2,, the emission from a compact emission region with strong losses is ν^{-1} up to ν^*. At higher frequencies inverse Compton emission may be relevant for values of the parameter a $\geq$ 1. For extended emission regions, with weak losses, the frequency ν_b remains close to ν^* and the total emission is, again for p=2, $\nu^{-0.5}$ over most of the frequency range below ν^*.

Given a flow along a jet outwards it appears natural to assume that repeated shocks, standing or moving in the observers frame, slow down the motion. Hence it is likely that the cutoff frequency diminishes along a jet and so the optical spectrum steepens. This has indeed been observed in the M 87 jet and is consistent with our model (Pérez-Fournon et al. 1987).

5.3 The Sharpness of the Cutoff

The steepness of the observed cut-off has also been subject of considerable discussion (cf. Rieke et al. 1982). A discussion of this phenomenon has been given within the framework of the present model by Fritz (1987) who notes that a positive dependence of mean free path on particle energy steepens the cutoff as compared to the results of Webb et al. (1984).

5.4 Variable Cutoffs

A final point of direct comparison with observations of AGNs arises from the behaviour of the cutoff frequency with photon energy density or - for a given

region - luminosity. From equation (4) it follows that increases in luminosity would be accompanied by a hardening of the visible wavelength spectrum provided $a<1$; this appears to be observed in several sources, for example 0235+23 (Rieke et al. 1976). For more extreme values of luminosity ($a>1$) the reverse behaviour is possible, although values of a greatly exceeding unity are unlikely since they imply very large luminosities in high energy photons.

The model also offers an explanation of the different spectral behaviour observed among BL Lac objects. Most such sources exhibit steep optical spectra but some of which (e.g. 0735+178) appear to maintain relatively flat spectra into the ultraviolet. The latter class would be associated with higher values of a.

5.5 X-Ray Emission

Note that as a consequence of the predicted correlation between a and the optical/infrared spectrum, it follows that higher Compton X-ray luminosity is expected for non-thermal (polarized) sources with relatively flat optical spectra. This is consistent with the observations of Ledden and O'Dell (1985) who noted that for radio samples, the X-ray to radio luminosity is generally lower for optically quiet (i.e. steep optical spectrum) sources.

While we recognize that the interpretation is complicated by the effects of possible relativistic beaming, the present model does appear to be capable of explaining a number of well established observational phenomena.

5.6 Quasars versus BL Lacs

Condition (5) on the density of the thermal gas may be only marginally fulfilled in the broad-line emission region of quasars. In that case, as Klemens (1987) has demonstrated, primary electron acceleration as well as secondary pair-production as a result of p-p collisions both have to be considered. Klemens (1987) modelled the shock-acceleration with a Monte-Carlo-code, calculating the electron/positron energy spectrum with the p-p collisions providing a source at all locations and momenta. Since p-p collisions (between very energetic protons and thermal protons) result in very high energy electron/positron pairs, X-rays can be emitted as synchrotron radiation. With two free parameters, the shock density ratio and the ratio of primary to secondary energetic electron densities, Klemens (1987) can readily fit mm to X-ray spectra of the objects 0521-365 and 1219+305 which have fairly good spectral data.

5.7 High Energy Particles

There are in addition a number of further consequences of this model which, while not yet observed, may have observable consequences. For example, it follows from canonical estimates of magnetic fields in jets (c.f. Bridle and Perley 1984) that the protons will be accelerated to energies of order 10^{12}GeV in such sources. Jets with $a<<1$ and $\nu^{*}\simeq 3\ ß10^{14}$Hz may therefore be sources of extremely energetic particles as

found in cosmic rays (see Fermi 1949, 1954; Hillas 1984). Again for active galactic nuclei in which Aa>1, a very substantial fraction of the luminosity may be emitted in high energy neutrons which would decay after travelling a distance λ_n, depositing inter alia high energy electrons and positrons in that region. The proton-photon collisions also lead to considerable neutrino emission which may be observable in neutrino-telescopes (Eichler 1979, Gaisser and Stanev 1985). The total luminosity in these emission components is not specified uniquely within the present model since it depends on the ratio of total energy in protons to that in electrons.

We may have thus identified regions where collisional processes in the high TeV range might lead to observable consequences.

ACKNOWLEDGEMENTS

Dr. P. A. Strittmatter has been a partner in this work and I am grateful for many years of fruitful collaboration. I wish to thank Drs. L. O'C. Drury and G. Webb for their invaluable assistance with the diffusive shock acceleration theory and Dr. D. Montgomery for enlightening correspondence on MHD turbulence. Helpful discussions with Messrs. K.-D. Fritz, Y. Klemens, W. Krülls, and T. Schmutzler are also acknowledged. Finally I wish to thank Drs. R. Schickeiser and W. Kundt for discussions of controversial aspects of shock acceleration theory.

REFERENCES

Angel, J.R.P., Stockman, H.S.: 1980, Ann. Rev. Astron. Astrophys. 18, 321

Armstrong, T.A., Hogg, W.R., Lewis, G.M., Robertson, A.W., Brookes, G.R., Clough, A.S., Freeland, J.H., Galbraith, W., King, A.F., Rawlinson, W.R., Tait, N.R.S., Thompson, J.C., Tolfree, D.W.L.: 1972, Phys. Rev D. 5, 1640

Axford, W.I.: 1981, Proc. of Workshop on Plasma Astrophysics, ESA SP-161, p. 425

Ballmoos, P.v., Diehl, R., Schönfelder, V.: 1987 Astrophys. J. 312, 134

Bezler, M., Kendziorra, E., Staubert, R., Hasinger, G., Pietsch, W., Reppin, C., Trümper, J., Voges, W.: 1984, Astron. Astrophys. 136, 351

Biermann, P.L., Strittmatter, P.A.: 1987a Astrophys. J. (Nov. 15)

Biermann, P.L., Strittmatter, P.A.: 1987b Proc. IAU Symposium No. 129 (in press)

Björnsson, C.-I.: 1985, Monthly Notices Roy. Astron. Soc. 216, 241

Blandford, R.D., Königl, A.: 1979 Astrophys.J. 232, 34

Bregman, J.N., Lebofsky, M.J., Aller, M.F., Rieke, G.H., Aller, H.D., Hodge, P.E., Glassgold, A.E., Huggins, P.J.: 1981, Nature 293, 714

Bregman, J.N.: 1985, Astrophys. J. 288, 32

Bridle, A.H., Perley, R.A.: 1984, Ann. Rev. Astron. Astrophys. 22, 319

Brodie, J., Königl, A., Bowyer, S.:, 1983, Astrophys. J. 273, 154

Cambridge Bubble Chamber Group: 1967, Phys. Rev. 155, 1477

Cayatte, V., Sol, H.: 1987 Astron. & Astrophys. 171, 25

Crane, P., Stockton, A., Saslaw, W.C.: 1987 Astron. & Astrophys. (in press)

Drury, L. O'C.: 1983, Rep. Prog. Phys. 46, 973
Eckart, A., Witzel, A., Biermann, P., Pearson, T.J., Readhead, A.C.S., Johnston, K.J.: 1985, Astrophys. J. Lett 296, L23
Eichler, D.: 1979, Astrophys. J. 232, 106
Fermi, E.: 1949, Phys. Rev. 75, 1169
Fermi, E.: 1954, Astrophys. J. 119, 1
Fritz, K.-D.: 1987, in preparation
Gaisser, T.K., Stanev, T.: 1985, Phys. Rev. Letters 54, 2265
Goldstein, M.L., Burlaga, L.F., Matthaeus, W.H.: 1984, J.G.R. (A) 89, 3747
Gwinn, G.R., Moran, J.M., Reid, M.J., Schneps, M.H.: 1987 Astrophys. J. (In press)
Heavens, A.F., Meisenheimer, K.: 1987 M.N.R.A.S. 225, 335
Heisenberg, W.: 1948, Z. Physik 124, 628
Hillas, A.M.: 1984, Ann. Rev. Astron. Astrophys. 22, 425
Keel, W.C.: 1986 Astrophys. J. 302, 296
Kellermann, K.I., Pauliny-Toth, I.I.K.: 1981, Ann. Rev. Astron. Astrophys. 19, 373
Klemens, Y.: 1987 Ph. D. Thesis, University of Bonn
Kolmogorov, A.N.: 1941, C.R. Acad. URSS 30, 201
Kraichnan, R.H.: 1965 Phys. of Fluids 8, 1385
Ledden, J.E., O'Dell, S.L.: 1985, Astrophys. J. 298, 630
Matthaeus, W.H., Goldstein, M.L., Smith, C.: 1982 Phys. Rev. Letters 48, 1256
Matthaeus, W.H., Goldstein, M.L.: 1982, J.G.R. (A) 87, 6011
Meisenheimer, K., Heavens, A.F.: 1986 Nature 323, 419
Meisenheimer, K., Röser, H.-J.: 1986, Nature 319, 459
Montgomery, D., Brown, M.R., Matthaeus, W.H.: 1986 preprint
Montgomery, D.: 1986, priv. comm.
Pérez-Fournon, I.: 1984, Proc. of Manchester Conference on Active Nuclei
Pérez-Fournon, I.: 1985, Ph.D. Thesis, Tenerife
Pérez-Fournon, I., Biermann, P.: 1986 Proc. IAU Symposium No. 119, p. 405
Pérez-Fournon, I., Colina, L., González-Serrano, J.I., Biermann, P.: 1987 Preprint
Rickett, B.J., Coles, W.A., Bourgois, G.: 1984 Astron. & Astrophys. 134, 390
Rieke, G.H., Grasdalen, G.L., Kinman, T.D., Hintzen, P., Wills, B.J., Wills, D.: 1976, Nature 260, 754
Rieke, G.H., Lebofsky, M.J., Kinman, T.D.: 1979, Astrophys. J. Lett. 232, L151
Rieke, G.H., Lebofsky, M.J.: 1980, IAU Symp. No. 92, 263
Rieke, G.H., Lebofsky, M.J., Wisniewski, W.Z.: 1982, Astrophys. J. 263, 73
Röser, H.-J., Meisenheimer, K.: 1986, Astron. Astrophys. 154, 15
Röser, H.-J., Meisenheimer, K.: 1987 Astrophys. J. 314, 70
Rybicki, G.B., Lightman, A.P.: 1979, Radiative Processes in Astrophysics, Wiley-Interscience, New York
Sagdeev, R.Z.: 1979, Rev. Mod. Physics 51, 1
Schlickeiser, R.: 1984, Astron. Astrophys. 136, 227
Simkin, S.: 1978 Astrophys. J. 222, L 55

Simkin, S.: 1986 Astrophys. J. 309, 100

Sitko, M.L., Stein, W.A., Zhang, Y.-X., Wisniewski, W.Z.: 1983, Publ. Astron. Soc. Pacific 95, 724

Smith, C.W., Goldstein, M.L., Matthaeus, W.H.: 1983, J.G.R. (A) 88, 5581

Stecker, F.W.: 1968, Phys. Rev. Lett. 21, 1016

Stocke, J.T., Rieke, G.H., Lebofsky, M.J.: 1981, Nature 294, 319

Sulentic, J.W., Arp, H., Lorre, J.: 1979, Astrophys. J. 233, 44

Warren-Smith, R.F., King, D.J., Scarrott, S.M.: 1984 M.N.R.A.S. 210, 415

Webb, G.M., Drury, L. O'C., Biermann, P.: 1984, Astron. Astrophys. 137, 185

Webb, G.M.: 1985a, San Diego, 19th International Cosmic Ray Conference, Proc. Vol. 3, 107 (OG 8.1-1)

Webb, G.M.: 1985b, San Diego, 19th International Cosmic Ray Conference, Proc. Vol. 3, p. 246 (OG 8.3-8)

Webb, G.M.: 1985c Astrophys. J. 296, 319

Webb, G.M.: 1987 Astrophys. J. 319, 215

Zdziarski, A.A. 1986 Astrophys. J. 305, 45

The Masses of Active Galactic Nuclei

Liu Ru-liang

Purple Mountain Observatory, Academia Sinica,
Nanjing, People's Republic of China

1. Correlation Analysis.

The study of the variability of the X-ray radiation in AGNs on time scales of 10^2 to 10^5sec is of fundamental importance for understanding the nature of their central source. Recently Wandel & Mushotzky [1] determined the masses of AGNs from their X-ray variability assuming the emitting region of X-rays to be five Schwarzschild radii. However, this size estimate of the X-ray emitting regon is conjectured only. So, a statistical analysis and some new considerations are needed for getting reasonable estimates of the size of the X-ray emitting region. On this basis, we would determine the masses of AGNs from Δt_x is the shortest time scale of the X-ray variability in sec, and Lx is the luminosity in the X-ray band between 2 and 10 keV. All the X-ray data (Δt_x, L_x) are taken from WANDEL et al [1], BARR et al [2], COURVOISIER [3] and WAN BOQI [4] (H_0=50km/sec/Mpc, q_0=0), and the number of sample 1528, we find the regression equation as follows

$$\log L_{x\,(2-10Kev)} = a + b \lg\Delta t \ , \qquad (1)$$

where a=39.00±1.10, b=0.94±0.24. The correlation coefficient is 0.79. Within the permissible deviation we suppose the regression equation between L_x and Δt to be linear

$$\lg L_x = \lg\Delta t + 40 \ . \qquad (2)$$

Assuming that the energy spectral index α_x is constant ≈ 0.7 [6] from hard X-ray to 1MeV soft γ-ray of Seyfert and quasars we can get

$$L_{x\,(0.5-1\,MeV)} = K \int_{500/h}^{10^6/h} \nu^{-\alpha_x}\, d\nu \ , \qquad L_{x\,(2-10KeV)} = K \int_{2\times10^3/h}^{10^4/h} \nu^{-\alpha_x}\, d\nu$$

From the above equations we have $L_{total} \cong L_{x+\gamma} \cong 10 L_{x\,(2-10Kev)}$.

We estimate the total luminosity and the X-ray variability scale as

$$L_t = \lambda L_{Edd} = 1.3 \times 10^{38} \lambda (M/M_\odot) \quad \text{ergs/sec.} \qquad (3)$$

$$\Delta t_x = R_x/c = \tau\, Rg/c = \frac{2\tau GM}{c^3} \qquad (4)$$

where G is gravitational constant, Rg is the Schwarzshild radius, λ and τ are parameters. Comparing (3), (4) with (2), we have the following relation

$$\lg(\tau/\lambda) = 2.1 \qquad (5)$$

2. The determination of the parameters λ and τ and the mass of AGNs?

We shall consider a stationary and electromagnetically neutral thick disc in mechanical equilibrium of a vorticity-free configuration rotating around a compact object M (e.g. a black hole), which is much greater than the disc mass. The only forces considered are caused by gravitation, rotation and radiation pressure gradients (neglect gas pressure, Fig. 1). the maximum flux F_{max} that can be emitted from a disc surface is achieved when the effective gravity $\vec{g}_{eff}$ is balanced by the gradient of radiation pressure alone.

$$\vec{F}_{max} = -\frac{c}{\kappa\rho}\Delta P_{rad}, \qquad \frac{1}{\rho}\Delta P_{rad} = -\Delta\Phi + \Omega^2\vec{R} = \vec{g}_{eff} \qquad (6)$$

where κ is the opacity per unit mass, $\vec{g}_{eff}$ is the effective acceleration, which is the vector sum of gravitation and centrifugal acceleration. Thus $\vec{g}_{eff}$ must be orthogonal to isobaric surfaces (free surface) of the thick disc. The maximum luminosity

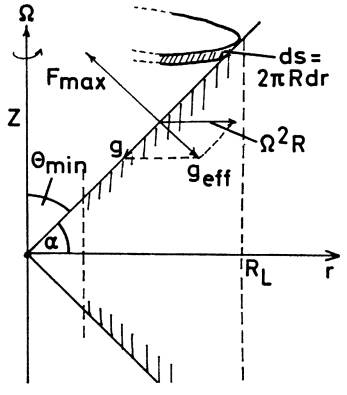


Fig. 1 Balance of forces in the thick disc

From Fig. 1 we obtained $F_{max} = (c/k)g_{eff} = (c/k)g\,tg\alpha$, and

$$\lambda = L_{max}/L_{Edd} = \sin\alpha \ln(R_2/R_1) \ , \qquad (7)$$

where $g=GM/r^2$, $R=r\cos\alpha$, and $L_{Edd}=(4\pi GMc)/K$ is the Eddington luminosity. R_2 and R_1 is the inner and outer radius of the disc, respectively. In the case of the high luminosity object, it tends to the large α, R_2/R_1, and L_{max} may exceed $L_{Edd}(\lambda \geq 1)$.

We follow FRANK et al [7], and take $l=R^2\Omega$ = constant, where l is the specific angular momentum. We define an effective potential $\Phi_{eff}=-GM/r+l^2/2R^2$. Φ_{eff} turns out to be the energy per unit mass. Given ℓ, one can define a Keplerian radius R_K: $l^2=GMR_K$. We may express the specific energy E of the fluid in units of the binding energy of a Keplerian circular orbit as

$$E = 2R_K\Phi_{Eff}/GM = (R_K/R)^2 - 2R_K/(R^2+Z^2)^{1/2} . \quad (8)$$

And $-1 \leqslant E < 0$. An important feature of the thick disc models is the existence of a narrow axial funnel. Matter in steady rotation supported by pressure can "fill" all equipotential surfaces up to a small negative value of E. The free surface of the disc is the equipotential E, and it can be proved that for higher luminosity the funnel is narrower with smaller negative E and a smaller cone of aperture Θ_{min}. Using the definition R=rsinΘ (Fig. 1) and equ. (8), we obtain

$$\sin^2\Theta = R_K^2/(2\gamma R_K + Er^2) . \quad (9)$$

After differentiating (9) with respect to r and finding that the minimum value of $\sin^2\Theta$ occurs at $r=-R_K/E$ for a given (negative) value of E, we can estimate the aperture of the funnel

$$\sin^2\Theta_{min} = -E . \quad (10)$$

It can be shown from (9) that a given equipotential surface extends between the inner and outer equatorial radii which may be given by solving (9) with $\sin^2\Theta=1$

$$R_{in}=R_K\left[1+(1+E)^{1/2}\right]^{-1} , \qquad R_{out}=R_K\left[1-(1+E)^{1/2}\right]^{-1} . \quad (11)$$

For a given E, we have $R_{in}=R_1$, $R_{out}=R_2$ from (11). Substituting R_2, R_1 and Θ_{min} from (10) into (7), we can plot the relation of E and λ (Fig. 2). Although we still do not know the exact relation between L_t and E of AGNs, values can be estimated to be between 10^{-1} and 10^{-2} from observation [8] and recent theoretical calculations [9] for AGNs. Therefore, the value of luminosity L_x could be divided into three regions: low luminosity $\lg L_x \leqslant 42.5$, median luminosity $42.5 < \lg L_x \leqslant 44.5$ and high luminosity $\lg L_x \geqslant 44.5$ (ergs/sec), and the available value of λ can be taken as λ=0.05, λ=0.1 and λ=1.0 respectively. Then, the values of τ are obtained from eq. (5). Finally, we can determine the masses of the central object from the value τ and the variability Δt_x as $M(\Delta t_x)=(c^3/2\tau G)\cdot\Delta t_x$.

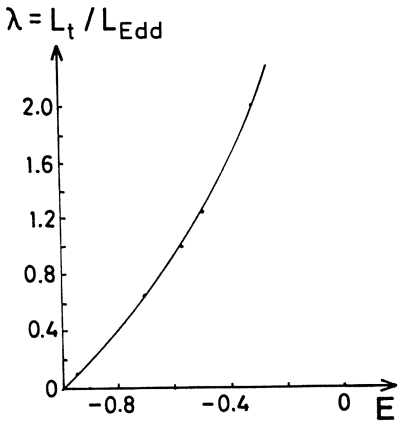


Fig. 2 The relation of the specific energy E and $\lambda=L_t/L_{Edd}$

3. The dynamic mass

For comparison, assuming that the emission-line width is produced by gravity only, we can give the dynamic masses of AGNs in case of the broad H_β in the broad line region (BLR):

$$M \propto G^{-1} V^2 R \qquad \text{or} \qquad M = \xi V_{eff}^2 R/G \tag{12}$$

where V is the characteristic velocity of the discrete gas clouds, V_{eff} is the projection of V on the line of sight, and ξ is a transformation factor from V into V_{eff}. ξ depends on the model of BLR, and R is the size of BLR, $R=[3L_{H\beta}/4\pi f_v](H_\beta)$ is the volume emissivity in H_β, f_v is the value filling factor of the emitting clouds. While DIBAI [8] and the author [10] arbitrarily set $f_v \sim 10^{-3}$, we estimate it from the covering factor f_a, which may be obtained observationally, and $f_v = 4/3(N_{H\beta}/n_e \cdot R) f_a] \cong 10^{-5}-10^{-6}$.

From (12), we obtain the masses of 1 type Seyfert galaxies, if the values of the full width zero intensity ($FWZI=2V_{eff}$) and $L(H_\beta)$ have been given. Similarly, the masses of AGNs evaluated from the forbidden line [1][O III] can be estimated. For 12 of type 1 Seyfert galaxies and 22 of Seyfert galaxies and quasars for which $FWZI(H_\beta)$ and the [O III] line width was available, their masses are comparable with that obtained from Δt_x. The samples show a good correlation between the two kinds of mass (Fig. 3).

4. Concluding remarks

(1) The fact that high luminosity objects do not vary rapidly [6] is not caused by either sensitivity or sampling bias. The absence of long time-scale variability for low

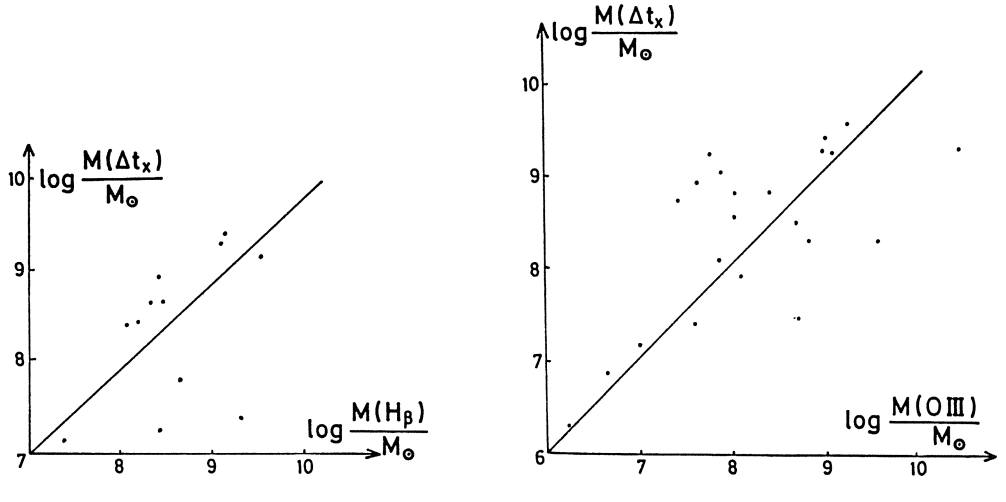


Fig. 3 the $M(\Delta t_x)$ mass vs. the H_β and [O III] dynamical mass

luminosity objects may be a selection effect because objects that are not observed to vary have been omitted. It is still true, however, that only low-luminosity objects vary rapidly.

(2) Typical values $M(\Delta t_x)$ are 10^7–10^8 $M_\odot$ for Seyfert and 10^9 $M_\odot$ for quasars, consistent with the current values. So the massive object confined within the small size of $R_x \leq c \cdot \Delta t_x$ (10^{-6}–10^{-2}pc) may be a black hole.

(3) The size of the X-ray emitting region R_x is about 5–10Rg for low and median luminosity objects, and 100 Rg for high luminosity objects. The size of the X-ray emitting region increases with L_x or M.

(4) Relativistic effects must be considered. If the source is moving toward us at relativistic speeds, then the observed luminosity should be higher than the intrinsic luminosity by a factor of δ^4, while the observed time for variability Δt should be shorter than the true time scale by a factor of δ where $\delta = \frac{\Gamma}{1-\beta\cos\theta}$ ($\gamma = 1/\left(1-\beta^2\right)$, θ is the angle between the direction of motion and the direction along the line of sight), which is called the relativistic Doppler factor. So, the dynamic mass calculated from broad lines (H_β) or narrow lines [O III] must be affected by nongravitational forces. It may be the reason why there is scattering in Fig. 3.

REFERENCES

1. A. Wandel, B. F. Mushotzky: Ap. J., Vol. 306, L61 (1986)
2. P. Barr, B. F. Mushotzky: Nature Vol. 320, 421 (1986)
3. T. J. L. Courvoisier: ESO preprint No. 476, Dec. (1986)
4. Wan Boqi: ISAS Research Note, Vol. 319, Feb. (1986)
5. L. Bassani, A. Deen: Astr. Astrophys., Vol. 122, 83 (1983)
6. A. F. Tennant, B. F. Mushotzky: Ap. J., Vol. 264, 92 (1983)
7. J. Frank et al: Accre tion Power in Astrophysics Cambridge Astrophysics Series 1985
8. E. A. Dibai: Soviet Astro. Letters Vol. 7, 248 (1981)
9. P. Moskalik, M. Sikora: Nature Vol. 319, 649 (1986)
10. Liu Ru-liang: Acta Astrophysica Sinica Vol. 3, 113 (1983)

The Study of X-Ray Radiation from AGNs

Liu Ru-Liang[1], *Zhang He-qi*[1], *and You Jun-han*[2]

[1]Purple Mountain Observatory, Academia Sinica,
Nanjing, People's Republic of China
[2]Center of Astrophysics, University of Science and Technology of China

This is a summary report based on several recent papers devoted to the study of X-ray emission in active galactic nuclei (including quasars) [1],[2],[3].

1. The properties of AGNs Spectral Luminosity

First, using the available monochromatic radio, optical and X-ray luminosities ℓ_r(5GHz), ℓ_o (2500A) and ℓ_x (2Kev) of 107 quasars, we have analysed the hardness of the spectrum of quasars. The evolution of the spectral properties of quasars with their luminosities may be studied statistically from Figure 1 and Figure 2, which show the relationships between lg ℓ_r and lg ℓ_o and between lg ℓ_x and lg ℓ_o respectively. It is seen from Fig.2 that for optically faint sources (lg ℓ_o^c <31.2 erg/sec Hz), lg ℓ_x and lg ℓ_o have almost approximately a linear relation with a slope of unity. It implies that X-ray and optical radiation increase similarly. On the contrary, one can see from Fig.1 that there exists a relation $\ell_r \sim \ell_o^5$ at optically faint sources, while for the optically bright ones the X-ray radiation tends to be constant and $\ell_r \sim \ell_o^{0.8}$. This may give some hints to understand the radiation mechanisms of quasars. However, the radio emission of AGNs is complex, it depends on the radio frequency. We shall study the relation of X-ray and radio emission of AGNs further.

2. Further study on the correlation between ℓ_x and ℓ_o

Such a study has been carried out through a statistical analysis recently. this relationship is important for understanding the structure and radiation processen in

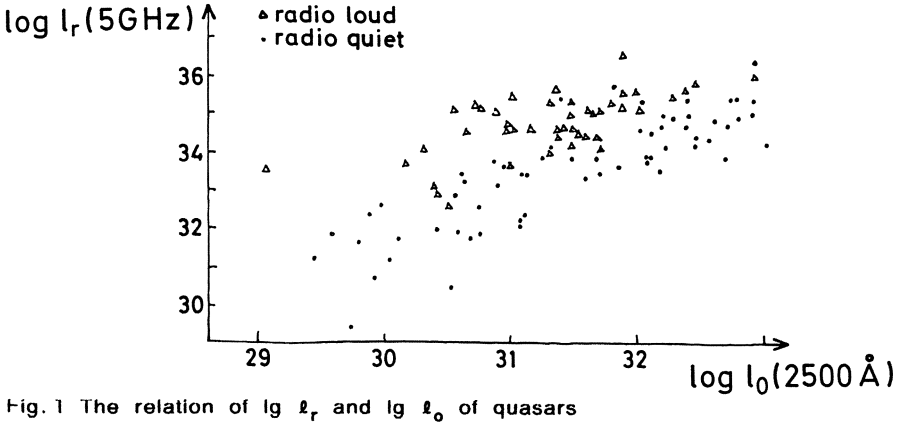


Fig. 1 The relation of lg ℓ_r and lg ℓ_o of quasars

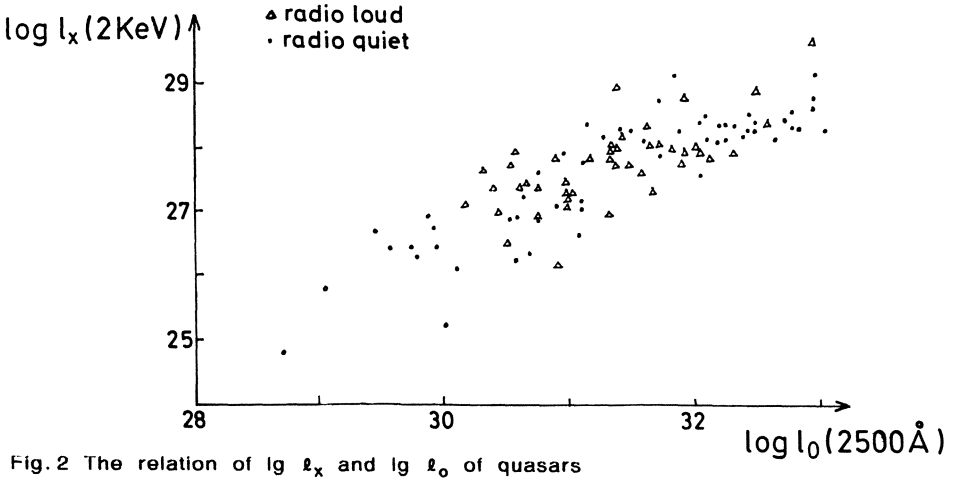


Fig. 2 The relation of lg ℓ_x and lg ℓ_o of quasars

AGNs. The main defect in the previous work is the lack of low-luminosity objects, causes a large scatter in the statistics for optically faint sources. However, 30 new type I Seyfert galaxies for which the absorption of low energy was considered are supplied in order to reduce this scatter, and the homogeneous complete set of 117 samples (including 87 quasars) is used in the new statistical research.

How do we determine the critical point in the ℓ_x-ℓ_o plot (Fig. 2) quantitatively? We chose an arbitrary point as an assumed critical point which divides the whole sample of observed luminosities ℓ_o into two parts - the optically faint region and optically bright region. Then we calculated the mean-square-errors for each region by the regression method. By using a so-called the quasi-Monte-Carlo method (step by step in the computer), the true critical point ℓ_o^c, which makes the total error minimal, is finally determined, and the result is lg ℓ_o^c =31.11 ergs/sec Hz. On the basis of this, the following relation is obtained

$$\ell_x \propto \ell_o^b \ . \qquad (1)$$

where

$$b=\begin{cases} 0.93 \pm 0.18 \text{ for } n=66,\ \gamma_c=0.68 \text{ (optically faint region } \ell_o<\ell_{o_c}^c) \\ 0.28 \pm 0.18 \text{ for } n=51,\ \gamma_c=0.41 \text{ (optically bright region } \ell_o>\ell_o^c). \end{cases}$$

n is the number of sample and γ_c is the regression coefficient. The proportional relation (when $\ell_o<\ell_o^c$) implies that X-ray and optical radiation originate by the same mechanism. Obviously, a reasonable model must explain the following observed facts simultaneously: a) The nearly proportional relation optically faint region, m = $M/M_{cr}<<1$) ; b) The shorter time scale of X-ray variability than that of optical, $\Delta t_x<<\Delta t_o$; c) The power law spectrum of optical and X-ray radiation. Several plausible radiation mechanisms have been considered such as :

1) Type I synchrontron-self-compton process (SSC1), i.e. the near UV-optical radiation around 2500 A comes from synchroton emission but the 2 Kev X-rays from

self-compton processes. If so, we gel $\ell_x \propto \ell_o^{1.5}$, contradicting fact a). Another problem in SSCI is that the variability $\delta t_x \simeq \Delta t_o$, disagrees with b). 2) Type II synchrotron-self-compton process (SSC II). the synchrotron process produces radio photons but the self-compton process produces the X-ray and optical photons. However, SSC II disagrees with fact b). Besides, the necessary condition for SSC is: all AGNs must be strong compact radio sources, which is not our samples (optically selects quasars and type I Sy).

3) Thermal radiation from a thin disk. The standard thin disk produces no X-ray emission. some modifications of a thin disk model have been proposed by PAYNE [4], TSURUTA [5]. The X-ray comes from the comptonization of optical photons in the hot coronae around the central compact objects. In this model $\ell_x \propto \ell_o^{1.5}$, and the spectral shape is $\ell_o(\nu) \propto \nu^{1/3}$.

4) Optical thermal radiation from a thick disk: $\ell_x \propto \ell_o^{0.75}$ and $\ell_o(\nu) \propto \nu^{1/3}$.

5) Optical radiation wind: $\ell_x \propto \ell_o^{0.75}$ and $\ell(\nu) \propto \nu^{1/2}$.

Obviously, the facts a) and b) reject the thermal radiation mechanism 3), 4) and 5). Only the synchrotron radiation in the UV-optical-X-ray region seem to be acceptable, because it is suitable for facts a), b), and c).

Non-proportional relation $\ell_x \propto \ell_o^{0.28}$. It comes the drag effect, i.e. the X-ray photons produced in the neighbourhood of a central massive black hole will be dragged by infalling dense gas, and swallowed by the central black hole. For simplicity, a spherical accretion is assumed. From the infall velocity of the accreting gas $V_{in}(R)$ we can estimate the time scales of collapse of gas in the X-ray emission region with radius $R_x \Delta T_{in}(R)$.

$$\Delta T_{in}(R) = 3.5 \times 10^3 \eta^{-1} m_8 \left[\left[R_x/R_o \right]^{3/2} - 1 \right], \qquad (2)$$

where $\eta<1$ is viscosity parameter, $m_8 = M/10^8 M_o$, $R_o = 3R_5 = (6GM)/c^2$

$= 9 \times 10^{13}\ m_8$ cm.

In order to discuss the drag effect, two assumptions are used in our work: a) Only X-ray photons can be dragged. This is because the X-ray emission region is much closer to the central black hole. b) If the escaping time of X-ray photon ΔT_{es} is much longer than the collapsing time of the gas in the X-ray emission region $T_{in}(R_x)$ [2], these X-ray photons will be dragged.

Further, the escaping time ΔT_{es} for X-ray photons from Rc to the outer edge Rx (Ro$\leqslant$ Rc $\leqslant$Rx) can be estimated, according to the theory of random walk and the formula of the free path of scattering. Our results are that $\ell_x \propto \ell_o^{0.30}$ for in>1 (optically bright region), very close to the observational result.

3. The properties of X-ray selected QSOs

It has been noted that the relation ℓ_x and ℓ_o for X-ray selected quasars is quite different from optically and radio selected quasars. We have analysed 109 X-ray selected quasars. On the histograms of their redshift, optical luminosity and average

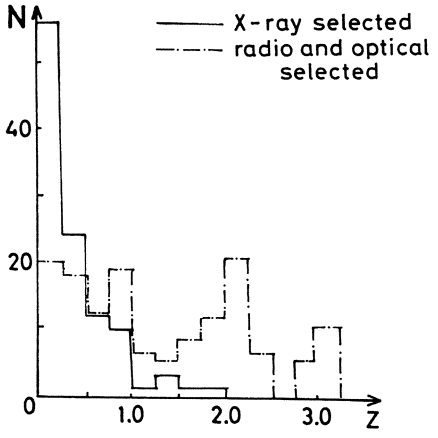


Fig. 3 The redshift distribution histogram of X-ray selected and radio-optically selected QSOs

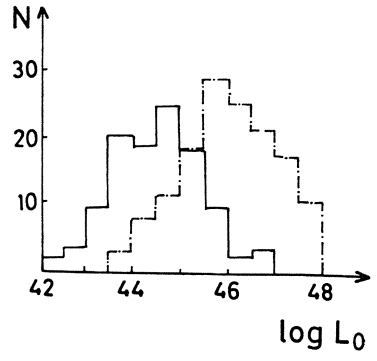


Fig. 4 The optical luminosity L_o distribution histogram of the two kinds of QSOs

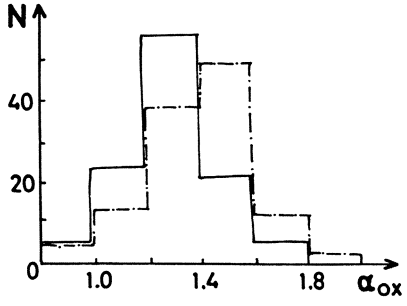


Fig. 5 The α_{ox} distribution histogram of the two kinds of QSOs

optically-X-ray spectral index ($\alpha_{ox} = -\frac{\lg \ell_x/\ell_o}{\lg \nu_x/\nu_o}$), it is found that X-ray selected quasars are distributed at the low ends of these parameters (Fig. 3, Fig. 4, Fig. 5). On the plot of ℓ_x (2kev) and ℓ_o(2500A), most of the X-ray selected quasars are situated in optically faint regions ($\lg \ell_o^c$ <31.11 ergs/sec Hz). In this region, the radio and optically selected quasars have a slope b≃1.0 ($\ell_x \propto \ell_o^b$), while the value of b of the X-ray selected quasars is only 0.78. It implies that they may be a subtype of quasars.

4. The L_x-Z relation for AGNs

Fig. 6 is the L_x-Z relation of 273 AGNs (235 quasars and 38 type Seyfert galaxies). The L_x values in the band of 0.5-4.5 kev are from the Einstein Observatory. The linear regression line for those objects is $\lg L_x$=41.2 + 1.51gZ and the correlatien coefficient is γ_c=0.905, implying a strong correlation. It can be seen from Fig. 6 that the 273 AGNs form a belt with clear boundaries, although there is some scatters in it. The lower envelope may originate from the limiting sensitivity of the

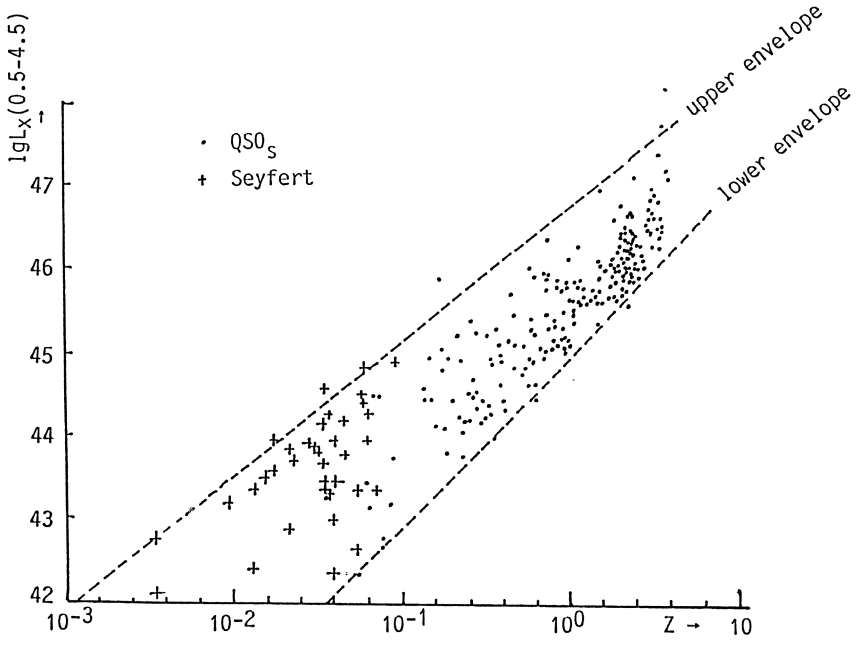


Fig. 6 The L_x–Z relation of QSOs and Seyfert

receiving instrument pointed out by Tananbaum [6], but he could not find the existence of the upper envelope because of the small size of his sample. The upper envelope in Fig. 6 is an important statistical result about the relation between L_x and Z of AGNs. Recently, GIOVANNELLI and POLCARO took BL Lac objects, elliptical and spiral galaxies as samples and obtained the relation $Lx \sim Z^{1.5}$ [7]. All these results reflect the fact that the strength of the X-ray radiation of AGNs in the early epoch of the universe was higher than those of at later stages statistically. The range of L_x from quasars with the highest redshift to Seyfert galaxies with low redshift is more than 6 orders of magnitude. This sequence is not dependent on the width of L_x. this shows some characteristics of cosmical evolution in the belt, but a simple analysis suggests that it is impossible that this sequence is an evolutionary track of AGNs from QSOs to Seyfert. Firstly, from Fig. 6 and the Eddington luminosity limit, it is easily found that the mass of the central black hole of QSOs in an early epoch is $M \geqslant 10^{10}$ $M_\odot$, and M decreases with redshift z up to the stage of Seyfert, when M is about 10^7 $M_\odot$. If a QSO evolves along the belt, M can only increase due to accretion. Secondly, evolution along the belt will lead to a very small density of accreting matter, which disagrees with observation.

5. Summary

(a) The X-ray radiation of AGNs constitutes an important part of their radiation. It is tightly related to the properties of the central black hole. Therefore, the study of X-ray radiation from AGNs is an extremely useful method to explore the properties of the central power of AGNs.

(b) It seems that there exist two evolutionary stages for AGNs: optically bright (larger Z) and optically faint (smaller Z), with a division point at lg ℓ_o^c =31.11 ergs/sec Hz in an ℓ_x-ℓ_o plot. In each stage the characteristic parameters of the typical properties of AGNs are very different. In the early stage of the universe, the high-redshift QSOs have higher density of the surrounding matter, which absorbs the X-ray radiation. According to the random walk theory, the X-ray photons are dragged by infall dense gas, and swallowed by the black hole. So, this is the reason why the X-ray emission tends to be flat on an ℓ_x-ℓ_o plot in the optically bright region. For the optically faint region, the density of the surrounding matter of QSOs and Seyfert is rather low, and the linear relation between ℓ_x(2kev) and ℓ_o(2500A) strongly implies that the two kinds of radiation mainly originate from the same mechanism and the mechanism is very possibly non-thermal synchrotron radiation.

(c) The properties of X-ray selected QSOs are quite different. They might be a subtype of QSOs, and evolve differently. Perhaps, some selection effects related to the sensitivity of the instrument and the sample bias, cause the difference. This needs to be studied further.

(d) On the L_x-Z plot of AGNs, there seems to exist a natural physical sequence, which connects all kinds of active galaxies including QSOs, Seyferts and others. Perhaps it is due to the mass distribution of the central massive objects [7]. It will be evidence of a physical continuity of AGNs, even between different classes of compact extragalactic X-ray sources.

REFERENCES

1. Zhang He-qi, Liu Ru-liang, Yang Hai-show: Chin. Astron. Astrophys., Vol. 7, 224 (1983)
2. Liu Ru-liang, Yang Hai-show: Chin. J. Space Sci., Vol. 5, 84 (1985)
3. Liu Ru-liang, Cao Li-hong, You Jun-han, Cheng Fu-zhen: Acta Astronomica Sinica Vol. 27, 343 (1986)
4. D.G. Payne, D.M. Eardley: Astrophys. Lett. Vol. 19, 39 (1977)
5. S. Tsuruta: Astr. Astrophys. Vol. 61, 647 (1977)
6. H. Tananbaum: X-ray Astronomy, ed by R. Gaiacconi and G. Setti, (Proceedings of the NATO Advanced Study Institute held at Erice, Sicily, July 1-14, 1979), p. 311
7. F. Giovannelli, V.F. Polcaro: Mon. Not. R. Astr. Soc., Vol. 222, 619 (1986)

The Starburst Model for AGNs

J. Melnick[1] *and R. Terlevich*[2]

[1]European Southern Observatory
[2]Royal Greenwich Observatory

1. Introduction

One of the "central dogmas" in astronomy is that nuclear activity in galaxies is powered by massive collapsed objects often called *monsters*, the most popular types of *"monsters"* being black holes and spinors. However, the strong similarities between the spectra of narrow line Seyfert galaxies and normal HII regions, led the first workers in the field to associate the Seyfert phenomenon with *starburst* activity in the nuclear regions of galaxies [1,2]. Figure 1 presents a comparison between the spectrum of a typical Seyfert 2 galaxy, NGC 3081, to that of the high excitation HII galaxy, C1543+091. The figure illustrates the similarities, but also the differences, between the *emission line* spectra of Seyferts and HII regions.

In fact these differences - i.e., the wider range of ionization displayed by Seyfert galaxies, historically led to the dismissal of the *starburst* model as the root of nuclear activity in spiral galaxies. But, using recent developments in the theory of stellar evolution, we have shown that the spectra of Seyfert galaxies can indeed be reproduced extremely well by stellar photoionization models [3]. This has led us to a critical revision of the need of monsters to explain the Seyfert phenomenon [4].

