Physics – Uspekhi **44** (1) 1 – 51 (2001)

REVIEWS OF TOPICAL PROBLEMS PACS numbers: **04.30. – w**, 04.80.Nn, 95.55.Jm, 95.85.Sz

Gravitational wave astronomy: in anticipation of first sources to be detected

L P Grishchuk, V M Lipunov, K A Postnov, M E Prokhorov, B S Sathyaprakash

DOI: 10.1070/PU2001v044n01ABEH000873

Contents

Abstract. The first generation of long-baseline laser interferometric detectors of gravitational waves will start collecting data in 2001 – 2003. We carefully analyse their planned performance and compare it with the expected strengths of astrophysical sources. The scientific importance of the anticipated discovery of various gravitational wave signals and the reliability of theoretical predictions are taken into account in our analysis. We try to be conservative in evaluating both the theoretical uncertainties in the parameters of the source and the prospects of its detection. Upon considering many possible sources, we place our emphasis on (i) inspiraling binaries consisting of stellar mass black holes and (ii) relic gravitational waves. We conclude that inspiraling binary black holes are likely to be detected by the early ground-based interferometers first. We estimate that the first interferometers will see 2 – 3 events per year from black hole binaries with component masses of $10-15M_{\odot}$, with a signal-to-noise ratio of about 3, in a network of detectors consisting of GEO, VIRGO and two LIGOs. It appears that other possible sources, including coalescing neutron stars, are unlikely to be detected by the early instruments. We also argue that relic gravitational waves may be discovered by space-based interferometers in the frequency interval $2\times10^{-3}-10^{-2}$ Hz, at a signal-to-noise ratio level of about 3.

L P Grishchuk, V M Lipunov, K A Postnov, M E Prokhorov Sternberg Astronomical Institute, M V Lomonosov Moscow State University, Universitetskiĭ prosp. 13, 119899 Moscow, Russian Federation
Tel. (7-095) 939 50 06. Fax (7-095) 932 88 41
E-mail: pk@sai.msu.ru
V M Lipunov, K A Postnov M V Lomonosov Moscow State University, Physical Department,
Vorob'evy Gory, 119899 Moscow, Russian Federation
L P Grishchuk, B S Sathyaprakash Cardiff University, P.O. Box 913, Cardiff, CF2 3YB, Great Britain
E-mail: grishchuk@astro.cf.ac.uk

Received 28 August 2000
Uspekhi Fizicheskikh Nauk **171** (1) 3 – 59 (2001)
Translated by L P Grishchuk, K A Postnov, B S Sathyaprakash; edited by A V Getling

1. Introduction

The goal of this review article is quite ambitious. We want to foretell the parameters of the first gravitational wave signals that will be seen by sensitive detectors, some of which are currently in the final stage of construction. The detectors will start collecting data within a couple of years. Obviously, we present a subjective point of view. It is based on our evaluation of what we consider the best theoretical knowledge available today, for the expected sensitivity of the instruments. Possibly, other authors would regard other sources as more promising, and would place their bets on something else. It is also possible that our view is biased because it is partially guided by the work we have personally been involved in. We will not be very disappointed if we are proved wrong. Nature may have many surprises in store for us. It is important, however, that for the first time in the long history of gravitational wave research, conservative astrophysical estimates overlap with the detecting capabilities of real instruments. Thus, it is now an appropriate time to prepare strategies for the search and analysis of signals that appear to be more probable than others.

The general theory of gravitational radiation is well understood and is described in textbooks [1 – 3]. The status of gravitational wave astronomy has been regularly reviewed [4 – 6], including papers in *Phys. Usp.* [7 – 10]. Here, we will only remind the reader that gravitational waves (GW) are an inescapable consequence of Einstein's general relativity and, indeed, of any gravitational theory which respects special relativity. Gravitational waves are similar to electromagnetic waves in some aspects. They propagate with the velocity of light c, have two independent transverse polarization states, and exhibit some analogs with the action of electric and magnetic components on masses. Gravitational waves carry away from a radiating system its energy, angular momentum, and linear momentum. The gravitational-wave field is dimensionless, and its strength is characterized by a single quantity — the GW amplitude h. The amplitude falls off in the course of propagation from a localized source, in inverse proportion to the distance traveled: $h \propto 1/r$. The difficulty of direct detection of GWs can be seen from the fact that the amplitude h of the signal from realistic astronomical sources is expected to be exceedingly small on the Earth, of the order of or smaller than 10^{-21}. The conceivable amplitudes from laboratory sources are even smaller. This small number h enters any possible scheme of detection of GWs and makes the detection difficult to achieve. For instance, GWs cause a tiny variation Δl of the distance l between two free masses: $\Delta l = hl$. In an interferometer with a 1 km arm length the variation of the distance between the two end mirrors would be of the order of $\Delta l = 10^{-16}$ cm. This tiny variation is supposed to be measured and distinguished against the background of noise. However, in cosmos, GWs are an important factor of cosmic evolution. Gravitational waves are routinely taken into account in the study of the orbital evolution of close pairs of compact stars [11]. The measured secular change of orbital parameters in the binary system of neutron stars which includes the pulsar PSR $1913 + 16$ agrees with the GW prediction of general relativity to within a 1% accuracy [12]. For the study of pulsars and this discovery, Hulse and Taylor were awarded a Nobel prize in physics in 1993.

Like any other observational science, gravitational wave astronomy deals with sources, detectors, data analysis, and interpretation. In what follows, we devote some discussion to each of these notions. However, we are not aiming to review all interesting astrophysical theories and all possible signals and detection techniques. We concentrate on sources, which, we believe, rest on the most solid theoretical foundation, are scientifically important, and involve minimal additional hypotheses.

To be interesting from the point of view of its detection, a source should be sufficiently powerful, should fall in the frequency band of the detector, and should occur reasonably often during the life-time of the instrument. The frequency range of the discussed signals is determined by the frequency intervals of the detectors' sensitivity. The currently operating bar detectors are sensitive at frequencies of about 10^3 Hz. The ground-based laser interferometers are sensitive in the interval $10 - 10^4$ Hz. The space-based laser antennas will be sensitive in the interval $10^{-4} - 1$ Hz. Great expectations are related to the forthcoming sensitive instruments. Japanese scientists have already built a 300-m laser interferometer called TAMA. A British-German collaboration is in the phase of completion of a 600m laser interferometer called GEO600 [13]. A French-Italian collaboration is building a 3-km interferometer called VIRGO [14]. The American project LIGO is building two interferometers of a 4-km arm length [15]. It is expected that these instruments will become operational in 1 – 2 years. A proposal to build a Laser Interferometer Space Antenna (LISA) [16] has been tentatively approved by the European Space Agency and NASA, and LISA may be launched in 2010 or so. There also exist plans for advanced ground-based interferometers, such as LIGO-II [17].

The ability of a given instrument to detect a signal depends on the nature of the signal. Burst sources, which accompany cosmic catastrophes, emit gravitational radiation at some characteristic frequency during only a few cycles. They may be inherently powerful, but their event rate is very low. It is very unlikely that such an event should happen in our own Galaxy during, say, a 1-year observational run. To see a few events per year, one needs to survey a large (cosmological) volume of space and, hence, to possess a sensitive instrument capable of detecting sources from this volume. Quasi-periodic astrophysical sources are expected to be more frequent than burst sources, but they produce signals much weaker in terms of h. However, the amount of energy radiated during some long time T may not be much smaller than that of a burst source. If one knows or can model the temporal structure of the signal, one can monitor the detector's output during many cycles within the observation time T. This can make a weak periodic signal not much more difficult to detect than a burst signal. Some rare but reliable astrophysical sources, such as binary neutron stars or black holes at their latest stages of evolution, emit a quasi-periodic gravitational wave signal at the inspiral phase and a signal more similar to a burst signal in the last moments of their coalescence. The stochastic backgrounds of gravitational waves are typically weak and difficult to distinguish from the instrumental noise. However, if one can cross-correlate the outputs of two or more instruments over a long integration time, the stochastic background can also be measured. Fundamentally important relic GWs form a sort of a stochastic background. They are the only direct probe of the evolution of the very early Universe, to the Planck era and Big Bang. It would be extremely valuable, even if difficult, to detect relic gravitational waves.

The balance between the expected scientific payoff and theoretical likelihood of various astrophysical sources versus their detectability by the forthcoming and planned instruments is the major thrust of this paper. Upon analyzing many possible sources of gravitational waves and taking all the factors into account, we place our emphasis on compact binaries (neutron stars and black holes) and relic GWs. In fact, we argue that inspiraling black holes, formed as a result of stellar evolution, are the most likely sources to be seen first by the forthcoming sensitive instruments. Also, we think that relic GWs are likely to be detected by the advanced ground-based and space-based laser interferometers. To justify our point of view, we go into detail in describing compact binary stars and relic gravitons.

Section 2 is devoted to the formation and evolution of binary systems. Binary stars are as numerous as single stars. Binaries emit gravitational radiation at twice their orbital frequency. To radiate large-intensity GWs at frequencies accessible to ground-based interferometers, the objects forming a pair should be massive and should orbit each other at very small separations — a few hundred kilometers. According to the existing views, such massive objects can only be the end-products of stellar evolution — neutron stars and black holes. Because of the loss of the angular momentum due to GWs, such binary objects are at their inspiral phase during the last thousands of cycles. They are only tens of minutes away from the final coalescence, i.e., from the formation of a black hole or, possibly, from another spectacular event, a gamma-ray burst. The central question is how many such close systems exist in our Galaxy and at cosmological distances. This determines the event rate — the number of coalescence events that can occur in a given volume of space during, say, 1 year. A detector, sensitive enough to see the most distant objects in this volume, will detect all of them. A detector of lower sensitivity is capable of seeing coalescing systems at shorter distances and, hence, will register a smaller number of such systems, or will not be expected to see them at all during a 1-year interval of observation.

In Section 2.1 we review all observational data on binary neutron stars. Even these data alone allow one to estimate the rate of neutron star coalescences. So far, there is no observational evidence of binaries consisting of a neutron star and a black hole or two black holes. However, we certainly do not see all products of stellar evolution in binary systems. We need to take into account the predictions of a theory which successfully explains the formation and relative abundance of various populations of observed binaries consisting of normal stars and neutron stars. Such a theory predicts the existence of close binaries including neutron stars and black holes, as the outcomes of the processes following certain evolution channels of binaries. It is these channels of evolution that are most important for gravitational wave astronomy.