In this contribution we would like to present an overview of the properties of nuclear starbursts and a qualitative description of their evolution as the ionizing stars evolve through the WARMER to the Supernova phases. We will show that nuclear emission line regions evolve

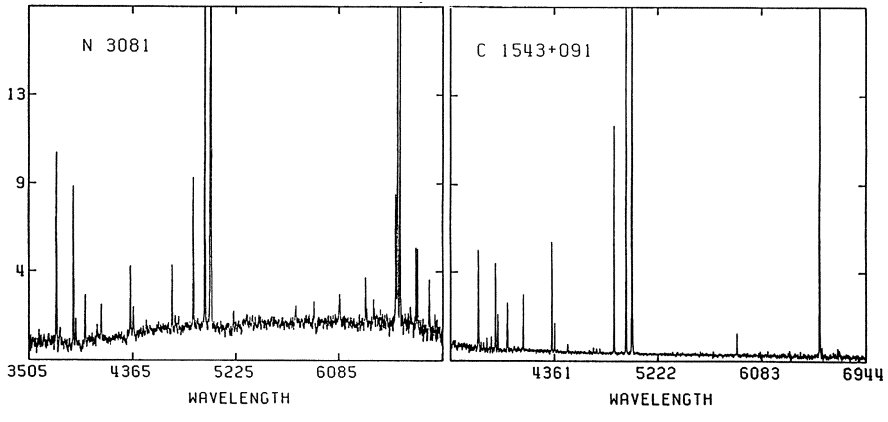


Fig. 1. Comparison between the spectra of a narrow line Seyfert galaxy (NGC 3081) and a high excitation HII region-like galaxy (HII galaxy), C1543+091.

from HII regions to Seyferts and Liners following the evolution of the ionising stars. In a separate contribution (Terlevich and Melnick, this volume) we will present a more quantitative comparison between theory and observation concentrating on the problem of variability.

2. The evolution of massive stars: WARMERS

A priori the most important difference between nuclear and extranuclear *starbursts* is that the former are significantly more metal rich. Thus, the nuclei of galaxies are probably the only places where large numbers of super metal-rich massive stars form; circum-nuclear HII regions are too small to contain appreciable numbers of very massive stars. Because of the influence of metallicity on mass loss the evolution of massive metal rich stars is very different from that of stars of similar initial masses but lower metal content. This difference is schematically illustrated in figure 2 where the evolution of a star that loses no mass (a very metal poor star) is compared to the evolution for a large value of the mass loss rate, $\dot{M}$.

After leaving the main sequence, stars with $\dot{M} \sim 0$ rapidly evolve to the red supergiant phase (RSG) near which they explode *pressumably* (we should not forget SN1987a!) as type II supernovae. If $\dot{M}$ is very large, as would be the case for metal rich stars, mass loss rapidly peels the envelopes leaving exposed the nuclear-burning regions. Thereafter the evolution proceeds at roughly constant bolometric luminosity but towards *increasing* effective temperatures until they reach the He-ZAMS where they explode as supernovae.

We have introduced the term WARMERS to characterize massive stars near the He-ZAMS (observationally WARMERS correspond to early WC and WO stars). Little is known theoretically about the ultimate fate of WARMERS but there is some observational evidence linking WARMERS to type Ib supernovae [5] so, in what follows, we will assume that massive metal-rich stars end their evolution in SNIb explosions.

3. The evolution of nuclear starbursts

Because of their large effective temperatures and luminosities WARMERS have a dramatic effect in the ionisation of nuclear HII regions. This is illustrated in figure 3 where we show a

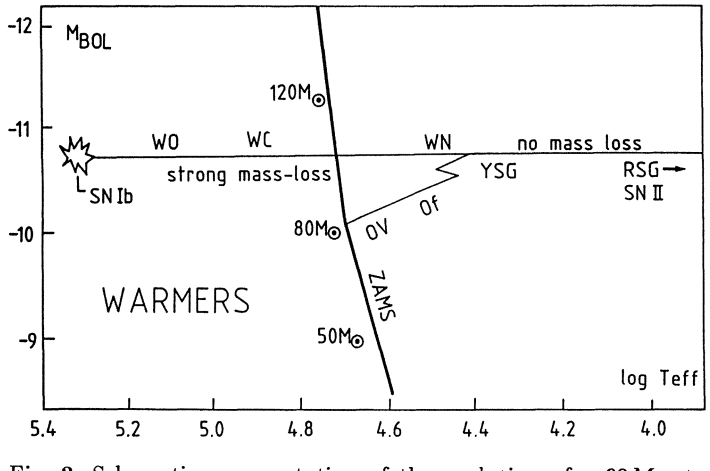


Fig. 2. Schematic representation of the evolution of a $60M_\odot$ star with and without mass loss.

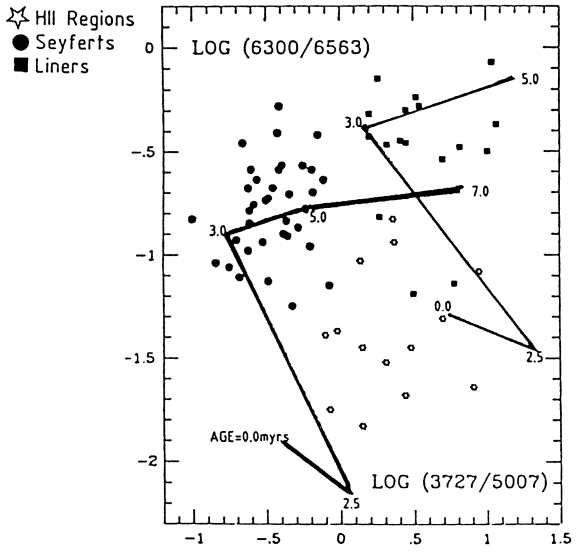


Fig. 3. The evolution of a metal rich nuclear *starburst* in a typical BPT diagnostic diagram for two values of the total mass differing by a factor of 10. Only massive *starbursts* develop Seyfert characteristics while small clusters evolve directly to the LINER phase.

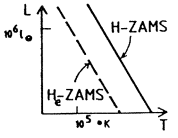
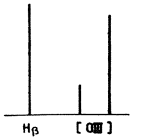
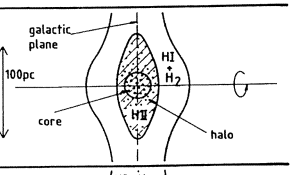
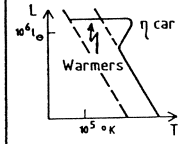
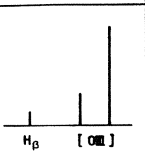
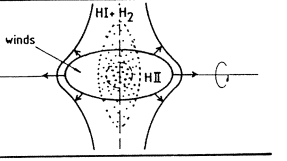
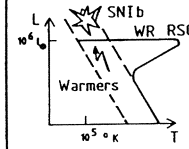
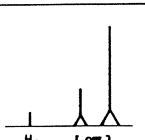
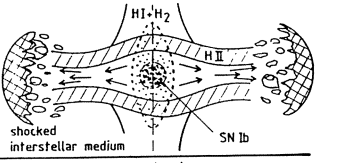
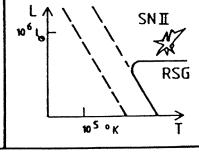
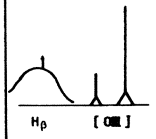
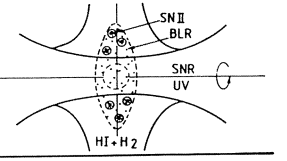
typical diagnostic diagram for emission line objects (after Baldwin, Phillips and Terlevich, [6]) to which we have superimposed starburst photoionization models including WARMERS [3].

The effect of WARMERS is to transform HII regions into Seyfert 2s and/or LINERS. After the onset of WARMERS the spectral evolution is so rapid that very few objects are expected to have spectral characteristics intermediate between HII regions and Seyferts or LINERS. Since the metamorphosis of nuclear starbursts from HII regions to Seyferts or LINERS depends on the existence of a WARMER phase, the starburst model predicts that only metal rich nuclei should harbor AGNs while "normal" nuclear HII regions should be most common among low metallicity nuclei, the cutoff abundance being close to the solar value. This provides a natural explanation to the fact that the nuclear emission line regions of early type spirals (SO, Sa, Sb), which are metal rich, are mostly Seyferts or LINERs while those of late type spirals (Sbc, Sc), which have bulge abundances lower than solar, are mostly *normal* HII regions [4,7].

Our models clearly show that the observations of optical emission line ratios can be explained by stellar photonionization models and therefore that the arguments which historically led to the dismissal of starbursts as the origin of nuclear activity are outdated and should be forgotten. Since the discovery of the AGN phenomenon, however, a wealth of new information at all observable wavelengths has accumulated on nuclear active regions so their existence as a distinct class is no longer based only on the optical spectra but also on aspects such as non-thermal radio emission, X-ray luminosity, variability etc. We believe that most of these properties can be explained within the framework of *starburst* models and we are at present constructing detailed

TABLE 1

The evolution of metal-rich nuclear Starbursts

AGE (10^6yrs) Type	evolution of starburst cluster	Lyman continuum	Optical continuum	Optical emission lines	Dust	radio emission	X- ray emission	Variability	nuclear morphology
0-3 HII region		O-B stars	young O+B stars and old bulge RGs		normal dust to gas ratio	thermal from H II regions	none	none	
3-4 Sy 2 liner		Warmers OB stars	young OB+RSG stars and old bulge RGs		large amounts of new dust from WR and η car-phases	thermal from H II regions	none	none	
4-8 Sy 2/1 liner		Warmers SN Ib OB stars	young OB BSG, RSG and old bulge RGs		new dust from WR and RSG-phases	non-thermal from SN Ib	some from SN Ib's; variable	short term "broad line" variability	
8-20 Sy 1		SNR SNII	young BSG+RSG and old bulge RGs		little dust; most swept or destroyed by SN and Warmers	non-thermal from SN II and SNRs	thermal from SNR inside hot wind blown bubbles; variable	in scale of: weeks (SN) 10^{49-50} ergs years(SNRs) 10^{51-52} ergs	

models to confront the observations. However, because very little is known theoretically and observationally about the evolution of massive stars beyond the WARMER phase our present picture remains qualitative.

Table 1 summarizes schematically the basic properties of the *starbusrt* scenario in four phases following the evolutionary phase of the ionising stars. The basic features of each phase are: phase 1) Nuclear HII regions; ionization by normal OB stars. Phase 2) appearance of the most massive WARMERS and metamorphosis from HII region to Seyfert 2/LINER. Phase 3) WARMERS explode as type Ib supernovae in a common wind blown cavity around the cluster core. These supernovae produce considerable non-thermal radio emission but low X-rays. Phase 3 is also characterized by a strong *bipolar* flow that produces jet-like radio structures and shocks the surrounding interstellar medium. Thus, the model predicts that collimated radio emission should mostly occur in Seyfert 2 galaxies. The last phase (4) is characterized by supernova explosions in wind blown bubbles surrounding individual stars. These (type II) supernovae produce variable thermal X-rays but low radio emission. Further details are given in [3,4] and in the contributions by Terlevich and Melnick and by Tenorio-Tagle et al. in this volume.

Epilogue

It is clear that the observations of a wide range of AGNs from Seyferts to LINERS are consistent with the *starburst* model. This suggests that perhaps the AGN phenomenon itself is linked to violent star formation and that monsters may not exist. We do not pretend, however, to explain the observations of *all* AGNs with our model. In particular, we do not see (at least at present) how to explain observations like superluminal motions or some properties of extreme BL Lac objects with only *starbursts*. Thus, if any, these objects are the best candidates where nuclear monsters might exist.

REFERENCES

1. Pronik, I.I., 1973, *Soviet Astr.*, **16**, 628.
2. Condon, J.J., et al., 1982, *Ap. J.*, **252**, 102.
3. Terlevich, R., Melnick, J., 1985, *M.N.R.A.S.*, **213**, 841.
4. Terlevich, R., Melnick, J., Moles, M., 1987 in IAU symposium **121**, *Observational Evidence of Activity in Galaxies*, eds. E.Ye. Khachikian, K.J. Fricke and J. Melnick, p. 499.
5. Porter, A.C., Filippenko, A.V., 1987, *A. J.*, **93**, 1372.
6. Baldwin, J.A., Phillips, M.M., Terlevich, R., 1981, *Pub. A.S.P.*, **93**, 5 (BPT).
7. Melnick, J., Terlevich, R., Moles, M., 1985 in IAU symposium **116**, *Luminous Stars and Associations in Galaxies*, eds. C.W.H. De Loore, A.J. Willis and P. Laskarides, p. 505.

Starburst Galaxies as Recently Formed Companions

H. Arp

Max-Planck-Institut für Physik und Astrophysik, Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-8046 Garching, Fed. Rep. of Germany

Direct photography and spectra of the companion galaxy NGC 2777 show that it is composed predominantly of young stars and that it exhibits outflow of hot gaseous material. A hydrogen filament leads back to a larger nearby galaxy. It is proposed that this smaller, currently star forming companion, originated recently from material ejected from the larger galaxy.

It is shown that NGC 2777 is typical of companion galaxies as a class. Unrelaxed morphology, young stellar spectra, "hot" infrared colors and small sytematic redshifts characterize this class. The usual assumption that collisional or merger events are responsible for the starburst is discussed. It is argued that exterior perturbations generally do not promote star formation and that in any case collision/mergers do not occur with any significant frequency. The disturbed structure and starburst characteristics found among companion galaxies is attributed to their recent formation.

The problem of the origin and nature of companion galaxies is an intriguing one. Study of the nearest galaxy groups such as the Local Group and the M81 group reveals that there are smaller galaxies distributed around the larger, dominant S_b spirals. These companion galaxies often have unusual properties. There is a tendency for peculiar shapes, young stellar spectra, anomalies in their gas and dust and systematic redshift peculiarities. The conventional scenario of galaxies all condensing at the same time, $\sim 2 \times 10^{10}$yrs ago from some primeval medium, does not explain the observed characteristics of these companion galaxies. In the following paper we discuss the preliminary findings from a detailed large telescope study of one such companion, NGC 2777. We relate these results to the general class of companion galaxies and draw some conclusions as to the recent origin of this class of companion galaxies.

Fig 1 shows a CCD image of the companion galaxy NGC 2777 as recorded with the 4 meter reflector at KPNO. (Arp and Sulentic, to be submitted). It is obvious that this galaxy has an irregular, amorphous shape which permits no presently relaxed, dynamical state.

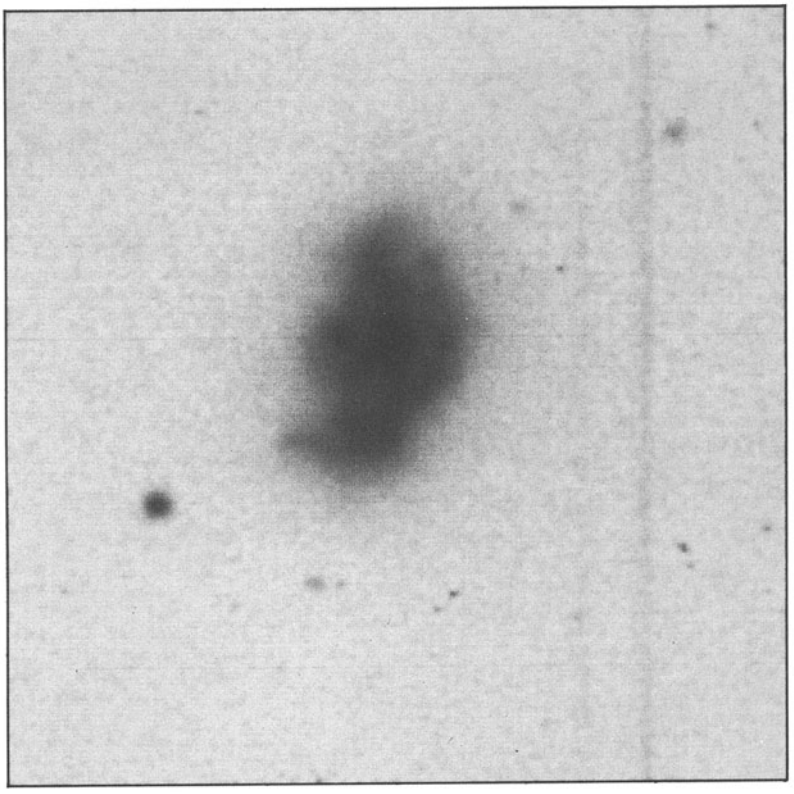


Fig 1. Direct image of NGC 2777 with CCD detector at Cass. focus of the 4 meter, KPNO reflector. (From Arp and Sulentic, in preparation)

It is clear that a recent perturbation must have disturbed this galaxy and that perturbation must either be from the exterior or the interior. In order to investigate what interior disturbance could be present we consult the spectrum of the galaxy.

Fig 2 shows a spectrum of NGC 2777. Aside from an emission line of excited oxygen emission ([O II]) the major features apparent are strong absorption from the Balmer series of hydrogen. The only other feature ia a very weak K line of Ca II. This an extremely unusual spectrum because larger, normal members of groups have spectra dominated by metal aborption lines characteristic of a later type, older stellar population. The spectral type exhibited by the spectrum in Fig. 2 is characteristic of a young stellar population in the range of 10^8 to 10^9 years age (Arp 1963; Sandage 1963). Arp and Sulentic were able to obtain spectra further toward red wavelengths (not shown here) with the cryogenic spectrographic on the 4 meter KPNO telescope. Those spectra revealed that there is a strong, broad, blue wing on the H alpha emission line coming from the body of NGC 2777. The

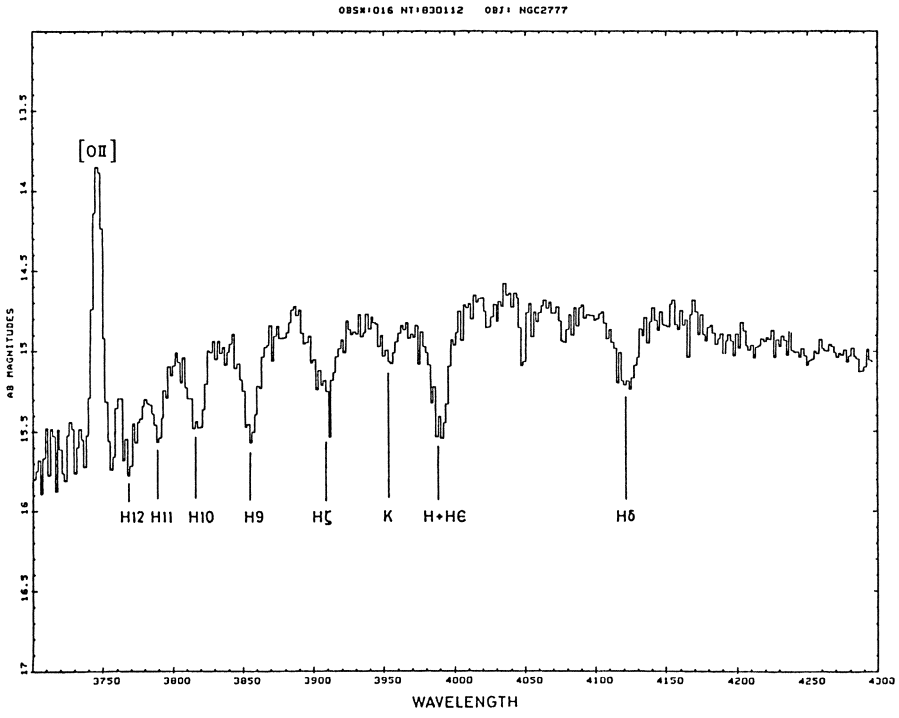


Fig. 2 Spectrum of NGC 2777 with double spectrograph and "2D FRUTTI" detector on 5 meter Palomar reflector.

implication of this is that excited hydrogen gas is flowing out from the galaxy. The gas on the far side of the galaxy which is flowing away from the observer is obscured by dust so we see only the approaching velocities. Rapid star formation accompanied by an outflowing wind from the central regions of the galaxy NGC 253 was first noted by M.-H. Ulrich (1978).

The Hydrogen Content of NGC 2777

It turns out that this companion galaxy has an unusually large amount of H I for its size. Measures with the Arecibo radio telescope give a very large signal (Arp and Sulentic 1985). Even more interesting is the fact that this H I is distributed in a large tail which reaches far outside NGC 2777. The picture derived from radio hydrogen mapping is shown in Fig. 3 (from Haynes 1981). It is clear that the tail of hydrogen reaches directly back to the larger galaxy, NGC 2775.

The fact that the hydrogen tail goes directly back to the center of NGC 2775 and not tangential to the side, is evidence against a tidal encounter between the two galaxies and evidence for the ejection origin of the companion from the center of NGC 2775. Moreover the morphology of NGC 2775 is not disturbed as in a recent encounter. The sharp circular absorption arc in NGC 2775 opposite the

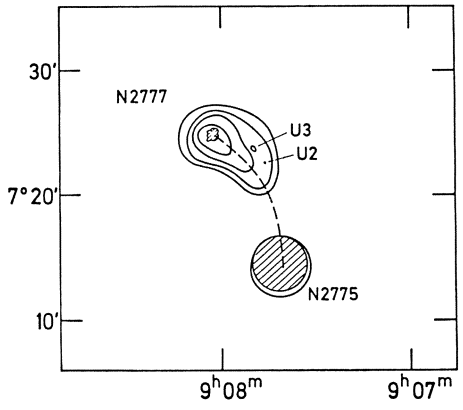


Fig. 3 A hydrogen (H I) map with Arecibo antenna by Haynes (1981). Hydrogen contours around NGC 2777 reach back toward center of NG 2775. Some ultra violet objects in the vicinity are marked as U2, U3.

direction to the companion would be more compatible with an ejected jet of material in the direction of the companion. Finally the redshift of the companion is +139 km/sec with respect to NGC 2775, indicating on an encounter interpretation, that NGC 2777 is on the far side of NGC 2775 and trailing its tail away from us. But the H I in the tail is 50 km/sec greater redshift than NGC 2777! The tail cannot be preceding the companion in an encounter!! In the ejection hypothesis material emitted later could be moving faster in this direction and there are also non-velocity redshifts indicated for companion galaxies as a class (Arp 1987b).

Infrared Measures and Hot Dust

NGC 2777 has a very high measure of 60 micron infrared radiation compared to 100 micron radiation. The ratio R=60μ/100μ is R=0.58 as table 1 shows. R≥0.5 is typical of well known star burst galaxies like M82 and NGC 5253. It is generally accepted that dust is associated with star formation processes and that the newly formed stars warm the dust so that it has the "hot" 60/100μ ratio of emissivity. Burbidge (1986) has proposed some of these infrared galaxies are very young. Table 1 attempts to list all the brightest apparent magnitude, best known galaxies which have properties like NGC 2777. These properties are:

1) Morphological class irregular or peculiar (usually classified amorphous)

2) Spectra show early type absorption features (usually emission lines and also red colors)

3) "Hot" infrared colors (R≥0.5)

Inspecting the most conspicuous examples of this classes as listed in Table1 leads to the conclusion that these galaxies are also:

4) Companions to larger mass galaxies which have older stellar population types.

5) These companions characteristically have excess redshift with respect to the dominant galaxy. (Δz is usually accurately quantized).

Table 1

Galaxies Similar to NGC 2777

Galaxy	Class	Spectra	60/100μ	Companion Status cΔz from main Galaxy	
NGC 2777	Am	v. early abs.+em	0.58	+139±9kms^{-1}	NGC2775
M82	Am	early abs.+em	1.02	+286±5	M81
NGC3077	Am	early abs.+em	.59	+57±6	M81
NGC404	SO pec	early abs.+em	.49	+228±10	M31
NGC5253	Am	early abs.+em	1.06	(-104±9)	NGC5128
NGC5195	SBO(Am)	A-F	1.7	+11±8	M51
Compⁿ B	triple compact	very early abs.	—	5,400	NGC685
Compⁿ SW	compact pec	very early abs.+em	—	10,600	NGC1199
				Possible Companions	
NGC520	Am	A-F	.66	v_o=2,350	M31?
NGC1569	S_mIV(Am)	very early+em	.91	+157±8	M31
NGC1808	S_{bc}pec	———	.72	v_o=820	?
		Dominant Galaxies in Groups (for Comparison)			
M81	S_bI-II	normal	.27		
NGC2683	S_bI-II	normal	.20		
NGC7331	S_bI-II	normal	.24		

The currently fashionable explanation for characteristics 1) through 3) is that encounters or collisions of two galaxies leads to enhanced star formation. (Joseph and Wright 1985; Telesco, Wolstencraft and Done 1987) There are two major criticisms of this hypothesis:

A. Collisions and Tidal Encounters are commonly supposed to stimulate star formation by shocking or tidably perturbing molecular clouds within a galaxy. But, I would suggest that if a cloud has its temperature raised by a shock wave or its macroscopic internal motions increased by a tidal pull, it will be less likely to

form stars than before this event. For example, Solinger, Morrison and Markert (1977) conclude that shocking the peripheral gas in M82 would end star formation.

I would argue that the critical requirement for star formation is instead to keep the material constrained while cooling takes place. There is an obvious mechanism for containment and that is the magnetic field we observe to run lengthwise along jets. If companion galaxies originated by recent ejection from larger galaxies (Arp 1987b), the containment within the observed kinds of jets would be a favorable site to initiate large amounts of star formation.

As empirical support for the conclusion that tidal encounters are not responsible for star formation we can cite the many cases, like M82, NGC5195, NGC2777, where star bursts are occurring in the companion but not in the larger of the interacting pair. The dominant galaxy has even more pre stellar fuel to be condensed and has suffered just as much shock or tidal pull, yet is not "starbursting". In addition the spiral arms in the dominant galaxy, where star formation is taking place, are not extremely deformed and their linear form as well as star formation rate are the same on the side nearest the companion as on the side furthest from the companion.

B. Mergers are currently supposed to account for a whole range of observed phenomena from starburst galaxies, to shells and generally any pairs of adjacent, disturbed galaxies. I would like to express the opinion, however, that mergers almost never occur and are not a realistic explanation for the observed phenomena. The first argument is that the Hubble flow of galaxies outward in space is supposed to be very quiet; peculiar motions of <50 kms^{-1} (Yahil, Tamman and Sandage 1980). Therefore there is no significant chance of a galaxy wandering in from an unrelated region of space to generate an encounter. Even within groups of galaxies, however, the dispersion velocities are very small. In addition, quantization of redshifts within groups reduces the allowable peculiar motion of galaxies to an even smaller value. For example within the Local Group $c\Delta z \leqslant \pm 17$ kms^{-1} (Arp 1986) or perhaps even $\leqslant \pm 8$ kms^{-1} (Arp 1987a).

Secondly, even if one had a gravitationally bound group where companions orbited around a central galaxy, only occasionally would one interact with the central galaxy. For every current interaction one would have to have a large surrounding cloud of companions, analagous to the Oort Cloud which furnish a reservoir for solar system comets. But a populous cloud of companions is clearly ruled out by the observations. (see for egs. Sulentic, Arp and di Tullio 1978 for optical surveys around large galaxies, Roberts 1987 for H I in groups of galaxies)

Of course there are cases of two galaxies obviously interacting but there is no reason why they were not born in close proximity. Close doubleness is a characteristic of extragalactic objects from quasars to small "isolated extragalactic H

II regions". The latter systems tend to be characterized by young stars and energetic events which cannot be very old (Arp 1987b). In any case it would surely be absurd to consider blue, compact-dwarf-double galaxies, isolated as they are in space, to be products of recent collision.

In summary we have seen that dynamically disturbed, star forming galaxies characteristically tend to be small companions of larger nearby galaxies of older stellar population. I have argued that the activity in these companions is not a consequence of tidal encounter. The only possibility which is left is that they are recently created galaxies, where star formation is stimulated by the recent ejection from the parent galaxy. Some older stars and stellar material may be carried out from the larger galaxy in this ejection. More observations of this class of galaxies is needed and close physical reasoning is required to understand further this process of galaxy formation and its implication for cosmology.

References

Arp, H. 1963, I.A.U. Symposium 20, 220.

Arp, H. 1986, Astron. Astrophys. 156, 207.

Arp, H. 1987a, I. Astrophys. Astron. (India) 8, 241.

Arp, H. 1987b "Quasars, Redshifts and Controversies", publ. Interstellar Media, 2153 Russel St., Berkeley 94705, Calif.

Arp, H. and Sulentic, J.W. 1985, ap.J. 291, 88.

Burbidge, G.R. 1986, P.A.S.P. 98, 1252.

Haynes, M.P. 1981, A.J. 86, 1139.

Joseph, R.D. and Wright, G.S. 1985, M.N.R.A.S. 214, 87.

Roberts, M.S. 1987, Symposium New Ideas in Astronomy, Venice, Cambridge Univ. Press.

Sandage, A.R. 1963, ap.I. 138, 870.

Solinger, A., Morrison, P. and Markert, T. 1977, ap.J. 211, 715.

Sulentic, J.W., Arp, H. and di Tullio, G. 1978, ap.J. 220, 47.

Telesco, C.M., Wolstencroft, R.D. and Done, C., "The Enhancement of Infrared Emission in Interacting Galaxies" M.N.R.A.S. in press.

Ulrich, M.-H. 1978, ap.J. 219, 424.

Yahil, A., Tamman, G. and Sandage, A. 1980, Physica Scripta 21, 320.

Collision of Giant Molecular Clouds

Guo-xuan Song

Shanghai Observatory,
Chinese Academy of Sciences, Shanghai, People's Republic of China

ABSTRACT. High mass star formation is numerically simulated by considering the collision between Giant Molecular Clouds with and without spiral perturbation in the Galaxy. The result reveals: 1. the collision rate due to orbit crowding in the presence of spiral perturbation is 2.5 times higher than that without spiral perturbation ($1/3.27\times10^5$ yr and $1/8.23\times10^5$ yr, respectively), 2. with spiral perturbation most collisions between GMCs are in spiral arm regions, 3. on average every collision between GMCs with spiral perturbation transforms 2.6% of two GMCs mass to high mass stars, i.e., every collision forms high mass stars of $2.04\times10^4 M_\odot$.

1. Introduction

Since the last decade star formation has been extensively studied because of the detection of molecular clouds. It is generally accepted that star formation occurs in molecular clouds, especially in Giant Molecular Clouds. From the Initial Mass Function it seems that the formation of high mass stars is different from that of low mass stars [1]. Recently, Scoville et.al. [2] found that the rate of formation of OB stars depends on $\langle n_{H_2}\rangle^{1.9}$ where $\langle n_{H_2}\rangle$ is the local mean density of H_2 averaged over 300 pc scale lengths. In addition the efficiency of high mass star formation (per unit mass of H_2) is a decreasing function of cloud mass in the range $2\times10^5 \backsim 3\times10^6$ $M_\odot$ so that they suggested that high mass star formation in the galactic disk is initiated by cloud-cloud collisions which is more frequent in the spiral arms due to orbit crowding.

In this paper a numerical simulation on the collision between GMCs is made, in which the rate of collisions is derived. This gives us a quantitative estimate about the formation of OB associations by the collision of GMCs.

2. Model

From the observation of molecular clouds in the Galaxy, a molecular cloud ring exists between 3 kpc and 7 kpc where the total H_2 mass is about 1.8×10^9 $M_\odot$. The mass in the disk of the Galaxy is 2×10^{11} $M_\odot$. Therefore, the total H_2 mass in the ring is two orders of magnitude smaller than the disk mass in the Galaxy. Due to this reason, when the motion of molecular clouds in the ring is studied, it can be accepted that the gravitational potential is fixed, which is induced just by the star

disk as adopted by Roberts and Hausman [3]. Most parameters in this simulation are the same as in the above mentioned paper except that A = 0.1 and $i_0 = 20^o$ are taken. In this case the ratio of the perturbing surface density to the average surface density is approximately 0.055. The loosely wound spiral is taken as $i_0 = 20^o$, because we are interested in late type spirals.

The mass spectrum of molecular clouds is generally

$$N(M) \propto M^{-1.61}$$

The Giant Molecular Clouds are defined as the mass of molecular clouds between 10^5 $M_\odot$ and 10^6 $M_\odot$, from which the ratio of the mass in GMCs to the total H_2 mass is $\sim$0.65. The central subject of this paper is to deal with the collision between GMCs. Actually, collisions between GMCs and smaller molecular clouds will occur. But the variation in the motion of GMCs is still small because the difference in mass of the colliding pair is significant. Thus in this simulation only the Giant Molecular Clouds are considered only and smaller molecular clouds are neglected.

Scoville & Sanders [4] listed the properties of a typical GMC: the diameter is 40 pc, the mass is 4×10^5 $M_\odot$ and the number of GMCs is nearly 3000.

In the simulation 3000 particles, representing GMCs and having the same mass, are randomly distributed in a ring at the initial moment. These particles are moving and colliding with each other in the given gravitational field. We assume that every collision between two GMCs will cause the formation of high mass stars.

We have run the simulation in two cases: one is without spiral perturbation, the other is with spiral perturbation. Every run corresponds to 2×10^8 yr, i.e., almost one revolution around the centre axis. In these simulations both the formation and the disruption of GMCs are not treated and only the collision between existent GMCs are considered. As we know, the lifetime of GMCs is between 4×10^7 and 2×10^8 yr so that this simplification is reasonable.

The timestep is to correspond to 10^5 yr and in every run 2000 steps are carried out. At every timestep the collision between two GMCs is checked. If the quantity $|S-D|$ is less than ϵ, where S is the separation between the centres of two GMCs, D = 40 pc is the diameter of GMC and ϵ = 0.7 pc, this pair of GMCs is identified as a colliding one. The colliding coefficient in the normal direction is taken as -0.5, that in the tangential direction is 1. Because the dispersion velocity for GMCs is less than 5 km/s, the relative displacement for GMCs during every timestep is about 0.5 pc. In that case this criteron for checking collisions between GMCs is applicable.

3. Result

The evolution of the distribution of GMCs without spiral perturbation at $t = 1.0\times10^8$ yr and $t = 2.0\times10^8$ yr is illustrated in Fig. 1. It is quite clear that the distribution of GMCs keeps a nearly random distribution in the evolution as no spiral perturbation is applied.

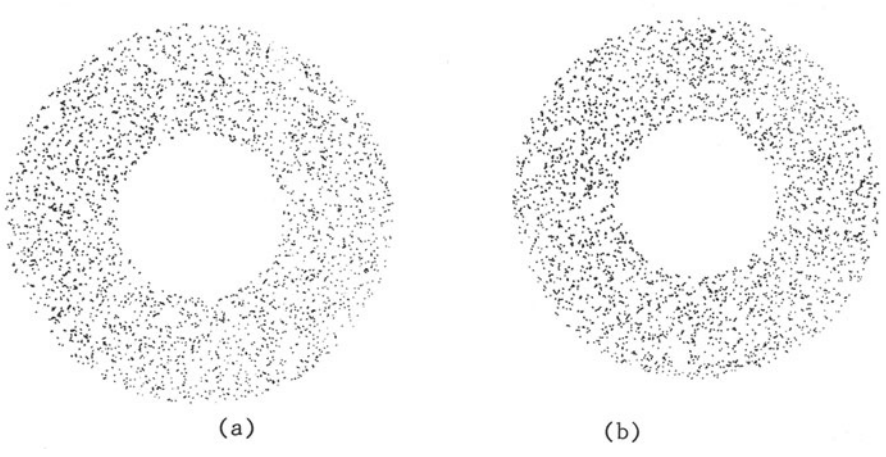


Fig. 1. Distribution of GMCs without spiral perturbation at (a) t = 1.0×10^8 yr, (b) t = 2.0×10^8 yr

Fig. 2. Collision points between GMCs without spiral perturbation during 2.0×10^8 yr

Figure 2 gives the sites of collision between GMCs during 2.0×10^8 yr, i.e., the birthplaces of high mass stars. From this plot, we can see that the collisions between GMCs occur randomly in the disk and the high mass star associations, generated by the collision between GMCs are distributed randomly in the molecular ring.

In considering the spiral perturbation, it is shown in Fig. 3 that at about half a revolution, the GMCs are crowding in the spiral arm region. As mentioned before, although the loosely wound spiral model with 5.5% perturbation is adopted in this simulation, the orbit crowding is prominent and rapid so that it enhances the collision rate greatly.

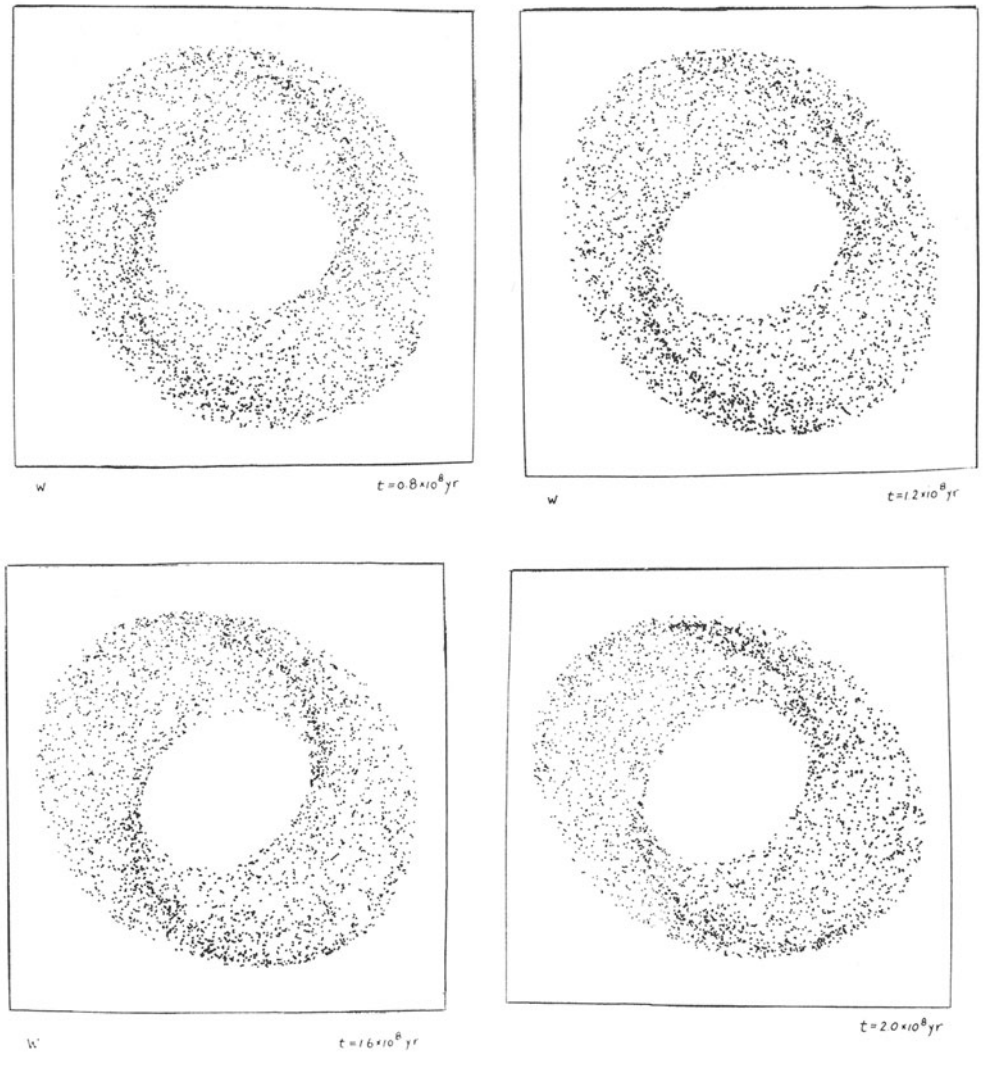


Fig. 3. Distribution of GMCs with spiral perturbation at different epochs

The sites of collision between GMCs during 2.0×10^8 yr is shown in Fig. 4. It is evident that most collisions occur in the spiral arm region because GMCs are crowding there. The detection of most OB associations in the spiral arm region is well accounted for by this simulation.

Figure 5 is the histogram of the rate of collision between GMCs with and without spiral perturbation, respectively. In dealing with the spiral perturbation, the number of collisions during 2.0×10^8 yr is 611 and the collision rate is $1/3.27 \times 10^5$ yr. If there is no spiral perturbation, the number of collision is 243 and the collision rate is $1/8.23 \times 10^5$ yr. From this it is derived that the collision rate with spiral perturbation is about 2.5 times higher than that without spiral perturbation.

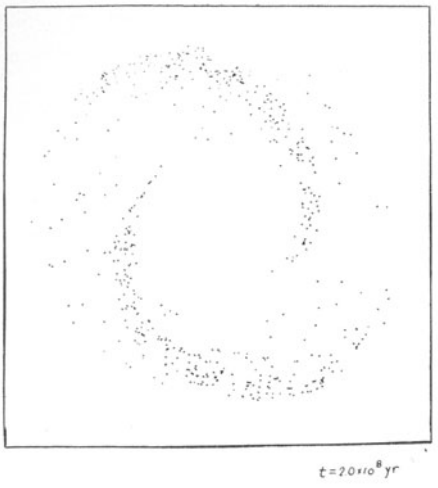


Fig. 4. Collision points between GMCs with spiral perturbation during 2.0×10^8 yr

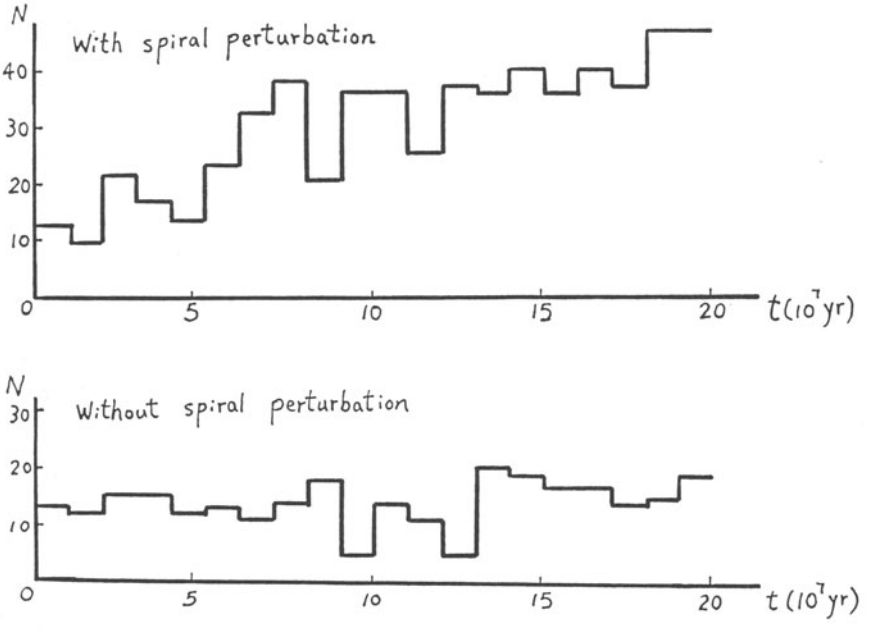


Fig. 5. Histogram of rate of collision between GMXC (a) without spiral perturbation (b) with spiral perturbation

The high mass stars are defined as those with masses between 10 $M_\odot$ and 50 $M_\odot$. If we take the three segment law of the IMF [1] ratio of the mass in massive stars to the total mass (stars with mass between 0.1 $M_\odot$ and 50 $M_\odot$) is

$$f = \frac{M(10 \leqslant \frac{M}{M_\odot} \leqslant 50)}{M(0.1 \leqslant \frac{M}{M_\odot} \leqslant 50)} = 0.0127$$

If the rate of star formation in the Galaxy is taken as 5 $M_\odot$/yr, the rate of high mass star formation is 0.0635 $M_\odot$. Therefore, every collision should form stars with a high mass of

2.08×10^4 $M_\odot$ with spiral perturbation

5.23×10^4 $M_\odot$ without spiral perturbation

In every collision, the ratio of the massive star formed to the mass of two GMCs is 2.6% and 6.54% with and without spiral perturbation, respectively.

4. Discussion

The collision rate between GMCs and the efficiency of massive star birth in collision are estimated by using simulations. In this approach it is adopted that the molecular clouds are distributed in a thin disk. From the observation the height scale is estimated. In that case a question can be raised: what is the effect of thickness on the estimate of the collision rate and the efficiency of star birth. Qualitatively, it will decrease the collision rate between GMCs. We leave this question to be solved later.

Acknowledgment. This work was supported by Chinese Academy of Sciences. I am grateful to Prof. G. Boerner and the Max-Planck Gesellschaft for hospitality during this workshop.

References

1. G.E. Miller, J.M. Scalo: Ap. J. Suppl. Series, 41, 513(1979)
2. N.Z. Scoville, D.B. Sanders, D.P. Clemens: Ap. J., 277, L77(1986)
3. W.W. Roberts Jr., M.A. Hausman: Ap. J., 277, 744(1984)
4. N.Z. Scoville, D.B. Sanders: In Interstellar Processes, eds. by H. Thronson & D. Hollenbach (1986)

Part III

Cosmology

Cosmic Strings and Galaxy Formation

N. Straumann

Institut für Theoretische Physik, Universität Zürich,
Schönberggasse 9, CH-8001 Zürich, Switzerland

INTRODUCTION

Recently it has repeatetly been claimed [1] that there is now a successful "standard model" for galaxy formation. This model involves: (i) cold dark matter with $\Omega_0 = 1$, (ii) primordial Gaussian fluctuations with a Harrison-Zel'dovich spectrum, and (iii) biasing, which presumes that galaxies only form in the rare peaks of density perturbations.

Beside many successes there are important shortcomings of this model. First of all, there is no really convincing dark matter candidate, in spite of the long list of suggested possibilities. Furthermore, the physical origin of biasing is not clear. (I have not understood what 'natural' biasing really means.) Finally, it is not yet clear whether the cold dark matter scenario can reproduce the main observations. In particular, the large peculiar velocities on scales up to $\sim$ 100 Mpc may cause great difficulties.

I think it must be admitted that we do not really know how galaxies and their associations have been formed. We cannot even answer in any convincing manner simple qualitative questions such as: What determines the mass scale of typical galaxies ?

Some progress has, however, been made. We know now better what does not work. An important constraint to theory is, of course, the enormous isotropy of the background radiation, the limits on which have reached an impressive and disturbing level. Hopefully, some fluctuations will eventually be seen.

In this somewhat desperate situation it is important to work out different scenarios, in order to see what has a chance to survive. Cosmic strings provide an interesting - albeit highly speculative - possibility for generating the necessary density fluctuations for galaxy formation [2]. In contrast to some other suggestions it can be ruled out - or confirmed - observationally.

Since this talk addresses a general audience of astrophysicists, I will go in some detail through material which is well known to the experts.

1. PHYSICS OF VACUUM STRINGS

As the universe cooled from incredibly high temperatures it certainly went through many symmetry breakings and a variety of topological defects may have been produced. Cosmic strings - normal or superconducting - are defects of dimension one and may stretch across a galaxy or perhaps the whole sky. Except for a few remarks in Sect. 5 we consider in this talk only normal strings.

1.1 String Solutions in Gauge Theories

Vacuum strings are one kind of 'soliton solutions' of certain spontaneously broken gauge theories. The prototyp are the Abrikosov vortices of type II superconductors [3], which have been rediscovered by Nielsen and Olesen [4] in the Abelian Higgs model, where the Higgs field plays the role of the superconducting order parameter. I recall briefly the string solutions in the context of this field theoretical model.

The Lagrangian density of the Abelian Higgs model, with gauge group U(1), is given by

$$\mathcal{L} = D_\mu \phi^* D^\mu \phi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{2}\lambda(\phi^*\phi - \eta^2)^2 , \qquad (1.1)$$

where $D_\mu = \partial_\mu - ie A_\mu$, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ and e is the coupling constant. The minimum of the Higgs potential $V(\phi) = \frac{1}{2}\lambda(|\phi|^2 - \eta^2)^2$ is assumed on the vacuum manifold $M_0 = \{\phi : |\phi| = \eta\}$, which is in this simple example a circle.

Let us look at configurations for which the fields are static and constant in one direction taken to be the x_3-direction. Then we can choose a gauge with $A_0 = A_3 = 0$ and are thus led to study regular finite energy solutions of the 2-dimensional Yang-Mills-Higgs-Lagrangian (1).

For a string solution, ϕ approaches the "Higgs-vacuum" far away from the vortex tube, i.e. $|\phi| \longrightarrow \eta$ and $|D\phi| \longrightarrow 0$ for $|x| \longrightarrow \infty$. The latter condition, together with $F_{\mu\nu} \longrightarrow 0$, guarantees that the energy density vanishes asymptotically outside the string core. Since ϕ lies asymptotically in the vacuum manifold M_0 , the Higgs field defines a mapping $f: S^1 \longrightarrow M_0$, which associates to each direction in (ordinary) space the asymptotic value of the Higgs field in the vacuum manifold M_0 , which in our example is again a circle. The winding number (or degree) n of the map f

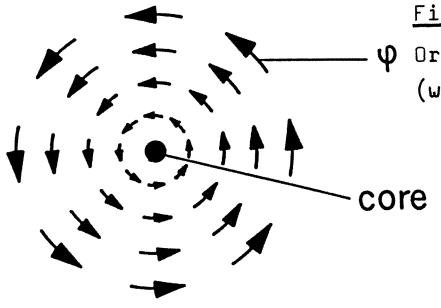


Fig. 1.
φ Order parameter ($\mathrm{Re}\Phi$, $\mathrm{Im}\Phi$) near the core (with extension m_H^{-1}, m_A^{-1}) of a string.

is an integer and characterizes the homotopy class (i.e. the element in $\pi_1(M_o)$) of the mapping. This is a simple example of a topological 'quantum number' (charge).

It is easy to show [5] that the magnetic flux of the string is n times the elementary magnetic flux $2\pi/e$. An elementary string corresponds to $n = \pm 1$ and is topologically stable. Strings with $|n| \geqslant 2$ are probably unstable and decay into elementary ones.

The thickness of the string is determined by the Compton wavelengths of the Higgs particle and the vector mesons, m_Φ^{-1}, m_A^{-1}, where $m_\Phi = \sqrt{2\lambda}\eta$, $m_A = \sqrt{2}\, e\eta$. Usually $m_\Phi < m_A$ and the string has an inner core of 'false vacuum' with linear mass density $\mu_\Phi \sim (\lambda\eta^4)\cdot m_\Phi^{-2} \sim \eta^2$, and a tube of magnetic field of radius m_A^{-1} with $\mu_A \sim B^2 m_A^{-2} \sim \eta^2$ (because the flux is $2\pi/e$). Thus the total mass of the string per unit length is roughly

$$\mu \sim \eta^2 \quad (\sim 10^{-6}\, m_{Pl}^2 \text{ in GUTS}) . \tag{1.2}$$

(A more accurate treatment can be found in books on superconductivity.)

More complicated strings are expected to occur, whenever $\pi_1(M_o)$ is nontrivial. In the general case of a symmetry breaking $G \longrightarrow H$ the vacuum manifold is G/H. If $\pi_1(G) = \pi_o(G) = I$ (G is simply connected) one knows from homotopy theory that $\pi_1(G/H) = \pi_o(H)$. So H should contain a discrete symmetry. This shows that string solutions are expected to occur for example in the scheme $SO(10) \longrightarrow SU(5) \times Z_2$. Strings form in many grand unified theories [2], as well as in the low energy sector of superstring theories [6], because these often have an extra U(1) factor.

1.2 Effective Action

The internal structure of the string is often unimportant and quantities like the energy momentum tensor $T_{\mu\nu}$ can be averaged over cross sections. For a static straight string along the z-axis one

finds from Lorentz invariance and the conservation law $T^{\mu\nu}{}_{,\nu} = 0$ easily [2]

$$\int T_{\mu\nu}\, dxdy = \mu(1,0,0,-1) \,. \tag{1.3}$$

Note that

$$\mu = G\,\frac{m_{Pl}}{\ell_{Pl}} = 2.2\,\left(\frac{G\mu}{10^{-6}}\right)\,10^{10}\,M_\odot/\,\mathrm{kpc}\,. \tag{1.4}$$

The center of a string, where $\phi = 0$, sweeps out a world sheet (Fig. 2) which we parametrize by $x^\mu = f^\mu(\tau,\sigma)$, where τ is a timelike and σ a spacelike parameter. If the string is straight on the scale of its width we can find a simple approximate action by performing first a transverse integration. In order to do this we introduce the coordinates $(\tau,\sigma,\rho_1,\rho_2)$ in a neighborhood of the world sheet by setting $x^\mu = f^\mu(\tau,\sigma) + \rho_1 n_1^\mu + \rho_2 n_2^\mu$, where n_1, n_2 are two orthogonal vector fields normal to the sheet. Close to the sheet the spacetime metric $g_{\mu\nu}$ has in these coordinates the form

$$(g_{\mu\nu}) = \left(\begin{array}{c|c} \gamma_{ab} & 0 \\ \hline 0 & 1_2 \end{array}\right) + O(\rho)\,,$$

where (γ_{ab}) is the induced metric.

For a 'long' string we find for the action, if there is no structure along the string,

$$\begin{aligned} S &= \int \mathcal{L}\sqrt{-g}\; d\tau\, d\sigma\, d\rho_1\, d\rho_2 \\ &= -\mu\int \sqrt{-\gamma}\; d\tau\, d\sigma + O\left(\frac{\text{width}}{\text{radius of curvature}}\right). \end{aligned} \tag{1.5}$$

Thus, to a very good approximation we obtain the Nambu-Goto action

$$S_{eff} = -\mu\int_{\text{world sheet}} dA(\gamma)\,, \tag{1.6}$$

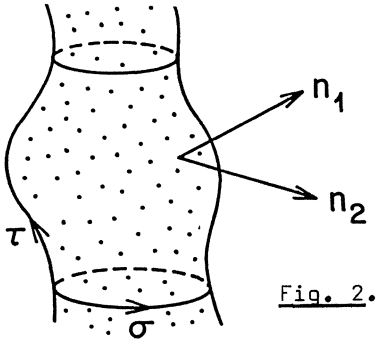


Fig. 2. World sheet of a (closed) string

where $dA(\gamma)$ is the surface area belonging to the induced metric γ_{ab}. The correction term in (11.5) is only important where the dynamics determined by S_{eff} produces cusps. For superconducting strings the longitudinal dependence of the fields introduces important additional terms to the action. In the approximation (1.6) the Lagrangian for the equations of motion for a string is

$$L = \sqrt{(\dot{f}\cdot f')^2 - \dot{f}^2 f'^2} \quad ; \quad \dot{f}^\mu := \frac{\partial f^\mu}{\partial \tau} \ , \ f'^\mu := \frac{\partial f^\mu}{\partial \sigma} \ . \qquad (1.7)$$

Note that the equations of motion are independent of μ. They depend only on the spacetime metric g. In differential geometric terms the equations of motion belonging to (1.6) just state that the mean curvature normal for the embedding of the world sheet vanishes. (See, e.g., the general variation of volume formula derived in Ref. [7].)

Equivalent Formulation

Sometimes it is useful to have the following alternative description. Beside γ_{ab} we may introduce a 'dynamic' metric h_{ab} which is, a priori, independent. If $f: \Sigma \longrightarrow M$ is the embedding of the world sheet Σ in spacetime it is natural to consider the action

$$S(f,h) = \text{const} \int \| df \|^2 \, dA(h) \ , \qquad (1.8)$$

where $\| df \|^2$ is the following natural coordinate independent norm of the differential of f :

$$\| df \|^2 = g_{\mu\nu} \, h^{ab} \, \partial_a f^\mu \, \partial_b f^\nu = h^{ab} \gamma_{ab} \ .$$

This is just the action functional of the bosonic Polyakov string. Variation with respect to f says that f is a harmonic map, satisfying the equation

$$\Box_h f^\mu + h^{ab} \Gamma^\mu_{\alpha\beta} \, \partial_a f^\alpha \partial_b f^\beta = 0 \ .$$

($\Gamma^\mu_{\alpha\beta}$ are the Christoffel symbols for the spacetime metric g.) Variation of h gives on the other hand for the pull-back of g under f

$$f^* g = \tfrac{1}{2} \| df \|^2 \, h \ ,$$

showing that h is proportional to the induced metric. Since (1.8) is invariant under conformal changes of h, the two formulations are indeed equivalent.

Variation of g in (1.6) gives for the energy-momentum tensor

$$T^{\mu\nu} = \mu \int \frac{\delta^4(x - f(\tau,\sigma))}{\sqrt{-g}} \, \gamma^{ab} \, \partial_a f^\mu \partial_b f^\nu \sqrt{-\gamma} \, d\tau \, d\sigma \ . \qquad (1.9)$$

In a conformal gauge for which the induced metric is conformally flat (and which according to Gauss, always exist), we have

$$T^{\mu\nu} = \mu \int \frac{\delta^4(x-f(\tau,\sigma))}{\sqrt{-g}} \left(\dot{f}^\mu \dot{f}^\nu - f'^\mu f'^\nu\right) d\tau\, d\sigma . \qquad (1.9')$$

Example

Let us consider as an important case a flat Friedman universe with metric

$$g = dt^2 - a^2(t)\, d\underline{x}^2 .$$

We choose a parametrization of the world sheet for which the variable τ is just the conformal time of this metric, i.e., $dt = a(\tau)\, d\tau$. The other variable σ can be restricted such that $\dot{\underline{x}} \cdot \underline{x}' = 0$ holds. Then the Lagrangian (1.7) reduces to

$$L = a^2(\tau) \cdot \sqrt{\underline{x}'^2 (1-\dot{\underline{x}}^2)} .$$

The corresponding equations of motion are convenient for numerical integrations [8]. We consider also the energy $M(\tau)$ of a string. From (1.9) one finds easily

$$M(\tau) = \mu\, a(\tau) \int \frac{|\underline{x}'|}{(1-\dot{\underline{x}}^2)^{\frac{1}{2}}}\, d\sigma .$$

If the string is practically at rest relative to comoving frames ($\dot{\underline{x}} \simeq 0$) the total mass grows like the scale parameter. For a rapidly moving string the growth rate is reduced as a result of the redshift of the kinetic energy. (In flat spacetime the total mass of the string would, of course, remain constant.)

1.3 Damping of Strings

For 'normal' (non-superconducting) strings the dominant energy-loss mechanism turns out to be gravitational radiation.

Very roughly this can be estimated as follows. Loops of size R have frequencies $\omega \sim R^{-1}$ and the quadrupole formula gives for the energy loss

$$\dot{M}_g \sim -G(MR^2\omega^3)^2 \sim -G\mu^2 , \qquad (1.10)$$

where $M \sim \mu R$ is the mass of the loop. The energy loss is thus independent of the size of the loop ! The lifetime of the loop is

$$\tau_g = \frac{M}{|\dot{M}_g|} \sim \frac{R}{G\mu} \qquad (1.11)$$

and the energy loss fraction in one oscillation period is

$$\frac{|\dot{M}_g|}{M} \cdot \frac{2\pi}{\omega} \sim G\mu . \qquad (1.12)$$

In other words, the total number of oscillations is $\sim (G\mu)^{-1}$.
We recall that

$$G\mu \sim G\eta^2 = (\frac{\eta}{m_{Pl}})^2 \sim 10^{-6}$$

for a typical grand unification scale $\eta \sim 10^{16}$GeV .

More accurate calculations can be performed in linearized theory, using the energy-momentum tensor (1.9), for explicit loop solutions [9]. If we set (see eq. (1.10))

$$\dot{M}_g = -\gamma_g G\mu^2 \ , \tag{1.13}$$

then the dimensionless factor γ_g depends on the details of the loop solution, but is typically $\sim$ 50-100.

Numerically the energy loss is huge

$$\dot{M}_g = -10^{49} (\gamma_g/50)(G\mu/10^{-6})^2 \text{ erg/s} \tag{1.12'}$$

(For reference, quasars typically emit 10^{46} erg/s.)

We sketch now why the gravitational radiation is more important than the electromagnetic radiation for macroscopic loops, if the electromagnetic gauge group is not spontaneously broken in the string. (We do not have to consider other energy losses, because macroscopic loops can only emit massless particles.) In this case the Higgs field of the string must be electrically neutral and thus does not couple to the electromagnetic field A_μ . In the non-Abelian case there may, however, be couplings to the expectation value of a heavy neutral gauge field in the string through gauge self couplings [10]. Dimensional analysis shows that the radiation loss rate $\dot{M}_\gamma$ is, to leading order, proportional to R^{-2}, with a dimensionless proportionality factor which is independent of the gauge coupling constant, since the gauge field of the string is proportional to the inverse gauge coupling constant [11]. We conclude that $|\dot{M}_g| \gg |\dot{M}_\gamma|$ for $G(\mu R)^2 \sim GM^2 \gg 1$, i.e. for $M \gg m_{Pl}$.

1.4 Intercommuting

When two strings intersect, they can change partners (intercommute), as shown in Fig. 3a. This phenomenon has been observed in superconductivity. Double intersections and self-intersections can result in the formation of closed loops (Figs. 3b, c). Numerical solutions [12] of the full nonlinear field equations describing colliding strings suggest that intercommuting occurs (for normal strings) with high

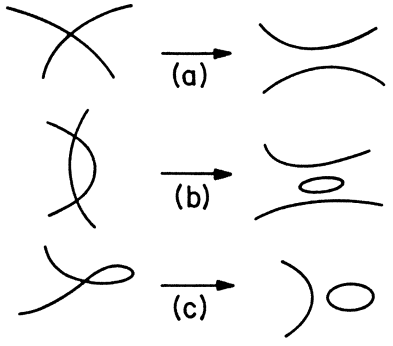


Fig. 3. Intercommuting of strings

probability. (Since this is a very local problem the simulations can be done in flat spacetime.)

1.5 Gravitational Field of Strings

Eq. (1.3) shows that the string tension μ is equal to the energy per unit length of the string. This implies that the gravitational field is very different from that of massive rods. Indeed, the correct Newtonian limit for a quasistatic matter distribution is [13]

$$\varphi = 8\pi G \, (T_{00} - \tfrac{1}{2} \, \mathrm{tr} \, T)$$

and the right-hand side vanishes for (1.3).

For a straight string it is not difficult to find the exact solution of the Einstein equations if the string is idealized as a homogeneous energy-momentum distribution. If the string lies along the z-axis the metric outside the string is in cylindrical coordinates [14]

$$ds^2 = dt^2 - dz^2 - dr^2 - (1-4G\mu)^2 \, r^2 d\varphi^2 \, . \qquad (1.14)$$

This metric is <u>locally</u> flat, because the coordinate transformation $\varphi' = (1-4G\mu)\,\varphi$ brings it locally to Minkowskian form. Since φ' varies from 0 to $(1-4G\mu)\,2\pi < 2\pi$, the metric (1.14) describes a 'conical space', that is, a flat space with a wedge of angular size $8\pi G\mu$ taken out (see Fig. 4). The two faces of the wedge must be identified [15]. It is obvious from Fig. 4 that a string would produce double images of equal brightness of quasars.

The typical separation angle between two images is $\sim 4\pi G\mu \sim$ 3-30 arcsec for $G\mu \sim 10^{-6}$-10^{-5}. This is a kind of classical analogue of the Aharanov-Bohm effect. There is no space-time curvature outside the string, but the metric is not globally Minkowskian.

The conical nature of space around a string has another interesting cosmological consequence. A wake is formed behind a relativistically moving string that has the shape of a wedge with an ope-

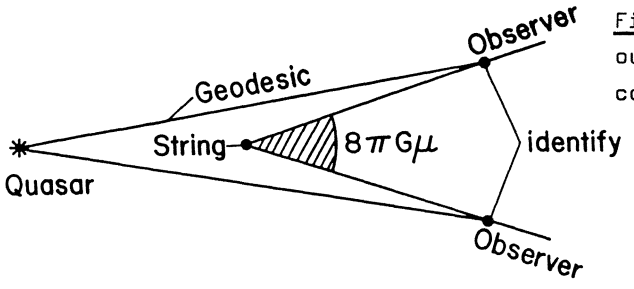


Fig. 4. The conical space outside a straight string core.

ning angle $\sim 8\pi G\mu$ and a density contrast $\delta\rho/\rho \sim 1$. This effect may be important for the formation of large scale structure in 'cold' matter scenarios (see Sect. 3).

In order to avoid possible misunderstandings, we emphasize at this point that the long range time-averaged field of a loop is just that of a point mass with mass equal to that of the loop. Loops could thus have served as seeds for the primordial density fluctuations.

Linearized Theory

Since this last point is sometimes questioned I present a formal argument within the linearized approximation for the gravitational field [16,17].

In flat spacetime eq. (1.9') reduces for the choice $\tau = t\ (= x^0)$ to

$$T^{\mu\nu} = \mu \int (\dot{f}^\mu \dot{f}^\nu - f'^\mu f'^\nu)\, \delta^{(3)}(\underline{x}-\underline{f}(\sigma,t))\, d\sigma \ . \qquad (1.15)$$

Recall that this expression holds for a conformal gauge, for which in flat spacetime

$$\dot{f}\cdot f' = 0 \ , \quad \dot{f}^2 = -f'^2 \ . \qquad (1.16)$$

The retarded solution of the linearized Einstein equation in harmonic coordinates is easily found to be

$$h_{\mu\nu}(\underline{x},t) = -4G\mu \int d\sigma \, \frac{F_{\mu\nu}(\sigma,t_{ret})}{|\underline{x}-\underline{f}(\sigma,t_{ret})|} \, \frac{1}{1-\underline{n}\cdot\dot{\underline{f}}(\sigma,t_{ret})} \ ,$$

where $t_{ret} = t - |\underline{x}-\underline{f}(\sigma,t)|$ is the retarded time and

$$F_{\mu\nu} = \dot{f}_\mu \dot{f}_\nu - f'_\mu f'_\nu - \eta_{\mu\nu}\dot{f}^2 \ , \ \underline{n} = \frac{\underline{x}-\underline{f}(\sigma,t_{ret})}{|\underline{x}-\underline{f}(\sigma,t_{ret})|} \ .$$

The time average for a loop with length L over a period T can be written in the form

$$\langle h_{\mu\nu} \rangle = -\frac{4G\mu}{T} \int_0^T dt \int_0^L d\sigma \, \frac{F_{\mu\nu}(\sigma,t)}{|\underline{x}-\underline{f}(\underline{x},t)|} \ . \qquad (1.17)$$

Now the acceleration a^i of a nonrelativistic particle in the weak-field limit is $a^i = -\frac{1}{2} h_{oo,i} + h_{oi,o}$ and thus the time average is determined by $\langle h_{oo} \rangle$:

$$\langle \underline{a} \rangle = -\frac{1}{2} \underline{\nabla} \langle h_{oo} \rangle \quad . \tag{1.18}$$

We conclude that this is just the acceleration due to the Newtonian potential for a surface matter density $2\mu F_{oo}(\sigma,t)/T = (2\mu/T)\dot{\underline{f}}^2$. It is not difficult to show [16], by making use of (1.16) and the string equations of motion, that the total mass of the surface is equal to the total mass of the string.

This shows that we can approximate the gravitational effect of a loop on surrounding matter by a spherical shell of mass $2\pi R\mu$.

2. FORMATION AND EVOLUTION OF COSMIC STRINGS

As the Universe cools through the critical temperature, fluctuation regions with $\langle \phi \rangle \neq 0$ develop, in which the directions of $\langle \phi \rangle$ are correlated over a correlation length ξ . For a second-order transition, the correlation length is $\sim T_c^{-1}$, but can be much larger for a first-order transition.

Since the free energy tends to be minimized, a slowly varying $\langle \phi \rangle$ is preferred. Much of the initial chaotic variations will therefore quickly die out in the course of further evolution. However, if $\pi_1(M_o)$ is non-trivial, the formation of topological defects is unavoidable.

One expects that the defect lines have the shape of random walks of step given by the correlation length ξ and typical distances between neighboring string segments also of this magnitude. This is confirmed by Monte Carlo simulations [18] which show also:
(i) About 80 % of the total string length is due to 'infinite' strings.
(ii) The remaining strings are closed loops with a scale-invariant distribution $dn \propto R^{-4} dR$, where R is the characteristic size of the loop.

The evolution of a network of cosmic strings is complicated in view of the physics discussed in the previous section. Soon after formation the strings become essentially linear objects which can be described by the Nambu-Goto action plus reconnection.

Qualitative discussions (see, for instance, the excellent review article of Vilenkin [2]) and numerical studies [8] lead to the following conclusions [19] :

(1) Expansion of the Universe straightens out long strings on scales smaller than the horizon and conformally stretches them on scales greater than the horizon.

(2) There are just a few segments of 'infinite' strings per horizon volume at any time.

(3) Loops surviving the decay due to gravitational radiation at time t have sizes greater than $\sim G\mu$ times the horizon length. In particular, the smallest surviving loops today have a size $\sim G\mu t_o \sim$ few kpc, for $G\mu \sim 10^{-6}$.

(4) During the radiation and matter dominated eras, the length distributions of closed loops are, respectively [20]

$$\frac{dn}{dR} \sim t^{-3/2} R^{-5/2} , \quad \frac{dn}{dR} \sim (tR)^{-2} . \tag{2.1}$$

Today, the smallest surviving loops have masses $M \sim R\mu \sim (G\mu/10^{-6})\ 10^{10}\ M_\odot$ and they are typically separated by ~ 10 Mpc.

(5) During all evolutionary phases, the energy density due to strings is always much smaller than the total energy density. Intercommuting and gravitational radiation is crucial for this. It turns out that the smallest surviving loops dominate the string energy density.

3. STRING SCENARIO FOR GALAXY FORMATION

Rapidly moving open strings and oscillating closed loops produce density perturbations which begin to grow at the time t_{eq} of equal matter and radiation densities if cold or hot dark matter dominates, respectively at the time t_{rec} of recombination for a baryon dominated universe. In the former case there is little growth before t_{eq} due to the Mészáros-Zel'dovich effect [21] and in the latter case there is strong Silk damping before t_{rec}

Larger loops produce larger potential wells, accreting larger masses. These contain smaller loops and matter around them. In this way a hierarchy of objects - smaller ones inside bigger ones - are produced, which may be identified with galaxies and their associations.

3.1 Determination of $G\mu$ from the Properties of Abell Clusters

BRANDENBERGER and TUROK have argued [22] that the positions of Abell clusters should accurately trace the positions of the very large loops that produced them, because there has been hardly any gravitationally induced motion of matter since the perturbations started to grow. They determined the dimensionless parameter $G\mu$ by requiring that loops with the separation of Abell clusters virialized the observed Abell overdensity inside an Abell radius by the present time. Following TUROK [23], we present now the detailed analytic determination of $G\mu$ by using the spherical accretion model. (These calculations

should be supplemented by N-body simulations.) Since it turns out [22] that $G\mu$ has to be too large for a baryon-dominated universe with $\Omega_0 \simeq 0.1$, we assume an $\Omega = 1$ universe which is dominated by some collisionless form of matter. We give here the details for a cold-dark-matter-dominated universe and only summarize briefly the main differences for hot dark matter [24].

The analysis involves the following four steps:

1) Mean separation of loops

The loops which started the accretion process were formed in the radiation-dominated era. Their number density distribution $n(R,t)$, as a function of the rms radius R at time t, has been found from numerical simulations [8] to be given by (see eq. (2.1)):

$$n(R,t) = \nu R^{-5/2} t_i^{-3/2} (a(t_i)/a(t))^3 , \qquad (3.1)$$

where $t_i \ll t_{eq}$ is some initial time and $\nu \simeq 0.01$, with a relatively large uncertainty.

The scale factor $a(t)$ satisfies the Friedman equation

$$(\frac{\dot{a}}{a})^2 = \frac{8\pi G}{3} (\rho_m)_{eq} (\frac{1}{a^3} + \frac{1}{a^4}) , \qquad (3.2)$$

where ρ_m denotes the matter density. We normalize a such that $a_{eq} = 1$. Then the Hubble parameter at t_{eq} is given by $H_{eq}^2 = (16\pi G/3)(\rho_m)_{eq}$, and is related to its present value H_0 through

$$H_0^2 = \frac{8\pi G}{3} (\rho_m)_0 = \tfrac{1}{2} H_{eq}^2 \frac{1}{(1+z_{eq})^3} . \qquad (3.3)$$

For three massless neutrino flavors we have $z_{eq} = 2.5\times10^4 \Omega_0 h_0^2$, and thus for $\Omega_0 = 1$, $z_{eq} = 6.25\times10^3 \cdot h_{50}^2$.
This gives

$$H_{eq}^{-1} = 8.6\ h_{50}^{-4}\ \text{kpc} . \qquad (3.4)$$

Introducing the conformal time τ by $dt = a(\tau)\, d\tau$, we find easily the solution of (3.2):

$$a = \frac{1}{\sqrt{2}} H_{eq}\tau + \frac{1}{8} (H_{eq}\tau)^2 , \quad t = \frac{1}{2\sqrt{2}} (\tau H_{eq})\tau + \frac{1}{24} (H_{eq}\tau)^2 \tau .$$

Using this for $t_i \ll H_{eq}^{-1}$ in (3.1) leads to

$$n(R,t) = 2^{3/4} \nu R^{-5/2} H_{eq}^{3/2} a(t)^{-3} .$$

This gives for the number density $n_>(R,t)$ for loops with radius greater than R

$$n_{>}(R,t) = \int_{R}^{\infty} n(R,t)\, dR = \frac{2}{3}\, 2^{3/4}\, \nu\, H_{eq}^{3/2}\, a(t)^{-3}\, R^{-3/2} \ .$$

The corresponding mean separation $d(R)$ today, defined by $n_{>}(R,t_o) = d(R)^{-3}$, satisfies

$$R = \sqrt{2}\, (\tfrac{2\nu}{3})^{2/3}\, H_{eq}\, d^2/(1+z_{eq})^2 \ . \qquad (3.5)$$

For Abell clusters the mean separation is $d_{cl} \simeq 110\, h_{50}^{-1}$ Mpc .

Thus (3.5) gives, together with (3.4) for the radius of the loops which produced them

$$R_{loop} = 1.8\, \nu_{.01}^{2/3}\, d_{110}^2\, h_{50}^{-2}\ \text{kpc} \ . \qquad (3.6)$$

Note that R_{loop} is not much smaller than the Hubble radius H_{eq}^{-1} at t_{eq} (eq.(3.4)). Thus the cluster loops have been laid down close to t_{eq} .

2) Loop mass

We write for the loop mass $M = \beta\mu R$. The numerical simulations [8] give $\beta \simeq 10$. With (1.4) and (3.6) we find

$$M_{loop} = 3.8\times 10^{11}\, h_{50}^{-2}\, \beta_{10}\, \nu_{.01}^{2/3}\, d_{110}^2\, \mu_6\, M_\odot \ , \qquad (3.7)$$

with $\mu_6 := G\mu/10^{-6}$.

3) Accreted mass

Next we determine the mass accretion of a loop in the spherical model. Consider a seed mass m at $r = 0$ and assume that at some initial time t_i we have an unperturbed Hubble flow, for which the radius r_b of a given mass shell satisfies

$$\tfrac{1}{2}\dot{r}_b^2 - \frac{G\bar{M}}{r_b} = 0 \ , \qquad \bar{M} = \frac{4\pi}{3}\, r_i^3 (\rho_b)_i$$

giving

$$r_b = r_i (t/t_i)^{2/3} \ , \qquad \rho_b = \frac{1}{6\pi G t^2} \ . \qquad (3.8)$$

The perturbed flow is determined by

$$\dot{r}^2 = \frac{2G\bar{M}}{r} - \frac{2G\bar{M}}{r_i}\, \delta_i \ , \qquad \delta_i := \frac{m}{\bar{M}} \ . \qquad (3.9)$$

This Friedman equation has the following well known solution in parametric form ($\delta_i \ll 1$)

$$r = \frac{r_i}{2\delta_i}(1-\cos\theta) , \qquad t = \frac{3t_i}{4\delta_i^{3/2}}(\theta-\sin\theta) . \tag{3.10}$$

Let r_m denote the maximal radius reached at time t_m. We have [25]

$$\frac{r_m}{r_i} = \delta_i^{-1} , \qquad \delta_i(t_m/t_i)^{2/3} = (\frac{3\pi}{4})^{2/3} . \tag{3.11}$$

This gives $\frac{\rho_m}{\rho_i} = \delta_i^3 = (\frac{3\pi}{4})^2(\frac{a_i}{a_m})^3$ or the well known result

$$(\frac{\rho}{\rho_b})_m = (\frac{3\pi}{4})^2 = 5.55 .$$

After turnaround the shell collapses and is thereby virialized. The virial theorem and energy conservation for dissipationless matter imply that the virialized radius r_f is equal to $r_m/2$.
Later on we have

$$\rho/\rho_b = (\frac{3\pi}{4})^2 \times 8 \times (a/a_m)^3 ,$$

where the last factor comes from the dilution of the background density. This gives

$$1 + z_m = (\rho/\rho_b)_o^{1/3} \tfrac{1}{2} (\frac{4}{3\pi})^{2/3} . \tag{3.12}$$

On the other hand, we have $m = \delta_i M_{ac}$, where M_{ac} is the accreted mass. But from (3.11) $\delta_i = (3\pi/4)^{2/3}(1+z_m)/(1+z_i)$, and thus the seed mass is related to the accretion mass by (using (3.12)):

$$m = \tfrac{1}{2}\frac{1}{1+z_i}(\rho/\rho_b)_o^{1/3} M_{ac} . \tag{3.13}$$

This is not quite the mass of the loop since the calculation assumed matter domination. But the cluster loops have been laid down close to t_{eq}, as was already emphasized. Thus expression (3.13) for $z_i \simeq z_{eq}$ gives a good estimate for the loop mass:

$$M_{loop} = M_{ac}(\rho/\rho_b)_o^{1/3}\frac{1}{1+z_{eq}}(\frac{\xi}{2}) . \tag{3.14}$$

Here, the additional correction factor ξ takes into account that in reality the accretion at $\sim t_{eq}$ is somewhat less efficient. A reasonable value for cluster loops should be $\xi \simeq 2$.

4) Masses and Overdensities of Abell Clusters

The mass of an Abell cluster is estimated with the help of the virial theorem $M_{cl} = 3\sigma^2 R_A/G$ for an Abell radius $R_A = 3\ h_{50}^{-1}$ Mpc

and a one dimensional dispersion $\sigma \sim 700$-1000 km/s. Numerically we find

$$M_{cl} = 10^{15}\, h_{50}^{-1}\, \sigma_{700}^{2}\, M_\odot \quad . \tag{3.15}$$

This gives, with $(\rho_b)_o = \rho_{crit} = 7\times10^{10}\, h_{50}^{2}\, M_\odot (\mathrm{Mpc})^{-3}$, an over-density of about $130\,\sigma_{700}^{2}$. Inserting this into (3.14) and setting $M_{ac} = M_{cl}$ yields for the mass of cluster loops

$$M_{loop} = 8\times10^{11}\, h_{50}^{-3}\, \sigma_{700}^{8/3}\, M_\odot \quad . \tag{3.16}$$

Comparing this with the independent estimate (3.7) allows us to compute $G\mu$:

$$\boxed{\mu_6 \simeq 2\, h_{50}^{-1}\, \beta_{10}^{-1}\, \nu_{.01}^{-2/3}\, \sigma_{700}^{8/3}\, d_{110}^{-2}} \quad . \tag{3.17}$$

The dependence of μ_6 on the various parameters, as well as the rough accretion model which has been used, make it obvious that $G\mu$ remains quite uncertain. The uncertainty is at best an order of magnitude. Presumably, however, cosmic strings would not help galaxy formation for $G\mu \lesssim 10^{-7}$. Observations of the millisecond pulsar 1937+21 may soon establish such an upper bound (see Sect. 4.3).

String theory makes also definite predictions for the correlation function. Scale invariance implies that it must be a dimensionless function $\xi(r/d)$, where d is the mean separation, with no free parameters. From numerical simulations TUROK [26] derived the result $\xi(r/d) \simeq 0.3\,(r/d)^2$, which agrees well with observations [27], both in amplitude and in shape.

Recently it has been shown [24] that cluster formation in a neutrino dominated universe with cosmic strings is rather similar to that for cold dark matter. Galaxy masses come out lower than with strings and cold dark matter and the spectrum of galaxies turns out to be flatter.

4. OBSERVATIONAL EFFECTS

Even if strings have nothing to do with galaxy formation, they might produce some unique observational effects.

4.1 Lensing

One expects [28] that out of $\sim 10^4$ quasars, one will be doubled by a string. It is, however, difficult to prove that a particular pair of quasar images is due to a cosmic string. PACZYNSKI [29] has,

therefore, suggested that observations of very distant galaxies with the Space Telescope may be a better way for discovering cosmic strings. A galaxy may appear cut by a sharp edge if there is a cosmic string between the galaxy and the observer. Since the probability for this is quite small, an extensive observational program would be necessary to see this.

4.2 Discontinuities in the 3K-Background

KAISER and STEBBINS [30] have pointed out that rapidly moving strings would produce steplike discontinuities in the 3K-background which might be observable.

The discontinuity can easily be determined. In a frame in which the string is at rest and we are moving with four velocity $U = \gamma\,(1,\underline{v})$, the frequency ratio ω_1/ω_2 of two light rays passing on different sides of the string is

$$\frac{\omega_1}{\omega_2} = \frac{U\cdot K_1}{U\cdot K_2} \ ,$$

where $K_{1,2}$ are the wave four vectors indicated in Fig. 5. Clearly, $K_{1,2} = \omega(1,\underline{n}_{1,2})$, where ω is the unperturbed frequency and $\underline{n}_{1,2}$ are the unit directions of the two rays at the position of the observer. Thus $(\beta = \frac{v}{c})$

$$\frac{\omega_1}{\omega_2} = \frac{1-\underline{v}\cdot\underline{n}_1}{1-\underline{v}\cdot\underline{n}_2} = \frac{1-\beta\,\sin\alpha/2}{1+\beta\,\sin\alpha/2} \simeq 1-\beta\alpha = 1-8\pi G\mu\beta \ . \qquad (4.1)$$

This shows that the relative temperature jump in the background radiation is $\Delta T/T = 8\pi G\mu\beta$, which may typically be $\sim 10\,G\mu$. Present observational limits are consistent with $G\mu \lesssim 10^{-5}$.

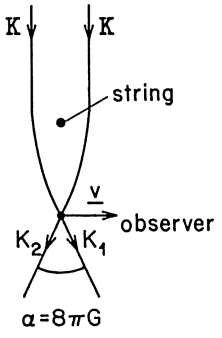


Fig. 5. Diagram for frequency jump induced by a moving cosmic string

4.3 Stochastic Gravitational Wave Background and Mean-Square Timing Residuals of Millisecond Pulsars

Gravitational waves produced by oscillating loops add up to a stochastic gravitational background which would induce a 'timing noise' of millisecond pulsars.

A quantitative analysis allows one to relate the density parameter Ω_g of the stochastic background radiation to the mean square timing residuals [31,32] . This analysis goes as follows. Let $z_g(t)$ be the redshift of a pulsar signal received at time t . In linearized theory we have

$$z_g(t_R) = -\tfrac{1}{2}\, n^i n^j \int_{t_s}^{t_R} \partial_t\, h_{ij}\, dt \,, \tag{4.2}$$

where $\underline{n}$ is the unit vector in the direction of propagation of the pulsar signal and $h_{\mu\nu}$ are the metric perturbations induced by the stochastic background in the traceless transverse gauge [5]. The integral is along the electromagnetic path from the source (emission time t_s) to the receiver.

We consider $z_g(t)$ as a stationary (wide sense) stochastic process. Its spectral density $S_{z_g}(\omega)$, which is the Fourier transform of the correlation function

$$\langle z_g(t)\, z_g(t+\tau)\rangle = 2\int_0^\infty \cos(\omega\tau)\, S_{z_g}(\omega)\, d\omega \,, \tag{4.3}$$

can easily be related to the spectral density $\rho_g(\omega)$ of the gravitational background, assuming isotropy and random phases and polarisations. One finds from (4.2) [32]

$$S_{z_g}(\omega) = \frac{8\pi}{3}\, \frac{\rho_g(\omega)}{\omega^2}\, B(\omega) \,, \tag{4.4}$$

where $B(\omega)$ is an explicit distance-dependent function, which is close to 1 for ω^{-1} much smaller than the light distance (which is always true in practice).

The redshift $z_g(t)$ is simply related to the phase residuals $\Phi_R(t)$:

$$z_g(t) = \frac{P}{2\pi}\, \frac{d\Phi_R}{dt} \,, \tag{4.5}$$

where P is the pulsar period. We are interested in the mean-square phase residuals $\langle R^2(T)\rangle$,

$$R^2(T) = \frac{1}{T}\int_0^T \Phi_R(t)^2\, dt \ ,$$

where T is the observation time. Using (4.5), (4.3), and (4.4) one finds

$$\langle R^2(T)\rangle = \frac{32\pi}{3}\left(\frac{2\pi}{P}\right)^2 \int_{\alpha/T}^{\infty} d\omega\, \frac{\rho_g(\omega)}{\omega^4}\left(1 - \frac{\sin \omega T}{\omega T}\right) . \qquad (4.6)$$

The lower limit in the integral cuts off low frequencies which play no role for a finite observation time. The precise value of α depends on the actual form of the low-frequency spectrum. A reasonable choice is $\alpha = 2\pi$.

Let us take as an example a constant value for

$$\Omega_g(\omega) := \omega\, \rho_g(\omega)/\rho_c \ , \qquad (4.7)$$

where ρ_c is the critical density. Then

$$\Omega_g = \langle R^2\rangle\, T^{-4}\, \alpha^4 H_o^2 \qquad (4.8)$$

or numerically

$$\Omega_g = 10^{-7} h_o^{-2} \alpha^4 T_y^{-4}\, R^2_{\mu s} \ . \qquad (4.9)$$

Here T is measured in yr and R in μs .

Taylor and coworkers [33] have now followed the pulsar arrivals times of the millisecond pulsar 1937+21 for more than four years with an impressive precision. Fig. 6 shows their post-fit arrival-time residuals. These imply the safe upper limit [33]

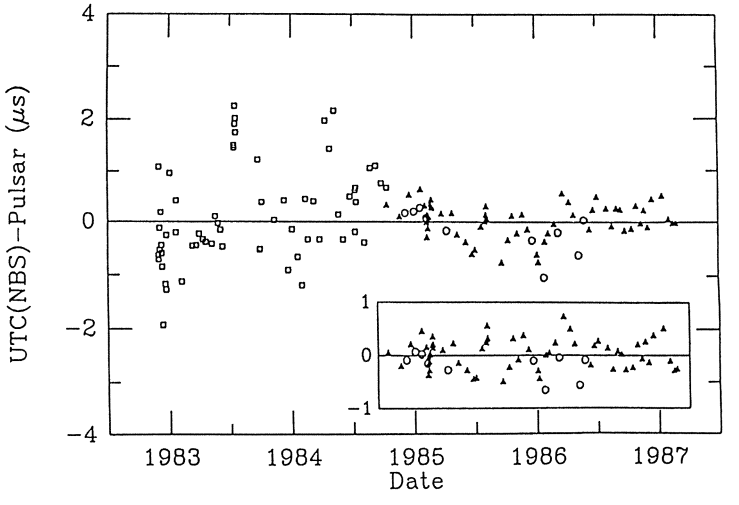


Fig. 6. Post-fit arrival-time residuals for PSR 1937+21. The inset shows the remarkable accuracy of the recent observations (Ref.[33])

$$\Omega_g < 4\times10^{-7} \qquad (4.10)$$

at a frequency $\nu \simeq 0.23$ cy y^{-1} (with a bandwidth of a factor of e). This limit is already close to what one expects [34] from cosmic strings [35]:

$$\Omega_g^{strings} \simeq 4\times10^{-8}\, h_o^{-2}\Omega_o^{-1}\,\beta_{10}\,\nu_{.01}\,\hat{\gamma}_5\,\mu_6^{\frac{1}{2}}\,, \qquad (4.11)$$

where $\hat{\gamma}_5 = \hat{\gamma}/5$, $\hat{\gamma} = \gamma/\beta$. It may well be that the pulsar observations will exclude (or confirm) cosmic strings of the type which would be interesting for galaxy formation in just a few years.

5. CONCLUDING REMARKS

The string scenario for galaxy formation has not yet been worked out in sufficient detail. Numerical simulations are needed (and are underway) in order to answer questions such as:

(i) Can string-type models account both for galaxy and cluster formation ?

(ii) Is it possible to explain the formation of superclusters and voids ?

(iii) What about peculiar streaming motions on large scales ?

(iv) Why is the inferred value of Ω_o on the scale of clusters $\lesssim 0.2$, in spite of $\Omega_o = 1$?

Apart from such potential problems, there is of course also the question of how to reconcile strings with inflation. While this may not be impossible, suggested models need fine tuning in order to produce two phase transitions with the right properties.