Section 2.2 is devoted to the population synthesis method of describing the continuous birth and future fate of binary stars. The purpose of this analysis is to find a statistically expected number of massive and sufficiently close binaries, which could be in their final stage of inspiral at the present cosmological time. This means that we are interested only in those binaries whose expected total lifetime, from the formation to the coalescence, is shorter than the Hubble time. As usually, the results of evolution depend on the initial conditions and on the physical processes along the evolutionary path. We combine the well-established observational facts with reasonable theoretical assumptions. Two parameters are especially important — the kick velocity $\mathbf{w}$ imparted to a newly born neutron star during a supernova explosion, and the mass fraction $k_{\rm BH}$ of a pre-collapse massive star that goes into the resulting black hole. The formation of a black hole can also be accompanied by imparting some kick velocity to it. A large kick velocity can either disrupt a binary system — a would-be powerful source of GWs, or, on the contrary, make the binary orbit more eccentric, thus increasing the gravitational wave luminosity. In our evolutionary calculations we vary $\mathbf{w}$ and $k_{\rm BH}$ within the observationally allowed limits. We also take into account the stellar wind and the loss of mass as factors of binary evolution. The kick velocity as a factor of binary evolution is so important that we devote a separate Section 2.3 to its analysis.

The results of the population synthesis are summarized in Section 3. These results are at the same time our predictions for the detection rate of various compact binary inspiral signals. In a given cosmological volume of space, the estimated event rate for coalescing black holes is about 10 times lower than that for coalescing neutron stars and neutron star – black hole systems. However, since the masses of black holes are significantly larger than the masses of neutron stars, they are more luminous GW sources than pairs of neutron stars. Hence, a given detector can observe inspiraling black holes at greater distances than pairs of inspiraling neutron stars. We conclude that a network of first-generation laser interferometers are likely to see black hole inspirals more often than neutron star inspirals, and as often as 2 – 3 events per year.

Section 4 is devoted to transient and periodic sources. They include supernovae explosions, various unstable modes in rapidly rotating neutron stars, and quasi-normal modes of black hole perturbations. All these sources are interesting and potentially detectable. However, we do not place them at the beginning of our priority list. Asymmetric supernovae explosions, as well as merging events of binaries, can produce powerful bursts of gravitational radiation, but the estimates of their performance rely on factors which are not well understood theoretically and do not have much observational evidence. A merging event will probably be seen as a confirming signature of the inspiral phase, but one cannot rely on this event alone. However, if we err in our priorities, a special kind of hypernovae explosion may top the list. As for the unstable modes in rotating neutron stars, they require quite sophisticated mechanisms of excitation and can be hampered by viscosity and other physical processes. The collision of black holes and quasi-normal modes of newly born black holes is an intriguing possibility, but should probably be treated as something to be discovered by gravity wave observations, rather than reliably calculated on purely theoretical grounds.

In Section 5 we review stochastic GW backgrounds of astrophysical origin. These are overlapping signals from many individual sources. The populations of sources that we consider include unresolved binary white dwarfs in our Galaxy and at cosmological distances, and the population of rotating neutron stars. The detection of such backgrounds would carry some scientific information on its own, but it is also necessary to study these sources for another reason. Such stochastic backgrounds set a confusion limit for the detection of more interesting signals by space-based and ground-based interferometers. We conclude that LISA will be free of GW

noise from unresolved binaries at frequencies near and higher than 2×10^{-3} Hz. This noise is mostly from unresolved binaries in our Galaxy, whereas the extragalactic binaries contribute only about 10% to this noise. Any detected stochastic background at frequencies above 2×10^{-3} Hz in the LISA window of sensitivity is expected to be of primordial origin. The population of non-axisymmetric rotating neutron stars could potentially blur the view of the ground-based interferometers. However, we find that this background is below the instrumental noise level of the first-generation interferometers, and can possibly present a problem only for advanced LIGO. We estimate other stochastic backgrounds of astrophysical origin to be weaker than those we have considered.

Section 6 is devoted to relic gravitational waves. In contrast to all other sources, which are based on classical physics, the generation mechanism of relic gravitons includes some elements of quantum physics. It is the inevitable zero-point quantum oscillations of gravitational waves amplified by the strong, variable gravitational field of the early Universe that end up as the stochastic background of relic GWs measurable today. Although the existence of such a GW signal involves an extra element — quantum physics, it is no less reliable than many other sources. The generation of relic gravitons relies essentially only on general relativity and the basic principles of quantum field theory. Since the same mechanism is thought to be responsible for the generation of the primordial density perturbation seeding the formation of galaxies, we present a qualitative picture of this mechanism in Section 6.1. The calculation of the expected relic gravitational wave background is given for a class of cosmological models supported by other observations. In particular, we use data on the measured microwave background anisotropies. The results of this analysis are presented in Section 6.6. We find that, in the most favorable case, the detection of relic GWs can be achieved by the cross-correlation of outputs from two ground-based laser interferometers of the first generation. In a more realistic case, an increased sensitivity of advanced ground-based and space-based instruments will be needed. We also discuss the specific statistical properties of relic gravitons, associated with the phenomenon of squeezing. This phenomenon is also known in formal quantum mechanics and quantum optics. The manifestations of squeezing could potentially help in further improving the signal-to-noise ratio (SNR).

The problems of detectability are systematically referred to throughout the paper. However, a rigorous discussion of detectors and data analysis is concentrated in Section 7 and Section 8. Whenever we qualify a source as detectable or undetectable, we base our conclusions on the more detailed treatment presented in these sections. Section 7 gives a general description of detectors and their sensitivity curves. An important notion concerning the detection of a signal with a known or suspected temporal structure is the notion of a template. A template allows one to use the matched filtering technique (Section 8.1) in order to increase the SNR. This method will be indispensable in the search for signals from inspiraling binaries. The related issues are the practically accessible number of templates, their overlap in the parameter space, the computational cost, etc. These issues are important not only for the confident detection of a signal, but also for the extraction of astrophysical information from a signal (Sections 8.4, 8.5) — the ultimate goal of gravitational wave astronomy.