Superconducting strings, which behave as superconducting wires and can produce large electromagnetic effects, have recently become very fashinable [36]. The maximal current can be gigantic. For bosonic charge carriers one finds [36]

$$J_{max} \sim e\sqrt{\mu} \sim 10^{20}\ \text{amps}\ \mu_6\,. \qquad (5.1)$$

The magnetic dipole radiation is roughly

$$\dot{M}_\gamma \sim -(JR^2\omega^2)^2 \sim -J^2 \sim -e^2\mu\ (J/J_{max})^2$$

or

$$\dot{M}_\gamma \sim -\alpha\mu\,(J/J_{max})^2 = -2\times10^{51}\mu_6^2\,(J/J_{max})^2\ \text{erg/s}\,. \qquad (5.2)$$

This electromagnetic radiation cannot propagate through the cosmic medium since the plasma frequency is very much larger than the wave frequencies. Like a giant pulsar, each loop will blow a bubble around itself. OSTRIKER, THOMSON, and WITTEN [37] have suggested that the

observed voids in the galaxy distribution may have been created in this manner and that galaxies have been formed on the boundaries of these huge bubbles.

Unfortunately, time does not allow me to discuss the interesting microphysics of superconducting strings and the various recent suggestions about their astrophysical applications [38].

Acknowledgements

I would like to thank G. Boerner for organizing this interesting workshop at such a beautiful place. I have profitted much from discussions with R. Brandenberger and N. Turok and I am grateful to J. Taylor for showing us in Zürich his latest results on the millisecond pulsar. I thank the Swiss National Science Foundation for financial support.

REFERENCES AND FOOTNOTES

[1] See, e.g., the contributions of M. Davies, C. Frenk, and M. Rees at the IAU Symposium Nr. 130, Balatonfüred (14.-21. 6.1987), to be published.

[2] For an earlier review, see A. Vilenkin, Phys.Rep. 121, 263 (1985).

[3] A.A. Abrikosov, Zh.Eksp.Theor.Fiz. 32, 1442 (1957) [Sov.Phys. JETP 13, 451 (1961)].

[4] H.B. Nielsen and P. Olesen, Nucl.Phys. B61, 45 (1973).

[5] See, for example, N. Straumann, in Proceedings of the 1982 SIN Spring School: Beyond the Standard Model.

[6] E. Witten, Phys.Lett. 153B, 243 (1985).

[7] M. Spivak, "Differential Geometry", Vol. IV, Chap. 9, Publish or Perish 1970-75.

[8] A. Albrecht and N. Turok, Phys.Rev.Lett. 54, 1868 (1985), and in preparation.

[9] T. Vachaspati and A. Vilenkin, Phys.Rev. D31, 3052 (1985).

[10] Quartic terms induce in general an effective position- and polarisation-dependent photon mass term, which vanishes outside the string and whose Compton wave length within the string is of the order of the string width.

[11] For further details see T. Vachaspati, A.E. Everett, and A. Vilenkin, Phys.Rev. D30, 2046 (1984). To our knowledge, there does not yet exist a satisfactory detailed calculation for the electromagnetic radiation.

[12] E.P.S. Shellard, Nucl.Phys. B283, 624 (1987).

[13] See, e.g., N. Straumann, "General Relativity and Relativistic Astrophysics", Texts and Monographs in Physics, Springer (Berlin), 1984.

[14] J.R. Gott, Ap.J. 288, 422 (1985).

[15] It has been shown recently (R. Gregory, DAMTP preprint, 1987) that the full coupled gravity-string field equations imply an asymptotically conical structure for a general local string.

[16] N. Turok, Phys.Lett. B126, 437 (1983).

[17] T. Vachaspati, Phys.Rev. D35, 1767 (1987).

[18] T. Vachaspati and A. Vilenkin, Phys.Rev. D30, 2036 (1984).

[19] The numerical simulations have not yet reached a satisfactory status and there are still some controversies going on.

[20] The number density $R\, dn/dR \sim R^{-3}$ of loops of sizes $\sim R$ in the interval $\Delta R \sim R$ is diluted by the expansion with the factor $(a(R)/a(t))^3$.

[21] P.J.E. Peebles, "The large-scale structure of the universe", Princeton Univ. Press 1980 (§ 12).

[22] N. Turok and R. Brandenberger, Phys.Rev. D33, 2175 (1986).

[23] N. Turok, CERN/ESO Winter School on Cosmology and Particle Physics, 1987, Imperial/TP/86-87/23.

[24] R. Brandenberger et al., DAMPT preprint, July 1987.

[25] Treating the spherical model in perturbation theory, one finds easily for $\delta\rho/\rho = (\bar{\rho} - \rho_b)/\rho_b$, where $\bar{\rho}$ is the average density of a given mass shell, for the same initial conditions:
$$\frac{\delta\rho}{\rho} = \delta_i \left[\frac{3}{5}\left(\frac{t}{t_i}\right)^{2/3} + \frac{2}{5}\left(\frac{t}{t_i}\right)^{-1} - 1\right] .$$
Thus the growing mode reached at t_m the value $\frac{3}{5}\left(\frac{3\pi}{4}\right)^{2/3} = 1.06$.

[26] N. Turok, Phys.Rev.Lett. 55, 1801 (1985).

[27] S. Shechtman, Ap.J.Suppl. 57, 77 (1985); N. Bahcall and W. Burgett, Ap.J. (Letters) 300, L35 (1986).

[28] A. Vilenkin, Ap.J. 282, L51 (1984).

[29] B. Paczynski, Nature 319, 567 (1986); see also: C.J. Hogan, 13th Texas Symposium on Relativistic Astrophysics, p. 324 (World Scientific 1987).

[30] N. Kaiser and A. Stebbins, Nature 310, 391 (1984).

[31] B. Mashhoon, Mon.Not.R.astr.Soc. 199, 659 (1982).

[32] B. Bertotti, B.J. Carr, and M.J. Rees, Mon.Not.R.astr.Soc. 203, 945 (1983); C.J. Hogan and M.R. Rees, Nature 311, 109 (1984).

[33] L.A. Rawley, J.H. Taylor, M.M. Davis, D.W.Allan, Princeton preprint, 1987.

[34] A rough estimate is obtained as follows: Loops decaying at time t during the radiation era have size $R \sim G\mu t$ and produce gravitational waves of frequency $\sim R^{-1}$ and energy density $\rho_g \sim \mu R n_>(R) \sim (G\mu)^{\frac{1}{2}}(Gt^2)^{-1}$. Since ρ_g is redshifted as $\rho_g \simeq (30\ Gt^2)^{-1}\Omega_\gamma$, we find $\Omega_g \sim 30(G\mu)^{\frac{1}{2}}\Omega_\gamma \simeq 0.6\times10^{-6}\mu_6^{\frac{1}{2}} h_o^{-2}$.

[35] R. Brandenberger, A. Albrecht, and N. Turok, Nucl.Phys. B277, 605 (1986).

[36] E. Witten, Nucl. Phys. B246, 557 (1985).

[37] J. Ostriker, C. Thompson, and E. Witten, Phys.Lett. B180, 231 (1986).

[38] E. Witten, 13th Texas Symposium on Relativistic Astrophysics, World Scientific 1987.

The Correlation Strength of Clusters in the Cosmic String Model

Jia-lu Zhang and Yi-peng Jing

Center for Astrophysics, University of Science and Technology of China, Hefei, People's Republic of China

It is known that the two-point correlation functions of galaxies and of clusters of galaxies have three systematic characteristics. 1. all correlation functions have power laws with approximately the same index -1.8, i.e. $\xi(r)=\alpha r^{-1.8}$; 2. the correlation strength α increases with the increase of the mean separations d of the systems considered and 3. α also increases with the increase of the mean richness of the systems (the Bahcall-Soneira relation).

A plausible model for these features is suggested by SZALAY et. al. (1) They supposed that the large scale structure of the universe is scale free with fractal correlations $\xi_{frac}(r)=0.3\left[\frac{d}{r}\right]^{1.8}$, the observed correlations ξ_{obs} of galaxy systems are a superposition of the fractal and gravitational (ξ_{grav}) correlations

$$1+\xi_{obs}=(1+\xi_{frac})(1+\xi_{grav}),$$

and the correlation functions of galaxies and of clusters can be produced by introducing different ξ_{grav}. However, this picture involved only the linear sizes of systems, so that only the first two observational facts mentioned above can be explained.

The suggestion of SZALAY et. al. can be developed in the cosmic string framework. TUROK (2) has shown that the two-point correlation function of loops with similar sizes can be represented as

$$\xi_{loop-loop}\left[\frac{d_1}{r}\right]=0.2\left[\frac{d_1}{r}\right]^2.$$

Because galaxies and clusters of galaxies are formed by accretion onto different size loops we infer the scale invariance of the large scale structure from the scale invariant property of loops. So, the suggestion of SZALAY et. al. can naturally be introduced in the cosmic string picture. Correlation functions of clusters represent the correlation of loops (2) and the correlation function of galaxies can be explained satisfactorily by gravitational amplifications (3) (4). Whether the observed relation between α and richness N can be obtained in this picture model is an unsolved problem, but it is still a very interesting and important problem. Now we begin to study this problem.

First of all, we plot the observational data of N and d in figure 1 for clusters with different richness, and find that the relation between N and d can be well represented by $d\propto N^{2/3}$.

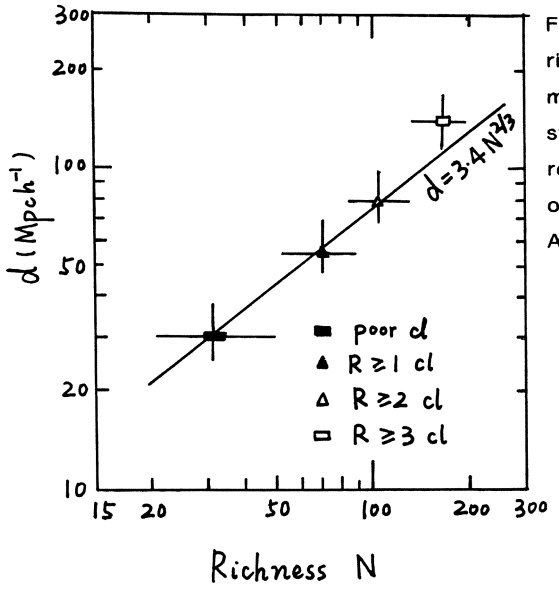


Figure 1 The relation between the richness N of clusters and their mean separation d. The straight line represents the relation $d=3.4N^{2/3}$. The observational data are taken from ABELL [12] and SHECTMAN [9]

Why there exists this relation between N and d has never been solved in any usual galaxy formation theory, even at a very crude level. Now we will consider this relation in the cosmic string model.

In the string theory, clusters of different richness are formed by accretion of galaxies onto the loops with different sizes and different mean separations. Here we assume that the loops of radius R, form clusters of richness N. We suppose that the cosmological model is Einstein-de-Sitter universe with $\Omega=1$ dominated by cold dark matter, and the accretion of strings began at t_{eq}, the epoch when matter and radiation density were equal.

Because Abell clusters usually can be considered to be virialized, the shell with radius r_A should have the maximum size at the time t_{max}, which should be less than the age of the Universe t_0.

According to the spherical accretion model, the dynamical equations of the r_A shell (assuming its initial radius is γ_i) are [5]

$$r = \frac{r_i}{2\delta_i}(1-\cos\theta), \qquad t=\frac{3t_i}{4\delta_i^{3/4}}(\theta-\sin\theta) ;$$

and the maximum radius of the shell is $r_{max}=\frac{\gamma_i}{\delta_i}$. Later, the shell would have δ_i collapsed and virialized with radius $\frac{1}{2}\, r_{max} \equiv r_A$.

Using the initial perturbation expression $\delta_i = \frac{3\beta\mu R_1}{4\pi r_i^3 \rho_i}$, where ρ_i is the mean density of the Universe at the initial time, we can get the initial radius of the r_A shell $r_i = \gamma_i = \left(\frac{3R_1 r_A \beta\mu}{2\pi\rho_i}\right)^{1/4}$. Then the number of galaxies in the Abell cluster formed from the R_1 loop is

$$N = \frac{4\pi}{3}(1+z_i)^3\, n_g(0)\left[\frac{3\beta\mu R_1 r_A}{2\pi\rho_i}\right]^{3/4} .$$

where $n_g(0)$ is the mean number density of galaxies nowadays and $n_g(0) = 8.0 \times 10^{-3}$ $(Mpch^+)^{-3}$, $(a+z_i) = 1+z_{eq} = 2.5 \times 10^4 h^2$, $\rho_i = \frac{3(1+z_i)^3}{32\pi G t_o}$.

From the distribution of loops, we can easily get the expression of R_1 in terms of the mean separation d of clusters formed around the R_1 loops

$$R_1 = \left[\frac{2\nu}{3}\right]^{2/3} \frac{1}{(1+z_i)^{1/2}} d^2 t_o^{-1} .$$

Therefore

$$N = \frac{4\pi}{3} (1+z_i)^{3/8} n_g(0) \left[16\beta G u \left[\frac{2\nu}{3}\right]^{2/3} \right]^{3/4} (t_o r_A)^{3/4} d^{3/2} .$$

where β, ν, μ have the usual meanings in the cosmic string theory. By substituting the values of z_i, $n_g(0)$, β, ν, r_A and $t_o = \frac{2}{3}H_o^{-1} = 2000$ $Mpch^{-1}$ into the equation above, we can find $N = 2100(G\mu)^{3/4} d^{3/2} h^{9/4}$.

If we take $G\mu = 3.2 \times 10^{-6} h^{-1}$ and express d in units of h^{-1}Mpc, then the relation between d and N

$$d = 3.4\ N^{2/3}$$

can be obtained. We plot this relation in Figure 1, and find that it fits well to the observational data. The value of $G\mu$ we have taken is reasonable. It is consistent with the demands from other observational facts such as the microwave background anisotropy and so on.

On the other hand, from the correlation function of loops, the relation between α and the mean separation d of clusters can be expressed by $\alpha = 0.2\ d^2$. Therefore we have the relation between α and N

$$\alpha = 2.3\ N^{1.3}$$

We can easily show that this equation can only be applied to clusters with $N \gtrsim 20$. Because the smaller richness clusters are formed from the accretion on to smaller loops, the spheres of radius 1.5 h^{-1}Mpc around the loops may have not collapsed and virialized. Also the gravitational amplifying effects on the correlation function become more important due to their smaller separations. Both these effects make α decrease with the decrease of N more slowly than the power law $N^{1.3}$.

BAHCALL et. al. [6] have suggested that superclusters might also have a invariant correlation function $\xi_s = 1500\ \gamma^{-1.8}$. But because the number of superclusters used in their statistical study is so small, the uncertainty is very large. Especially, KALINKOV et. al. [7] recently have reported that there are no correlations for another supercluster sample. So the $\xi_s = 1500\ \gamma^{-1.8}$ for superclusters is very doubtful. Moreover, the definition of the superclusters is very ambiguous. It is not necessary to include superclusters here.

BAHCALL et. al. [6] also suggested the relation $\alpha = 20N^{0.7}$ for galaxies, clusters (not included the R≥3 Abell clusters) and supercluster. FANG [8] has proposed that in the cosmic string model this relation (and the first two observational properties listed at the head of this paper) can be explained in a phenomenological model with strings distributed with fractal dimension to be 1.4.

We plot our result in Figure 2, and find that it is completely consistent with the statistical results of SHECTMAN [9] (for poor clusters), and BAHCALL et. al.]10] (for R≥1, R≥2 Abell clusters). It is shown that the model we used here can explain the correlation-richness dependence of clusters. We should point out here that not only $\xi_s = 1500\ \gamma^{-1.8}$ for superclusters is doubtful, but also the comparison of correlation-richness relations of clusters and of superclusters is difficult and problematic because of the very different definitions of their richness. Therefore our discussion excludes superclusters. The statistical study of R≥3 Abell clusters by GAO [11] showed that the correlation function can be approximately represented by $\xi_{R\geq3} = 1800\ \gamma^{-1.8}$. It is interesting that this result just agrees with our Fig. 2.

Finally, we make our conclusions: we have analysed the observational data of clusters and found that there exists the relation $d \propto N^{2/3}$ between the richness of clusters N and their mean separations d. We explain this fact and the Bahcall-Soneira relation in the cosmic string scenario. The value of $G\mu = 3.2 \times 10^{-6}h^{-1}$ we have adopted is consistent with other requirements and with some other observed facts.

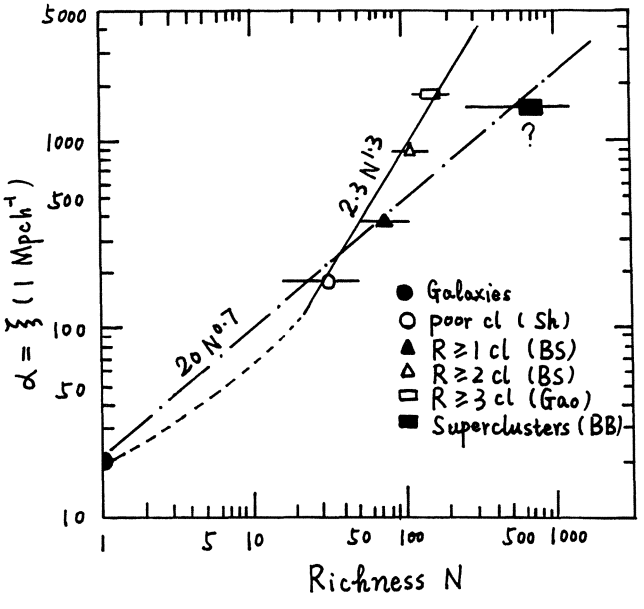


Figure 2. The dependence of the correlation function on the mean richness of the system. The straight line represents $\alpha = 2.3\ N^{4/3}$. The dashed line is the qualitative trend of α after considering the univirialized and gravitational effects.

REFERENCES

1. A. Szalay, D. Schramm: Nature 314, 718 (1985)
2. N. Turok: Phys. Rev. Lett. 55, 1801 (1985)
3. N. Turok, R.H. Brandenberger: Phys. Rev. D33, 2175 (1986)
4. Y.P. Jing, J.L. Zhang: Acta Astrophys. Sinica (in press)
5. P.J.E. Peebles: The large-scale structure of the Universe (Princeton University Press, Princeton 1980)
6. N.A. Bahcall, W. Burgett: Astrophys. J. 300, L35 (1986)
7. M. Kalinkov, I. Kuneva: M.N.R.A.S. 218, 49P (1986)
8. L.Z. Fang: "Fractals of the large-scale structure of the Universe", USTC-preprint (1986)
9. S.A. Shectman: Astrophys. J. Suppl. 57, 77 (1985)
10. N.A. Bahcall, R.A. Soneira: Astrophys. J. 270, 20 (1983)
11. Y. Gao: A report on Yellow Mountain International Summer School on Physics and Astrophysics No.1 (August, 1987)
12. G.O. Abell: Astrophys. J. Suppl. 3, 211 (1958)

Fractals in the Distribution of Galaxies

L.Z. Fang

Beijing Astronomical Observatory, Academia Sinica,
Beijing, People's Republic of China

The cosmic string model [1] pictures the formation of galaxies as an accretion process by density perturbation seeds of cosmic strings which are topologically stable, macroscopic defects formed during a phase transition of grand unified symmetry breaking in the early universe. The system of cosmic strings consists mainly of closed loops, which were produced by a self-intersection process from parent strings. The formation rate of such loops is about one or few per horizon volume per expansion time. In this case, the number density of loops with size R can be described by

$$n(R) \backsim R^{-D} \tag{1}$$

where D=5/2. The meaning of D can also be obtained from the expression of the total number of R loops in a volume with length scale r, it is

$$N_R(r) \backsim B_R r^D \tag{2}$$

where B_R is constant for a given R.

Since n(R) and $b^D n(bR)$ are identical in distribution for every ratio b>0, the distribution of loops is diagonally self-affine [2]. Namely, the distribution of galaxies should be represented by a self-affine system with local fractal dimension D'=1 and global fractal dimension D=5/2. [3]

It can be shown that the two-point correlation function of objects with mean separation d is given approximately by [4]

$$\xi(r) = \frac{1}{4\pi^2 n(R) R^3} \frac{1}{(r/R)^2 [1 - (r/2R)^2]^{1/2}} \tag{3}$$

where R is equal to about (2~3)d. Eq. (3) is available only for r<R. When r is larger than R, the correlation function is

$$\xi(r) = (DL^{3-D}/3) r^{D-3} - 1, \tag{4}$$

where L is the radius of the volume of the sample considered.

A series of statistical features of the distribution of galaxies can be explained by the above-mentioned fractal description of the cosmic string model.

1. $r^{-1.8}$ Law

From Eq. (3) one can find that the dependence of $\xi(r)$ with r is given by the term $1/(r/R)^2\,[1-(r/2R)^2]^{1/2}$, which can be fitted by a power law of r with index −1.8 in the range r<R. In fact, this index is determined by the local fractal dimension D′, namely,

$$1.8 \backsim 3 - D'. \tag{5}$$

This is just the reason for that the correlation functions of all types of objects have the same index 1.8 [5].

2. Dependence of Correlation Strength on the Scales of Objects

In the correlation function $\xi(r)=\alpha r^{-1.8}$, the dimensional correlation strength α is found to be [4,5]

$$\alpha \backsim d^{1.4}. \tag{6}$$

From Eq.(3), the dimensional correlation strength α is given by

$$\alpha \backsim 1/4\pi^2 n(R) R. \tag{7}$$

Substituting (1) into (7), one finds

$$\alpha \backsim R^{1.5} \backsim d^{1.5}, \tag{8}$$

in which we considered R∝d.

3. Dependence of Correlation Strength on Richness

The richness of objects related to a loop of size R can be defined as the total number of galaxies in a volume with scale R. Therefore, from Eq.(2), one finds that the richness <N> is

$$\langle N\rangle \backsim N_g(R) \backsim R^{2.5}. \tag{9}$$

Considering Eqs.(8) and (9), we have

$$\alpha \backsim \langle N\rangle^{0.6}. \tag{10}$$

A statistical result is [5]

$$\alpha \backsim \langle N\rangle^{0.7}. \tag{11}$$

4. Flattened Tail in Correlation Function

It has been found that for r>10 Mpc the correlation function of galaxies has a flattened tail, for which the power index is about 1. Recently, CfA redshift survey showed that 1) almost all galaxies are distributed on the surfaces of bubble like structures; 2) the edges of the bubbles are remarkably sharp; 3) the bubbles have a typical diameter of about 25 Mpc. [6] If we consider that

galaxies distribute homogeneously on thin shells with radius R, the correlation function should be given by

$$\xi(r) = 1/8\pi R^2 r\mu, \tag{12}$$

where μ is the mean number density of the shells. In the case of a close-packed distribution of the shells $\mu=1/(2R)^3$, we have

$$\xi(r) = 4r^{-1}, \tag{13}$$

where we take R=12.5 Mpc and r is in Mpc. Comparing Eq.(13) with the usual result of the correlation function of galaxies $\xi(r)=20r^{-1.8}$, one finds $\xi(r)$ to have a break at $r_b=(20/4)^{1/0.8}=7.5$ Mpc. When $r>r_h$, the power index becomes approximately 1. Namely, the $r^{-1.8}$ law can be used only in the range of $r<(2\sim3)d$, the index drops to about 1 when $r>>(2\sim3)d$ the correlation function is then

$$\xi(r)\sim r^{D-3}\sim r^{-0.5}. \tag{14}$$

This result is roughly in agreement with the bubble like structure having a flattened tail in their correlation function.

5. Dependence of Correlation Length on the Size of Sample

The correlation length r_o is defined by $\xi(r_o)=1$. Therefore, from Eq.(4) we have

$$\log r_o = \log L - \frac{1}{3-D}\log\frac{6}{D} = \log 2L - 1.06. \tag{15}$$

This means that the correlation length increases with the scale of the volume of the sample considered. This result is consistent precisely with the observation, which is [7]

$$\log r_o = \log 2L - 0.96. \tag{16}$$

6. Magnitude Count of Galaxies

The number of galaxies with absolute magnitude in M→M+dM, distance in r→r+dr and direction in $d\Omega$ is given by

$$dN = \Phi(M)\frac{dN_R}{dr}\,dMdr\frac{d\Omega}{4\pi}, \tag{17}$$

where $\Phi(M)$ is the luminosity function of galaxies. Considering (visual magnitude) m=M+5 log r+25 and Eq.(2); the above formula can be rewritten as

$$dN = \Phi(m-5\log r - 25)(DB_R/4\pi)r^{D-1}drd\Omega dm. \tag{18}$$

Integrating over r from 0 to infinity and replacing r by $r=s10^{0.2m}$, leads finally to

$$\frac{dN}{dm d\Omega} = (DB_R/4\pi)\, 10^{0.2mD} \int_0^\infty s^{D-1} \Phi(-5 \log s - 25)\, ds. \qquad (18)$$

The observed value of $\frac{dN}{dm d\Omega}$ in the range of m=14 to 19 was found to be [8]

$$\frac{dN}{dm d\Omega} \sim 10^{0.44m} \qquad (19)$$

Comparing Eq.(19) with Eq.(18), we have D=2.2. The deviation of $\frac{dN}{dm d\Omega}$ from the Euclidean result $10^{0.6m}$ is due mainly to non-Euclidean geometry of the cosmological space, because most galaxies used in the count possess redshifts smaller than 0.1, in this range the effect of the curved spacetime on the count is not large. Therefore, D=2.2 is most probably from the self-affine distribution of galaxies.

References

1. L.Z. Fang: Lecture at the Yellow Mountain Summer School, 1987.
2. B.B. Mandelbrot, in Fractals in Physics, eds. L. Pietronero and E. Tosatti (North-Holland, Amsterdam, 1986) p.3.
3. L.Z. Fang, H.J. Mo and H.G. Bi, Mod. Phys. Lett. A2, 473 (1987)
4. L.Z. Fang, Mod. Phys. Lett. A1, 601 (1986)
5. N. Bahcall and W.S. Burgett, Astrophys. J. 300, L35 (1986)
6. V. De Lapparent, M.J. Geller and J.P. Huchra, Astrophys. J. 302, L1 (1986)
7. J. Einasto, A.A. Klypin and E. Saar, Month. Not. R.A.S. 219, 457 (1986)
8. P.J.E. Peebles, in Ninth Texas Symposium on Relativistic Astrophysics (Academy of Sciences, New York, 1980).

Fractal Turbulence and Origin of Large Scale Structure

Liu Yong-zhen and Deng Zu-gan

Physics Dept., Graduate School, University of Science and Technology of China, Beijing, People's Republic of China

1. Introduction

Observations show that the large scale distribution of galaxies is like a selfsimilar sponge /1/. The term "sponge" is due to most of volume observed in the Universe being devoid of galaxies /2/, there are many voids (diameter at least 60 Mpc) in the Universe /3/. The "selfsimilarity" is due to the density - radius relation /4/:

$$\text{Log } \bar{\rho}(r) \backsim 15.2 - \gamma \text{ Log } r, \qquad \gamma \backsim 1.7 - 1.8, \tag{1}$$

where $\bar{\rho}(r) = Mg\bar{n}(r)$, $\bar{n}(r)$ is the mean number density of galaxies in the aggregative region of galaxies of radius r; or the galaxy two-point correlation function /5/.

From (1), the total mean number in the region of the scale r is

$$N(r) \propto r^{3-\gamma} \equiv r^{D_f}, \qquad D_f \equiv 3 - \gamma \backsim 1.2 - 1.3. \tag{2}$$

From (2), the geometry of the mean distribution of galaxies has been regarded as a fractal with the fractal dimension D_f /6/. Here, we emphasize that this fractal should be a "selfsimilar sponge" which is called "Sierpinski sponge" in fractal geometry.

What physicl process in the early universe might produce the selfsimilar sponge structure observed today?

We have suggested a tentative scenario /1/, in which the early perturbations are assumed to be adiabatic. Principal difference between our and usual scenarios is that the perturbations which come within the horizon will undergo different hydrodynamic processes. In usual scenarios they are supposed to undergo acoustic oscillations, but in our scenario they might form the fractal turbulence.

2. Fractal Turbulence

Geometric aspects of turbulence, especially its intermittent behaviour can be described as fractals /6,7 and 8/. In this paper, if the geometry of an intermittent turbulence may be described as a fractal, we shall call it "fractal turbulence" (FT).

As is known to all, a turbulent motion may be divided into interactin motions on various length-scales, and this is usually expressed by talking of "eddies of different scales". If u_i is the characteristic velocity of an eddy of scale r_i (the velocity fluctuation on the region of size r_i), one may define a Reynolds number,

$$Re_i = u_i r_i / \nu, \tag{3}$$

to explain the properties of turbulence, where ν is the kinematic viscosity of fluid. In the turbulent cascade process, the Kolmogorov law holds over the inertial range /9/, i.e.,

$$u_i^3/r_i = u_j^3/r_j, \qquad r_o > r_i > r_j > r_d \ , \tag{4}$$

where r_o is the basic scale (or the scale of the largest eddy) of the turbulence, r_d is the dissipation scale corresponding to the $Re_d \backsim 1$. Note that in (4) one has practically assumed the turbulence to be filling space, i.e., the number $n_{j,i}$ of eddies of smaller scale r_j generated by an eddy on larger scale r_i is just equal to the ratio of their volume

$$N_{j,i} = (r_i/r_j)^3. \tag{5}$$

Thus, the Kolmogorov law can be rewritten as

$$r_i^3(u_i^3/r_i) = n_{j,i}\, r_j^3(u_j^3/r_j). \tag{6}$$

Equation (6) gives us a well-marked form that can be generalized to the case of intermittent turbulence.

Because an intermittent turbulence is not filling space, in the cascade process from r_i or r_j the number $n(j,i)$ of the "active" eddies of scale r_j which are generated form an active eddy of scale r_i is only a fraction β of the number $n_{j,i}$, i.e.,

$$n(j,i) = \beta n_{j,i} = \beta(r_i/r_j)^3 \equiv (r_i/r_j)^{D_f} \ , \quad D_f < 3. \tag{7}$$

And other $(n_{j,i} - n(j,i))$ eddies of r_j are "passive". The passive eddies do not take part in the cascade process. Using $n(j,i)$ in place of $n_{j,i}$ in (6), we then get the rate of energy transfer of the intermittent turbulence among the active eddies of size r_i and r_j.

If β is constant for any scale in the inertial range, D_f defined by (7) will be a constant. In this case, the intermittent turbulence will be a FT with the fractal dimension D_f. Thus, for a FT we have

$$u_i^3/r_i = (r_i/r_j)^{(D_f-3)} u_j^3/r_j = (r_i/r_j)^{-\gamma} u_j^3/r_j \ , \qquad \gamma = 3 - D_f \ . \tag{8}$$

As can be seen from (7) that if an intermittent turbulence is a FT (D_f = constant), its structure is like a selfsimilar sponge.

The D_f of a FT depends on the practical conditions that produce it. If the Reynolds number of the basic scale r_o, $Re_o = u_o r_o/\nu$, is very large, the turbulence will be of filling space, i.e., $n(j,i) = n_{j,i}$; in that case the fractal structure will not be able to occur. The fractal structure occurs, in general, at the initial stages of the transition to turbulence. Thus when the fractal structure occurs, Re_o should be large enough, but does not need to be very large.

As an example, we consider the wake of a moving object in a fluid. For simplicity, assume the object be a cylinder (or a sphere) of diameter d with velocity u. Then the Reynolds number is $Re = ud/\nu$. Experiments show /10/ that

the behavior of the wake arising from the drag depends on Re. When Re is greater than about 100, the attached eddies are periodically shed from the cylinder to form the Karman vortex street. Concentrated regions of rapidly rotating fluid form two rows on either side of the wake. An important transition occurs at a Reynolds number a little below 200. Above this, the vortex street breaks down and produces ultimately a turbulent wake. At Re >> 10^2, the wake will become fully turbulent. As has been pointed out before, the fractal structure occurs, in general, at the initial stages of transition to the turbulence. So, when the basic Reynolds number Re is the order of 10^2, the FT might occur in the vorices of the wake of a moving object. If Re >> 10^2, the vortices are not fractal, but fully turbulent. This example is very important for us because it gives us a very simple mechanism in which the vortices of a rapidly rotating fluid are generated by the drag, and tells us a condition Re $\backsim$ 10^2 on which the vortices might be fractal turbulent.

Similarity of the geometry between the FT and the galaxy distribution prompts us to consider if the origin of the large scale structure had been started in a process of the FT in the early universe. It is this clue that led us to suggest the following scenario which includes a process of the FT/1/.

3. Origin of Large Scale Structure

This scheme involves the following basic assumptions and processes:

(a) We assume that the density perturbations are adiabatic in the early universe, and the spectrum of perturbations has the form

$$|\delta_k|^2 \propto k^n, \qquad k = 2\pi/\lambda. \tag{9}$$

When the index n = 1, (9) gives the scale-free Zeldovich spectrum. Since the amplitude of the early perturbations should be very small, the cosmic fluid can be regarded as being incompressible. The spectrum of velocity perturbatins for the wavelength λ has the form

$$(u_\lambda/u_s) \backsim (\delta\rho/\rho)_\lambda^{1/2} \propto \lambda^{-n/4}\ , \tag{10}$$

where u_s is the sound speed in the cosmic fluid.

(b) At time t before recombination, the perturbations on the scales $\lambda < \lambda_h = ct$ (the horizon) will undergo dissipative process. Therefore, for the adiabatic perturbations the cosmic fluid can be regarded as a viscous fluid with the kinematic viscosity

$$\nu = (3/5)(u_s/c)^2(cm_p/\sigma_T\rho_m), \qquad \rho_m/m_p \backsim n_p \backsim n_e\ . \tag{11}$$

where σ_T is the Thomson cross-section, n_e is the free electron density. Thus, each perturbation with the characteristic length $\lambda < \lambda_h$ and velocity u_λ can be associated with a Reynolds number

$$Re_\lambda = u_\lambda\lambda/\nu\ . \tag{12}$$

Only when $Re_\lambda \backsim 1$, the viscous dissipation is important. Since the time-scale of energy dissipation is about $t_\lambda \backsim \lambda^2/\nu \backsim \nu/u_\lambda^2$, all perturbations on the scales λ that satisfy $t_\lambda \backsim t$ do not appear in the cosmic fluid at time t, which are dissipated by the viscosity. Then, from $t_\lambda \backsim t$ we obtain a smallest scale λ_d (i.e., the dissipation scale) at time t and a corresponding velocity u_d:

$$\lambda_d \backsim (\nu t)^{1/2} \quad \text{and} \quad u_d \backsim (\nu/t)^{1/2} . \tag{13}$$

On the dissipation scale the Reynolds number $Re_d = u_d\lambda_d/\nu = 1$.

Assume that the velocity spectrum of the perturbations within the horizon still retains the form of (10), we then can get the amplitude of the velocity spectrum from λ_d and u_d at that time. That is,

$$u_\lambda \backsim u_d(\lambda_d/\lambda)^{n/4} , \qquad \lambda_d < \lambda < \lambda_h . \tag{14}$$

The characteristic velocity u_h corresponding to the largest scale λ_h within the horizon is

$$u_h \backsim u_d(\lambda_d/\lambda_h)^{n/4} . \tag{15}$$

Thus the Reynolds number on the scale λ_h can be worked out by

$$Re_h = u_h\lambda_h/\nu \backsim (u_d\lambda_d/\nu)(\lambda_h/\lambda_d)^{(1-n/4)} , \quad \text{from (15)}, \tag{16}$$

$$\backsim (\lambda_h/\lambda_d)^{(1-n/4)} \backsim (c/u_d)^{(1-n/4)} , \quad \text{from (13)}.$$

One of the advantages of using the Reynolds number is that we can get some valuable information concerning the evolution of perturbations in the early universe by means of dynamical similarity. In the dynamical similarity the Re_h which corresponds to the largest scale of the onset of fluid dynamical behavior at time t is the most important, because it determines the basic behavior of the perturbations in the cosmic fluid. Clearly, the Re_h depends on the cosmological model and the spectrum index n of adiabatic perturbations.

(c) A perturbation region of scale λ may be approximately treated as a "sphere" (or a "cloud") with the characteristic size λ und u_λ. Once this sphere comes within the horizon the interaction between it and the surroundig background universe will start the onset. In this sense, an adiabatic perturbation of scale λ can be described as a moving sphere with radius λ and the velocity u_λ in a viscous fluid.

According to dynamical similarity, if the Reynolds numbers are the same for two geometrically similar systems, the same flow behavior occurs. Thus, from the value of the R_h we can judge if the perturbation sphere described here might produce the Karman vortices of the wake as mentioned in Section 2, these vortices arising from the drag are fractal of fully turbulent. The condition for which the vortices might be fractal is that the Re_h should be of the order of 10^2.

If the above perturbation sphere might produce the vortices of FT, then for a vortex the rate of energy transfer in the cascade process from larger scale λ_i to smaller scale λ_j should satisfy (8), i.e.,

$$u_i^3/\lambda_i = (\lambda_i/\lambda_j)^{(D_f-3)}\ u_j^3/\lambda_j, \qquad \lambda_h > \lambda_i > \lambda_j > \lambda_d,$$

$$= (\lambda_i/\lambda_j)^{-\gamma}\ u_j^3/\lambda_j, \qquad \gamma = 3 - D_f\ . \tag{17}$$

On the other hand, assuming that the velocity spectrum in the cascade process still retains the form of (14), we can rewrite it as the form

$$u_i^3/\lambda_i = (\lambda_i/\lambda_j)^{-(3n/4+1)}\ u_j^3/\lambda_j. \tag{18}$$

If we compare (17) with (18), we then get

$$\gamma = 3 - D_f = (3n/4 + 1). \tag{19}$$

Within a fractal vortex the cascade process occurs over a range of scales from $\lambda \backsim \lambda_h$ up to $\lambda \backsim \lambda_d$. The eddies on scales around about λ_d have no longer an interior structure, for them the energy extracted from larger scale eddies is directly dissipated in the background fluid, owing to the viscosity. We call these smallest eddies "elementary eddies". Therefore, a vortex of FT can be regarded as a fractal aggregation of elementary eddies, which is like a selfsimilar sponge. Clearly, it is the elementary eddies that should be expected to form a typical kind of astronomical objects. It is natural to identify the elementary eddies with the protogalaxies in this scheme. We have taken the dissipation scale λ_d corresponding to the Reynolds number $Re_d = 1$. If we take the scale of elementary eddies λ_{pg} over a narrow range near λ_d (say, corresponding to the Reynolds number $Re_{pg} \backsim 1\text{-}5$), we can get the associated velocity range u_{pg}. And then, we can obtain the ranges for the characteristic mass and angular momentum of the protogalaxies by

$$M_{pg} = (4\pi/3)\lambda_{pg}^3\ m \quad \text{and} \quad J_{pg} = (2/5)(Mu\lambda)_{pg}\ . \tag{20}$$

Within a fractal vortex, one active eddy of scale λ_i only includes $n(j,i)$ active eddies of scale λ_j, and the $(n_{j,i} - n(j,i))$ other eddies are passive (i.e., they do not take part in the energy cascade). The elementary eddies (the smallest scale active eddies whose scales are about λ_d) only occur in larger scale active eddies. In other words, the protogalaxies only may be formed and concentrate in the regions of larger scale active eddies. Therefore, it is reasonable to regard the passive eddies within each vortex of FT as the protovoids. Thus, in a region of galaxies of scale λ_i the probability of finding voids of scale λ_j is about

$$P(j,i) = (n_{j,i} - n(j,i))/n_{j,i} = (1 - (\lambda_i/\lambda_j)^{-\gamma}). \tag{21}$$

(d) For a given cosmological model, we can work out values of the basic quantities in this scenario. Using the present values of the Hubble parameter,

$$H_o = 100(h)\ \text{km s}^{-1}\ \text{Mpc}^{-1}, \tag{22}$$

and the density of matter,

$$\rho_o = \Omega\rho_c = 1.88\text{x}10^{-29}(\Omega h^2)\ \text{g cm}^{-3}\ , \tag{23}$$

we obtain the following results just before recombination (taking the redshift z_r = 1480):

$$\lambda_{h,r} = ct_r = 1.62\times10^{23}\ (\Omega h^2)^{-1/2}\ \text{cm}, \tag{24}$$

$$\lambda_{d,r} = (\nu t)_r^{1/2} = 2.50\times10^{20}\ (\Omega h^2)^{-5/4}\ \text{cm}, \tag{25}$$

$$u_{d,r} = (\nu/t)_r^{1/2} = 4.62\times10^{7}\ (\Omega h^2)^{-3/4}\ \text{cm s}^{-1}\ . \tag{26}$$

From (20) and taking $Re_{pg,r}$ = 1 - 5, we get

$$M_{pg,r} = 2.01\times10^{9} - 1.26\times10^{12}\ (\Omega h^2)^{-11/4}\ M_\odot\ . \tag{27}$$

$$J_{pg,r} = 1.85\times10^{70} - 5.77\times10^{73}\ (\Omega h^2)^{-19/4}\ \text{g cm}^2\ \text{cm}^{-1} \tag{28}$$

And then, from (16) and (26) we can get the basic Reynolds number $Re_{h,r}$ and the fractal dimension D_f (from (19)) for different values of the index n in the spectrum (9). Several typical results are

$$Re_{h,r} = 25(\Omega h^2)^{3/8},\quad D_f = 0.5,\quad \gamma = 2.5,\quad \text{for } n = 2, \tag{29}$$

$$Re_{h,r} = 129(\Omega h^2)^{9/16},\quad D_f = 1.25,\quad \gamma = 1.75,\quad \text{for } n = 1, \tag{30}$$

$$Re_{h,r} = 3\times10^{3}(\Omega h^2)^{15/16},\quad D_f = 2.75,\quad \gamma = 0.25,\quad \text{for } n = -1. \tag{31}$$

As seen from the results above, for n = 1 the $Re_{h,r}$ is just the order of 10^2. This result is highly important for our picture, because it means that the adiabatic perturbations whose λ_r are somewhat below $\lambda_{h,r}$ might produce the Karman vortices before recombination, and the vortices might be fractal turbulent, owing to the drag. If the n is not 1, the conclusion will be very different. For n = 2, $Re_{h,r}$ = 25, it means that the Karman vortices cannot occur yet; if n = -1, $Re_{h,r} = 3\times10^3$, it means that the vortices are not fractal, but fully turbulent. The above results are sensitive to the index n in (9).

It can be seen from (27) and (28) that the characteristic mass and angular momentum are comparable with the estimated values of the observed galaxies. So, it is reasonable to assume the elementary eddies to be the protogalaxies.

Thus, in this scenario we can draw the conclusion that if the spectrum of the early perturbations is the scale-free Zeldovich spectrum, they might generate the vortices of the FT, the structure of each vortex is like a selfsimilar sponge which is composed of the protogalaxies (the elementary eddies). And the values of $M_{pg,r}$, $J_{pg,r}$ and D_f have been comparable with the present observations. The basic features of the galaxies and the large structure have been formed before recombination.

(e) With the onset of recombination the free-electron density n_e decreases rapidly, and then the viscosity ν increases and the Reynolds number Re_h decreases rapidly to $Re_h << 10^2$. That leads the process which might generate the vortices of FT to stop immediately. That is, new fractal vortices cannot occur again after recombination.

Owing to the same condition $Re << 10^2$, the cascade process within each vortex of the FT which have been formed before recombination is turned off. Therefore, after recombination these vortices are no longer turbulent, but their fractal structure is still left. That is to say, the fractal structure of each vortex is frozen in the cosmic background, which will expand with the Universe up to the present. Each fractal structure is like an expanding selfsimilar sponge which is composed of the protogalaxies. In this scenario, the upper-limit scale of the inhomogeneity of the galaxy distribution observed today is expected to be about

$$2\lambda_{h,r}(1 + z_r) \backsim 4.7\times10^{26}(\Omega h^2)^{-1/2} \text{ cm} \backsim 150(\Omega h^2)^{-1/2} \text{ Mpc}. \qquad (32)$$

This means that the distribution of the observed galaxies should be homogeneous on the scale $> 150(\Omega h^2)^{-1/2}$ Mpc; but on the scale $< 150(\Omega h^2)^{-1/2}$ Mpc, it is inhomogeneous and fractal with the fractal dimension $D_f = 1.25$, and the index γ in the two-point correlation function and in the distribution of voids expected from (21) should be 1.75.

4. Discussion

The principal differences between our scenario and the existing scenario lies in that we assume that the adiabatic perturbations which come within the horizon might produce the Karman vortices, and the vortices might be fractal turbulent. It should be recognized that a number of problems in theory remain unsolved, especially, the description of the early perturbations appears to be simplistic. But, disregarding there open questions, for the moment, we have found that for the scale-free Zeldovich spectrum the theoretical results coincide with the observations. That is, the typical mass and angular momentum of galaxies, the geometry of the galaxy distribution (the selfsimilar sponge structure with the fractal dimension $D_f = 1.2 - 1.3$) can be explained in a summary fashion. It implies that this picture might include some factors which are responsible for the origin of galaxies and large scale structure.

In this picture the voids are corresponding to the regions of the "passive eddies" in which the protogalaxies could not occur. Therefore it is explained that there are no observable galaxies in the voids. But then the voids still contain much matter, in which the upperlimit of the mean matter density is about $\rho_o = \Omega\rho_c \backsim 10^{-29}(\Omega h^2)$ g cm^{-3} at present. This is an open question, because one does not know what it is (black holes, hydrogen clouds, or other things?). To answer this question, more observations and detailed studies in theory are needed.

References

1. Yong-zhen Liu, Zu-gan Deng: Ap. S. S. (1987) in press
2. M.S. Longair: In Observational Cosmology, ed. by A. Hewitt, G. Burbidge, L-z Fang (D. Reidel Pub. Co., Holland 1987)
3. V. de Lapparent, M.J. Geller, J.P. Huchar: Ap.J. 302, L1 (1986)
4. G. de Vaucouleurs: Science 167, 1203 (1970)
5. P.J.E. Peebles: The Large Scale Structure of the Universe (Princeton Univ. Press 1980)
6. B. Mandelbrot: The Fractal Geometry of Nature (Freeman, S. Francisco 1982)
7. U. Frisch, P. Sulem, M. Nelkin: J. Fluid. Mech. 87, 719 (1978)
8. G. Paladin, A. Vulpiani: In Fractals in Physics (L. Pietronero, E. Tosatti 1986) p.313
9. L.D. Landau, E.M. Lifshitz: Fluid Mechanics (Pergammon, New York 1959)
10. D.J. Tritton: Physical Fluid Dynamics (Van Nostrand Reinhold Company 1977)

The Two-Point Correlation Function of Groups in CFA Survey

Zou Zhen-long and Liu Da-qing

Academia Sinica, Beijing Astronomical Observatory,
Beijing, People's Republic of China

The power law form of the two-point spatial correlation function of groups of galaxies is observed among a statistically homogeneous group sample based on the CfA redshift survey. The correlation range is from 5 to 40 Mpc (H_0=100 km/sec/Mpc). The power index is 1.8, and the correlation strength is 45 which is consistent with its dependence on the mean separation and richness of galaxy systems. This correlation of groups suggests the existence of structure in the intermediate scale between galaxies and clusters of galaxies.

1. Introduction

Calculation of the correlation functions of galaxies and galaxy systems is an important method to study the large scale structure of the universe. Several papers cover this topic and arrive at the same power law result. The two-point correlation functions of galaxies, e.g., GROTH and PEEBLES [1], SHANKS et. al. [2], DAVIS and PEEBLES [3], clusters of galaxies, e.g., BAHCALL and SONEIRA [4], and superclusters, e.g., BAHCALL and BURGETT [5], all have the power law form $\xi(r)=Ar^{-1.8}$ but different correlation strength for different systems, and the correlation scale increases with the richness of galaxy system. Bahcall and co-workers discovered that there exists power law dependence of the correlation strength on the mean richness and the mean separation of objects in the system.

Considering groups of galaxies as a stage in the clustering of galaxies it is interesting to calculate the group correlation function, and check whether the correlation strength of groups is consistent with the dependence on the richness and the mean separation of groups.

In a recent paper, ZOU and LIU [6], we calculate the spatial two-point correlation function of groups using the Geller-Huchra catalog (hereafter GH catalog), GELLER and HUCHRA [7], based on the CfA redshift survey, HUCHRA et. al. [8], and found a marginal power law form of the correlation function in the range less than 40 Mpc (H_0=100 km/sec/Mpc). Moreover, the correlation strength is consistent with the dependence on the richness and the mean separation of groups.

2. Groups of Galaxies in the CfA Redshift Survey

The GH catalog is a homogeneous sample based on the CfA redshift survey with the magnitude limit B_{lim}=14.5 mag. Groups in it are selected according to the galaxy density enhancements of a factor larger than 20 in redshift space. There are 176 groups in the GH catalog, 123 of them are in the angle area b>40° and δ>0° (NGP area), and others are in the SGP area. We only consider groups in the NGP area. The radial velocities of all groups are less than 8000 km/sec in the GH catalog.

The number of galaxies in each group varies very greatly, from 3 to 248. The group No. 94 contains 170 galaxies and the group No. 106 contains 248 galaxies in the GH catalog, other groups contain the number of galaxies limited in 30. The large number of galaxies in group No. 94 and No. 106 may be due to the high galaxy density in the Virgo cluster and may have some influence on the statistics, especially on the determination of the mean richness of the groups. In order to avoid this influence, we cut out these two groups in our work.

Figure 1 shows the number-radial distance distribution for 121 groups in the NGP area. In order to determine the group correlation function, the expected group distribution in the same range with that of the real sample is necessary. But we can see in Fig. 1 that the number of groups in the bin from 10 to 20 Mpc (Group No. 94 and No. 106 are also in the range of this bin) is much larger than that in other bins, this is due to the Virgo cluster and will have some influence on the determination of the expected distribution of groups in the radial direction. In the range larger than 20 Mpc, the fluctuation of the number of groups in each bin around the dashed line in Fig. 1 is nearly less than the square root of the number of groups in that bin, then we suspect this dashed line is the expected distribution of 88 random groups in the range from 20 to 80 Mpc. Thus in our determination of the correlation function the dashed line distribution is used to generate the random groups.

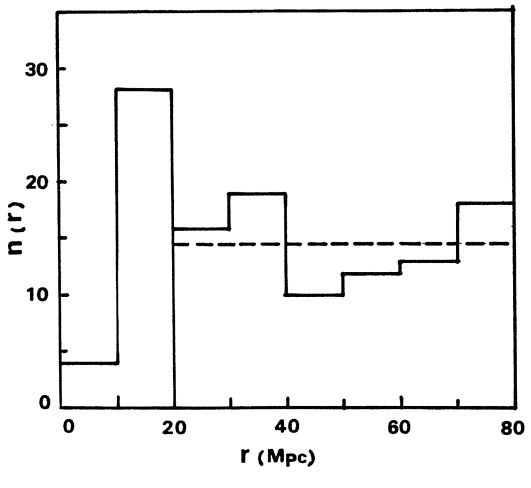


Fig. 1. The distribution of group distances. The solid line is the number-radial distance relation for groups in the range from 20 to 80 Mpc, and the dashed line was used to generate random groups. Group No.94 and 106 are in the bin from 10 to 20 Mpc and not in the range we considered.

The mean separation of groups in the range we consider (from 20 to 80 Mpc in the radial direction, totally 88 groups, and named as the sub-GH catalog hereafter) is about 15 Mpc, and the mean richness of groups is about 6, which is less than that of poor clusters.

3. Estimation of the Correlation Function of Groups

The spatial two-point correlation of groups $\xi_{gr}(r)$ was estimated from the real and the random frequency distribution ($F(r)$ and $F^r(r)$), their relative difference gives $\xi_{gr}(r) = F(r)/F^r(r) - 1$. The real frequency distribution $F(r)$ is obtained by calculating the spatial separation of each group pair in the sub-GH catalog and counting the number of pairs in each r - r+dr interval (dr is the bin size). In order to calculate $F^r(r)$ in the same way (this can have the boundary and other effects in the calculation of $\xi_{gr}(r)$ removed), we generate 1000 random catalogs, in which the random groups distribute in the angle area $b>40°$ and $\delta>0°$ uniformly and in the radial range from 20 to 80 Mpc under the dashed line distribution shown in Fig. 1.

The group selection procedure in the GH catalog paper shows that no meaningful correlation should be expected within about 5 Mpc. So the calculation of $F(r)$ and $F^r(r)$ begins from r (group pair separation) = 5 Mpc.

The peculiar velocity of groups should exist, but their values can not be determined. This makes the uncertainties of the center position of groups and has some influence on the calculation of $F(r)$ and $F^r(r)$. Using the velocity dispersion of each group listed in the sub-GH catalog, we get that the mean value of the velocity dispersions of 88 groups is 370-380 km/sec (or 3.7-3.8 Mpc), so the suitable bin size for the calculation of the frequency distribution should be about 7.5 Mpc. This bin size can partly decrease the uncertainties of $F(r)$ and $F^r(r)$, and the bin size smaller or larger than 7.5 Mpc is also available. But too small a bin size may increase the uncertainties of $\xi_{gr}(r)$, and too large a bin size may smooth out the useful correlation information.

Changing the bin size from 5.0 to 15.0 Mpc (according to 5.0 + 0.5 i, i = 1,2,...,20) to calculate $\xi_{gr}(r)$, we find that there always exists the power law form of correlation (within S.D. of ξ_{gr} = 1 σ) on the scale less than 40 Mpc. From least square fitting, we find that the power index γ varies mainly between 1.6 and 2.0 for different bin size, and the mean value is about 1.8. Fixing γ=1.8 to determine the correlation strength A (in order to check whether A is consistent with the dependence on the mean separation and richness), we get that A varies mainly between 35 and 55 (unit of r in Mpc) for different bin size, and the mean value is about 45 (when using the least square fitting to determine and A, if γ is about 1.8 for some bin size, the corresponding A is also about 45). Thus the power law form of the correlation function of groups is stable.

4. Discussion

We get the correlation strength about 45, which is the value of $\xi_{gr}(r)$ at r = 1 Mpc. In the sub-GH catalog, the mean spatial separation of groups is about 15 Mpc and the mean richness of them is about 6 ± 4. These values are consistent with the dependence of the correlation strength on the mean separation and the richness of the galaxy systems discovered by Bahcall and co-workers. This result illustrates that groups of galaxies may be a clustering stage and there may exist structure in the intermediate scale between galaxies and clusters.

We list some two-point correlation parameters of galaxy systems in Table 1. It is remarkable that all power indices γ are about 1.8. It is very clear that the correlation strength A (or r_0) and the correlation range increase with the richness of galaxy system (Col. 1 in Table 1). It is interesting to note that our group correlation parameters are just between those of galaxies and clusters of galaxies (Col 3 or 4, and Col 5 in Table 1). This also shows that the group of galaxies may be the intermediate clustering stage between galaxies and clusters of galaxies. The same power law form of the two-point correlation functions of galaxy systems (including group) gives a new evidence that there exists self-similar structure in the large scale distribution of the universe. The power law form of group correlation function requests that any theory (trying to explain the correlation phenomena of galaxy systems) should permit the existence of the group clustering stage and also explain why the group clustering function has the same power law form as other galaxy systems.

Table 1. Parameters of Correlation Functions of Galaxy Systems

galaxy system	γ	A	r_0	effective range
Galaxies	~ 1.8	~ 20	~ 5.3 Mpc	< 10 Mpc
Groups	~ 1.8	~ 45	~ 8.3 Mpc	< 40 Mpc
Clusters	~ 1.8	~ 360	~ 26 Mpc	< 100 Mpc
Super-clusters	~ 1.8	~1500	~ 60 Mpc	< 150 Mpc

Notes: values of parameters are cited from the corresponding paper (about the correlation function of galaxy system) mentioned in § 1 except those of groups.

We are indebted to Professors Z. G. Deng, J. S. Chen, and L. Z. Fang for their helpful discussion.

References

1. E. J. Groth, P. J. E. Peebles: Ap. J., 217, 385 (1977)
2. T. Shanks, A. J. Bean, G. Efstathiou, R. Fong, B. A. Peterson: Ap. J., 274, 529 (1983)
3. M.Davis, P. J. E. Peebles: Ap. J., 267, 465 (1983)
4. N. A. Bahcall, R. M. Soneira: Ap. J., 270, 20 (1983)
5. N. A. Bahcall, W. S. Burgett: Ap. J. Lett., 300, L35 (1986)
6. Z.-L. Zou, D.-Q. Liu: submitted to Ap. J. (1987)
7. M. J. Geller, J. P. Huchra: Ap. J. Suppl., 52, 61 (1983)
8. J. P. Huchra, M. Davis, D. Latham, J. Tonry: Ap. J. Suppl., 52, 89 (1983)

Large Scale Distribution of Galaxies with Different Luminosity

G. Börner[1], *Z.-G. Deng*[1,2], *and X.-Y. Xia*[3]

[1]Max-Planck-Institut für Physik und Astrophysik, Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-8046 Garching, Fed. Rep. of Germany
[2]Department of Physics, Graduate School, Chinese Academy of Sciences, P.O. Box 3908, Beijing, People's Republic of China
[3]Department of Physics, Tianjin Normal University, Tianjin, People's Republic of China

1. Introduction

The approximately uniform distribution of matter in the early universe and the high clustering of luminous galaxies have both been confirmed by observations. A theory about the formation and evolution of galaxies has to reconcile these two aspects. In addition, the estimated mean mass density from derived masses of clusters of galaxies is lower than the critical value by a factor 3∽5, as pointed out by Davis and Peebles /1,2/, but the inferred mean mass density from the inflation model /3/ would be equal to the critical value. Because the inflation model is able to solve many cosmological problems, many theorists are willing to accept this idea.

To coordinate these different aspects, astrophysists are invoking the non-baryonic matter predicted by some particle theories. If dark matter follows the same clustering distribution as luminous glaxies, the dynamical methods should include its effects in the inferred masses. The only way to avoid the discrepancy between the theoretically inferred value and the observationally estimated value of the mean mass density is to suppose that the luminous galaxies do not trace the distribution of matter. That is, the overall distribution of matter is approximately uniform in the universe, but the galaxies are formed only in some regions which have certain preferable environment. The dynamical method cannot find out any uniformly distributed matter.

Various mechanisms of biased galaxy formation have been suggested /4/. All of them suggested that the formation of galaxies is significantly affected by one or more environmental conditions which segregate galaxies from the underlying mass. DEKEL and SILK /5/ have argued that in a "bottom-up" model the normal bright galaxies are likely to originate from high-density peaks (2σ-3σ) in the initial fluctuation field; typical ($\backsim 1\sigma$) peaks cannot make a luminous galaxy at all, because the gas is too hot and too dilute to cool in time. This would lead the bright galaxies to be biased towards the clusters and the dwarf galaxies may trace the mass and provide an observational clue to the distribution of dark matter. But, the analyses for the segregation between the high and low luminosity galaxies have not got a decisive conclusion yet. EDER et. al. /6/ have studied the distribution of dwarfs around a nearby void. They found that neither normal bright galaxies, nor low surface brightness galaxies nor dwarfs inhabit the void. DAVIS and DJORGOVSKI

/7/ try to investigate the difference between the distribution of high and low surface brightness galaxies. But, their analysis has several problems which have been pointed out by BOTHUN et. al. /8/.

But, there is another way which can also provide information about the reality of biased galaxy formation. In this work, we present the results of two-point correlation analysis for galaxies with different luminosity. The samples are sorted out from data provided by the Cfa /9/ redshift survey and should be statistically homogeneous and complete. If this kind of selective bias does exist, we would be able to obtain a systematic variation of the correlation functions along with the luminosity of galaxies.

2. Samples

To analyse the distributions of galaxies with different luminosity the first important thing is sorting out statistically homogeneous and complete subsamples in given luminosity intervals. Up to now, the available data supplied by the largest field redshift survey is the CfA catalog /9/. This sample contains 2401 galaxies which satisfy $m_{pg} \leq 14.5$ and $b_{II} \geq 40^{\circ}$, $\delta \geq 0^{\circ}$, or $b_{II} \leq -30^{\circ}$, $\delta \geq -2^{\circ}.5$. Every galaxy in this catalog has a measured redshift. We use the galaxies in the northern field, i.e. with $b_{II} \geq 40^{\circ}$ and $\delta \geq 0^{\circ}$, as the parent sample of our analysis.

The local motion of the earth will distort the Hubble flow of the observed galaxies. To correct this effect we transform the coordinates and redshift of every galaxy into galactic center coordinates, then use the formula given by GELLER and HUCHRA /10/,

$$V_{LSC} = V_{in}(\sin \delta_i \sin \delta_v + \cos \delta_i \cos \delta_v \cos(\alpha_i - \alpha_v)), \qquad (1)$$

to correct the effect of the dipole Virgocentric flow. V_{LSC} is the correction of the galaxy redshift for a galaxy with right ascension α_i and declination δ_i, α_v and δ_v are the corresponding coordinates of the Virgo center. V_{in} is the infall velocity to the Virgo center which has been assumed to be 300 km/s.

The peculiar motion of galaxies will also distort the redshift-distance relation. The value of the peculiar motion is difficult to measure for every galaxy, but the generally accepted value is a few hundred kilometers per second. To reduce the relative importance of the peculiar motion we discard the galaxies with corrected redshift smaller than 1000 km/s from our sample. For galaxies with corrected redshift $\geq$1000 km/s, we simply assume the corrected redshift as the indicator of their distances. The possible effect of this approximation will be discussed in the last section.

According to Hubble's law

$$M = m - 25 - 5 \lg(v/H_0), \qquad (2)$$

we can find out the absolute magnitude of each galaxy in the CfA catalog. In (2), H_0 is the Hubble constant and has been taken to be 50 km/s Mpc throughout this work. We use the following procedure to sort out the subsamples which are complete and statistically homogeneous in a given interval of absolute magnitude and space region.

If we want to sort out a complete and statistically homogeneous subsample which contains galaxies with absolute magnitude between M_1 and M_2, and suppose that M_1 is the absolute magnitude of a galaxy corresponding to the lower limit luminosity of the absolute magnitude range, we can imagine that this galaxy should be observed as a galaxy with apparent magnitude $14^m.5$ if its corrected redshift is

$$v_{max} = 10^{\frac{2.01 + M_1}{5}} \qquad (3)$$

Because the CfA survey is considered to be complete to $14^m.5$ in the given field, all galaxies with higher luminosity should be brighter than $14^m.5$ if their redshifts are less than v_{max} and will be discovered by the CfA survey. So, if we restrict the analysis to galaxies with solute magnitude in the given range and at the same time to galaxies with corrected redshifts less than v_{max}, we could obtain a complete and statistically homogeneous subsample in this given luminosity interval.

3. Two-point correlation analysis of galaxies with different luminosity

Two-point correlation analysis is a powerful tool for investigating the clustering properties in the distribution of given objects. The two-point spatial autocorrelation function is defined by /11/

$$\xi(r) = \frac{n_s(r)}{n_r(r)} - 1. \qquad (4)$$

Here, $n_s(r)$ is obtained from the analysed sample. It is the number of galaxy pairs in which the distances between the two galaxies of each pair are in the interval $r - r+\Delta r$. $n_r(r)$ is the corresponding mean number of galaxies given by many times Monte Carlo sampling which are produced in the same region and with the same total number of galaxies as that of the sample. Because the sample is complete and statistically homogeneous in the region, the Monte Carlo sampling should produce uniformly distributed random points. Because the region in which the analysis has been performed is not very large, the spatial geometry can be approximated by a Euclidean geometry.

From the CfA redshift survey catalog we sort out four subsamples with the method mentioned in the last section. These four subsamples are in the absolute magnitude intervals $-18.5 - -19.5$, $-19.5 - -20.5$, $-20.5 - -21.5$ and $-21.5 - -22.5$, and are called subsamples, A, B, C, and D, respectively. All these subsamples are complete and statistically homogeneous in the given absolute magnitude intervals, the field of the CfA survey, and with redshift less than the maximum redshifts determined by (3). In table 1, we give the maximum redshifts v_{max} (and the corresponding maximum distance D_{max}), and the numbers of galaxies for each subsamples.

The number of Monte Carlo samplings is increased until it gives a statistically stable result for the given subsample. It varies from 25 for the faintest subsample to 500 for the brightest subsample.

Table 1. The values of maximum redshifts v_{max}, the corresponding maximum distances D_{max} and the galaxy numbers with redshifts less than v_{max} but larger than 1000 km/s for each subsample

sample	v_{max} (km/s)	D_{max} (Mpc)	range of M	number of galaxies
A	1986	39.72	-18.5 - -19.5	143
B	3148	62.96	-19.5 - -20.5	290
C	5012	100.24	-20.5 - -21.5	273
D	7907	158.14	-21.5 - -22.5	198

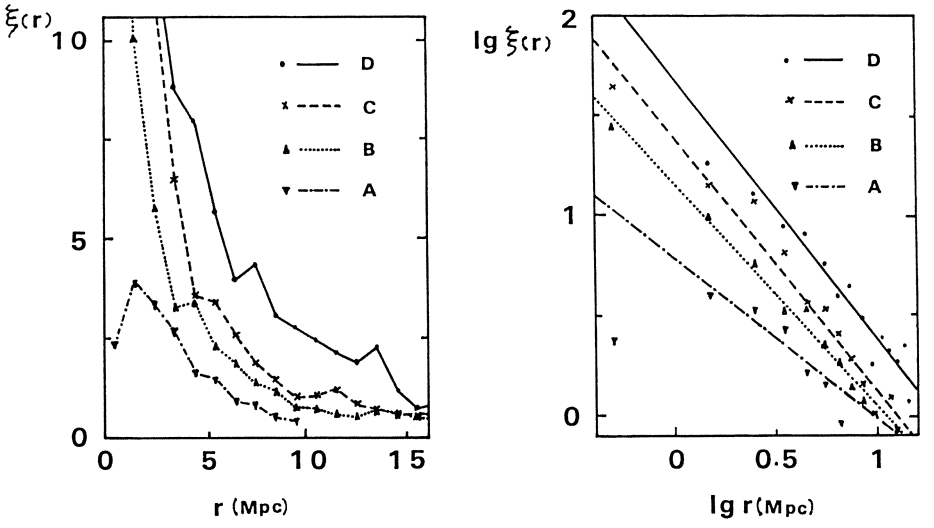


Fig. 1a) The spatial two-point autocorrelation function for subsamples A, B, C, and D; b) The two-point autocorrelation functions of subsamples A, B, C, and D, on the lg $\xi(r)$ - lg r diagram. The straight lines are obtained from linear regression.

Figure 1a gives the resulting two-point autocorrelation functions for these four subsamples. The corresponding lg$\xi(r)$ - lg r diagrams are presented in fig. 1b.

Figure 1 shows that all the two-point autocorrelation functions of these subsamples can be approximated by a power law, i.e.

$$\xi(r) = A\, r^{-\alpha} = (r/r_0)^{-\alpha}. \qquad (5)$$

The correlation strength A is the interpolated value of the autocorrelation function at r = 1 Mpc, and the correlation scale r_0 indicates the value of r for which the autocorrelation function is equal to 1. Among A, r_0 and the power index there is a relation

$$A = r_o^{\alpha} . \qquad (6)$$

The values of A, r_o and α can be found by linear regression. We do the regression analysis in the range from the minimum r to the r at which the correlation functions have the least values larger than 1 for subsamples B, C, and D. Since the first point of subsample A deviates from the straight line determined by the other points, we start the regression from the second minimum distance r for sample A. The results obtained from linear regression are presented in table 2, in which we have also given the correlation coefficients γ of the regression analysis to show how good the correlation function can be approximated by a power function.

Table 2. The power indices α, correlation strengths A and correlation scales r_o of the four subsamples obtained from regression analyses. The correlation coefficient γ of every linear regression is also presented.

sample	α	A	r_o(Mpc)	γ
A	0.810	6.145	9.44	0.952
B	1.116	14.642	11.07	0.993
C	1.268	24.717	12.54	0.982
D	1.304	48.507	19.62	0.983

We can see from table 2 that all correlation functions of these four subsamples can be approximated by power laws with high accuracy. But, the most striking fact revealed from this analysis is that the character of clustering for galaxies with different luminosity is obviously different and shows systematic changes. All the power indices of these four autocorrelation functions are much less than the value 1.8, which is generally accepted. Moreover, the power index α of the correlation function is increasing as the luminosity of the subsample is increasing, i.e. the autocorrelation function for the more luminous galaxies is steeper. The values of A and r_o for the more luminous galaxies are also larger than those of the fainter galaxies. The more luminous galaxies cluster stronger and with larger clustering scale. In fig.2, we give the plots of A, r_o and α as functions of absolute magnitude from table 2.

To completely understand the distribution of galaxies with different luminosity we have also investigated the cross-correlation among galaxies with different luminosity. The cross-correlation function between samples 1 and 2 in the same region is defined by

$$\xi_c(r) = \frac{N_{12}^s(r)}{N_{12}^r(r)} - 1. \qquad (7)$$

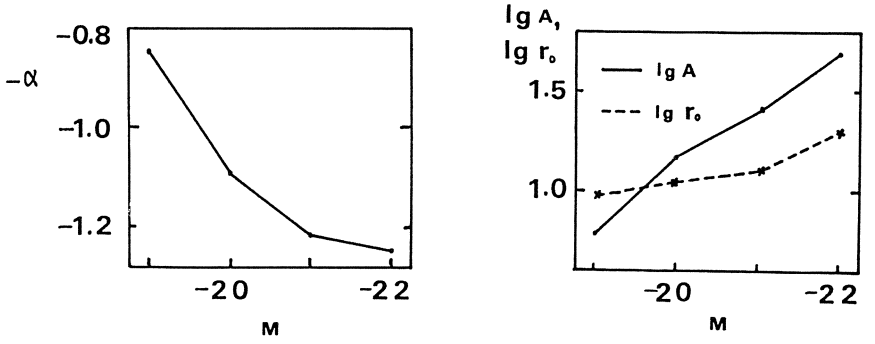


Fig. 2a) Diagram for -α vs. M; b) The diagrams for lgA and lgr_0 vs. M.

N^s_{12} is the number of galaxy pairs in which the distances between two galaxies, one from each sample, are in the range of r - r+Δr. N^r_{12} is the mean number obtained from many times Monte Carlo sampling.

To do the cross-correlation analysis the two samples have to be both complete and statistically homogeneous in the same region. To this end, we would restrict the v_{max} or D_{max} for both samples with the value given by the fainter one of these samples. To assure that there are enough numbers of galaxies in each cross-correlation analysis we use an alternative division of the luminosity intervals. We restrict the analsis to galaxies in the absolute magnitude range -19 - -22, and divide it into three luminosity intervals each spanning one absolute magnitude. So, we obtain three subsamples in absolute magnitude intervals -19.0 - -20.0, -20.0 - -21.0, and -21.0 - -22.0, and call them subsample A_c, B_c and C_c, respectively. In table 3, we give the v_{max}, D_{max} and the galaxy number for each subsample involved in every cross-correlation analysis.

The resulting cross-correlation functions and the lg $\xi_c(r)$ - lg r diagrams are presented in fig. 3. These cross-correlation functions can also be approximated by power functions. The straight lines in fig.3 b), are obtained from linear regression.

Table 3. The v_{max}, D_{max}, and galaxy numbers for each cross-correlation analysis

cross correlation	v_{max}(km/s)	D_{max}(Mpc)	galaxy numbers
C_c - B_c	3970	79.4	C_c: 91
			B_c: 304
C_c - A_c	2500	50.0	C_c: 42
			A_c: 223
B_c - A_c	2500	50.0	B_c: 145
			A_c: 223

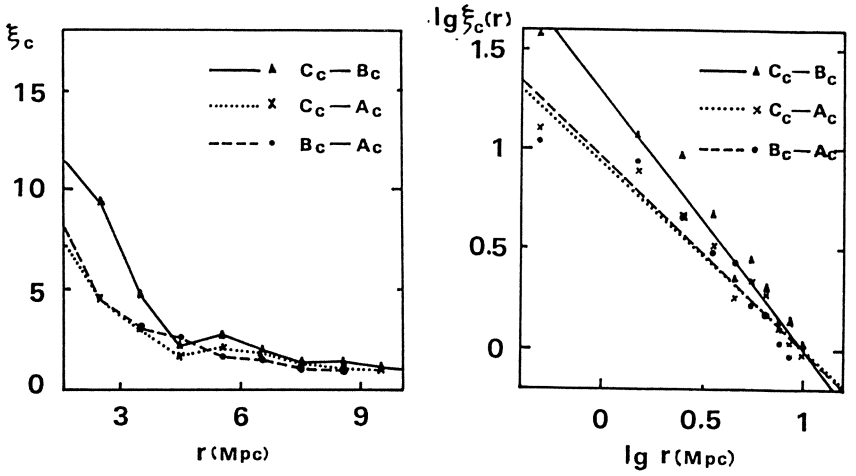


Fig. 3a) The cross-correlation functions between galaxies with different luminosity; b) The lg $\xi_c(r)$ - lg r diagrams for these cross-correlation functions. The straight lines are obtained from regression analyses.

Similar to the autocorrelation analysis, we can determine the cross-correlation strength A_c, cross-correlation scale r_{co} and the power index α_c of each cross-correlation function from the regression. The results are presented in table 4.

Table 4. The parameters of cross-correlation functions and the correlation coefficients of linear regressions for the cross-correlation analysis between galaxies with different luminosity

cross correlation	α_c	A_c	r_{co}(Mpc)	γ
C_c - B_c	1.266	19.636	10.50	0.985
C_c - A_c	0.911	8.606	10.62	0.974
B_c - A_c	0.926	8.527	10.11	0.958

We can see from these results that the correlation between the two most luminous subsamples is obviously stronger than the other two correlation. The result of the cross-correlation analysis is consistent with the result of the autocorrelation analysis, i.e. the more luminous galaxies are clustering stronger. In the luminosity range considered there is no trace showing that the most luminous galaxies are surrounded by faint galaxies.

Combining the results given by auto- and cross-correlation analyses we can conclude that the clustering of galaxies with different luminosity is different. More luminous galaxies cluster stronger and with larger clustering scale. The more luminous galaxies have also a steeper correlation function than the fainter galaxies.

4. Conclusions and discussions

Our analysis about the large scale distribution of galaxies with different luminosity can propose some new problems on the current models of galaxy formation, but can also supply some new clues about the formation and evolution of galaxies.

The results given by our analyses show that the more luminous galaxies have steeper correlation function, larger correlation strength and scale. All of these properties cannot be explained by the most popular "top-down" and "bottom-up" models, even could not be explained by the newest cosmic-string model, because all these models disregard the differences among galaxies with different luminosity.

The flatter correlation functions for galaxies in any luminosity interval will make the time difficulty of the top-down model even worse. The "bottom-up" model seems to have more difficulty to explain the differences of power indices in the correlation functions of galaxies with different luminosity. The cosmic-string model has problems to provide a reason why the more luminous galaxies are more likely to cluster with each other than with the fainter galaxies.

On the other hand, the results given by these analyses have also supplied some useful information on galaxy formation. All results can easily be understood from the idea of biased galaxy formation.

In the linear regime of the early universe, there are fluctuations with various scales which are randomly superposed. The larger scale fluctuations contain more mass and will supply a suitable environment for galaxy formation. Firstly, the gravitation of a large scale fluctuation is likely to stop the expansion caused by the expansion of the universe and leave the smaller scale fluctuations to survive and grow. Secondly, it supplies a preferrable evironment for cooling which is an important condition of galaxy formation.

Based on these arguments we might suggest that in the early universe the large scale fluctuation might supply a suitable background in which the galaxies can easily form. In areas which lack this condition, the galaxy formation will be surpressed and voids could evolve.

According to this scenario the galaxies would form at first in large scale fluctuation regions. So, galaxies are clustering from their birth. This model explains the results given by our analysis of the large scale distribution of galaxies with different luminosity.

In the analysis presented above, the effects of peculiar motion have been disregarded. The effect of peculiar motions would make the clustering of galaxies look more disperse than in reality. The more luminous galaxies are most likely to be in richer clusters. They would have larger mean velocity dispersion. Therefore, the effects of peculiar motion would cause the correlation function of the more luminous galaxies to become flatter, and weaken the differences between the correlation functions of galaxies with different luminosity. Thus, peculiar motion cannot be the reason of the systematic change which is revealed from our analysis.

References:

1. M. Davis, P.J.E. Peebles: Ann. Rev. Astr. Astrophys. 21, 109 (1983)
2. M. Davis, P.J.E. Peebles: Ap.J. 267, 465 (1983)
3. A. Guth: Phys. Rev. D23, 347 (1981)
4. A. Dekel, M.J. Rees: Nature 326, 455 (1987)
5. A. Dekel, J. Silk: Ap.J. 303, 39 (1986)
6. J. Eder, J.M. Schombert, A. Dekel, and A. Oemler: in preparation
7. M. Davis, S. Djorgovski: Ap.J. 299, 15 (1985)
8. G.D. Bothun, T.C. Beers, J.R. Mould, J.B. Huchra: preprint (1986)
9. J. Huchra, M. Davis, D. Latham, J. Tonry: Ap.J. Suppl. 52, 89 (1983)
10. M.J. Geller, J.P. Huchra: Ap.J. Suppl. 52, 61 (1983)
11. P.J.E. Peebles: The large Scale Structure of the Universe, (Princeton Univ. Press, 1980)

The Large Scale Distribution of Abell Clusters

G. Börner[1], *You-yuan Zhou*[1,2], *and Yu Gao*[2]

[1]Max-Planck-Institut für Physik und Astrophysik, Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-8046 Garching, Fed. Rep. of Germany
[2]Center for Astrophysics, University of Science and Technology of China, Hefei, Anhui, People's Republic of China

1. Introduction

The most common and reliable method of studying the large scale structure of the universe is through the space distribution of galaxies, and one of the most powerful tools in studying the large scale structure is the two-point correlation function of galaxies [1-5]. The study of the distribution of galaxies has confirmed the existence of superclusters and revealed the string-void structure (e.g. from the cfA survey with a limiting magnitude in $14^{m}.5$) or the bubble-void structure (e.g. from cfA survey down to $15^{m}.5$ [6]). All the results and conclusions are very important for the understanding of the cosmological evolutional scenario and the theory of galaxy formation and evolution. [1,8]

But the existence of the supercluster-void structure has also drawn attention to the problem about the super large scale structure of the universe, i.e. the existence of super-superclusters. It involves the understanding of the whole scenario of the universe and the transition of the clustering on scales about 100 Mpc to the uniformitiy of the universe as a whole, and it will influence existing theories of the galaxy formation. An important part among them is the correlation of superclusters. The galaxy distribution analyses cannot solve this problem, because there has not been complete galaxy survey on a large area with the depth of several hundreds Mpc as jet. One hope is to search for it in samples of objects, which have mean separations above 100 Mpc and which can mark the positions of supercluster.

Some tentative work has been done using QSOs and radio galaxies as markers of superclusters [9,10]. Other objects capable of marking superclusters are clusters of galaxies. On the base of the strong correlation of Abell clusters with distance class $D \leq 4$ [11,12] Baheall and Soneira determined a specific catalog of supercluster candidates [13] and found a correlation function for superclusters [14]. But in their study the sizes of the supercluster sample are 12-16, so it is valuable to confirm further the conclusions about the correlation of superclusters. This will be the main purpose of our work, and we also attempt to extend the correlation investigation of Abell clusters to the distance class D=6 in order to obtain more information about the dependence of the correlation function on distance class D, richness R and the morphological type of Abell clusters. We expect that all these type statistical

properties will be helpful to trace the large scale structure of the universe including superclusters and voids, search for the super large scale structure, estimate the peculiar velocities, investigate the relation between clusters and superclusters, and give the statistical basis for a theory of galaxy formation.

2. The Distance Indicator of Abell Clusters

The galaxy redshift has been accepted commonly as its distance indicator except in the case of nearby galaxies with relatively large peculiar velocities. So one extends it to the case of clusters and also uses the measured redshift value of galaxies within a cluster as the distance indicator of the cluster. On this basis many investigators made the three-dimensional analysis of clusters with $D \leqslant 4$ [11, 12, 15]. But when we attempt to extend this to the sample of Abell clusters with $D \leqslant 6$, we find that the number of Abell clusters with measured redshifts is about 650 which is only one third of the total number of Abell clusters. It makes the three-dimensional analysis difficult.

Furthermore, it is valuable to estimate the reliability of using the redshift of galaxies within a cluster as the distance indicator of the cluster. As we well know, the foreground and background galaxies severely affect the determination of the members and thus influence the measured parameters of clusters. The influences are not only on the apparent magnitude but also on the redshift. For the latter one generally accepts the redshift of two even only one galaxy within the cluster as the cluster redshift value. Some observations show that such a suspicion might be justified. For instance, after measuring the redshift of two galaxies for each cluster in a sample of 26 clusters, one found that in 6 of them the two galaxies have different redshifts with the differences larger than 2000 km/s, i.e. $\frac{\Delta z}{z} \geqslant 10\%$. In another redshift survey with 37 cluster postmanetal measured the redshifts of 3 galaxies for each cluster and found $\Delta z/z \geqslant 10\%$ [15]. We collect the redshift values of about 600 Abell clusters and find that 170 clusters among them have two measured values of redshift from different observers or two epoches of observations and 83 (about 50% of the test sample) have $\frac{\Delta z}{z} > 5\%$. That $\sigma\left[\frac{\Delta z}{z}\right] = 0.17$, i.e. $\sigma \log z = 0.07$, shows that due to this foreground effect there are comparatively large systematic errors in the redshift values of Abell clusters. We suppose that this value will be the criterion of using redshift-magnitude relation to estimate the redshift in our work.

3. The redshift-magnitude Relation of Abell Clusters

As for only about one third of the whole Abell clusters redshifts have been measured, and only clusters with $D \leqslant 4$ form a complete sample with z value. The most useful method to obtain statistical properties of an Abell sample with $D \leqslant 6$ is to take apparent magnitudes m instead of z as distance indicators. But existing z-m relations always have comparatively large scatters. For instance, the deviation of the relation obtained recently by Postman et al.[15] is $\sigma \log z = 0.14$, which is twice of

the errors of measured redshifts σ logz = 0.07. So we should first improve the z-m relation for Abell clusters.

The main reasons for the scatter of a z-m relation are the following two: there is the influence of the foreground effect, and there is the scatter of optical luminosity in whole Abell cluster sample, which is composed of clusters with different R and various Bautz-Morgon types and affected by the selection effect dependent on the D. In order to improve the z-m relation we suggest that for the former we discard about 50 clusters with quite different z from the sample, and for the latter we classify the sample into some subsamples according to their values of D, R, and their B-M types. After the necessary magnitude corrections, including A-correction, K-correction, Scott effect correction, Malguist effect correction, and B-M classification correction, we obtain the improved z-m relations for different D and R and various B-M types. All the deviations σ logz are between 0.09-0.12, which are smaller than the result of Postman et al without the classification. Although they are still larger than 0.07, as a tentative work we can use them to obtain compatatively reliable results and to discuss the three-dimensional correlation function of Abell clusters and the correlation of superclusters determined by using the spacis distribution of Abells.

4. Two-point Correlation Functions of Abell Clusters

1) D≤4

As a check we calculate the two-point correlation function of Abell clusters with D≤4, and compare it with the results of Bahcall et al. which were obtained by using the measured redshift values. From Fig. 1 we find that our results agree with those of Bahcall et al. except for very small distances. This means that the improved z-m relations are good enough for the calculation of spatial correlation functions at large separations.

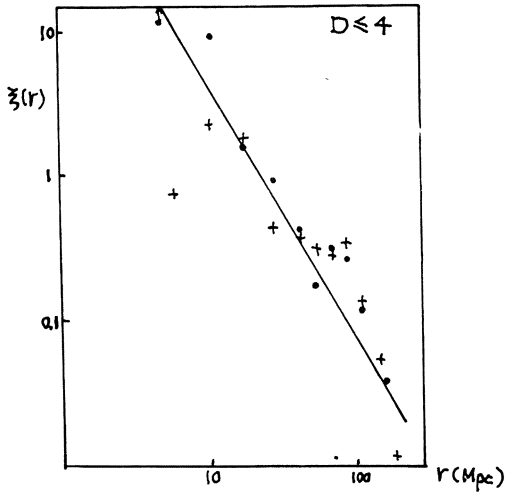


Fig 1. the comparison of correlation function with D≤4 calculated from the measured redshift(·) and the estimated redshift (+)

2) The dependence of the two-point correlation functions $\xi(r)$ on distance class D

Fig 2. gives the two-point spatial correlation functions with $D \leqslant 5$ and $D \leqslant 6$. From Fig 2 we find:

1/. As the regions covered by the samples are larger, the correlation is continous up to 300 Mpc. At separations of 100--300 Mpc the correlation curves become flat.

2/. For separations smaller than 100 Mpc $\xi(r)$ is not strongly dependent on D

3/. At large separations (>100 Mpc) the correlation functions increase obviously with D increasing.

How can one explain these results? Comparing with Fig 3, which has the corresponding angular correlation functions, we find that the curves of the two-point angular correlation functions also become flat at large angular separations. We suppose that the flatness of $\xi(r)$ at large r is not caused mainly by the estimated z.

So we can conclude that

1/. the samples with $D \leqslant 4$, 5, 6 are quite fair except for r>100 Mpc.

2/. the sample with $D \leqslant 4$ is deep enough to get mainly the correlation of clusters within one supercluster, but the sample s with $D \leqslant 5$ and $D \leqslant 6$ are deep enough for testing the correlation of clusters in different superclusters. In Fig 2 we also give the composite curves which represent the correlation of clusters located in one supercluster (straight line) and in different superclusters (dotted curve). The flatnaness for separations larger than 100 Mpc reveals the correlation of Abell clusters in different superclusters. So it is important to study the sample with $D \leqslant 5$ and $D \leqslant 6$ if we want to search for the correlation between superclusters, though for these sample we have not measured all the redshift.

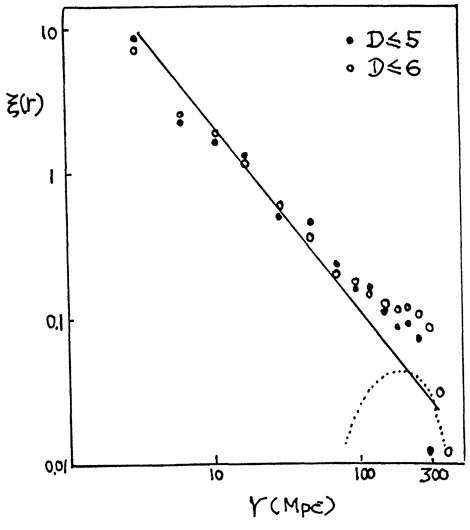


Fig 2. the spacial correlation functions with $D \leqslant 5$ and $D \leqslant 6$

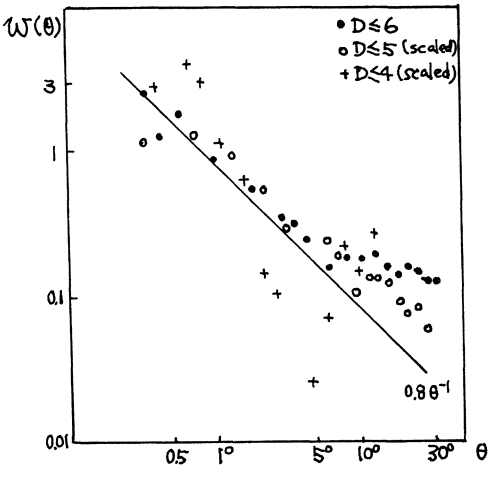


Fig 3 The angular correlation functions with $D \leqslant 4$, $D \leqslant 5$ and $D \leqslant 6$

3) The dependence of correlation functions $\xi(r)$ on richness R.

Fig 4 is the spatial correlation functions $\xi(r)$ for different R. We find that the correlation functions $\xi(r)$ increases as R increases. It is qualitatively comparable with the relation $\xi \cdot \alpha <N>^{0.7}$ which was obtained by Bahcall and Burgett[14]. For large distances or angular separations, respectively, spatial correlation functions $\xi(r)$ with R=1, 2, ⩾3 flatten, considerably as do the angular correlation functions W(O) (Fig. 5). This supports the explanation of the flattening for D⩽5 and D⩽6.

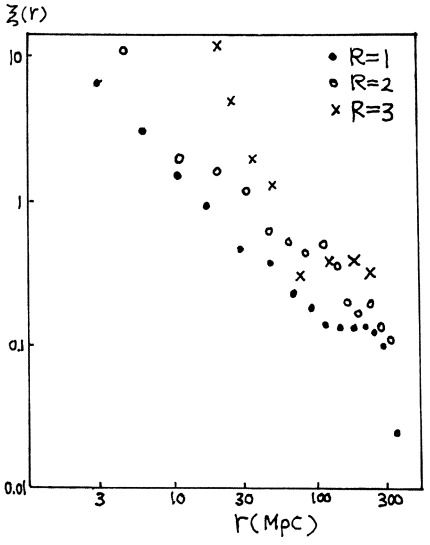


Fig 4. the spacial correlation function $\xi(r)$ with different R

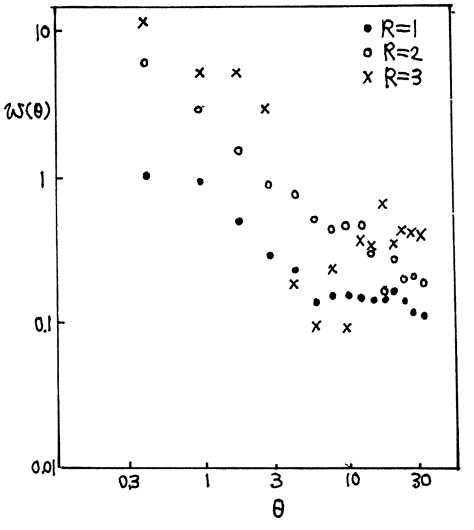


Fig 5. the angular correlation functions $\omega(\Theta)$ with different R

4) the relation between $\xi(r)$ and B-M types.