Some mathematical details on the Keplerian motion of a binary system and the gravitational reaction force are described in Appendix I. Appendix II contains some technical issues of the mass transfer modes and mass loss in binary stars. Appendix III gives post-Newtonian expressions for energy and gravitational wave flux. The main conclusions of the review are formulated in the Conclusion.

2. Astrophysical sources. Close binary neutron stars and black holes

In this section we discuss observational and theoretical estimates for the coalescence rate of close binary neutron stars and black holes. We start from a review of observational limits on the coalescence rate of binary neutron stars. Then, we describe the basics of the population synthesis of binary evolution, which allows one to theoretically predict the event rates for systems involving neutron stars and black holes. The role of the kick velocity in the binary evolution is discussed in Section 2.3. The expected detection rates in the forthcoming sensitive gravitational wave detectors are summarized in Section 3.

2.1 Observational limits on the binary neutron star coalescence rate

What do we know about compact binary stars and the rate of their mergings on *observational* grounds? More than a thousand of single neutron stars (NS) are currently observed as radio pulsars (see Ref. [18]; new data are being continuously added at *http://puppsr.princeton.edu*). In addition, about 30 NS are seen as X-ray pulsars and yet more than 100 NS are seen as burst and transient X-ray sources. These NS enter into binary systems with non-degenerate companions, that is, their companions are normal stars rather than neutron stars or black holes. Only six NS are known to enter into binary systems with another NS as a secondary component [1]. All these six systems belong to binary radio pulsars. The systems and some of their parameters are listed in Table 1. The orbital periods are given in days and masses, in units of the solar mass $M_\odot$. Three of these systems (namely, B1913 + 16, B1534 + 12 and B2127 + 11c) are close enough to merge due to GW emission in a time interval shorter than the Hubble time t_H. We loosely refer to binaries as coalescing or merging binaries if their expected lifetime until the coalescence, t_{coal}, is shorter than t_H. For numerical estimates we use the value $t_H \sim 12 \times 10^9$ yr.

Table 1. Binary PSR with NS secondaries (data from Ref. [23]).

PSR	P_{orb}, days	e	$M_1 + M_2$	M_1	M_2	t_{coal}, yr
J1518 + 4904	8.634	0.249	2.62	–	–	$\geqslant 3.6 \times 10^{12}$
B1913 + 16 †	0.323	0.617	2.8284	1.44	1.39	1.0×10^8
B1534 + 12 †	0.420	0.274	2.6784	1.34	1.34	1.0×10^9
B2127 + 11c †, ‡	0.335	0.681	2.712	1.35	1.36	8.0×10^7
B2303 + 46	12.340	0.658	2.60	–	–	$\geqslant 1.6 \times 10^{12}$
B1820-11 §	357.762	0.795	–	–	–	$\gtrsim 2.4 \times 10^{15}$

† Coalescing binary pulsars.
‡ Binary pulsar in a globular cluster.
§ The secondary companion may not be an NS.

[1] New binary pulsars were found in recent pulsar surveys (see e.g. Refs [19, 20]). However, a reliable determination of the component masses is only possible after sufficiently long observations.

Much less is known about black holes (BH). A dozen BH candidates participate in binary systems with non-degenerate companions. They are observed as persistent X-ray sources (like Cyg X-1) or X-ray transients (mostly X-ray Novae) (see Ref. [21] for a review). Neither single BH nor BH forming a binary with a radio pulsar or another BH have been found so far. The parameters of the BH candidates in binary systems are listed in Table 2. Note that, according to these data, the mean BH mass is $M_{\rm BH} \simeq 8.5 M_\odot$, i.e. notably higher than a typical NS mass $M_{\rm NS} \simeq 1.4 M_\odot$. (Of course, we mean a black hole in an astrophysical sense, i.e. as a highly compact gravitating object of a certain mass. The presence or absence of the event horizon is irrelevant to our discussion.) For a recent summary of NS mass determination see Ref. [22].

Table 2. Black hole candidates (from Ref. [21]).

System	Spectral class	$P_{\rm orb}$, days	$f_{\rm v}(m)$, $M_\odot$	$m_{\rm x}$, $M_\odot$	$m_{\rm v}$, $M_\odot$
Cyg X-1	O9.7 Iab	5.6	0.23	7–18	20–30
LMC X-3	B(3–6) II–III	1.7	2.3	7–11	3–6
LMC X-1	O(7–9) III	4.2	0.14	4–10	18–25
A0620-00	K(5–7)V	0.3	3.1	5–17	~ 0.7
GS2023 + 338	K0IV	6.5	6.3	10–15	0.5–1.0
GSR1121-68	K(3–5)V	0.4	3.01	9–16	0.7–0.8
GS2000 + 25	K(3–7)V	0.3	5.0	5.3–8.2	~ 0.7
GRO J0422 + 32	M(0–4)V	0.2	0.9	2.5–5.0	~ 0.4
GRO J1655-40	F5IV	2.6	3.2	4–6	~ 2.3
XN Oph 1977	K3	0.7	4.0	5–7	~ 0.8
Cyg X-3(?)					
Mean value of the BH mass				$\sim 8.5 M_\odot$	

There are two types of estimates of the binary NS coalescence rate. Estimates of the first type are derived directly from observations (see Table 3), while estimates of the second type are inferred from the theory of binary-star evolution (see Table 4). We will consider each of these estimates in turn.

Table 3. Observational estimates of binary NS coalescence rate.

Paper	Coalescence rate, $\rm yr^{-1}$
Phinney 1991 [24]	$1/10^6$
Narayan et al. 1991 [25]	$1/10^6$
Curran, Lorimer 1995 [26]	$3/10^6$
Van den Heuvel, Lorimer 1996 [27]	$8/10^6$
Bailes 1996 [28]	$< 1/10^5$
Arzoumanian et al. 1999 [29]	$< 1/10^4$

Table 4. Theoretical estimates of binary NS coalescence rate.

Paper	Coalescence rate, $\rm yr^{-1}$
Clark et al. 1979 [32]	$1/10^4 - 1/10^6$
Lipunov et al. 1987 [33]	$1/10^4$
Hills et al. 1990 [34]	$1/10^4$
Tutukov, Yungelson 1993 [35]	$3/10^4 - 1/10^4$
Lipunov et al. 1995 [36]	$< 3/10^4$
Portegies Zwart, Spreeuw 1996 [37]	$3/10^5$
Lipunov et al. 1996 [38]	$3/10^4 - 3/10^5$
Portegies Zwart, Yungelson 1998 [39]	$\sim 1/10^4 - 3/10^5$
Bethe, Brown 1999 [40]	$\sim 1/10^4$

The first-type estimates are based on the data on three binary radio pulsars which should merge within the Hubble time (see Table 1). These estimates use the following argumentation. The average coalescence time for these pulsars is approximately 3×10^8 years. So the binary NS merging rate based on these 3 pulsars should be approximately one event per 100 million years. As we observe only about 1% of the galactic volume, a *lower* limit for the binary NS merging rate becomes one every million years [24]. In fact, this estimate was obtained at a time when only two of the three presently known coalescing binary radio pulsars were discovered. Taking into account the spatial distribution of pulsars inside the Galaxy and the fact that a typical radio pulsar switches off long before the coalescence, the lower limit for the NS merging rate can be increased by almost an order of magnitude [27], thus reaching 10^{-5} per year.

An interesting *upper* limit, the so-called 'Bailes limit', was derived from independent arguments [28]. It was noted that the properties of the pulsars in the three coalescing binary radio pulsars (most of all, their surface magnetic fields) are quite different from those found in ordinary single radio pulsars. Since the number of single radio pulsars is about 1000, it was estimated that merging binary NS containing pulsars similar to single pulsars should form at least 1000 times more rarely than single radio pulsars. Taking the birth rate of single pulsars from a large sample of known pulsars, Bailes found an upper bound for the birth rate of binary NS to be (formation rate of single pulsars) $\times$ (number of pulsars with ordinary properties among binary radio pulsars) $= (1/60\ \rm yr^{-1}) \times (1/1000) \simeq 2 \times 10^{-5}\ \rm yr^{-1}$.

One should note, however, that both estimates — that based on the statistics of coalescing binary radio pulsars, and the other based on the Bailes limit — suffer from selection effects. They depend on the pulsar distances (in some cases known no better than to a factor of 2) [2], on the characteristic pulsar lifetime (known no better than to an order of magnitude), and on the differences between the properties of single and binary pulsars. So, one cannot infer from observations an absolutely reliable estimate of the binary NS merging rate. Indeed, a recent re-assessment of the Bailes limit [29] taking into account the current pulsar numbers and a reduction for search sensitivity to short orbital period binaries gave $10^{-4}\ \rm yr^{-1}$ as an upper limit for the binary NS galactic merging rate. An alternative way of deriving the upper limit based on the empirical pulsar birth rate and the theoretical understanding of binary NS formation was used in Ref. [31] to yield a few mergers per 10^5 years. The cited paper clearly demonstrates that (1) there is a steady tendency to *increasing* the empirical upper limit of the binary NS coalescence rate and (2) various selection effects generic to radio pulsar surveys and the lack of detailed knowledge of galactic pulsar population properties still prevent us from deriving a fully reliable estimates.

2.2 Population synthesis of coalescing binary NS and BH

Now we turn to estimates partially based on theoretical grounds. The merging rates of binary compact stars have been calculated by different independent research groups, mostly with the help of population synthesis numerical simulations (see Table 4). The reliability of these results depend on whether the binary evolution scenarios properly reflect different aspects of the real, observed populations. It is

[2] For example, recent observations of PSR1534 + 12 [30] suggest a distance two times larger than assumed previously, so the 'observational estimate' of binary NS mergings should be decreased by a factor of two.

encouraging that these independent calculations yield similar results.

Theoretical estimates of the double NS coalescence rate are systematically higher than the observational ones by, on average, an order of magnitude. This does not mean that the estimates are in conflict with each other. The main reason for the discrepancy is that the observational estimates directly refer only to the merging rates of binary NS in which one of the components is a radio pulsar. The participation of a pulsar in a binary system is of course not a *necessary* condition for the system to be interesting from the point of view of gravitational wave astronomy. For example, neutron stars could be born with weak magnetic fields and/or slow rotation and, hence, they would never manifest themselves as radio pulsars. Theoretical calculations provide a broader range of estimates because they depend on several evolutionary parameters which are not well known. However, the population synthesis calculations are the only way to estimate the coalescence rates of, so far unobserved, compact binaries consisting of two BH or an NS and a BH. Such systems are of great importance for gravitational wave astronomy. We will consider below, in Section 2.3, the results of the population synthesis for all possible pairs: NS + NS, NS + BH, and BH + BH.

Basics of population synthesis. The coalescence rates of different types of binary compact stars are calculated using the modern theory of binary star evolution (e.g. Ref. [41] and references therein). A full description of the method can also be found in Ref. [38]. Some key formulas for binary system evolution are summarized in Appendix II.