Fig 6 shows the spacial correlation functions for various B-M types. From it we find that the correlation functions of type III and type II-III are similar to that of whole Abell clusters, but $\xi(r)$ of type II, I-II, and I have different behavior. Fig. 7 shows $\xi(r)$ for type I, I-II and there the strange features are more obvious. It seems to have three peaks located at about 30 Mpc, 100-150 Mpc, and 300 Mpc. According to the Bautz & Morgan classification we know that type I clusters always have a very bright galaxy, a cD galaxy, in its centrical region the features of bright galaxies in type II clusters are between cD galaxies and normal ellipticals. Type III have no obviously bright galaxy. Among the 1683 Abells the total number of type I and I-II is 122, which is less 10% of the total number of Abell clusters, therefore, the mean separation of 180 Mpc is so large, that it can be influenced not only by the correlation of clusters within one supercluster but also of clusters located in different superclusters. The decline part between 30 Mpc - 60 Mpc represents the correlation of clusters within one supercluster as Bahcall et al have shown. The second peak between 100 Mpc and 150 Mpc, which is about the size of superclusters, reveals the

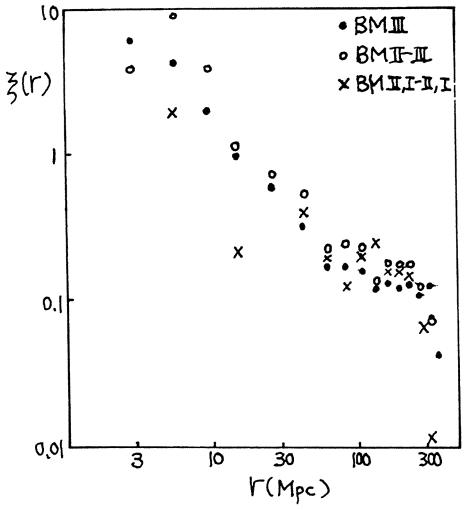


Fig 6 The correlation functions for various Bautz-Morgan types

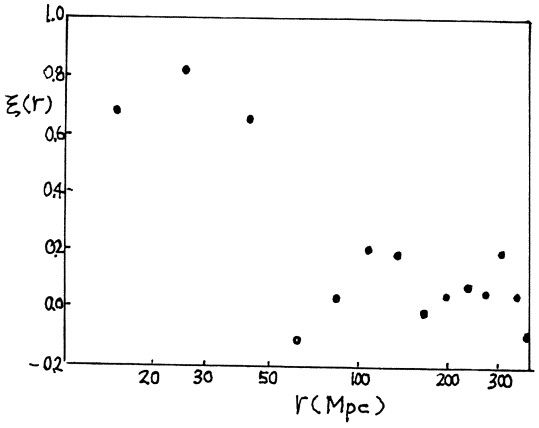


Fig 7 The correlation function of B-M type I and I-II

correlation of clusters in different superclusters. Furthermore, the cD galaxy always is a strong radio galaxy, so it is natural that the functional dependence shown Fig 7 is similar to $\xi(r)$ of bright radio galaxies[10].

5. The determination of supercluster candidates.

Bahcall and Soneira have determined 12-16 supercluster candidates [13] by using Abell clusters with D≤4. Bahcall and Burgett [14] further get one point at about 100 Mpc of supercluster correlation. In order to obtain more properties of the correlation function of superclusters above 100 Mpc we should use the sample of Abell clusters with larger D. Following the method in [13] we use

$$n \geq f\, n_0$$

as the criteria of superclusters, where n is the density of Abell cluster in a considered region, n_0 is the mean density of Abell clusters in the sample, and f is an enhancement factor for which we take 10, 20, 40... And we also make the following improvements:

1/. We assume the center of mass as the central position of the supercluster and take weight of different R as $m_{p=1} : m_{R=2} : m_{R=3} : m_{R=4} = 65 : 105 : 165 : 250$

2/. We restrict the size of superclusters to radii between 40 Mpc and 100 Mpc or slightly larger than 100 Mpc.

3/. We use $n_0(z)$ instead of n_0, where $n_0(z)$ is the mean density of clusters at z, which is obtained from the smoothed function of z distribution of Abell clusters.

4/. We define the richness Rsc of a supercluster as

$$Rsc = \Sigma\, Rc$$

and take the richness restriction Rsc≥3. We can form two kinds of samples. One is a magnitude limited sample. For the tenth bright magnitude of clusters $m_{10}\leqslant 17.^{m}0$

$$(D=4,\ m_{10}\leqslant 16.^{m}4;\ D=5,\ m_{10}\leqslant 17.^{m}2)$$

there are 75 and 37 superclusters for f=20 and 40, respectively. The other sample is a distance limited one. There are 45 and 86 superclusters for $z\leqslant 0.12$ (f=20) and $z\leqslant 0.14$ (f=20), respectively.

6. The Correlation of Superclusters

For the correlation of superclusters there are different conclusion. Bahcall and Burgett [14] have found the obvious correlation of superclusters determined by the agglomeration of Abell clusters with D≤4. Kalhikov and Kuneva have found that the distribution of superclusters is random[16]. Their conclusions were obtained from the distribution of cluster within the scale of about or slightly larger 100 Mpc. The boundaries and the properties of samples will affect the computational results. So it is necessary to extend the scale covered by the samples as far as possible. Recently Pietronero pointed out that the usually adopted $\xi(r)$ provides a consistent characterization of correlations only for systems in which these correlations decay at a characteristic length that is much smaller than the radius of the sample[17]. So we should use samples with larger scale to test the correlation of superclusters.

Fig. 8, 9 and 10 are the correlation functns $\xi ss(r)$ of superclusters which are obtained from the Abell cluster samples with the restriction of $m_{10}\leqslant 17$, $z\leqslant 0.12$ and $z\leqslant 0.14$, respectively. From it we can find that there are no obvious correlations of superclusters at the separation of 100 Mpc – 300 Mpc. The correlation function for all samples of $m_0\leqslant 17.^{m}0$, $z\leqslant 0.12$, and $z\leqslant 0.14$ at 100–300 Mpc is below 0.3 and within 1σ regions determined by the Monto-Carlo sampling. Even if we consider the correlation at the region from 30 Mpc to 300 Mpc and keep the $\gamma^{-1.8}$ low, we find that the correlation function $\xi ss(r)$ of superclusters satisfies 850 $\gamma^{-1.8}$, which correlation strength is still less than the result 1500 obtained in [14].

Although we find the correlation of clusters located in different superclusters in the correlation function of Abell clusters in deeper samples, for instance, D≤5, D≤6, different R and various B-M types, especially, type I, I–II and II, this means the existence of supercluster structure, but this does not show the space distribution of superclusters. The correlation function $\xi ss(r)$ will reveal the distribution and structure of superclusters. That there is no obvious correlation of superclusters shows that there may not exist a structure of the universe with a length scale that is much larger than the one of superclusters. Random distribution of superclusters favours the understanding of the transition from the clustering on the scale of superclusters to the uniformity of the universe. Otherwise, if the strength of clustering increases according to the law $\xi_0 \alpha <n>^{0.7}$, we may ask on which scale it will change and how it stops abruptly.

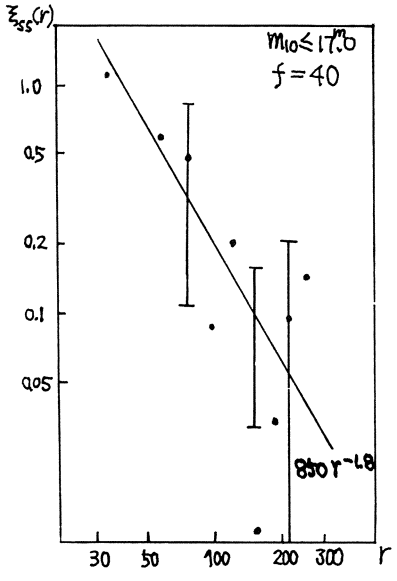


Fig. 8 The correlation function of superclusters determined by Abell clusters with $m_{10} \leq 17^m$

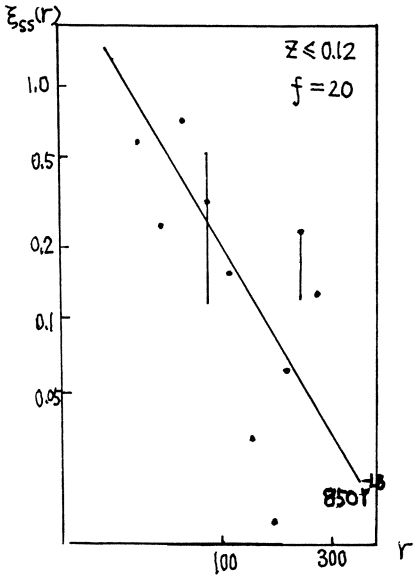


Fig. 9 The correlation function of of superclusters determined by Abell clusters with $z \leq 0.12$

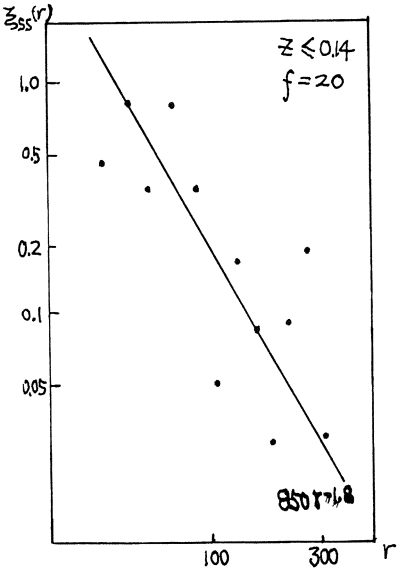


Fig. 10 The correlation function of superclusters determined by Abell clusters with $z \leq 0.14$

7. Summary

1/. Besides the space distribution of galaxies, we can study the large scale structure of the universe by using the marking object methods of the superclusters, including Qs0s, bright radio galaxies, and rich clusters. We have shown that Abell clusters and Bautz-Morgan type I and I-II are two kinds of marking objects with many available observable data.

2/. In order to study more statistical properties of correlation functions of Abell cluster and the correlation of superclusters we should extend our sample to $D \leqslant 5$ and $D \leqslant 6$.

3/. As the redshift values are not available for all Abell clusters to use the redshift-apparrent magnitude relation is a very useful method.

4/. The foreground effect will influence not only the apparent magnitude but also the measured redshift of Abell clusters.

5/. The improved z-m relation makes the deviation of estimatd redshift of cluster smaller. It is only slightly larger than the systematic errors in the redshift measurements of clusters.

6/. The computational correlation function of $D \leqslant 4$ from the improved z-m relation is very close to that from the measurd redshift except at very small separations.

7/. The correlation functions are independent of D for r less thn 100 Mpc, and $\xi(r)$ increases with D at $r > 100$ Mpc.

8/. The correlation functions increase with the richness R.

9/. Bautz-Morgan type I and I-II can reveal the correlation of clusters located in different superclusters. The size from the correlation agrees with the size of superclusters from direct observation.

10/. The correlation functions of superclusters are very small at 100-300 Mpc. This shows that superclusters are likely to be distributed randomly and there may exist no structure of the universe larger than superclusters.

References

[1]. J.H. Oort: Ann. Rev. Astron. Ap. 21 373 (1983)
[2]. M. Davis and P.J.E. Peeples: Ap.J. 267 465 (1983)
[3]. J.E. Felten: Comments Astrophys. 11 53 (1985)
[4]. J. Einasto, A.A. Klypin, E. Saar and S.F. Shandarin: M.N.R.A.S. 206 529 (1984)
[5]. P.J.E. Peeples: The Large Scale Structure of the Universe (Princeton University Press 1980)
[6]. V. de Lapparent, M.J. Geller and J.R. Huchra: Ap.J. Lett. 302 L1 (1986)
[7]. A. Dekel and M.J. Rees: Nature 326 455 (1987)
[8]. A. Dekel: Comments Astrophys. 11 235 (1986)
[9]. Y.Y. Zhou, D.P. Fang, Z.G. Deng and X.T. He: Ap.J. 311 578 (1986)
[10]. Y.Y. Zhou, Y.P. Jing Proceedings of IAU Symp. 130 (1987)
[11]. N.A. Bahcall and R.M. Soneira: Ap.J. 270 20 (1983)
[12]. A.A. Klypin and A.I. Kopylov: Soviet Astron. Lett 9 41 (1983)
[13]. N.A. Bahcall and R.M. Soneira: Ap.J. 277 27 (1984)
[14]. N.A. Bahcall and W. Burgett: Ap.J. Lett 300 L35 (1986)
[15]. M. Postman, M.J. Geller and J.P. Huchra: A.J. 91 1267 (1986)
[16]. M. Kalinkiv and I. Kuneva: M.N.R.A.S. 218 49p (1986)
[17]. L. Pietronero: Physica 144 257 (1987)

A Study of the Cosmic Constant H_0 with SN1

Wu Shimin, Guo Wanwen, and Li Zongwei
Astronomy Department of Beijing Normal University

To determine the accurate value of the Hubble constant H_0 is one of the most important things of cosmology. Since 1929 the value has been revised from 550 km/s. Mpc to 50-100 km/s. Mpc.

Sandage and Tammann obtained that the value of H_0 is about 50 km/s but de Vaucoulers obtained that it is about 100.

It is difficult to say which is the better, 50 or 100 km/s?

We use SN1 as standard candle to determine H_0 and obtain that it is about 50.

From the light curve of many SN1 (Fig. 1), demonstrated by Kowal (1968)[1], we can see that the dispersion of the absolute magnitude at maximum brightness is very small.

It shows that the luminosities of SN1 at maximum brightness reach unanimity. As a kind of standard candle SN1 has an advantage that its absolute magnitude can be calculated from theoretical models. We can obtain the Hubble distance directly and it is unnecessary to correct the distance from near to far step by step because the luminosity of SN1 is very high. So the problem will become very simple.

According to Hubble's law for a sample of SN1, we have

$$m = 5 \lg V + D$$

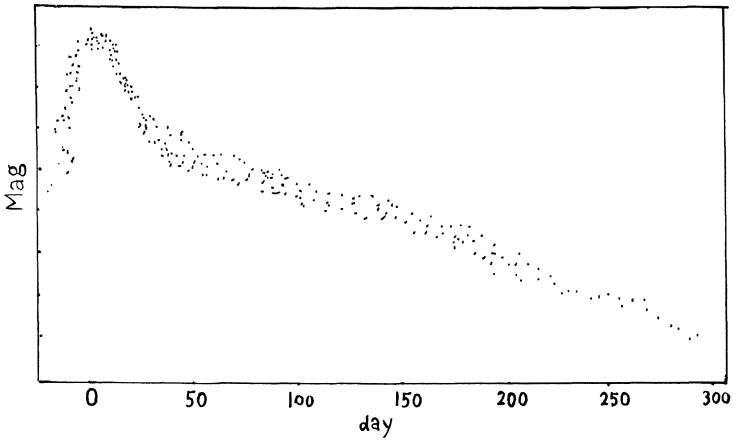


Fig. 1 Synthetic light curve of many SN1

m is the apparent magnitude at maximum brightness of SN1.

V is the radial velocity of the SN1 parent galaxy as seen in the frame of our Galaxy.

We use a Friedmann cosmological model without cosmological constant

$$m-M = 5\ \lg V-5\ \lg H_0+25$$

therefore the constant D can be written as

$$D = M+25-5\ \lg H_0$$

D can be given by a linear fit from the Hubble diagram.

M has to be determined by a method which is independent of the Hubble constant.

By means of a physical model some authors have obtained the value of M by different methods and different samples. Their results are nearly the same (Table 1). We take their average value

Table 1

Author	Method and sample	M_B(max)
Cadanau et. al (1985)	5 SN1 in Virgo cluster	-19.75±0.3
Tammann (1982)	Historical SN1 Tycho and Kepler	-20.00±0.60
Arnett (1982)	Model-dependent expansion Parallaxes of SN1	-19.8±0.7
Branch (1982)	SN1 model of exploding carbon-oxygen white dwarfs invoke the radio-active of decay N_i^{56}	-19.0
Sutherland et. al (1984)	the same as above	-19.6
Arnett et. al (1985)	Carbon deflagration model for SN1	-19.6±0.6
Sandage	1973c and 1954a	-19.7±0.19

M = -19.6±0.12 as the maximum luminosity SN1

We have 114 SN1 samples (Fig.2), most of them are adopted from "A Revised Supernovae Catalogue"[2] and 11 samples are newly discovered in the period from Oct. 1984 to June 1986.

The procedures used to determine the H_0 are as follows:

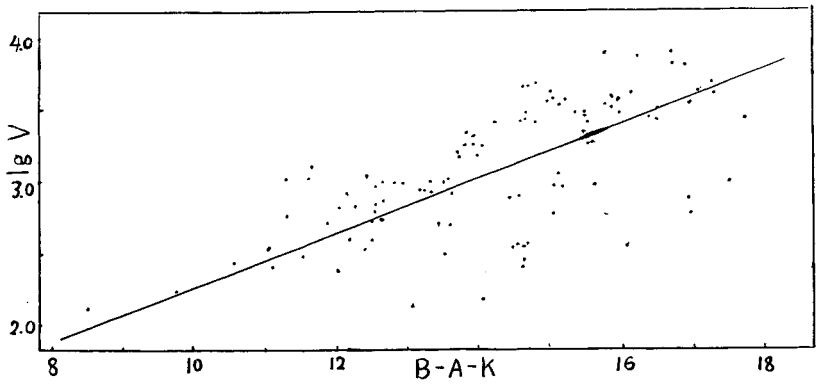


Fig. 2 Hubble diagram of 114 SN1 (after correction)

1. The values of magnitude are corrected to take account into the interstellar extinction in the Galaxy by

$$A_{pg} = 0.24\ \mathrm{cscb}^{[3]}.$$

2. Unify m_{pg}, V and B magnitude given in the catalogue by relations

$$m_{pg} + 0.4 = B^{[4]}$$

$$B-V = -0.15.$$

we change all m_{pg} and V into B.

3. K-correction[5]

4. In order to reduce the velocity dispersion we keep most of the samples with V>1000 km/sec and discard a few samples with V<1000 km/sec.

5. Classify the samples into four subsets and take their weight average (Table 2)

$$H_0 = \frac{\sum_i \frac{H_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}}$$

where σ_i is the standard deviation of the ith subset, and $\frac{1}{\sigma_i^2}$ is the weighted factor, of the ith subset. The error of H_0 is given by

$$\Delta H_0 = \sqrt{\frac{\sum_i (H_i - H_0)^2}{n-1}} = 5.5 \text{ km/s.mpc}$$

Finally, we obtain H_0=52±5.5 km/sec.mpc.

According to the obtained value H_0 the formal age of the universe would be $t=1/H_0 = 1.97\times10^{10}$ years, which agrees with the age of the globular clusters.

Table 2

Type	Number of samples	D	k	σ	H_o	Relative error of H_o
$SN1_a$	20	-3.38	0.15	0.69	57.0±15.4	27%
SN1 in elliptical galaxies	11	-3.27	0.15	0.88	54.2±20.2	37%
SN1 with unassured subtype	53	-2.87	0.15	1.05	45.1±11.8	26%
SN1 disregarded	82	-3.00	0.14	1.11	47.9±10.0	24%
I_b and I_{pec} all of SN1	86	-3.00	0.14	1.08	47.9±11.3	24%

If we take the value of H_o as 100 km/sec.mpc, the Hubble time will be less than the age of globular cluster.

Source of error

We think that the errors may come from following factors:

1. Extinction correction
2. Dispersion of velocities
3. Dispersion of absolute magnitude of SN1 at its maximum brightness
4. Magnitude transformation among m_{pg}, V and B

References

1. Kowal, G.J. A.J.73 (1968) 1021
2. Barbon, R. et.al, A.Ap.suppl.58 (1984) 735
3. Branch, D. and Bettis, G. A.J.83 (1978) 224
4. Bust, B.W., Thesis uniu illinois 1974; Barbon, R. et.al, Ap.J. 44 (1975) 267
5. Oke, J.B. and Sandage, A.Ap.J.154 (1968) 21

Editor's note: Unfortunately the authors were unaware of the work by Sandage and Tammann on type I SN to determine H_o.
I recommend to the interested reader consultation of , e.g.,
Sandage A., Tammann G., 1984, Nature 307, 326
and related papers.

The Large Scale Inhomogeneous Distribution of Ly-Alpha Absorption Lines in the Spectra of Quasars

Yaoquan Chu and Xingfen Zhu

Institute for Astrophysics, University of Bonn, Fed. Rep. of Germany
Center for Astrophysics, University of Science and Technology of China, Hefei, People's Republic of China

Using a quite resonable homogeneous sample of Ly-alpha absorption lines in quasars spectra, we find some evidence for existence of very large scale inhomogeneity in distribution of Ly-alpha absorption line redshifts. We find a significant peak at $Z \sim 2.9$ in the histogram of redshift distribution. The power-spectrum analysis shows that a periodic component at the length of 0.2 for the argument $x=\ln(1+z)$. The significance of confidence is not very high, about 83%, but the length of periods is just the same as for the periodicity of emission line redshifts of quasars. Moreover, the ratio $R=(1+Zem)/(1+Zab)$ for Ly-alpha absorption lines also shows several peaks in its distribution which is in good agreement with the prediction if the periodicity exists in both emission line and absorption line redshifts.

1. Introduction

In the early days of discovery of quasars, various astronomers noted that there might exist a large scale inhomogeneity in the space distribution of quasars. Burbidge (1968) discussed the presence of two peaks in the distribution of quasars` emission line redshifts at z=0.06 and z=1.95 and he also pointed out the probable existence of periodicity in that distribution. Subsequent investigations performed by Cowan (1969) and Burbidge and O'Dell (1972) confirmed further Burbidge's results, while Plegemann et. al. (1969) and Wills and Ricklefs (1976) could not detect any period. In 1972 Karlsson suggested that the periodicity is with respect to argument $x=\ln(1+z)$, but not to z. Using a much large number of quasars listed in Hewitt and Burbidge catalogue and statistical analysis of power-spectrum, we have cofirmed the periodicity in the distribution of quasars' emission line redshifts with the periodicity length of 0.205 in the variable $x=\ln(1+z)$. (Fang, Chu, et. al. 1982) As for the quasar's emission line redshifts, one could argue that the periodic distribution may be due to the vary many observational selection effects.(for example, the preference of identifications of certain emission lines in quasars' spectra, i.e. Ly-alpha, C IV, Mg II, ...). However if the periodicity distribution of quasars is a real cosmological effect and not due to the selection effect, it is reasonable to predict that a similar feature should also exist in the distribution of other high redshift objects. In the present paper we concentrate our discussion on the statistical analysis of the redshift distribution of Ly-alpha absorption lines and compare it with that of quasars' emission lines.

The spectra of high-redshift quasars show a great number of absorption lines on the short-wavelength side of the Ly-alpha emission lines (i.e. the Ly-alpha forest). In this spectral region only a small proportion of lines can be identified with elements heavier than hydrogen (or metal-line systems). Such systems are generally believed to be due to intervening galaxies at the distances indicated by the systems redshifts, and the forest lines are far more numerous.

While there is a point of view that these Ly-alpha absoption lines are formed by filamentary composition which ejected from the nuclei of the parent galaxies during active stage,(see Kundt and Krause, 1985.) it is now generally believed that these absorptions lines are due to Primordial Hydrogen Clouds (PHC), which are not associated with galaxies. If we accepted that these PHC are located at

cosmological distances given by their redshifts, the PHC become a new class of objects which have high redshifts comparable with quasars. So the study of their distribution would give us new test to confirm whether the periodicity distribution is a real feature of the large scale structure in the universe. Indeed, the results we present here show that there is some evidence in favor of the existence of periodicity in the distribution of redshifts of PHC.

2. Sample of Ly-alpha absorption lines

As is well known, an statistical study of the distribution of absorption line redshifts of quasars is difficult because there exist many selection effects in process of identifications of absorption lines in quasars' spectrum. The number of absorption lines in a quasar spectrum is a function of many variables: emission redshift, wavelength covarage, resolution, detector signal/ noise ratio, location of the continuum and equivalent-width limit. In order to get meaningful statistical results, it is necessery to have a carefully selected homogeneous sample in which all quasars are treated in the same way. Here we use a quite reasonable homogeneous sample of Ly-alpha absorption lines available in the literature compiled by Murdoch et. al. (1986), which was obtaited at high resolution (0.8 -1.5 A FWHM) and with the same type of detector (the image photon counting system). We also set a minimum intrinsic Ly- alpha equivalent width Wo=0.32A. The total sample includes 11 quasars and 277 Ly-alpha absorption lines. The observed absorption redshifts range from 1.510 to 3.715 which give the largest interval in redshift currently available from the ground observation. Relevant details for the Ly-alpha sample are given in Table 1.

In Table 1 we give the name of quasar, the emission line redshift Zm and the low limit of Ly-alpha absorption redshift Zo which we accept in this sample. In

Table 1. The Ly-alpha Sample

QSO	No. of Lines	Zo	Ze	Redshift of metal lines system	Reference
2000-330	73	3.033	3.780	3.182	1
				3.333	
				3.549	
2126-158	52	2.500	3.28	2.638	2
				2.769	
0002-422	34	2.200	2.763	2.302	2
				2.464	
PHL 957	29	1.900	2.690	1.929	2
				1.940	
				1.948	
				2.310	
0453-423	37	2.210	2.656	2.276	2
				2.397	
1225+317	20	1.700	2.200	1.796	2
				1.887	
				1.897	
				2.120	
0421+019	8	1.715	2.051		3
0119-046	10	1.682	1.937		3
0002+051	5	1.715	1.899		3
1115+080	2	1.682	1.725		3
0215+015	7	1.501	1.715		4

References for Table 1:
1. Hunstead, R. W., Murdoch, H. S., et. al. Ap. J. vol.301, 522.
2. Sargent, W. L. W., Young, P. J., Boksenberg, A., and Tytler, D., 1980, Ap. J., Suppl., vol. 42, 41.
3. Young, P., Sargent, W. L. W., and Boksenberg, A., 1982, Ap. J, vol. 252, 10.
4. Blades, J. C., Hunstead, R. W., Murdoch, H. S., and Pettini, M., 1985, Ap. J., vol. 288, 580.

Table 1 we also list the redshifts of Ly-alpha lines which belong to definite heavy-element system for first six quasars. These lines are formed by intervening galaxies and we will use these data to calculate the cross-correlation function between the galaxies and primordial hydrogen clouds.

3. The Redshift Distribution of PHC

The histogram of all Ly-alpha absorption lines is shown in Fig.1.

Before we try to interpret such distribution we must point out that there is a strong selection effect due to the different wavelength coverages of different quasars. A spectrum of one quasar can only contribute the Ly-alpha absorption line count between Zm and Zo for this quasars. Therefore we could correct the initial histogram by multiplying a factor that is proportional to the number of quasars' spectra which cover the redshift bin we study. Each bin of Z which is located between the redshift limit of a quasar's spectrum is given a "spectrum count" of +1, the Z bin outside the limit is given a spectrum count of zero. The procedure is prepared for all quasars in the sample. Finally for each Z bin we have the total raw line count N, the total spectrum counts S and the number of quasars n. We then have the corrected count Nc given by:

$$Nc(Z)=N(Z) * n / S$$

The corrected histogram is shown in Fig.2. Of course we have multiplied a factor to the counts of each Z bin, so only the relative numbers in the corrected histogram have physical meaning.

Several obvious features can be easily seen in Fig. 2.

First, there is a very significant peak at Z= 2.9. As we discussed before, a similar feature also exists in the distribution of emission line redshifts at Z= 2. The reality of the peak at Z= 2 in the distribution of emission line redshifts of quasars is doubted by many authors due to the fact that there is a strong selection effect in the identifications of quasars (i.e. preference for identification of Ly-alpha lines). But in the case of absorption lines, the peak appearing in the corrected histogram is a new result and we don't think it also could be interpreted using same selection effect. The similar feature in both distributions of quasars and the PHC may imply that the existence of a peak is a real feature, which reflects an interesting character of the large scale structure in the universe.

Secondly, we find that the number of PHC increases with increasing of redshifts. This result confirms the evidence that the number of lines per unit redshift increases strongly with redshifts Z (Peterson, 1978).

Thirdly, the number of PHC in the range of Z > 2.9 is singnificantly larger than that in the range of Z < 2.7. It means that the PHC is one kind of objects which mostly exist in the early universe. It reminds us that similar things also

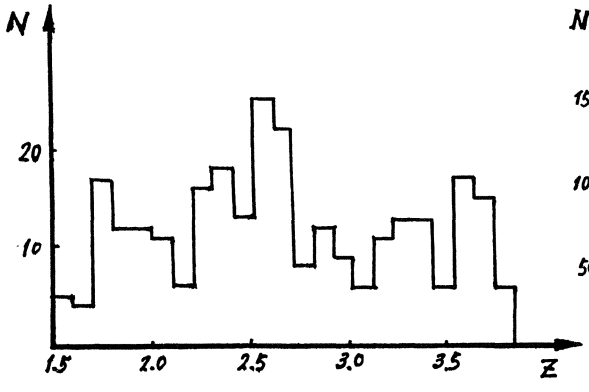


Fig. 1 Uncorrected distribution of redshifts of PHC

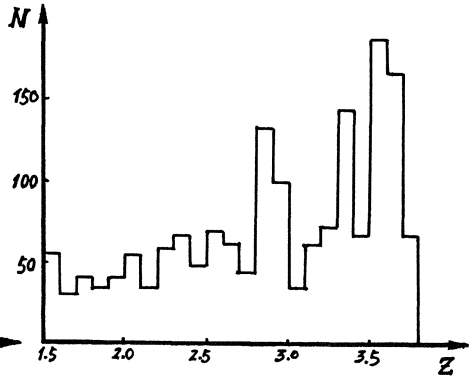


Fig. 2 Corrected distribution of redshifts of PHC

happen for quasars ,that is, the number of quasars also has an obvious change around Z= 2.5, but in opposite direction. (Veron, 1987)

4. Power-Spectrum Analysis

Besides the peak at Z= 2.9, there are also some other peaks in Fig. 2. So a natural question is that does the periodicity also exist in the distribution of redshifts of PHC?

We have found the periodicity component in the distribution of emission line redshifts for quasars, using the power-spectrum analysis which is a very effective method for a distribution of discrete points. (Fang, Chu, et. al. 1982). So, it is interesting to analyse the redshift distribution for PHC using the same method. The spectral power P(n) of PHC redshift distribution against x=Ln(1+z) is given by:

$$P(n)=\frac{1}{N}\left\{\left[\sum_{i=1}^{N}\cos\left(\frac{2\pi nXi}{Xm}\right)\right]^2+\left[\sum_{i=1}^{N}\sin\left(\frac{2\pi nXi}{Xm}\right)\right]^2\right\}$$

where N is the number of PHC, Xi=Ln(1+Zi), Zi the redshift for ith PHC, Xm the length of X range and n the wave number. It can be shown that the average of P(n) equals 1 for a random sample. The probability that a given peak in a random spectrum should have a power P > Po is given by S(Po)=exp(-Po).

Fig. 3 shows the power-spectrum for all PHC redshifts. The redshift range is 1.5 < Z < 3.7. We find a periodic component at n= 3, which corresponds to the length of period of 0.2. Although the significance of confidence is not very high, about 83% , but the length of period is just the same as we found in the distribution of emission line redshifts of quasars (0.205)! (Fang, Chu, et. al. 1982)

5. The Peaks in the Distribution of R

The feature in the distribution of PHC, which we disscussed above, should also appear in statistics of other observational quantities of PHC. One of such quantities for absorption lines is R=(1+Zem)/(1+Zab). The important point about the ratio R is that they are not grossly affected by small inhomogeneities in the spectra of the quasars. R is also related to the ejective velocity which has been widely discussed in the study of absorption lines.

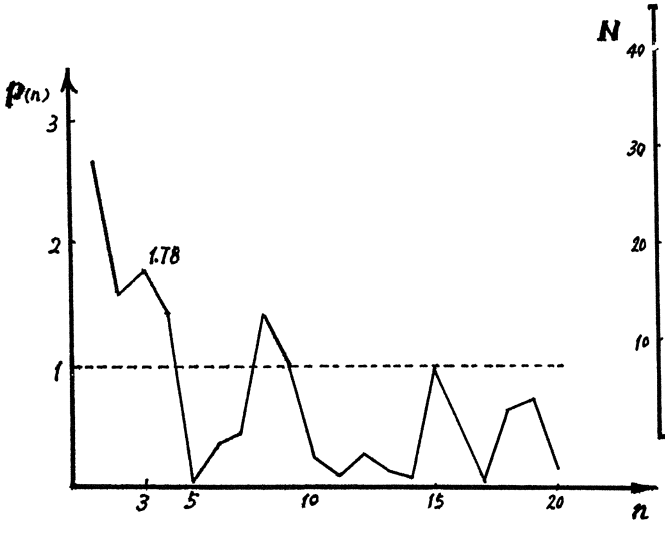


Fig. 3 Power spectrum for redshift distribution of PHC

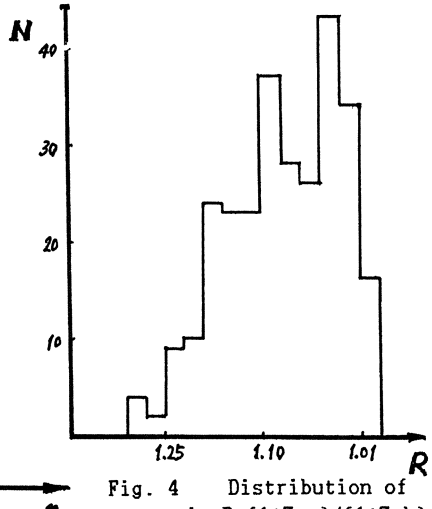


Fig. 4 Distribution of ratio R=(1+Zem)/(1+Zab)

In Fig.4 we show the histogram of R for all quasars in our sample. We can find that several peaks obviously appear in the distribution of R. Such peaks can be interpreted as follows:

If the periodicity is a real cosmological feature, it should exist in both quasars and PHC redshifts. As discussed by Chu, Fang and Liu(1984), for a set of objects located in the same direction, the distribution of redshifts should be periodic and have a series of peaks. The redshift values corresponding to the peaks are given by

$$Ln(1+Z) = An + C$$

Here A is a constant, C denotes inititial phase which is a function of direction of the object and n equals some positive integer.
For emission line redshifts of quasars we have:

$$Ln(1+Zem) = An + C1$$

for PHC we also have :

$$Ln(1+Zab) = An' + C2$$

Since the quasar and the PHC are located on the same line of sight, the initial phases are the same,i.e. C1= C2, so we have:

$$Ln\frac{(1+Zem)}{(1+Zab)} = A\Delta n$$

$$R = \frac{1+Zem}{1+Zab} = e^{A\Delta n}$$

where Δn= 0,1,2,3,.....
This indicates that the histogram of R should have a series of peaks. If we take the value of A as 0.05 we find:

Δn= 1, R= 1.05

Δn= 2, R= 1.10

Δn= 3, R= 1.16

These results are consistent with the feature in Fig.4. It is interesting to point out that a similar result has been found by Chu, Fang and Liu(1984) in the distribution of relative velocity , using a large sample of metal absorption lines in quasars from the Hewitt and Burbidge Catalog(1980).

6. Anti-association of PHC with Galaxies

Sargent et al (1980) suggested that the primordial hydrogen clouds which contribute to the Ly-alpha forest in quasar's spectra are not associated with galaxies. Due to the fact that some of the Ly-alpha lines in our quasars' sample are identified as belonging to definite heavy-element systems which are formed by intervening galaxies at the distance indicated by the redshifts (see Table 1), so it is possible to study the association between the PHC and galaxies using the cross-correlation function.

The statistical method of cross-correlation function has been used to study the possible association between quasars and galaxies by Chu, Zhu, Burbidge and Hewitt (1984). The basic idea is as follows: the cross-correlation function Wgh(s) is defined as the probability of finding galaxy-PHC pair, abve those expected from random coincidence, at the separation of s:

$$Wgh(s) = \frac{\bar{N}gh(\Delta s)}{\bar{N}hh(\Delta s)} - 1$$

where the $\bar{N}gh$ is the mean number of galaxies-PHC pairs at distance of s and $\bar{N}hh$ is the mean number of PHC at distance of s. Following Sargent et. al. (1982), the distance s for two objects with redshift Zi and Zj is calculated by:

$$S = \Delta Z(1+Z)^{-3/2}$$

where: $\Delta Z = |Zi - Zj|$

$$Z = 0.5(Zi+Zj)$$

In Fig.5 we show the cross-correlation function for six quasars in Table 1. We use the distance resolution S=0.004 and for all s<0.06. We find that all the cross-correlation function Wgh(s) are less than 1! Some of them are even negative value. These results provide another evidence that the PHC are not associated with galaxies.

In summary, from several statistical analyses of the distributions of redshifts for Ly-alpha absorption lines in quasars, we find very similar results as we found in the distribution of emission line redshifts of quasars, i.e. the peaks and periodicity distribution of the redshifts. Moreover, we find the same length of the periodicity (0.2 for X=ln(1+Z)) for both Ly-alpha absorption line and emission line redshifts. It seems very likely that the periodicity is one of the features of large scale structure in the universe.

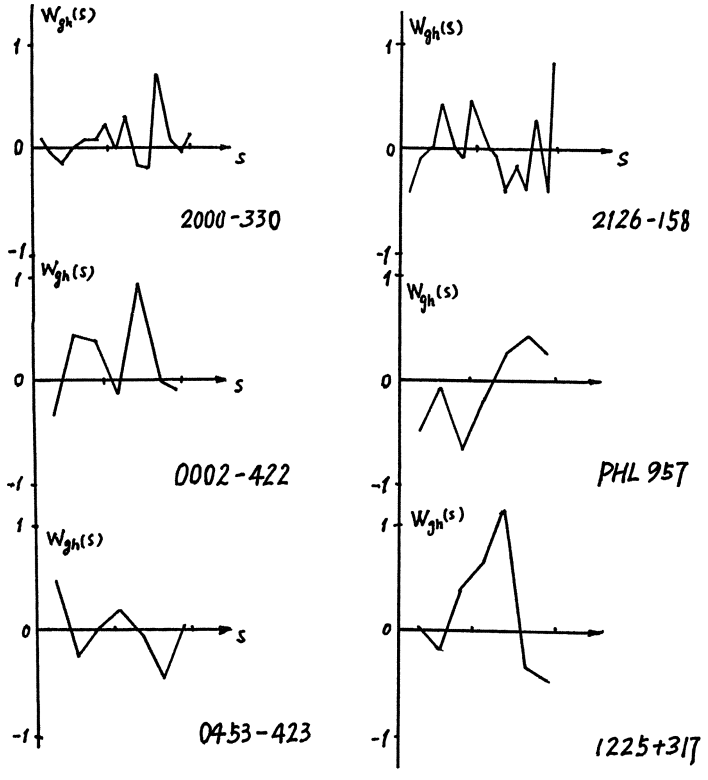


Fig. 5 The cross-correlation function for PHC and metal-line systems

ACKNOWLEDGMENT

The authors are particularly indebted to Prof. Dr. W. Priester and Prof. Dr. W. Kundt for their kind hospitality during their visit at the Institute for Astrophysics, University Bonn. Y. Chu would like to thank Alexander von Humboldt Fundation for financial support.
Both authors would like thank Dr. G. Boerner for his kind invitation to attend the Second Workshop on High Energy Astrophysics between Max-Planck Society and Chinese Academy of Sciences.

REFERENCE

Burbidge,G., 1968, Ap. J. letter, vol.154, L41.
Burbidge. G. R., O'Dell, S. L., 1972, Ap. J. vol. 178, 583.
Chu, Y., Zhu, X., Burbidge, G., and Hewitt, A., 1984, A. Ap. 138, 408.
Chu, Y., Fang, L.Z., and Liu, Y., 1984, Astrophys. letters, 24, 95.
Cowan, C. L., 1969, Nature, vol. 224, 655.
Fang, L.-Z., Chu, Y., Liu,Y., Cao, Ch., 1982, Astron. Astrophys. vol.106, 287.
Karlsson, K. G., 1971, A. Ap. vol.13, 333.
Kundt, W. and Krause, M., 1985, A. Ap. vol.142, 150.
Murdoch,H.S., Hunstead,R.W., Pettini,M., Blades,J.C. 1986, Ap.J. vol.309, 19.
Peterson, B. A. 1978, IAU Symp. 79, p.389.
Plagemann. S. H., Feldman, P. A., Gribbin, J. R., 1969, Nature, vol.224, 875.
Sargent, W. L. W., Young, P. J., Boksenberg, A., and Tytler, D., 1980, Ap. J. Suppl., 42, 41.
Sargent, W. L. W., Young, P. J., Boksenberg, A., 1982, Ap. J. 252, 54.
Veron,P. Astron. Astrophys. vol.170, 37.
Wills, D., Ricklefs, L., 1976, MNRAS, 175, 81p.

Clustering of Quasars and Its Evolution

Xingfen Zhu and Yaoquan Chu

Institute for Astrophysics, University of Bonn, Fed. Rep. of Germany
Center for Astrophysics, University of Science and Technology of China,
Hefei, People's Republic of China

The discovery of quasars, about 25 years ago, was one of the most exciting events in the history of recent astronomy. Despite the slow growth in our understanding of their physical nature, these objects, which have the highest known redshifts, still provide the best available probe of the most remote observable region of the universe. In recent years the investigation of clustering of quasars has become one of the most interesting problems in observational cosmology. This is mainly due to the development in observational technique and the theoretical understanding of the formation of galaxies and the evolution of the density fluctuation in the early universe.

Thanks to the modern high-efficient detectors, it is now possible to construct three-dimensional maps of large regions in the universe. A series of systematic measurements of galaxy redshifts have exposed some very interesting features in distribution of the matter in universe. Holes, voids, sheets and filaments are apparent in the distribution of galaxies. Very Large scale voids, about 60 Mpc and simply limited by the size of the survey, have been found in the CfA redshift surveys (De Lapparent, et. al. 1986). The supercluster-supercluster correlation function show that the structure on scale up to 150 Mpc maybe also exist (Bahcall and Burgett, 1986). Also the statistical results exhibite that the clusters of galaxies are more strongly correlated than the galaxies and the very largest scale structures more strongly correlated than clusters.(Bahcall and Soneira, 1983). In order to explore the structure of the universe on much large scale, it is reasonable to expect that the study of space distribution of quasars will give us more new information.

On the theoretical side, many important advances have been achieved recently. Several scenarios of clustering have been proposed. Many new ideas, such as inflational cosmology, cosmic strings, explosion...,have been introduced to interprete the observational results. In order to distinguish which scenario is the right one, we need information of the larger scale structure at different cosmological time, i.e., at different redshifts. However , up to now most of the information of the structure of the universe is limited to a very tiny fraction of observable universe, it comes from the study of distribution of galaxies, all of which with redshifts $Z < 1$. Only quasars provide us a chance to see the state of clustering at redshifts intermediate between those of the microwave background,which is very smooth, and the present, which is very clumpy on the scale up to at least those of superclusters. So the study of the distribution of quasars could give us some information on the evolution of clustering.

The clustering of quasars is also very important for understanding the nature of quasars itself. The physical quasar pairs or quasar groups would provide diagnostics of processes and insight into phenomena of quasars, e.g. the role of gravitational interaction in triggering and fueling of galactic nuclear activity. On the other hand, the possibility that there exists association of quasars with different redshifts would support the hypothesis of noncosmological redshift.

In this paper we would like to discuss a series of our statistical research on the distribution of quasars. The aim of our study mainly concerns to following subjects: to search for large or super-large scale structure; the evolution of the clustering and possible remains of physical processes at early stage of the universe.

2) Distribution of Quasars on the Sky

Progress in survey for quasars at frequencies from radio to X-ray have been rapidly made in last 25 years. Up to now , total number of confirmed quasars is more than 3000 and quasar candidates are about 10000.(Barbieri and chu, 1986, Hewitt and Burbidge, 1987). From these large samples of quasars, now it is valuable to carry out statistical analyses of distribution of quasars.