Binary stars are formed with different initial masses, semi-major axes, eccentricities, etc.. These initial parameters are deducible from certain distribution laws. Also, there are some other physical parameters important for binary evolution, such as the efficiency $\alpha_{\rm CE}$ of angular momentum removal in the common envelope stage, or the kick velocity distribution $f(\mathbf{w})$ for newly born neutron stars. This means that, in order to calculate the expected rate of binary star mergers, we need to derive the number of binaries formed in all appropriate regions of the parameter space, and then integrate over these regions. Some distributions can be deduced more or less accurately from astronomical observations. Among the are the initial stellar mass function and the distribution over binary semi-major axis. The distributions of other physical parameters can be found on theoretical grounds. We performed evolutionary calculations using the 'Scenario Machine' code — a version of the Monte Carlo method (see [38] for a review). In a typical numerical experiment, about 10^6 binary evolutionary tracks with different initial conditions are calculated. Similar approaches in binary evolution studies are used by other groups (e.g. Ref. [39]) and are commonly named 'population synthesis' methods.

Initial binary parameters. The initial components of a binary system are taken to be zero age main-sequence stars. The initial parameters which determine the subsequent binary evolution are the mass of the primary component M_1, the binary mass ratio $q = M_2/M_1 < 1$, and the orbital separation a. For sufficiently close binaries, which are capable of producing a merging NS or BH, we assume that the orbits are circular. This assumption is justifiable since the tidal interaction between the components is efficient enough to circularize the orbit.

The distribution of binaries over initial orbital separations is partially known from observations [42]:

$$f(\log a) = \text{const}, \text{ for } a \text{ such that}$$
$$\max\left\{10R_\odot, \text{ Roche Lobe } (M_1)\right\} < a < 10^7 R_\odot . \qquad (1)$$

We assume that the distribution of masses of the primary (more massive) components can be described by the Salpeter mass function found for the birth rates of main-sequence stars in the solar vicinity [43]:

$$f\left(\frac{M_1}{M_\odot}\right) = \left(\frac{M_1}{M_\odot}\right)^{-2.35} . \qquad (2)$$

The observed star formation rate in the galactic disk relates to this distribution as

$$\frac{{\rm d}N}{{\rm d}(M_1/M_\odot)\,{\rm d}t} = 0.9\ {\rm yr}^{-1} f\left(\frac{M_1}{M_\odot}\right),$$
$$0.1M_\odot < M_1 < 120M_\odot . \qquad (3)$$

Assuming that 50% of the total number of stars in the Galaxy reside in binaries, this distribution law predicts one massive star (with $M_1 > 10M_\odot$, capable of producing a compact remnant) to form in a binary system, approximately every 60 years. This estimate agrees with the binary birth rates derived from observations of binary stars [44].

The initial mass ratio q in a binary is crucial for its subsequent evolution [45] because it determines the mode of the initial mass transfer between the components. The initial distribution over q has not been reliably derived from observations due to various selection effects. Usually, one uses the 'zero-order assumption' according to which the mass ratio distribution is flat, i.e. low mass ratio binaries are formed as frequently as those with equal masses (e.g. Ref. [41]):

$$\frac{{\rm d}N}{{\rm d}q} = \text{const}, \qquad q \equiv \frac{M_2}{M_1} < 1 . \qquad (4)$$

In the calculations presented below in Section 2.2 we used this prescription. The effect of the initial q distribution on binary pulsar statistics was studied in Ref. [46].

Neutron star kick velocities. In the course of evolution of massive binary stars, one or two neutron stars can form. One of the most important parameters affecting the eventual binary NS coalescence rate is the kick velocity $\mathbf{w}$ imparted to an NS at its birth. There exists a plenty of observational evidence for the presence of kick velocities. The impact of a kick velocity of ~ 100 km s^{-1} explains the precessing binary pulsar orbit in PSR J0045-7319 [47]. Evidence for the presence of a kick velocity is seen in the inclined, with respect to the orbital plane, circumstellar disk around the Be star SS 2883 — an optical component to the binary pulsar PSR B1259-63 [48]. Further direct evidence comes from observations of the geodetic precession in the famous binary pulsar PSR 1913 + 16 [49, 50]. All these observations indicate that, in order to produce the observed misalignment between the orbital angular momentum and the neutron star spin, a component of the kick velocity perpendicular to the orbital plane is required. A non-zero kick velocity is also required in order to properly explain the observed pulsar velocity distribution [3].

[3] For an alternative point of view see Ref. [51] and for its criticism, [52].

Most likely, the origin of the kick velocity should be attributed to an asymmetry in the supernova explosion. Astrophysical evidence for the existence of a substantial kick velocity during supernova explosions was discussed in Ref. [53] and recently summarized in Refs [54, 55]. However, the nature of the particular physical mechanisms giving rise to the kick velocity is still unclear (see Ref. [56] for a recent review). A promising possibility capable of producing small and moderate kick velocities (up to 100 km s^{-1}) involves asymmetric neutrino emission in the strong magnetic field of a newly born NS [57, 58].

Spatial velocities of NS are usually derived from direct observations of proper motions of single radio pulsars [59, 60] or, with more uncertainty, from the observed offsets in the positions of young pulsars relative to the centers of their associated supernova remnants (e.g. Ref. [61]). Both methods determine, however, only the component of the space velocity transverse to the line of sight. One should also bear in mind the existing uncertainty in pulsar distances, which affects the evaluation of the kick velocity. On average, the uncertainty in the distance scale is $\sim 30\%$ [62] and can be as much as a factor of two in individual cases. In general, the observed distribution of transverse pulsar velocities is recognized to have (1) a high mean velocity value ($\sim 200-350$ km s^{-1}) and (2) a broad shape with a high-velocity tail stretching to 1500 km s^{-1}.

It is a difficult problem to derive the intrinsic kick velocity distribution from these data. If all the pulsars presently seen as single pulsars took their origin from single massive stars, their velocity distribution would exactly reflect the initial kick velocity distribution. This is because single massive stars have very small (of order 10 km s^{-1}) spatial velocities. However, when a supernova explosion occurs in a binary system, which can be disrupted as a result of the explosion, the neutron star can acquire a substantial spatial velocity, equal to the orbital velocity of the progenitor, even without any additional kick velocity. If a non-zero kick is present, it becomes virtually impossible to analytically solve the inverse problem for the intrinsic kick velocity distribution from the observed pulsar velocities.

So, the only way to check the very assumption of a non-zero kick is to find numerically the pulsar velocity distributions arising from various theoretical kick distributions, and compare the results with observations. This is usually done by Monte Carlo simulations of binary evolution. Presently, there is no general agreement with regard to the form of the kick distribution. A Maxwellian distribution for $w = |\mathbf{w}|$,

$$f_{\mathrm{M}}(w) \propto w^2 \exp\left(-\frac{w^2}{w_0^2}\right) \tag{5}$$

with $w_0 = 190$ km s^{-1}, was used by Hansen and Phinney [63] in order to fit the observed pulsar velocities. A different form of the kick velocity distribution was suggested in [46]. It was found that this latter distribution fits well with the observed 2D pulsar velocity distribution given by Lyne and Lorimer [59]. In contrast to a Maxwell-like distribution, the proposed distribution function has a power-law shape:

$$f_{\mathrm{LL}}(|\mathbf{w}|) \propto \frac{(|\mathbf{w}|/w_0)^{0.19}}{\left[1+(|\mathbf{w}|/w_0)^{6.72}\right]^{0.5}} \tag{6}$$

with $w_0 \approx 400$ km s^{-1}. In our calculations described below, we have used this form of the NS kick velocity distribution with w_0 treated as a free parameter. A study of the kick velocity effects on the binary neutron star merging rate can be found in Refs [64, 65]. We give a detailed analysis of these effects in Section 2.3.

Binary neutron star formation and merging. We are interested in evolutionary tracks which lead to the formation of a pair of coalescing NS. Detailed studies of possible evolutionary channels which produce merging binary NS can be found in the literature (e.g. Refs [35, 39, 64, 66, 67]). Usually, an evolutionary analysis is done in the following order: one starts from the observed parameters of the binary and tries to deduce the parameters of the supernova progenitor and then, the initial binary masses and orbital separation. In contrast, the Monte Carlo population synthesis method, which we apply, evolves a trial binary and looks for appropriate results by changing the initial parameters within their distributions. One such typical track calculated by us is sketched in Fig. 1, and we will explain it in detail.

Close NS binaries originate from two sufficiently massive main-sequence stars separated by a moderate distance of order 100 solar radii (1st row in Fig. 1). To get an NS in the course of evolution, the mass of the progenitor star must

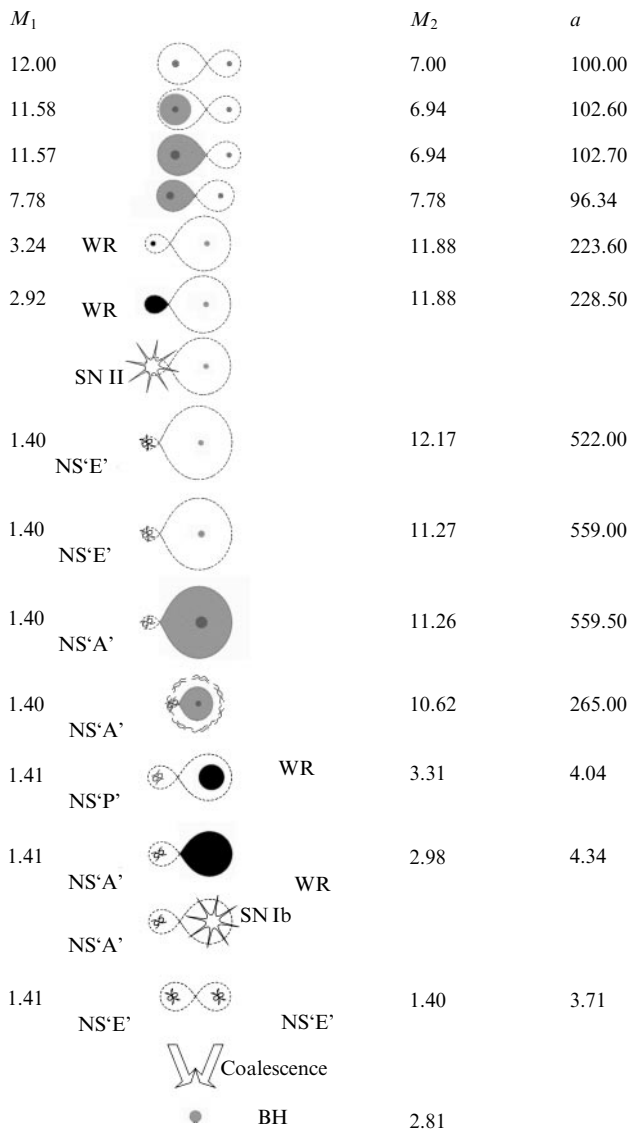

Figure 1. Evolutionary track of a massive binary star leading to the formation and coalescence of two NS.

initially be larger than $\approx 10M_\odot$ or, in any case, taking into account a possible mass transfer in a close binary, the mass should be $\approx 10M_\odot$ during the stage of nuclear burning. The more massive the primary star, the faster it evolves. For main-sequence stars, the time of core hydrogen burning is $t_{\rm nucl} \propto M^{-2}$. The star burns out its hydrogen in its central parts, so that a dense central helium core with a mass $M_{\rm He} \sim 0.1(M/M_\odot)^{1.4}$ forms by the time the star leaves the main sequence. The outer shell expands and the star moves towards the red super-giant region in the Hertzsprung – Russell diagram. At some stage of its evolution the star fills out its Roche lobe [see Eqn (A.42)] (3-rd row in Fig. 1). The hydrogen envelope starts outflowing onto the secondary, less massive star, which still resides on the main sequence. The primary star is continuously stripped off of its hydrogen envelope, until a naked helium core emerges. This core can be observed as a hot compact helium star, or, for more massive stars, as a Wolf – Rayet star with an intense stellar wind (5th row).