At first, we show a plot of all quasars from Hewitt and Burbidge catalogue (1987) on the sky using an Aitoff projection and supergalactic coordinates. From Fig.1 it can be seen that there is a wide variation in the density of identified quasars. Of course this is not a real clustering effect, but it illustrates that many of the surveys, particularly those using the objective-prism and grism technique, are confined to very small areas in which very large number of quasars have been identified.

Owing to that the quasars' surface density is very lower campared with that of galaxies, it is difficult to detect and to measure reliable clustering of quasars. Moreover, the different selection effects in quasar surveys make the statistical analyses complicate. There are mainly two approaches in the study of quasars clustering: with homogeneous samples, which are generally small (from slitless spectrum survey) or contain some fraction of non-quasars(from UVX survey);and with large quasar catalogues (for example, Hewitt and Burbidge catalogue and Veron-Veron catalogue), which are heterogeneous and much more attention have to be paid to the selection effects.

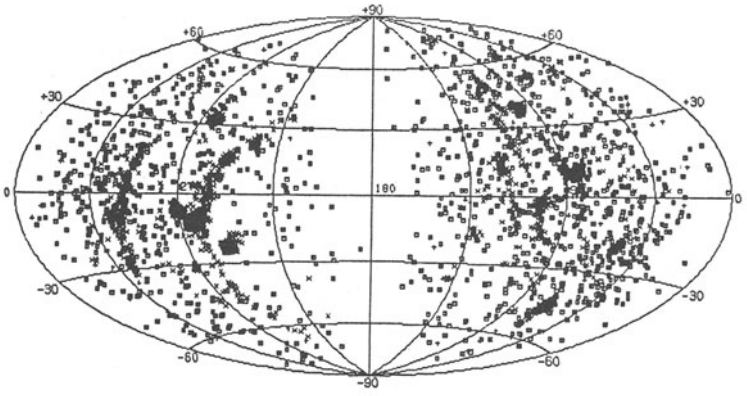


Fig. 1: Sky distribution of quasars from Hewitt and Burbidge catalogue (1987)

3) Quasar Pairs

It has been suggested by Dr. Shaver (1984) that clustering is expected to show up amongst pairs or groups of quasars which are close to each other in both redshift and position on the sky. By contrast, pairs which have large differences in either redshift or position are expected to be physically unassociated with each other, and should exhibit no clustering. The search for clustering can then be done by comparing these two groups. This has great advantage that selection effects (such as survey areas, redshift ranges, limiting magnitudes, sensitivity variations, edge effects, etc.) should cancel out, making possible to use of large and inhomogeneous quasars' catalogue for this purpose. With the Veron Catalogue (1984), possible clustering of quasars has been revealed by Shaver.

Recently we apply the same method to the Hewitt and Burbidge Catalogue(1987), which contain 3594 quasars with published redshifts(70 quasars in this catalogue without accurate positions have been excluded from our analyses.) The angle distance θ on the sky between two quasars is calculated by:

$$\cos\theta = \cos\delta_1 \cos\delta_2 \cos(\alpha_1 - \alpha_2) + \sin\delta_1 \sin\delta_2$$

where α_1, α_2 are right ascensions and δ_1, δ_2 are declinations.
The projected separation of quasar pair, ds, is calculated under the condition of a standard cosmological model with Qo=0.

$$ds=R(t)*r*\theta$$

$$R(to)=c/Ho=3000Mpc,\ (\ Ho=100\ km/(s*Mpc)\ )$$

$$r=Z*(1+0.5*Z)/(1+Z)$$

here $Z=0.5*(Z_1+Z_2)$ is the average of two redshifts.

In Fig. 2 we plot the ratio of the number of quasar pairs with small redshift difference (V < 2500 km/s) to the number with large redshift difference, against projected linear separation. It clearly shows an excess of quasar pairs with small separation and redshift difference. If there were no clustering, the ratio in Fig.2 should be constant as a function of separation difference.
The quasar pairs with small redshift difference from the Hewwit-Burbidge catalogue are listed in Table 1.

Some of the quasar pairs in Table 1 were previously identified as gravitatinal lenses, these pairs are irrelevant for clustering. The problem of distinguishing between gravitational lens and physical pair cannot easily be solved in individual case, particularly in the case that two quasars are seperated only few arcseconds. However, in our sample there are only few quasar pairs in which the gravitational lens interpretation is now reasonably established ,namely, 0957+561, 2016+112. Other three quasar pairs 1146+111, 1634+267 and 2345+006, which have no obvious visible lens objects, have been claimed to be physical pairs with small seperation (Bahcall, et. al. 1986). So the contamination due to lenses seems not likely to be statistically serious in our analyses.

Table 1. Quasar pairs with small separation and redshift difference

QSO	Other name	Z	θ (arc sec)	ds (Mpc)	V (Km/s)	Comments
0000-022		2.20	92	1.8	1869	Grens survey
0000-022		2.22				
0103-294		2.18	122	2.5	1892	Objective prism survey
0103-294		2.16				
0107-025	QS010	0.956	77	0.8	614	UVX object
0107-025	QS0 9	0.952				5 QSOs near NGC450
0120-378		2.15	191	4.0	1898	Objective prism survey
0120-378		2.17				
0131-401		(1.67)	71	1.2	2255	Objective prism survey
0131-401		1.65				
0307-195A	UM680	2.144	58	1.2	2106	Objective prism survey
0307-195B	UM681	2.122				
0952-698	HOAG 1	2.048	121	2.4	590	3 of 6 QSOs found on grism plate of M82 field
0952-698	HOAG 2	2.054				
0953-698	HOAG 1	2.048	214	4.2	788	
0953-698	HOAG 3	2.040				
0953-698	HOAG 2	2.052	128	2.5	1378	
0953-698	HOAG 3	2.040				

Table 1 (continued)

QSO	Other name	Z	θ (arc sec)	ds (Mpc)	V (Km/s)	Comments
0957+561A		1.405	6	0.1	0	Gravitational lens
0957+561B		1.405				
1146+111	C	1.01	157	1.7	0	Close group of 9
1146+110	A	1.01				QSOs from objective prism survey
1254+370	B 142	(0.28)	186	0.7	986	UVX object
1255+370	4C36.22	0.284				Radio source
1258+285	US136	1.355	272	3.8	2284	UVX object
1258+286	5C4.127	1.373				Radio,X-ray source
1259+357		2.03	252	4.9	988	Grens survey
1259+357	BF161	2.04				
1623+268	KP76	2.49	147	3.4	2397	Objective prism survey
1623+268	KP77	2.518				
1634+267A	KP86	1.961	4	0.1	0	Objective prism survey
1634+267B	KP87	1.961				Gravitational lens ?
1641+399	3C345	0.595	506	3.6	188	Radio,X-ray source
1641+399		0.594				Objective prism survey
2016+112	MG B	3.273	3	0.1	0	Gravitational lens ?
2016+112	MG A	3.273				
2203-187	F13.18	0.627	412	3.0	1664	Objective prism survey
2203-188	PKS	0.618				Radio source
2237+003		(2.2)	213	4.5	0	Objective prism survey
2238+004		(2.2)				
2340+011		2.35	206	4.6	896	Objective prism survey
2340+010		2.34				
2345+006	B	2.147	7	0.2	476	Gravitational lens ?
2345+006	A	2.152				

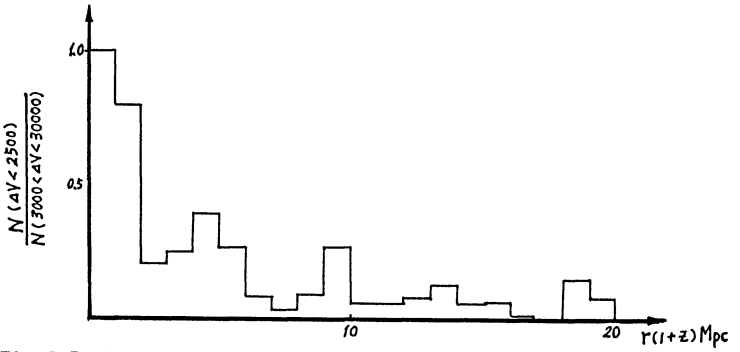


Fig.2 Ratio of the number of quasar pairs with small redshift difference to the number with large redshift difference, against projected linear separation

4) Clustering of Quasars from Homogenous Survey

Work on the quasar distribution is often hampered by the lack of a complete and unbiased sample. Most of homogeneous surveys for quasars list only bright quasars, or cover only very small areas.

In recent years a large number of quasars' surveys have been published. Slitless spectrum surveys remain one of the most efficient ways of finding quasars in large numbers. More than 10000 quasar canditates have been identified. In complement, the ultra- violent excess (UVX) method is best suited for quasars with smaller redshifts. With a large number of quasars and quasar canditates, many statistical analyses of quasar clustering have been done in recent few years. Thanks to advances and widespread availability of computers, the distribution on one, two and three dimensions of quasars and quasars' candidates have been investigated. The most popular method Correlation Function (CF), which are widely used in analyses of cluster and supercluster of galaxies, makes fairly easy to compare the results between quasar and galaxy distributions. Another powerful statistical method to detect week clustering is Nearest Neighbor Test (NNT). The basic idea of NNT is to compare the mean nearest neighbor distance between objects with that expected from random distributed sample. Beside that, power-spectrum is fourier or spherical harmonic transform of correlation fuction; the binning analysis is also known as the cluster cell method; the percolation test, which were suggested by Zeldovich and his colleagues, seems to be sensitive to the clustering pattern.

In the early stage, the statistical results from different researcher are sometimes not consistent with each other. Some of them are even contradictory. For example, with the same quasar sample in the CTIO survey, Arp (1980) has claimed that there is some evidence of quasar grouping, while Osmer(1981) and Webster (1982) have found that there is no clustering of quasars. Chu and Zhu (1983) analysed the three-dimension distribution of quasars in two $5^{\circ} \times 5^{\circ}$ survey done by Savage and Bolton (1979), and found weak clustering in one field (02h, -50°) but no clustering in another field(22h,-18°).

In the last few years, very rapid progress has been made in the optical surveys for quasars. With the automatic measuring machines (AQD in Edinberg and APM in Cambridge) and fibre-optic spectrograph (FOCAP), much larger and deeper samples of quasars are now being produced, yet so far the results for clustring of quasars are still contradictory: Boyle et.al. (1986) and Shanks et.al. (1987) find evidence for clustering in their UVX sample. He, Chu et. al. (1986), Kundth and Sargent (1986), Crampton et. al. (1987) and Clowes et. al. find no clustering amongest quasars from their optical-prism or grens survey.

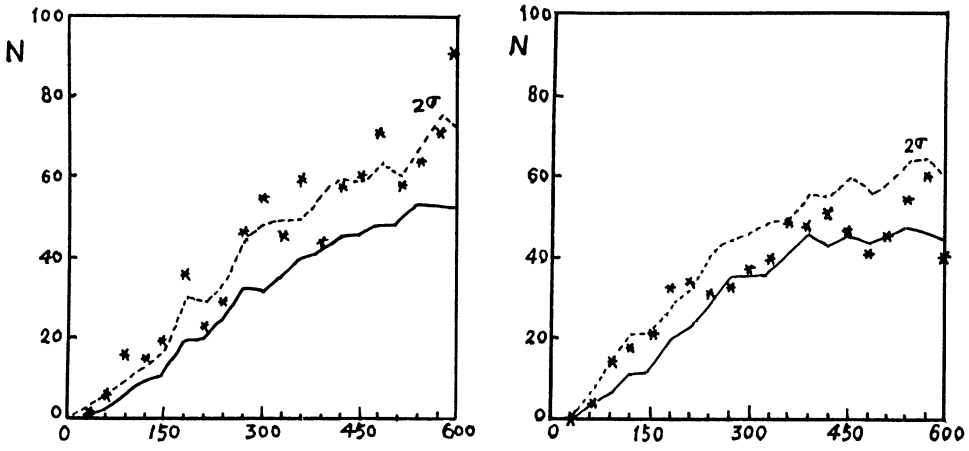


Fig. 3, The correlation function for quasars in the fields (02h,-50°) (left) and (22h, -18°) (right). Stars: observed values. The soild line for the mean distribution and the dashed line for the upper 2σ envelope of the population are derived from Monte Carlo simulations.

The differences among these results are due mainly to the samples,i.e., the numbers of quasars in these early surveys are not large enough to do statistical analysis with high signifecance. Meanwhile the criteria of clustering in different statistical methods may not be equivalent. Moreover different selection effects may play important role in different surveys,for example:

a), Variation of limiting magnitudes, it would cause the change of surface density in different samples.

b), Vignetted effect, the surface density even changes in one field, more quasars would be found in the central area.

c), Selection of emission lines, Slitless spectroscopy particularly efficient for quasars with redshifts about 2.

d), Systematically error in the measurement of redshifts from objective prism plate.

e), Misidentification of emission line, i.e. assummed Ly-alpha line could really be C IV or Mg II.

f), Non-quasar candidates, for quasars' candidates it is necessary to make follow-up spectra with large telescope in order to confirm and obtain accurate redshifts.

In spite of the differences in different statistical researches, a conclusion seems to already be acceptable: the clustering of quasars is rather week, at least weeker than that of galaxies.

5) Evolution of Quasars' Clustering

After carefully campared all the previous statistical results, we find a very interesting fact that all results from the samples containing UVX quasars(or quasar candidates) always show stronger clustering than others (see, Chu and Fang 1986). As we know that the redshifts of quasars found by objective prism and grism are generally higher than that of quasars found by UVX methods. Therefore the stronger clustering in UVX samples may imply that the strength of quasar clustering depends on the redshifts of quasars, namely, the clustering is evolving: the larger the redshifts, the weaker the clustering.

For showing the evolution of clustering, we should do the clustering analyses for quasars in different redshifts range. Fang, Chu and Zhu (1985) analysed the Savage-Bolton sample, which consists of two classes of quasars identified by objective prism technique and UVB two colour method and whose redshift spreads on more broad region than that of other samples at that time. When we began to study the evolution of quasars' clustering, it was the most convenient sample to do the comparision between quasars with larger and smaller redshifts.

We divided whole sample into two groups : Z>2 and Z<2, and applied the nearest neighbor test. In a standard cosmological model, if two quasars i and j have redshifts Zi and Zj and the angle distance betweem them is , the linear distance in space is given by:

$$D=R(t)*\left\{ r_j^2 \sin^2\theta + \left[r_i(1+r_j^2)^{1/2} - r_j\cos\theta(1+r_i^2)^{1/2} \right]^2 \right\}^{1/2}$$

where R(t) is the cosmic scale factor, at present time to, R(to)=c/Ho= 3000Mpc, Ho=100 km/s*Mpc, r_i and r_j are the dimensionless comoving radial coordinates of quasars i and j respectively. The main results are given in Table 2.

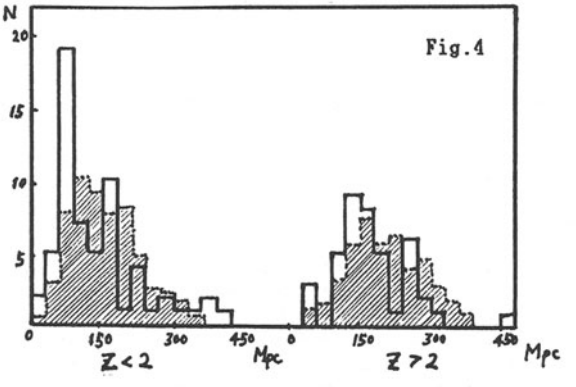
In Table 2, N is the number of quasars, $\langle D_< \rangle$ denotes the sample's mean of nearest neighbor separations, $D^*_<$ and $\hat{\sigma}$ are the mean and standard deviation from Monte Carlo sample respectively. In order to estimate the statistical sigificance, statistic δ is defined as:

$$\delta = N^{1/2} \frac{\langle D_< \rangle - D^*_<}{\hat{\sigma}}$$

For a random sample the distribution of δ is asymptotically normal with mean 0 and variance being 1. Therefore 1 - P(δ) is the probability of clustering to be found in the sample. Table 2 show apparent clustering at 95% significant level for quasars with Z<2 in both fields, and there is no evidence

Table 2: Nearest neighbor test for Savage-Bolton sample

Redshift	N	Quasar data $\langle D_\zeta \rangle$ Mpc	Monte Carlo data D^*_ζ Mpc	$\hat{\sigma}$	δ	1-P(δ)
		(02h ,-50°) field				
Z< 2	62	141.7	159.0	79.6	-1.72	96%
Z> 2	48	201.0	205.9	83.2	-0.40	66%
		(22h ,-18°) field				
Z< 2	57	146.7	165.8	77.9	-1.84	97%
Z> 2	26	207.1	193.0	75.9	> 0	---



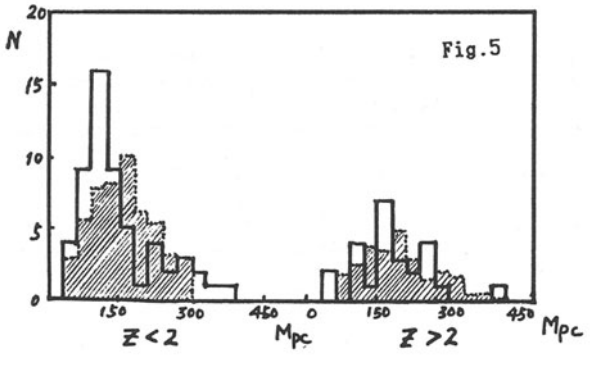


Fig.4 The nearest neighbor distribution for quasars with Z < 2 and Z > 2 from field (02h, -50°)

Fig.5 The nearest neighbor distribution for quasars with Z < 2 and Z > 2 from field (22h, -18°)

of clustering for quasars with Z>2 at all. These results can be more clearly seen in Fig. 4 and Fig. 5, in which the distribution of the nearest neighbor distance for each field are ploted. The observed Z<2 distribution (soild Line) deviate obviously from the Monte Carlo samples (dashed line) on the scale of about 100 Mpc. It means that the distribution of Z<2 quasars does have about 100 Mpc clustering. The Z>2 quasars distribution does not show the difference from that of random sample, namely, there is not distinguishable inhomogeneity.

More evidences for the evolution of quasar clustering have been found recently. Iovino and Shaver (1987) analyse three deep quasar samples (376 quasars from UVX and Grens survey) and find strong clustering (at 5σ significance level) for quasars with Z <1.5 and no clustering for quasars with Z > 1.5. Using the quasar pairs from Veron catalogue, Kruszewski (1986) also find strong redshift depence of quasars clusterimg

In one word, a common result in these studies on clustering of quasars from different samples is that the clustering exists in the distribution of Z< 2 quasars, but not of Z> 2 quasars.

ACKNOWLEDGMENT

The authors are greatly indebted to Prof. Dr. W. Priester and Prof. Dr. W. Kundt for their hospitality during their visit at the Institute for Astrophysics, University Bonn.

Y. Chu would like to thank Alexander von Humbolt Fundation for the research fellowship.

Both authors would like thank to Dr. G. Boerner for his kind invitation to attend the Second Workshop on High Energy Astrophysics between Max-Planck Society and Chinese Academy of Sciences.

REFERENCES

Bahcall, J. N., Bahcall, N. A., and Schneider, D. P., 1986, Nature, vol.323, 515.
Bahcall, N. A. and Burgett, W. S., 1986, Ap. J. vol. 300, L35.
Bahcall, N. A. and Soneira, R. M., 1983, Ap. J. vol. 270, 20.
Barbieri, C., Chu, Y., et. al., 1986, "The Asiago Catalogue of QSOs edition 1985", in "Quasars", IAU Symp. 119, ed. Swarup, G. ad Kapahi, V. K.
Burbidge, G. R., Narlikar, J. V. and Hewitt, A., 1985, Nature, vol.317, 413.
Boyle, B. J., et. al., 1986, in "Observational Cosmology", IAU Symp. 124, ed. Hewitt, A., et. al., 643.
Chu, Y.,and Zhu, X., 1983, Ap. J., vol.267, 4.
Chu, Y., and Fang, L., 1986,in "Observational Cosmology", IAU Symp. 124, ed. Hewitt, A., et. al.,
Clowes, R. G., Iovino, A. and Shaver, P., 1987, MNRAS, vol.227, 921.
Crampton,D., Cowley, A., P., and Hartwick, F. D. A., 1987, Ap. J., vol.314, 129.
De Lapparent, V., Geller, M. J., Huchra, J. P., 1986, Ap. J., vol.302, L1.
Fang, L., Chu, Y. and Zhu, X., 1985, Ap. S. S., vol. 115, 99.
He, X., Chu, Y., et. al., 1986, Acta Astronomia Sinica, vol.27, No.2, 144.
Iovino, and Shaver, P. A., 1987, in IAU symp. 130, Hungary.
Kunth, D., and Sargent, W. W., 1986, A. J., vol.91, 761.
Kruszewski, A., 1986, Princeton Observatory Preprint.
Osmer,P. S., 1981, Ap. J., vol. 247, 761.
Savage, A., and Bolton, J. G., 1979, MNRAS, vol.188, 99.
Shanks, T., Fong, R., et. al. 1987, MNRAS, vol. 227, 739.
Shaver, P. A., 1984, A. Ap., vol.136, L9.
Veron-Cetty, M.-P. and Veron, P., 1987, ESO Scientific Report, No. 5.
Webster, A.,1982, MNRAS, vol.199, 683.

Evolution in Quasar Clustering Properties

A. Kruszewski

Warsaw University Observatory, Al. Ujazdowskie 4,
PL-00-478 Warszawa, Poland and
European Southern Observatory, Karl-Schwarzschild-Str. 2,
D-8046 Garching, Fed.Rep.of Germany

Nine complete samples of quasars are analysed for correlations in positions. The detected correlations show apparent dependency on redshift.

The present results agree very well with those obtained using an incomplete general catalog of quasars (Kruszewski /1/).

1. Introduction

The studies of quasar clustering properties are difficult to carry on because quasars are rare objects usually situated far away from nearest neighbours, so that any pair statistics analysis is suffering from small number of relevant pairs. In addition the bulk of known quasars do not form a sample which could be considered complete in any sense. The early complete samples have been typically only around 100 objects rich. However, the observational data on quasars have been accumulating steadilly and the paper by Oort, Arp and de Ruiter /2/ has indicated that it may be already possible to investigate clustering among quasars. The complete sample which seemed then to be best suited for such work was that published by Osmer and Smith /3/. The three analyses performed on this sample have resulted with different conclusions. Osmer /4/ has found no any correlations, Webster /5/ has found correlations among objects with redhifts smaller than 1.0 but he has attributed it to the existance of a single group of 4 quasars with redshifts 0.37 or 0.36 (see also Arp /6/). Finally Fang, Chu and Zhu /7/ have concluded that the clustering is present there only in the range of redshifts smaller than 2.0.

Another sample used for such analysis was the sample published by Savage and Bolton /8/. Papers by Chu and Zhu /9/, Fang et al. /7/, and Zhou et al. /10/ contain marginal detections of correlations and also suggestion that the correlation is present only for redshifts smaller than 2.0 /7/.

Clowes /11/ and also Knuth and Sargent /12/ have used their new samples and concluded that there is no any correlation in quasar positions.

Thus the work done on complete samples prior to 1987 has generally resulted in marginal, inconclusive and sometimes contradictory results. This has prompted the use of incomplete samples which are more difficult to analyse properly but which should already be sufficiently rich for obtaining statistically significant results. The first step in that direction was made by Shaver /13/ who using the general catalog of quasars (Veron-Cetty and Veron /14/) has concluded that quasars treated together irrespective of redshifts cluster in a similar way like ordinary galaxies do. This discovery have been followed by the analysis of the same catalog divided into 7 subsamples according to redshift with the conclusion that there is a strong dependence of clustering amplitude with the redshift (Kruszewski /1/). The quasars seem to cluster in recent epochs as strongly as clusters of galaxies do, while at redshifts larger than 1.5 there are no detectable correlations in quasar positions.

The prospect of studying quasar clustering has served as an additional stimulus for creating more numerous complete samples of quasars. The work in that direction is being conducted presently at several observatories. One such sample have been completed recently and the quasar data have been published (Crampton et al. /15/). This sample is not sufficient by itself for drawing any definite conclusions about quasar clustering, yet after combining it with other complete samples it can be hoped that the informations contained in such combined material may be already sufficient for testing conclusions reached with the use of the incomplete sample of Veron-Cetty and Veron.

2. The Sample

The samples used for the present analysis have been chosen using the following criteria. First of all, a sample has to be complete in some well defined sense. The definition of completeness may be different for different samples. Next, all the objects should be confirmed and have its redshift determined with help of slit spectra. The sample should not be composed of less than 25 objects. And finally, a particular sample should have sufficiently wide sky distribution. Samples limited to a single Schmidt plate or smaller have not been considered.

The following 9 samples fulfilling the above criteria have been known to me at the time when this research were started. The first is the sample of Crampton et al. /15/. The sample objects have been identified in that case with help of the CFH telescope by means of grism technique. Eight fields with complete identifications have been used. The second sample is an objective prism sample of Osmer and Smith /3/. It covers a 9h long and 5° wide strip of the southern sky. The third is the Cerro Tololo 4-meter grism survey of Osmer /16/. It is composed of 17 prime focus fields in 7 separate regions superposed on the objective prism survey area. These three samples contain most of informations about short range correlations in quasar positions. The remaining samples are more uniformly distributed on the sky and therefore they are usefull mainly for investigating of clustering on large and very large scales. The Palomar-Green survey (Schmidt and Green /17/) covers high galactic latitude regions of the northern sky. The objects have been identified by presence of ultraviolet excess down to apparent magnitude 15.7. The 3C quasars listed in Hewitt and Burbidge /18/ catalog have formed the fifth sample. Four more samples are taken from the list of radio quasars published by Wills and Lynds /19/, there are two 4C and two PKS samples with different limiting flux.

Some objects belong to more than one sample. The method of analysis requires that objects are divided between samples so that each object should be labeled as a member of only one sample. Therefore in case of multiple membership an object is labeled as a member of the least numerous of relevant samples. The total number of objects is 629.

3. The Method

In order to infer about correlation in quasar positions one has to calculate the observed distribution of pair separations and compare it with the expected distribution. In an ideal case, when the characteristics of true sky and redshift distributions are known for each of considered samples, it is possible to simulate uncorrelated space distribution by assigning randomly redshifts and positions for all the objects. It is difficult to do this for real cases. The redshift distribution of objects is subjected to strong redshift dependent selection effects which depend strongly on the method used for identification of objects. The sky distribution is also affected by factors like telescope optics vignetting, absorption by

interstellar matter in our galaxy, or seeing and emulsion dependent variations in the plate limiting magnitude. One can try to determine empirically various selection factors and approximate them by some interpolation formulas, but especially in case of non numerous samples it is difficult to say what features in an apparent distribution are due to selections and where the real clustering comes in. As a consequence of approximations made in an attempt to simulate the selection process an apparent correlation may be underestimated or overestimated thus influencing the final conclusions.

Therefore it was decided to play it safe and make no assumptions about selection effects except that within a single sample the selections in redshift and sky distributions are mutually independent. The analysis is performed by means of reshuffling technique. The reshuffling is done for each sample separately. The sets of redshift values and object positions in a particular sample are kept unchanged. Only the assignment of particular redshift values to specific positions is changed randomly by reshuffling redshift values. The pair statistics is performed for all 9 samples together. Only the reshuffling is done separately for each of the samples. The reshuffling was repeated 100 times and the average of as many pair statistics was taken as a reference.

Such procedure removes not only selection effects but also real correlations existing in the projected distribution of the objects on the sky and in the projected redshift distribution. Therefore the obtained values of the correlation function will be systematically smaller. Such an approach have been chosen in order to be sure that there is no any effect that can produce spurious correlation.

The formula for calculating comoving separations between any two points in the expanding Universe was used after Osmer /4/. The calculations have been done only for the Einstein - de Sitter flat cosmological model. The Hubble constant was assumed equal to 100 km/s/Mpc.

4. The Results

The quasar pairs have been divided into redshift classes according to mean value of pair redshifts. The pair statistics and the evaluation of correlation function is then performed for each of classes separately.

Figure 1 shows the resulting correlation functions in case of division into three redshift classes. Bordering values of redshift

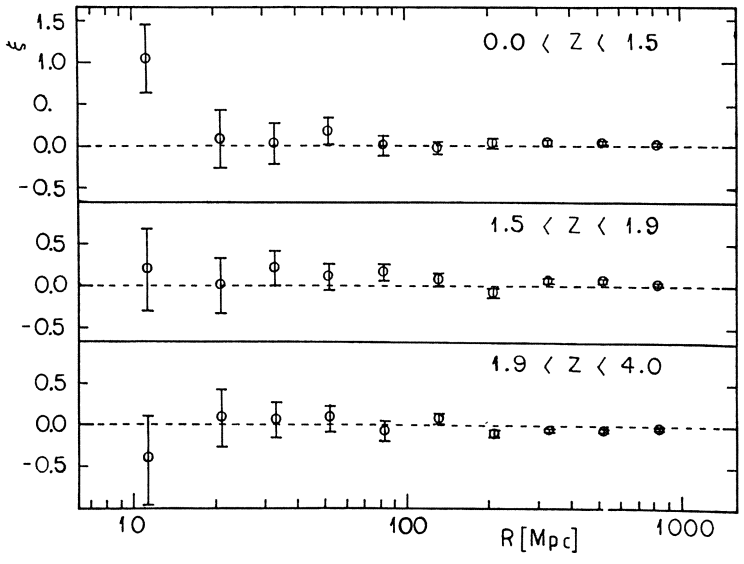


Figure 1. Quasar correlation functions obtained for three redshift ranges using 9 complete samples of quasars

have been chosen so that there is approximatele equal number of pairs and therefore comparable accuracy for each of classes. We can see that there is an evidence on about 2-sigma (or 95%) significance level for the existance of correlation at small redshifts and that the correlations are undetectable for larger redshifts. This evidence is limited to pair separations on the order of 10 Mpc. Fig. 1 sets also tight limits on possible correlations for larger separations.

It is interesting to see how the results of the analysis performed on the complete samples compare with similar results obtained with help of a large incomplete sample. Such comparison is presented in Fig. 2. The data obtained from 9 complete samples divided into five redshift classes is compared there with results of the analysis of the incomplete sample (Kruszewski /1/). The plotted values represent correlation function averaged over the range of separations between 0 and 16 Mpc. We can see that there is perfect agreement between two sets of data, supporting both the statistical significance of the claimed evolutionary effect and the correctness of the treatment applied to the incomplete sample.

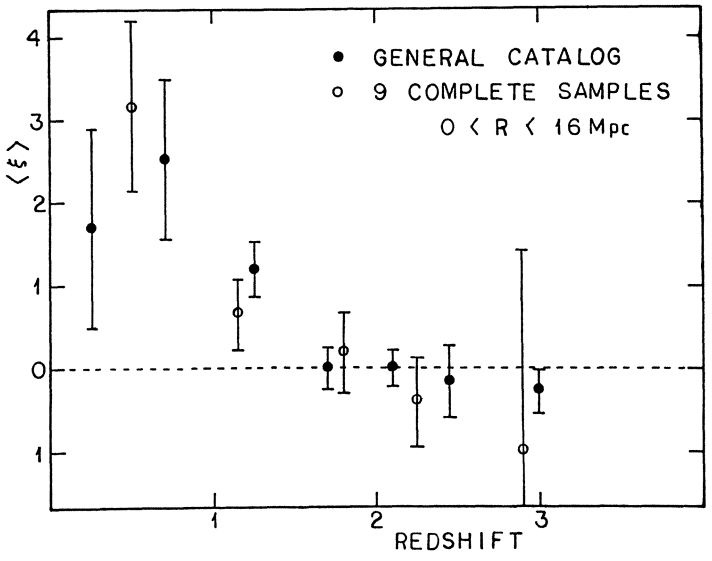


Figure 2. Comparison of average quasar correlation function for separations up to 16 Mpc as a function of redshift basing on 9 complete samples (open circles) and on incomplete general catalog of quasars (filled circles)

5. Conclusions

Recently some additional results have been obtained basing mainly on new samples of quasars. Shanks et al. /20/ have performed an investigation of quasar clustering basing on a new, unpublished redshift survey with results which agree well with that of **Sha**ver /13/. Clowes, Iovino and Shaver /21/ analysing another unpublished sample conclude that there is no detectable correlations for redshifts larger than 2.0. Iovino and Shaver /22/ have analysed the sample of Crampton et al. /15/ together with two new unpublished samples confirming the presence of clustering and suggesting presence of evolution. Shaver /23/ has summarized all these works including the present one and also his new analysis of the third edition of Veron-Cetty and Veron /24/ catalog concluding that quasar clustering is established at a high level of significance, and that there is strong evidence for its evolution.

Thus it is clear that the available observational data is already sufficient for analysing the evolution in quasar clustering. The influx of new data should help to obtain better accuracy and higher significance of results. It should also help in taking into account the possible effects of absolute magnitude.

More numerous data is also needed for analysing clustering at larger separations because the quasar clustering are more likely in that case to reflect the distribution of matter. A prospect for detection of transition to negative correlations at some large value of separation is very atractive because it would provide an excellent opportunity for cosmological tests. Negative values visible in Fig. 1 around 200 Mpc are not yet statistically significant. Clearly more numerous data is needed.

Another feature which is visible in Fig. 2 but is not statistically significant is the presence of negative correlation values for redshifts larger than 2.2. The correlation values presented in Fig.2 are averaged over separation range 0 - 16 Mpc, therefore the negative sign would imply that in the early Universe there is a deficit of quasars (and possibly also other objects) around a bright quasar in regions about 30 Mpc in diameter. This size is strikingly similar to the present day sizes of voids. A seemingly related feature is observed in the distribution of L_α clouds, which appear less frequently in close neibourghood of the quasar (Carswell et al. /25/, Carswell et al. /26/, Murdoch et al. /27/, Bajtlik et al. /28/). Therefore it is very interesting to check if the negative correlations seen in Fig. 2 are real.

Acknowledgments

I am indebted to Peter Shaver and Angela Iovino for communicating their results prior to publication and for helpful discussions. This research has been partly supported by grant RR I.11.

References

1. A. Kruszewski: submitted to Ap. J. (1986)
2. J.H. Oort, H. Arp, H. de Ruiter: Astr. Ap. 95, 7 (1981)
3. P.S. Osmer, M.G. Smith: Ap. J. Suppl. 42, 333 (1980)
4. P.S. Osmer: Ap. J. 247, 762 (1981)
5. A. Webster: M.N.R.A.S. 199, 683 (1982)
6. H. Arp: Ap. J. 239, 463 (1980)
7. L.Z. Fang, Y.Q. Chu, X.F. Zhu: Astrophys. and Space Sc. 115, 99 (1985)
8. A. Savage, J.G. Bolton: M.N.R.A.S. 188, 599 (1979)
9. Y. Chu, X. Zhu: Ap. J. 267, 4 (1983)
10. Y.Y. Zhou, D.P. Fang, Z.G. Deng, X.T. He: Ap. J. 311, 578 (1986)

11. R.G. Cloves: M.N.R.A.S. 218, 139 (1986)
12. D. Knuth, W.L.W. Sargent: A.J. 91, 761 (1986)
13. P.A. Shaver: Astr. Ap. 136, L9 (1984)
14. M.-P. Veron-Cetty, P. Veron: ESO Scientific Report, No. 1 (1984)
15. D. Crampton, A.P. Cowley, F.D.A. Hartwick: Ap. J. 314, 129 (1987)
16. P.S. Osmer: Ap. J. Suppl. 42, 523 (1980)
17. M. Schmidt, R.F. Green: Ap. J. 269, 352 (1983)
18. A. Hewitt, G. Burbidge: Ap. J. Suppl. 63, 1 (1987)
19. D. Wills, R. Lynds: Ap. J. Suppl. 36, 317 (1978)
20. T. Shanks, R. Fong, B.J. Boyle, B.A. Peterson: M.N.R.A.S. 227, 739 (1987)
21. R.G. Clowes, A. Iovino, P.A. Shaver: M.N.R.A.S. in press, (1987)
22. A. Iovino, P.A. Shaver: In Evolution of Large Scale Structures in the Universe, IAU Symposium No. 130, Balatonfured (1987)
23. P.A. Shaver: In Evolution of Large Scale Structures in the Universe, IAU Symposium No. 130, Balatonfured (1987)
24. M.-P. Veron-Cetty, P. Veron: ESO Scientific Report No. 5 (1987)
25. R.F. Carswell, J.A.J. Whelan, M.G. Smith, A. Boksenberg, D. Tytler: M.N.R.A.S. 198, 91 (1982)
26. R.F. Carswell, D.C. Morton, M.G. Smith, A.N. Stockton, D.A. Turnshek, R.J. Weyman: Ap. J. 278, 486 (1984)
27. H.S. Murdoch, R.W. Hunstead, M. Pettini, J.C. Blades: Ap. J. 309, 19 (1986)
28. S. Bajtlik, R.C. Duncan, J.P. Ostriker: preprint (1987)

The Observational Status of Super Clusters at High Redshifts

M.-H. Ulrich

European Southern Observatory, Karl-Schwarzschild-Str. 2,
D-8046 Garching, Fed. Rep. of Germany

This field is in a state of flux. The significance of the results on the distribution of objects depends on observations of large and homogeneous samples. The increasing interest in the field and the continual instrumental advances are such that more and better data become available every month. Here, I am presenting the status of the field as of July 1987.

1. Aim

One wants to determine the redshift or the redshift interval at which clustering of galaxies occurred and compare it to the predictions of cold dark matter models, string models and other models.

2. Approaches

2.1. The best approach is to measure the redshifts of galaxies from $z = 1$ to $z = 2$ (larger redshifts are out of reach for the foreseeable future unless there is unexpectedly strong evolution in magnitude and colour), and determine their three-dimensional distribution.

The galaxies at $z = 1$ and probably also at $z = 2$ are already being recorded on deep CCD images taken with existing large telescopes (e.g., TYSON [1]). The problem is to measure their redshift with sufficient accuracy to detect groups, clusters and superclusters, as is done currently at low redshifts, $z \lesssim 0.5$ (e.g. DE LAPPARENT, GELLER and HUCHRA [2]).

The feasibility of the spectrographic observations at large redshifts has been investigated by ULRICH [3] with regard to the scientific programs of the ESO Very Large Telescope Project. It is found that while it is within the reach of the next generation of telescopes to obtain velocities of a large number (i.e. large enough to form statistically significant samples) of galaxies at $z = 1$, the measurement of velocities of galaxies at $z = 2$ will be possible only if the technical capabilities of these new instruments are pushed to their limits and if there is in addition an evolutionary brightening of galaxies by ~ 2 magnitudes.

2.2. The next best approach is to study the distribution of the absorption lines of metals in quasar spectra. The numerous Lyα lines are unsuitable because the nature of the objects producing them is unknown. With the metal absorption lines,

we are rather certain to be dealing with galaxies. Results on the distribution of absorption lines and possible discrepancies from a Poissonian distribution are presented below.

2.3. The third approach is to study the distribution of quasars on the sky and in z and assume that quasars are, at all epochs (regardless of redshifts and absolute luminosities), markers of the distribution of galaxies.

3. Results on the distribution of quasars, and on the distribution of absorption lines in quasar spectra

From analysis of catalogues of quasars and of complete quasar samples, clustering on co-moving scale has been detected for z = 1.5 at the 5 σ level (SHANKS et al. [4]; CHU and FANG [5]). The clustering at $z \leq 1.5$ is found to be twice that at $z > 1.5$, with a significance of 2 σ (Kruszewski, this volume; Iovino and Shaver, preprint).

Several remarkable groupings at $z > 1.5$ are known. We outline here the characteristics of the most interesting ones.

3.1. The triplet of quasars near M82. The three quasars lie in a circle 3.6 arcminutes or 0.8 Mpc diameter and have z = 2.048, 2.054 and 2.040. The significance of having these three quasars so close in the sky and in redshift is highly significant. The probability of chance occurrence is $< 10^{-4}$ (BURBIDGE et al. [6]) and suggests that at least one supercluster existed at z = 2 (see also OORT [7]).

3.2. The field of Tololo 1037-27. The study of this field was initiated by the discovery of the remarkably rich absorption line spectrum of the brightest quasar in the field, Tol 1037-2704 (m = 17.4), by ULRICH and PERRYMAN [8], who suggested the possibility of a super cluster on the line of sight to this quasar. The emission redshift is 2.2 and 5 CIV absorption line systems with rest equivalent width larger than 0.3 Å, are found in the range 1.9 to 2.15.

Then the second brightest quasar, Tol 1038-2712 (z_{em} = 2.33) was observed spectroscopically and it was found that it also had five absorption line systems in the same redshift range. The two quasars are at 17:9 arc minutes apart or 4 Mpc (H = 100, q_o = 0.5). Considering that the average number of absorption lines with rest equivalent width (W_{rest}) larger than 0.3 Å at $z \sim 1.7$ is 1.5 per unit redshift interval, the number of absorption lines seen in each of the two quasars Tol 1037-2704 and 1038-2712 is highly improbable ($\sim 4\times10^{-4}$) (JAKOBSEN et al. [9]).

However, in both quasars the two absorption lines with highest redshifts are broad and complex. Thus perhaps, they are intrinsic to the quasars and represent material that each quasar has ejected, in which case they must not be included in the statistics of absorption lines assumed to represent intervening galaxies. On

the other hand, perhaps, the lines are broad and complex because they are in fact a blend of absorption lines caused by individual galaxies in a cluster.

If the two absorption lines with the highest redshifts in each of the two quasars are caused by ejected gas then there remain 3 absorption lines in each quasar spectrum with $W_{rest} \gtrsim 0.3$ Å in the internal $\Delta z = 0.15$. The probability of finding such a grouping is 10^{-2} in each line of sight. Other quasars in the same field have been observed and five more emission lines and absorption lines have been found in the same redshift range 1.85 to 2.15 where appear the absorption lines in Tol 1037-2704 and Tol 1038-2712. This is summarized in Table 1.

Table 1

Emission and Absorption Line Redshifts of Quasars
in the Field of Tololo 1037-2704 and Tololo 1038-2712
in the Range of Redshift 1.85-2.15

Number*	Name** Tololo ...	Ref.	z_{em}	z_{abs}
19	1037-2704	1,2,3,4	2.23	1.912 1.972 2.0289 2.08 complex 2.13 complex
22	1038-2712	1,2,3,4	2.32	(1.85) (1.89) 1.956 2.014 2.08 trough 2.146 broad
23	1038-2707	3,5	1.92	1.89 broad
21	1037-2742	3	1.88	spectrum of insufficient quality for absorption line detection
16	1035-2737	5	2.15	1.98 2.125
7	1033-2653	5	2.42	no absorption in fairly good spectrum
17	1036-2713	5	3.09	CIV in spectral range 1.85-2.15 falls in Lyα forest

*) Numbers refer to numbering sequence in BOHUSKI and WEEDMAN [10].

**) 1037-2704 means 10^h37^m -27°04'

References: 1. ULRICH and PERRYMAN [8]
2. JAKOBSEN et al. [9]
3. SARGENT and STEIDEL [11]
4. CRISTIANI et al. [12]
5. JAKOBSEN et al. [13]

If the quasars and the galaxies producing the emission lines and the absorption lines are physically associated, then they are in a supercluster whose dimension are: ~ 10 Mpc in projection on the sky, and ~ 40 Mpc along the line of sight (this latter figure coming from the spread in redshift 1.85-2.15 and H = 100 km s^{-1}/Mpc, $q_0 = 0.5$). These dimensions are consistent with the sizes of superclusters in our vicinity at the present epoch. The elongation of the supercluster along one dimen-

sion - here along the line of sight - can be understood if most galaxies are distributed in sheets forming the thin walls of vast adjacent bubbles and if we are looking along the interface between two bubbles (STEIDEL, SARGENT and BOKSENBERG [14]). Some evidence for such large bubbles have been found at low redshifts by DE LAPPARENT, GELLER and HUCHRA [2].

4. Clustering of absorption lines in quasar spectra

The CIV doublet at λ 1548, 1550 Å is the strongest metal absorption line observable in the optical range for redshifts larger than 1.5. The first step in a study of the clustering of metal absorption lines in quasar spectra is to obtain optical spectra of a large number of quasars, with a well defined limit for the rest equivalent width of the CIV doublet. In each quasar spectrum the useful range of wavelength is between the CIV emission line and the Lyα emission line since at wavelengths shorter than Lyα, the CIV doublets are practically lost in the dense Lyman-α forest. The next step is to compare the number of systems observed along each line of sight to that expected for a Poisson distribution with the same mean number of systems per line of sight. One can note that if galaxies are distributed in large structures which are not 3-dimensional - e.g., sheets or filaments - then one expects some occurrence of a large number of absorption systems along the line of sight corresponding to the cases where we are looking along such large structures, thus violating Poisson statistics. To apply the "Poisson" test one has to assume that, in first approximation, there is no evolution in the number density of absorption systems per unit redshift in the interval of redshift covered by the sample.

At present the largest and best quality sample of quasar spectra suitable for a study of the clustering of the metal absorption lines is the one which has been assembled and which is being analyzed by Sargent, Boksenberg and Steidel (1988, in preparation). Preliminary results can be summarized as follows: The sample contains 57 quasars which have all been observed spectrographically with a limiting rest equivalent width of the CIV λ 1548, 1550 doublet of 0.15 Å.

The interval of redshift adequately covered by the sample is $1.5 < z < 3.4$. The average number of absorption lines with $W_{rest} \geq 0.15$ Å per interval of redshift $\Delta z = 1$, is 2.85.

The number of absorption systems per quasar spectrum is consistent with a Poisson distribution except for a tail consisting of 2 quasar spectra which each have 6 systems. The number of such spectra expected with a Poisson distribution is 0.3, corresponding to a rejection of the Poisson distribution hypothesis with ~ 95% confidence.

Moreover, for the 2 spectra which have 6 absorption systems one has 5 systems within $\Delta z = 0.32$ and the other has two groups of 3 systems with $\Delta z = 0.17$ and

$\Delta z = 0.15$. These two distributions of absorption lines are similar to that which is found in Tol 1037-2704 and Tol 1038-2712, suggesting that such distributions are not as rare as might be expected if the absorbers were to be randomly distributed on large scales (STEIDEL, SARGENT and BOKSENBERG [14]).

References

1. J.A. Tyson: Astronom. J. 92, 691 (1986)
2. V. de Lapparent, M.J. Geller, J.P. Huchra: Ap.J. Letters 302, 11 (1986)
3. M.-H. Ulrich: in Second ESO Workshop on ESO's Very Large Telescope, ESO Conference and Workshop Proceedings No. 24, ed. by S. D'Odorico and J.-P. Swings, p. 15 (1986)
4. T. Shanks, R. Fong, B.J. Boyle, B.A. Peterson: in Quasars, ed. by G. Swarup and V.K. Kapahi (Reidel, Dordrecht), p. 37 (1986)
5. Y.-Q. Chu, L-Z. Fang: in Observational Cosmology, ed. by A. Hewitt, G. Burbidge and L.-Z. Fang (Reidel, Dordrecht), p. 627 (1987)
6. E.M. Burbidge, V.T. Junkkarinen, A.T. Koski, H.E. Smith: Ap.J. Letters 242, L55 (1980)
7. J. Oort: Annual Rev. Astron. & Astrophys. 21, 373 (1983)
8. M.-H. Ulrich, M.A.C. Perryman: M.N.R.A.S. 220, 429 (1986)
9. P. Jakobsen, M.A.C. Perryman, M.-H. Ulrich, F. Macchetto, S. di Serego Alighieri: Ap.J. Letters 303, L27 (1986)
10. T.J. Bohuski, D.W. Weedman: Ap.J. 231, 653 (1979)
11. W.L.W. Sargent, C.C. Steidel: Astrophys. J., in press (1988)
12. S. Cristiani, I.J. Danziger, P.A. Shaver: M.N.R.A.S. 227, 639 (1987)
13. P. Jakobsen, M.A.C. Perryman, S. Cristiani: STScI Workshop on QSO Absorption Lines, ed. by C. Blade, C. Norman, D. Turnshek, Poster Paper Volume, p. 110 (1987)
14. C. Steidel, W.L.W. Sargent, A. Boksenberg: 3rd IAP Workshop High Redshift and Primeval Galaxies, ed. by J. Bergeron, D. Kunth, B. Rocca-Volmerange (1987)

Exact Solutions of the Nonlinear Equations for Self-Gravitating Matter

G. Götz

Max-Planck-Institut für Physik und Astrophysik,
Institut für Astrophysik, Karl-Schwarzschild-Str. 1,
D-8046 Garching, Fed. Rep. of Germany

An important problem in the theories of galaxy formation is the evolution of density inhomogeneities in the regime where perturbation theory fails. Here, several exact solutions of the equations of Newtonian hydrodynamics for self-gravitating matter are discussed, that provide models for the "pancake"-collapse, for flat, one-dimensional structures with non-vanishing gas pressure and for a "pancake"-like collapse with pressure in two dimensions: The general exact solution for vanishing pressure and two vanishing eigenvalues of the expansion tensor is derived, modeling a highly anisotropic collapse on large scales. It is also shown that all nonlinear plane-wave solutions in an isothermal gas can be generated by a sine-Gordon equation, leading to soliton solutions for the gravity system. A two-dimensional solution with gas pressure is obtained by conformal transformations from the hydrostatic equilibrium showing a "pancake"-like collapse on certain hypersurfaces which are determined by the nodal lines of a two-dimensional Laplace equation.

I. Introduction

There are several scenarios to explain the evolution of structure in the universe ([1],[2],[3],[4],[5]) differing mainly by the assumptions on the dominant matter content in the universe (baryons or dark matter), the initial conditions and other physical processes which single out different length scales on which the clustering of matter occurs first. But all the scenarios have in common that gravity is the driving force for the evolution of clumpiness. In an early stage, prior to the recombination era, the matter is supposed to be almost homogeneously distributed with small wrinkles in the density superimposed on this homogeneous, isotropic background. In this stage for which the fractional density contrast δ ($\delta \equiv \frac{\rho-\rho_0}{\rho_0}$, where ρ is the true density and ρ_0 is the density of the background) is small compared to one, the growth of density fluctuations is well described by the linear theory of perturbations on a Friedman-Robertson-Walker background [5],[8], [9]. After recombination density inhomogeneities must reach a regime where $\delta > 1$, since for instance the density contrast on galactic scales today is about 10^5. For $\delta > 1$ perturbation theory clearly breaks down and another method has to be employed to describe the evolution taking account of the inherent nonlinearities of gravitational systems. Numerical N-body simulations are used to tackle this problem [6],[7]. There is also an analytic approach by Zeldovich ([10],[11],[12]) based on an approximate solution of the gravity system with zero pressure that lead to the so-called "pancake"-scenario, realized either by a baryon-dominated model or a model with hot dark matter [13],[16]. Zeldovich predicted that the evolution of density fluctuations on large scales generally gives rise to flat, "pancake"-like objects, for which the motion of matter is primarily in one direction. These analytic claims were corroborated by numerical calculations [14],[15].

Since velocities and gravitational fields involved in this process are comparatively small, Newtonian gravity as a limiting case of Einstein's theory of gravitation is appropriate and will be used for all calculations in the sequel. Denoting the velocity field, the density,

the pressure and the gravitational acceleration by $\vec{v}$, ρ, p, $\vec{g}$, respectively, the basic equations of Newtonian hydrodynamics read:

$$\rho_t + \nabla \cdot (\vec{v}\rho) = 0 \tag{1a}$$

$$\vec{v}_t + (\vec{v} \cdot \nabla)\vec{v} = \vec{g} - \frac{\nabla p}{\rho} \tag{1b}$$

$$\nabla \cdot \vec{g} = -\rho \tag{1c}$$

$$\nabla \times \vec{g} = 0 \quad \Rightarrow \quad \vec{g} = -\nabla\Phi \tag{1d}$$

$$\rho \geq 0 \tag{1e}$$

The hydrodynamic equations may be used for the description of baryonic matter as well as for weakly interacting particles (dark matter) on scales large compared to the mean free path, if one identifies the square of the speed of sound $\frac{\partial p}{\partial \rho}$ with the velocity dispersion. For small scales a refined description by means of distribution functions has to be adopted (Vlasov-equations) [17].

In this context we do not refer to a particular scenario of galaxy formation, but investigate the system (1) by constructing exact solutions based on assumptions which are supposed to reflect some generic features present in any scenario of gravitational collapse.

In section II. the fact that the gravitational collapse in its late stages is highly anisotropic is modelled by the idealized assumption of two vanishing eigenvalues of the expansion tensor. The general solution with $p = 0$ and rotation-free velocity field can be derived, whereby the density still depends on three spatial coordinates without being invariant under some symmetry transformations [19].

Plane-symmetry is an idealization suitable to describe objects for which the spatial variation of all quantities is large in one direction compared to the typical length scales in the other two directions. In the "pancake"-collapse pressure forces begin to counteract gravity in a stage where the typical scale of density variations in the direction of most rapid contraction is of the order of the Jeans-length. In section III. it is shown that all nonlinear plane-wave solutions of (1) with an isothermal equation of state can be generated by a sine-Gordon equation, by means of which it is possible to construct soliton solutions of (1) [32].

Section IV. deals with a two-dimensional solution of (1) with pressure exhibiting an interesting collapse behavior. The matter piles up only on certain hypersurfaces determined by the nodal lines of a two-dimensional Laplace equation.

II. An exact model of the "pancake"-collapse

In a baryon-dominated model or a universe filled with hot dark matter (e.g. massive neutrinos) damping scales (photon-viscosity in the baryonic case, "free-streaming" for neutrinos) determine the size of objects that collapse first after recombination. In both cases the typical mass associated with these damping scales is of the order of $10^{13} - 10^{15} M_\odot$ at recombination, much larger than the Jeans-mass which is $\approx 10^6 M_\odot$ at that time. This justifies the assumption of vanishing gas pressure for the description of objects of cluster and supercluster size. The rotation of the velocity field is also assumed to be zero, since there are no large scale rotations observed in superclusters of galaxies. Moreover, linear perturbation theory predicts initial vorticity to decay in the course of the evolution. The third assumption on which the following solution of (1) is based, is the vanishing of two eigenvalues of the expansion tensor θ_{ij} which is defined as the symmetric part of the velocity gradient

$$\frac{\partial v_i}{\partial x_j} = \omega_{ij} + \theta_{ij} = \omega_{ij} + \sigma_{ij} + \frac{1}{3}\delta_{ij}\theta \qquad (2).$$

ω_{ij} is the antisymmetric part and vanishes for zero vorticity. σ_{ij} is the trace-free part of θ_{ij} describing a volume preserving distortion of fluid elements and θ is the trace of θ_{ij}. Since the eigenvalues of θ_{ij} are the expansion rates along the eigendirections in each fluid element, this assumption means that every fluid element contracts/expands in one direction only which in general can vary from fluid element to fluid element. This is a possibility to model anisotropic motions dominant in a certain region in the late phase of the collapse. It is possible to derive the general solution of (1) under these three assumptions.

To obtain solutions of (1) with $p = 0$ it is useful to transform from Eulerian coordinates x_i to Lagrangian coordinates X_i $(i = 1,2,3)$ defined as the labels of the integral curves $\vec{x} = \vec{f}(\vec{X},t)$ of the vector field

$$\frac{d\vec{f}}{dt} = \vec{v}(\vec{f},t) \qquad f_i(\vec{X},0) = \overset{\vee}{f}_i \equiv X_i \qquad (3)$$

$\frac{d}{dt}$ or equivalently a "dot" denotes the total derivative along the path of a fluid element $\partial_t + \vec{v}\cdot\nabla$. The dynamics of the flow is completely determined by this map $f_i : x_i \mapsto f_i(\vec{X},t)$ between Eulerian and Lagrangian coordinates. The derivative $J_{ik} = \frac{\partial f_i}{\partial X_k}$ and its determinant $J = \det(J_{ik})$ obey the following evolution equations:

$$\dot{J}_{ik} = \frac{\partial v_i}{\partial x_j} J_{jk}\,, \qquad \dot{J} = (\nabla\cdot\vec{v})\,J \qquad (4).$$

With (4) the continuity equation (1a) is easily integrated

$$\rho = \frac{\rho_0(\vec{X})}{J} \qquad (5).$$

$\rho_0(\vec{X})$ is an arbitrary function specifying the initial $(t = 0)$ density distribution since $J(t = 0) = 1$ for $f_i(\vec{X},0) = X_i$. The Euler equations (1b) read

$$\dot{v}_i = \ddot{f}_i = g_i \qquad (6).$$

Then (1d) and the Poisson equation (1c) can be completely written in terms of the map f_i. Denoting the functional determinant of any three functions $h_1(\vec{X},t)$, $h_2(\vec{X},t)$, $h_3(\vec{X},t)$ by

$$D(h_1,h_2,h_3) \equiv \frac{\partial(h_1,h_2,h_3)}{\partial(X_1,X_2,X_3)} \qquad (7)$$

inserting (6) and (5) into (1c) and (1d) yields four equations that contain only f_i and ρ_0:

$$\epsilon_{pq[j} D(\ddot{f}_{i]}, f_p, f_q) = 0 \qquad (8a)$$

$$\frac{1}{2}\epsilon_{abc} D(\ddot{f}_a, f_b, f_c) = -\rho_0 \qquad (8b).$$

ϵ_{abc} is the totally antisymmetric tensor in three dimensions. (8a) may be considered as three evolution equations for f_i and (8b) as a constraint, since ρ_0 is arbitrary and has only to satisfy $\dot{\rho}_0 = 0$. From a solution of (8) one can recover the original variables g_i, v_i via the relations $g_i = \ddot{f}_i$, $v_i = \dot{f}_i$ and the density via (5) with $J = D(f_1,f_2,f_3)$. The transformation back to Eulerian coordinates is obtained by solving $x_i = f_i(X_j,t)$ for X_j which is possible as long as $J \neq 0$. The singularities of the transformation $J = 0 \Leftrightarrow \rho \to \infty$ are associated with the "pancakes" in Zeldovich's scenario [21].

In order to see what the assumption of two vanishing eigenvalues of θ_{ij} implies, one considers one fluid element labeled by X'_j whose path is

$$\vec{\gamma}: t \mapsto f_i(X'_j, t) \tag{9}.$$

In general the evolution equation for a vector n_i connecting two infinitesimally seperated fluid elements X'_i, $X'_i + \delta X_i$ along $\vec{\gamma}$ is ([18])

$$\dot{n}_j = (\omega_{jk} + \sigma_{jk} - (\sigma_{ab} n_a n_b)\delta_{jk}) n_k \tag{10}$$

Now, if n_j is an eigenvector of θ_{ij} and if $\omega_{ij} = 0$, then from (10) follows $\dot{n}_j = 0$. Thus, for rotation-free motions the eigenvectors of θ_{ij} do not change along a path $\vec{\gamma}$. This allows one to fix a coordinate system at $t = 0$ in which θ_{ij} is diagonal and remains diagonal for $t > 0$. Note that this coordinate system is different for different fluid elements. From (4) one deduces that J_{ik} remains also diagonal along $\vec{\gamma}$ if θ_{ij} is diagonal and if $J_{ik}(t = 0) = \delta_{ik}$. Then, it is easy to check that in this case the evolution equations (8a) for f_i are identically satisfied along $\vec{\gamma}$ if $\theta_{ij}|_{\vec{\gamma}} = \mathrm{diag}(\theta_1, 0, 0)$. (8b) yields

$$\ddot{f}_{11}|_{\vec{\gamma}} = -\rho_0(\vec{X}') \tag{11}$$

implying

$$\ddot{f}_{11} = 0 \tag{12}.$$

Since these considerations hold on each $\vec{\gamma}$ and the transformations relating different coordinate systems in which θ_{ij} is diagonal on different paths are linear, one has the general form for the time dependence of the map f_i

$$f_i = X_i + \breve{v}_i(\vec{X})\, t + \frac{1}{2}\breve{g}_i(\vec{X})\, t^2 \tag{13}.$$

$\breve{v}_i$, $\breve{g}_i$ are the initial conditions ($t = 0$) of v_i, g_i, respectively, which still are constrained by (8). The remaining task is to plug (13) into (8) and solve the resulting 17 equations for the six functions $\breve{v}_i$, $\breve{g}_i$. The equations and the lengthy calculation to solve them can be found in [19]. The result is the following:

$$\breve{v}_1 = F(\breve{g}_1),\ \breve{v}_2 = H(\breve{g}_1),\ \breve{v}_3 = K(\breve{g}_1) \tag{14a}$$

$$\breve{g}_1 = E(X_1 + A'X_2 + B'X_3),\ \breve{g}_2 = A(\breve{g}_1),\ \breve{g}_3 = B(\breve{g}_1) \tag{14b}$$

$$F'A' = H',\ F'B' = K' \tag{14c}.$$

F,H,K,E,A,B are arbitrary functions of one argument. Since (1d) guarantees the existence of a potential $\breve{\Phi}$ defined by $\breve{g}_i = -\breve{\Phi}_{,X_i}$ the constraints (14b) for $\breve{g}_i$ can be expressed in terms of equations for $\breve{\Phi}$ whose solution is

$$\breve{\Phi} = \alpha X_1 + A(\alpha)X_2 + B(\alpha)X_3 + C(\alpha) \tag{15a}$$

$$0 = X_1 + A'(\alpha)X_2 + B'(\alpha)X_3 + C'(\alpha) \tag{15b}.$$

For the prove of (15) see [19].

Finally one can write down the complete solution of (1) for $p = 0$, $\nabla \times \vec{v} = 0$ and two vanishing eigenvalues of θ_{ij}:

$$\rho = \frac{\rho_0}{1 + F'\rho_0 t - \frac{1}{2}\rho_0 t^2}\,, \qquad \rho_0 = \frac{-(1 + A'^2 + B'^2)}{A''X_2 + B''X_3 + C''} \tag{16a}$$

$$\begin{aligned} v_1 &= F(\alpha) - \alpha t, & g_1 &= -\alpha \\ v_2 &= H(\alpha) - A(\alpha)t, & g_2 &= -A(\alpha) \\ v_3 &= K(\alpha) - B(\alpha)t, & g_3 &= -B(\alpha) \end{aligned} \tag{16b}$$

$$H' = F'A'\,, \qquad K' = F'B' \tag{16c}$$

$$\begin{aligned} x_1 &= X_1 + Ft - \frac{1}{2}\alpha t^2 \\ x_2 &= X_2 + Ht - \frac{1}{2}At^2 \\ x_3 &= X_3 + Kt - \frac{1}{2}Bt^2 \end{aligned} \tag{16d}$$

$$0 = X_1 + A'(\alpha)X_2 + B'(\alpha)X_3 + C'(\alpha) \tag{16e}.$$

A particular solution is determined by a choice of the four functions A,B,C,F. (16e) gives $\alpha(\vec{X})$ and H,K are determined by (16c). The inversion of (16d) yields the solution in Eulerian coordinates $\vec{x}$.

Setting $\breve{v}_2 = \breve{v}_3 = 0 \Leftrightarrow H = K = A = B = 0$ leads to the general one-dimensional dust solution of (1) since C and F are still free functions of one variable that can be fixed by an arbitrary choice of ρ and v at $t = 0$ [20].

For $\rho_0 > 0$ there exists always a finite time $t_c = F' + \frac{1}{\rho_0}\sqrt{\rho_0^2 F'^2 + 2\rho_0}$ for which $J(t = t_c) = 0$. Therefore, every particle hits a hypersurface of infinite density in a finite time. For $F' = 0$ t_c is simply the free-fall time. The expansion $\theta = \mathrm{Trace}(\theta_{ij})$ measuring the rate of change of a fluid element's volume is given by

$$\theta = (F' - t)\rho \tag{17}$$

If $F' > 0$, during the time intervall $0 \leq t \leq F'$ the corresponding fluid element expands $\theta > 0$, but no matter how large F' is, this expansion turns always into a contraction at $t = F'$.

An important information on the density distribution (16a) is contained in the properties of the level surfaces

$$\rho = const. \tag{18}.$$

(18) may be viewed as an implicit representation of a two-dimensional hypersurface in Eulerian space. With a function $M(\alpha)$ defined by $M'(\alpha) \equiv 1 + A'^2 + B'^2$ (18) reads simply

$$M_x = M'\alpha_x = const. \tag{19}.$$

It turns out that the Gaussian curvature K_g of (18) is negative for any choice of initial conditions and for $t \geq 0$:

$$K_g = \frac{-\alpha_x^6 M_x^4}{M_{xx}^4}\left[B'''A'' - A'''B''\right]^2|_{\rho=const.} \tag{20}$$

III. Nonlinear plane waves and the sine-Gordon equation

Having collapsed down to an extent of the order of the Jeans-length, pressure forces can no longer be neglected in the evolving "pancake"-like objects. Since in this phase of the collapse the motion is already almost one-dimensional in a large part of space one can assume that all quantities ρ, p, $\vec{g} = (g(x,t),0,0)$, $\vec{v} = (u(x,t),0,0)$ depend on one spatial coordinate x and time t only. It is shown in this section that under these simplifying assumptions and for an equation of state

$$p = \gamma\rho\,, \quad \gamma = const. > 0 \tag{21}$$

the system (1) can be reduced to a sine-Gordon equation ([32]), which in turn is solvable by an "inverse-scattering" transformation [25],[26],[28]. The sine-Gordon equation has two salient features: the existence of solitons (almost elastically interacting localized humps) and a definite asymptotic behavior that is independent of the initial conditions. It turns out that for subsonic motions this asymptotic behavior carries over to the gravity system. Arbitrary non-characteristic initial data decay into a superposition of solitons for $t \to \infty$. All solutions of the 1-d version of (1) with (21) generated by all possible one- and two-soliton solutions of the sine-Gordon equation can be constructed explicitely. It has been conjectured by Liang and Barrow ([4],[24]) that such soliton structures might exist also in gravity systems (see also [22], [23]).

Firstly, one transforms (1) with all quantities depending on (x,t) to characteristic coordinates (λ,μ) defined as the labels of the two characteristics running through each point in the (x,t)-plane:

$$\begin{aligned} C_+ &: \ \frac{dx}{dt} = u + \sqrt{\gamma}, \ \ \lambda = const. \\ C_- &: \ \frac{dx}{dt} = u - \sqrt{\gamma}, \ \ \mu = const. \end{aligned} \tag{22}.$$

Transforming the gravity system to (λ,μ)-coordinates yields

$$x_\mu = (u + \sqrt{\gamma})t_\mu \tag{23a}$$

$$x_\lambda = (u - \sqrt{\gamma})t_\lambda \tag{23b}$$

$$\rho u_\mu + \left[\frac{1}{2\sqrt{\gamma}}g^2 + \sqrt{\gamma}\rho\right]_{,\mu} = 0 \tag{23c}$$

$$-\rho u_\lambda + \left[\frac{1}{2\sqrt{\gamma}}g^2 + \sqrt{\gamma}\rho\right]_{,\lambda} = 0 \tag{23d}$$

$$g_\mu = -\sqrt{\gamma}\,\rho\,t_\mu \tag{23e}$$

$$g_\lambda = \sqrt{\gamma}\,\rho\,t_\lambda \tag{23f}.$$

It can be shown in general ([31]) that the initial value problem for the 1-d version of (1) is equivalent to the corresponding initial value problem of the characteristic system (23). One of the equations (23e,f) is redundant since (23) are six equations for the five functions $\rho(\lambda,\mu)$, $u(\lambda,\mu)$, $g(\lambda,\mu)$, $x(\lambda,\mu)$, $t(\lambda,\mu)$. With a new variable $R \equiv \ln\rho$ the integrability conditions for $u(\lambda,\mu)$ and $t(\lambda,\mu)$ derived from (23c,d) and (23e,f) yield two second order equations for R and g

$$2g_{\lambda\mu} = R_\lambda g_\mu + R_\mu g_\lambda \tag{24a}$$

$$R_{\lambda\mu} = -\frac{1}{\gamma}e^{-R}g_\mu g_\lambda \tag{24b}$$

Multiplying (24a) with g_μ and g_λ, respectively, one gets two linear equations for g_μ^2 and g_λ^2 whose solutions contain two arbitrary functions of λ, μ. Since (23) is invariant under a relabeling of the characteristics $\lambda \to f(\lambda)$, $\mu \to h(\lambda)$ it is possible to choose the free functions in the solution of (24a) such that

$$g_\lambda^2 = \frac{1}{2}\gamma e^R[4 - R_\lambda^2] \tag{25a}$$

$$g_\mu^2 = \frac{1}{2}\gamma e^R[4 - r_\mu^2] \tag{25b}.$$

(25) pluged into (24b) yields one second order equation for R:

$$R_{\lambda\mu} = -\frac{1}{2}\sqrt{4 - R_\lambda^2}\sqrt{4 - R_\mu^2} \tag{26}.$$

To solve (26) one introduces two new dependent variables $\alpha(\lambda,\mu)$, $\beta(\lambda,\mu)$

$$R_\lambda = 2\cos(\frac{\alpha+\beta}{2}) \qquad R_\mu = 2\cos(\frac{\beta-\alpha}{2}) \tag{27}.$$

(27) imposes no restrictions on R since g and R have to be real, see (25), (26). From (26) one derives two first order equations for α, β:

$$\frac{1}{2}(\alpha_\mu + \beta_\mu) = \sin(\frac{\beta-\alpha}{2}) \tag{28a}$$

$$\frac{1}{2}(\beta_\lambda - a_\lambda) = \sin(\frac{\alpha+\beta}{2}) \tag{28b}.$$

The integrability conditions for α, β following from (28) are

$$\alpha_{\lambda\mu} = \sin\alpha\,, \qquad \beta_{\lambda\mu} = \sin\beta \tag{29}.$$

(28) is nothing else than the auto-Bäcklund transformation of the sine-Gordon equation (29) mapping solutions of the sine-Gordon equation onto itself ([27], [29], [30]).

The recipe to produce a solution to the gravity system in 1-d is as follows: start with any solution of the sine-Gordon equation , say α, generate a second one β via the Bäcklund transformation (28), then all the original quantities ρ, g, u are determined by the following quadratures:

$$R_\lambda = 2\cos(\frac{\alpha+\beta}{2})\,, \qquad R_\mu = 2\cos(\frac{\beta-\alpha}{2}) \qquad \rho = e^R \tag{30a}$$

$$g_\lambda = -\sqrt{2\gamma}e^{\frac{1}{2}R}\sin(\frac{\alpha+\beta}{2})\,, \qquad g_\mu = -\sqrt{2\gamma}e^{\frac{1}{2}R}\sin(\frac{\beta-\alpha}{2}) \tag{30b}$$

$$u_\lambda = \frac{1}{\rho}[\frac{1}{2\sqrt{\gamma}}g^2 + \sqrt{\gamma}\rho]_{,\lambda}\,, \qquad u_\mu = -\frac{1}{\rho}[\frac{1}{2\sqrt{\gamma}}g^2 + \sqrt{\gamma}\rho]_{,\mu} \tag{30c}$$

$$t_\lambda = -\sqrt{2}e^{-\frac{1}{2}R}\sin(\frac{\alpha+\beta}{2})\,, \qquad t_\mu = \sqrt{2}e^{-\frac{1}{2}R}\sin(\frac{\beta-\alpha}{2}) \tag{30d}$$

$$x_\lambda = t_\lambda(u - \sqrt{\gamma})\,, \qquad x_\mu = t_\mu(u + \sqrt{\gamma}) \tag{30e}$$

A disadvantage of solving differential equations in characteristic coordinates is the dependence of the transformation $(x,t) \leftrightarrow (\lambda,\mu)$ on the solution itself which makes it impossible in general to relate a region in (λ,μ)-space to the corresponding region in (x,t)-space wi-

thout already knowing the solution. Especially initial data for the gravity system can not be related to initial data of the sine-Gordon equation without having worked through all the quadratures (30). But fortunately the whole variety of solutions of the sine-Gordon equation can be divided into two classes: the solitons and the so-called radiation part which correspond to the discrete and continuous spectrum, respectively, of the eigenvalue problem associated with the sine-Gordon equation via the "inverse-scattering" procedure.

Another prominent feature of the sine-Gordon equation is its asymptotic behavior. The sine-Gordon equation divides the (λ,μ)-plane in two causally disconnected regions whose boundary is the "lightcone" $\hat{\lambda}^2-\hat{\mu}^2=0$ with $\hat{\lambda}=\lambda+\mu$, $\hat{\mu}=\lambda-\mu$. For $\hat{\mu}\to\infty$ the solution of the sine-Gordon equation decays into a superposition of solitons if the initial data are given in the region $\hat{\mu}^2-\hat{\lambda}^2<0$. Because of the symmetry $\alpha\to-\alpha,\hat{\lambda}\leftrightarrow\hat{\mu}$ this holds also for $\hat{\lambda}\to\infty$ if initial data are given in $\hat{\mu}^2-\hat{\lambda}^2>0$. For the gravity system one is interested in the behavior of the solution for $t\to\infty$ on curves $x=const.$ if initial conditions are prescribed on curves $t=const.$. Thus, the question is, under which conditions do the curves $x=const.$ and $t=const.$ lie in causally different regions in the $(\hat{\lambda},\hat{\mu})$-plane? The modulus of the tangent vector of the curves $x=const.$ in the $(\hat{\lambda},\hat{\mu})$-plane endowed with a Minkowski metric is

$$\hat{\mu}_t^2-\hat{\lambda}_t^2=\frac{2}{\sqrt{\gamma}}\frac{1}{J}(u^2-\gamma)|_{x=const.} \tag{31}$$

where J is the Jacobian of the transformation between (x,t) and (λ,μ)-coordinates. The modulus of the tangent vector of the curves $t=const.$ in the $(\hat{\lambda},\hat{\mu})$-plane is

$$\hat{\mu}_x^2-\hat{\lambda}_x^2=\frac{2}{\sqrt{\gamma}}\frac{1}{J}|_{t=const.} \tag{32}.$$

Therefore, the curves on which initial data are given and the curves on which the asymptotic behavior is investigated lie in causally different regions only if the sign of (31) and (32) is different. This is obviously true for subsonic flows $u^2<\gamma$. In this case the behavior of the physical quantities in the asymptotic region $t\to\infty$ is obtained by letting $\hat{\mu}\to\infty$ in the parametric representation of the solution of (30) provided that $(u^2-\gamma)\frac{1}{J}>0$. Otherwise $t\to\infty$ corresponds to $\hat{\lambda}\to\infty$. This means, for subsonic flows the solution of the gravity system in the region $t\to\infty$ is essentially generated by the asymptotic form of the solution of the sine-Gordon equation , namely a linear superposition of N solitons. In the discussion of particular solutions generated from all possible two-soliton solutions of the sine-Gordon equation it can be shown that a solution of two moving solitons of the sine-Gordon equation indeed produces a solution with two moving humps of matter. For the case of N solitons one can expect that this produces also a N-soliton solution of the gravity system. The sole imprint of initial conditions is the number of remaining solitons N and the amplitude distribution.

We now come to a brief discussion of *particular solutions* (see [32]) constructed from all the possible *two-soliton solutions* of the sine-Gordon equation . Since the associated eigenvalue problem in the "inverse-scattering" procedure is not self-adjoint there exist not only two solitons with non-vanishing relative velocity but also bound states of two solitons ("breather").

The one-soliton solution: Starting with the trivial solution $\alpha=0$ the Bäcklund transformation (28) produces a kink $\beta=4\arctan(e^{\mu+\lambda})$ and the quadratures (30) yield:

$$\rho = A\,\mathrm{sech}^2\left[\sqrt{\frac{A}{2\gamma}}(x - u_0 t)\right], \quad A = const.$$
$$u = u_0 = const. \tag{33}$$

g is obtained from $g_x = -\rho$. (33) is a single hump in the density moving with a constant velocity u_0. The typical nonlinear feature is the dependence of the width $\sqrt{\frac{2\gamma}{A}}$ on the amplitude A (this is similar to the KdV-solitons). The density in (33) is nothing else than the hydrostatic equilibrium in 1-d transformed into a frame moving with velocity u_0. An analysis of linear perturbations of (33) shows that this one-soliton solution is marginally stable for arbitrary amplitudes A (see [32]).

The two-soliton solution: If α is the one-soliton solution of the sine-Gordon equation, β is the two-soliton solution containing a new constant ϵ which is the relative speed of the two kinks in the (λ, μ)-plane. The corresponding solution of the gravity system can be constructed explicitly. Since it consists of very lengthy expressions it will not be written down in this summary (see [32]). An analysis of the local extrema shows that the two kinks indeed generate a solution with two humps in the density moving with different non-constant velocities. The speed of the minimum between the two humps is constant and determined by ϵ. The amplitude of the two solitons changes during their motion but reaches asymptotically a constant value. The relative velocity tends to zero asymptotically.

Since α and β do not enter the quadratures (30) and the Bäcklund transformation (28) symmetrically, an exchange of two solutions of the sine-Gordon equation that are related via (28) might produce a different solution to the gravity system. But (28) is invariant under

$$\alpha \leftrightarrow \beta, \quad \lambda, \mu \rightarrow -\lambda, -\mu \tag{34}.$$

(34) applied to the pair α, β which generated the single soliton (33) yields simply the *homogeneous 1-d solution* for which $\rho_x = 0$ ([32]).

Pulse on a homogeneous background: (34) applied to the two-soliton solution produces a solution of the gravity system which can be interpreted as a single pulse moving on a homogeneous background. The remarkable property is a diverging density $\rho \rightarrow \infty$ occurring after a finite time that depends on the background quantities and the amplitude of the pulse.

Breather I: Only the limiting case of the bound state of two solitons for which the oscillation frequency is infinite is related to a real solution of the sine-Gordon equation via (28). The corresponding solution of the gravity system represents a single hump at rest onto which two smaller pulses collide symmetrically. The three local maxima are separated by discontinuities in the density gradient.

Breather II: (34) applied to the set of solutions of the sine-Gordon equation that generated the "breather I" yields a completely different solution for the gravity system. It exists only in a finite range of the (x, t)-plane and may be interpreted as a layer of gas that is trapped between two surfaces of infinite density located at $x = \pm\sqrt{\frac{\gamma}{2A}}$. The local minimum between these two planes is $4A$, $A = const.$.

IV. Two-dimensional solution with pressure

In this paragraph a two-dimensional (translationally invariant in one direction), time-dependent solution of (1) is generated from the hydrostatic equilibrium by means of time-dependent conformal transformations. It exihibits an interesting collapse behavior on

special hypersurfaces determined by the initial conditions. It might serve as a model for collapsing, elongated structures with non-vanishing pressure. The assumptions about the symmetry imply $\vec{v} = \big(u(x,y,t), v(x,y,t), 0\big)$, $\rho_z = 0$, $g_z = 0$. The equation of state shall be again (21).

We start with the hydrostatic equilibrium in two dimensions determined by

$$\triangle\Phi = c\, e^{-\frac{1}{\gamma}\Phi}\,, \qquad \rho = c e^{-\frac{1}{\gamma}\Phi}\,, \quad c = const., \quad \vec{v} = 0 \tag{35}.$$

(35) is a Liouville-equation for Φ whose solutions can be mapped to the solutions of a two-dimensional Laplace equation. On account of $\rho > 0$ the solutions can be represented as

$$\rho = 2\gamma \frac{(h_x^2 + h_y^2)}{\cosh^2(h)}\,, \qquad \triangle h = 0 \tag{36}.$$

Next, we apply a time-dependent conformal transformation

$$\hat{x} = \Omega(x,y,t)\,, \qquad \hat{y} = \Psi(x,y,t)$$

$$\Omega_x = \Psi_y\,, \qquad \Omega_y = -\Psi_x \tag{37}$$

$$\Longrightarrow \quad \triangle\Omega = \triangle\Psi = 0$$

to the Liouville equation (35) which is invariant if Φ is transformed like $\Phi \to \Phi + \gamma \ln J$ where J is the Jacobian of (37): $J = \Omega_x^2 + \Omega_y^2$. If the equilibrium is assumed to be maintained under (37) ρ has to be transformed like

$$\rho = 2\gamma J \frac{(h_{\hat{x}}^2 + h_{\hat{y}}^2)}{\cosh^2(h)} \qquad \text{with} \quad h_{\hat{x}\hat{x}} + h_{\hat{y}\hat{y}} = 0 \tag{38}.$$

(38) inserted into the continuity equation yields three equations for Ω, Ψ and J which are all satisfied if the velocity field is given by

$$u = \frac{1}{J}\big[\Omega_y \Psi_t - \Omega_t \Psi_x\big]\,, \qquad v = -\frac{1}{J}\big[\Omega_t \Psi_y + \Psi_t \Omega_x\big] \tag{39}$$

These expressions for u, v plugged into the remaining Euler equations determine finally the form of the conformal transformation (37). It is much easier to work in the complex coordinates $z = x + iy$, $\bar{z} = x - iy$. Then the solution of the Cauchy-Riemann equations (37) is

$$\Omega = \alpha_1(z,t) + \alpha_2(\bar{z},t)\,, \qquad \Psi = i(\alpha_2 - \alpha_1) \tag{40}$$

and (39) reads

$$u = -\frac{1}{2}\Big[\frac{\dot{\alpha}_1}{\alpha_1'} + \frac{\dot{\alpha}_2}{\alpha_2'}\Big]\,, \qquad v = \frac{i}{2}\Big[\frac{\dot{\alpha}_1}{\alpha_1'} - \frac{\dot{\alpha}_2}{\alpha_2'}\Big] \tag{41}.$$

("prime" and "dot" denote derivatives with respect to z or $\bar{z}$ and t, respectively). The Euler equations are satisfied for (41) if $\alpha_1(z,t)$ fulfills the following evolution equation:

$$\alpha_1'' \dot{\alpha}_1^2 - 2\dot{\alpha}_1 \alpha_1' \dot{\alpha}_1' + \ddot{\alpha}_1 \alpha_1'^2 = 0 \tag{42}$$

Since the conformal transformation must be real α_2 is simply the complex conjugate of α_1. It can be shown with the help of a Legendre transformation that the general local solution to (42) is given implicitly by

$$0 = z + F(\alpha_1)t + G(\alpha_1) \tag{43}$$

F, G are arbitrary. The complete solution contains three free functions F, G and $h(\hat{x},\hat{y})$ as a solution of a Laplace equation. A special choice of F, G determines α_1 via (43), this in turn fixes u, v via (41) and ρ is given by (38) with $J = 4\alpha_1'\alpha_2'$, $\alpha_2 = \bar{\alpha}_1$.

All the free functions involved in this solution are solutions of a two-dimensional Laplace equation. To extract general properties it is more useful to represent these free functions as Fourier serieses:

$$h = \sum_{i=1}^{\infty} A_i \cosh(\kappa_i\hat{x})\sin(\kappa_i\hat{y}) + B_i \cosh(\omega_i\hat{x})\sin(\omega_i\hat{y}) + \ldots\ldots \tag{44}$$

$$A_i, B_i, \ldots\kappa_i, \omega_i, \ldots = const.$$

If α_1 is expanded like (44) then it is easy to see from (38) that the density diverges for $x, y \to \pm\infty$ on the lines where $h = 0$. To avoid such a behavior one uses simply a Fourier expansion for α_1 which is deduced from an expansion in terms of products of sinh, cosh, sin, cos by an inversion with respect to the unit circle in the (x,y)-plane:

$$\alpha_1 = \sum_{i=1}^{\infty} a_i \sinh(\frac{k_i}{z}) + b_i \cosh(\frac{l_i}{z}) + c_i \cos(\frac{m_i}{z}) + d_i \sin(\frac{n_i}{z}) \tag{45}$$

where the coefficients $a_i, \ldots k_i, \ldots$ can still depend on time. Such a α_1 produces a ρ (38) that is well behaved in the whole (x,y)-plane. The time dependence of the coefficients is determined by the evolution equation of α_1 (42). (45) inserted into (42) yields an infinite set of ordinary differential equations that are solved by

$$\ddot{k}_i = \ddot{l}_i = \ddot{m}_i = \ddot{n}_i = 0 \tag{46a}$$

$$k_i \propto l_i \propto m_i \propto n_i \tag{46b}$$

$$a_i,\ b_i,\ c_i,\ d_i = const. \tag{46c}$$

Since all the coefficients k_i, l_i, m_i, n_i grow linearly in time, α_1 and h and its derivatives grow exponentially, but ρ (38) decreases for $t \to \infty$ at all points where $h \neq 0$ because of the additional exponential in $\cosh^{-2}(h)$. But at all points where $h = 0$ $\cosh(h) = 1$ and the growth of α_1 is not compensated by the more rapidly decreasing $\cosh^{-2}$. In general this solution describes the collaps of matter on particular hypersurfaces given by $h = 0$. The density is finite in the whole (x,y)-plane and decays for $t \to \infty$ at all points where $h \neq 0$.

This interesting behavior is best illustrated by means of a special example. Let

$$F = -2c\alpha_1\ ,\quad G = -2\alpha_1 \qquad c = const.$$

$$h = a\cosh(q_{\hat{x}})\sin(q_{\hat{y}})\ ,\quad q = \frac{1}{2}\, b\, \ln(\hat{x}^2 + \hat{y}^2) \qquad a, b = const.$$

then the density in x, y, t coordinates reads:

$$\rho = 2\gamma(ab)^2 \frac{(ct+1)^2}{(x^2+y^2)^2}\left[\sinh^2\Big(\frac{bx(ct+1)}{x^2+y^2}\Big) + \cos^2\Big(\frac{by(ct+1)}{x^2+y^2}\Big)\right] \times$$

$$\times \operatorname{sech}^2\left[a\cosh\left(\frac{bx(ct+1)}{x^2+y^2}\right)\sin\left(\frac{by(ct+1)}{x^2+y^2}\right)\right] \tag{47}$$

On the hypersurface $y = 0$ (47) is

$$\rho(x, y=0, t) = 2\gamma(ab)^2\,\frac{(ct+1)^2}{x^4}\cosh^2\left(\frac{b}{x}(ct+1)\right) \tag{48}.$$

For $x, y \to \pm\infty$ ρ falls off to zero and decays for $t \to \infty$ everywhere except at $y = 0$. (48) shows that $\rho(x, 0, t)$ grows like

$$\rho \longrightarrow t^2 e^{\frac{2bct}{x}} \quad .$$

In three dimensions this means that the matter piles up only on the plane $y = 0$.

Literature:

1. Efstathiou G., Silk J. (1983): Fund. of Cosmic Physics 9, 1-138
2. Shandarin S., Doroshkevich A., Zeldovich Y.B. (1983): Sov. Phys. Usp. 26, 46
3. Zeldovich Y.B.,Novikov I.D. (1983): The structure and evolution of the universe, Relativistic Astrophysics Vol.2, Univ. of Chicago Press
4. Barrow J.D. (1980): Phil. Trans. R. Soc. London A 296, 273-289
5. Peebles P.J.E. (1980): The large-scale structure of the universe, Princeton Univ. Press
6. Aarseth S., Gott J., Turner E. (1979): ApJ 228, 664 and ApJ 228, 684
7. Davis M., Efstathiou G., Frenk C., White S. (1985): ApJ 292, 371
8. Bardeen J.M. (1980): Phys. Rev. D22, 1882
9. Lifshitz E.M., Khalatnikov I.M. (1964): Sov. Phys. Usp. 6, 495
10. Zeldovich Y.B. (1970): AA 5, 84
11. Doroshkevich A., Ryabenkii V., Shandarin S. (1975): Astrophysics 9, 144
12. Shapiro P.R.(1984): Pancakes and galaxy formation, in: Clusters and groups of galaxies, ed. by Mardirossian, Reidel, Dordrecht
13. Bond J.R., Szalay A.S. (1983): ApJ 274, 443
14. Shapiro P.R., Struck-Marcell C., Melott A. (1983): ApJ 275, 413
15. Melott A. (1983): ApJ 264, 59
16. Faber S. (ed.) (1987): Nearly normal galaxies, Santa Cruz Summer Workshop, Springer, New York
17. Saslaw W.C. (1985): Gravitational physics of stellar and galactic systems, Cambridge Univ. Press
18. Ellis G.F.R. (1971): Relativistic Cosmology, in: Proc. of Enrico-Fermi School, Course XLVII, ed. by B.K. Sachs, Academic Press, New York
19. Buchert T., Götz G. (1987): to appear in J. Math. Phys.
20. Novikov E.A. (1970): Sov. Phys. JETP 30, 512
21. Arnold V., Shandarin S., Zeldovich Y.(1982): Geophys.Astrophys. Fluid Dyn. 20, 111
22. Asano N. (1974): Suppl. of the Progr. of Theor. Phys. 55, 52
23. Kates R. (1986): AA 168, 1
24. Liang E.P. (1976): ApJ 204, 235
25. Ablowitz M.J., Segur H. (1981): Solitons and the inverse scattering transform, SIAM, Philadelphia

26. Calogero F., Degasperis A. (1982): Spectral transform and solitons I, North-Holland, Amsterdam
27. Dodd R.K., Bullough R.K. (1976): Proc. R. Soc. London A351, 499
28. Ablowitz M. et al. (1973): Phys. Rev. Lett. 30, 1262
29. Lamb G.L. (1980): Elements of soliton theory, Wiley, New York
30. Eilenberger G. (1981): Solitons, Springer, Berlin
31. Courant R., Hilbert D. (1968): Methoden der mathematischen Physik, Springer, Heidelberg
32. Götz G. (1987): Solitons in Newtonian gravity, preprint

Index of Contributors