While the mass of the primary star reduces, the mass of the secondary star increases, since the mass transfer at this stage should be quasi-conservative. For not-too-massive main-sequence stars, $M \lesssim 20M_\odot$, no significant mass loss due to stellar wind occurs, which could, otherwise, remove matter from the binary. The secondary star acquires a large angular momentum due to the infalling material, so that its outer envelope can be spun up to an angular velocity close to the limiting (Keplerian orbital) value. Such massive, rapidly rotating stars are observed as Be-stars. During the conservative stage of mass transfer, the semi-major axis of the orbit first decreases, reaches a minimum, when the masses of the binary components become equal, and then increases. Such a behavior is dictated by the angular momentum conservation law [Eqn (A.26)]. After the completion of the conservative mass transfer, the star that was initially more massive becomes less massive than its initially lighter companion. The parameter $q = M_2/M_1$ becomes larger than 1. In a short time, typically of 10% of the hydrogen burning time, the nuclear evolution of the helium star is completed and, provided its mass is larger than $2-3M_\odot$, the star explodes as a type-II core-collapse supernova leaving a neutron star as its remnant.

Even for asymmetric supernova explosions, most such binaries do not get disrupted. This is because the mass ratio $q = M_2/M_1$ of the pre-supernova binary becomes generally high, $q \sim 3-5$. After the first SN explosion, the binary system consists of a Be-star and an NS in an elliptical orbit (8th row). The orbital evolution following the SN explosion is described in more detail in the Appendix [Eqns (A.35) – (A.40)].

Be-stars have very rapidly rotating envelopes but in other respects they do not differ from ordinary main-sequence stars. After the completion of hydrogen burning in the core, a Be-star starts expanding until it fills out its Roche lobe while passing through the periastron of an elliptical orbit (10th row). This initiates the second episode of mass transfer, which takes place on the thermal scale of the Be-star, typically $10^{-6}M_\odot$ per year. However, this mass transfer is qualitatively different from the first one, since the mass is now transferred onto a compact star. Once the accretion rate exceeds the value which provides a luminosity equal to the Eddington luminosity limit near the NS surface ($\sim 10^{-8}M_\odot$ per year), the NS cannot accrete all the infalling matter. The so-called common envelope stage sets in (11th row), during which the neutron star finds itself inside quite dense outer layers of the companion star. Numerical hydrodynamic calculations [68, 69] show that the dynamical friction of the orbiting NS leads to an efficient transfer of orbital angular momentum to the common envelope, thus dispersing this envelope on a very short timescale (typically, $10^3 - 10^4$ yr). The semi-major axis of the binary system reduces dramatically [Eqn (A.41)], which results in the formation of a close binary system consisting of an NS and a WR star (12th row). Alternatively, the NS can sink into the center of its red giant companion (a so-called Thorne – Zytkow object; not shown in this figure).

In a short time ($\lesssim 10^5$ yr), the companion WR star explodes as a type-Ib supernova, thus producing a second neutron star. During this explosion, the system is more likely to be disrupted than during the first SN explosion, since the exploding star is now more massive than its companion. The surviving systems form close highly eccentric NS binaries, similar to the NS binary PSR1913 + 16. The orbital parameters of such binaries change exclusively due to the emission of GWs (see Appendix I). If the neutron stars are close enough, they coalesce in a time shorter than $t_{\rm H}$.

Black hole formation parameters. So far, we have considered the formation of individual NS and their binaries. It is believed that very massive stars end their evolution with the formation of stellar mass black holes. We will discuss now the formation of an individual black hole.

In the analysis of BH formation, new important parameters appear. The first one is the threshold mass $M_{\rm cr}$ beginning from which a main-sequence star, after the completion of its nuclear evolution, can collapse into a BH. This mass is not well known, and different authors assume different values: $40M_\odot$ [70], $60M_\odot$ [71], or more than $20M_\odot$ [72]. A simple physical argument usually cited in the literature is that the mantle of the main-sequence star with $M > M_{\rm cr} \approx 30M_\odot$ is bound before the collapse with a binding energy well above 10^{51} ergs (typical observed supernova energy), so that the supernova shock is not strong enough to expel the mantle.

The second parameter is the mass $M_{\rm BH}$ of the formed BH. There are various studies of what the mass of the BH should be (e.g. Refs [40, 73 – 75]). In some papers a typical BH mass was found not to be much higher than the upper limit for the NS mass (the Oppenheimer – Volkoff limit $\sim 1.6 - 2.5M_\odot$, depending on the unknown equation of state for the neutron star matter), even if fallback accretion onto the supernova remnant is allowed [73]. However, observations strongly indicate much higher masses of BH candidates, of the order of $6 - 10M_\odot$ (see Table 2). To obtain such BH masses, it is sometimes assumed [40] that $M_{\rm cr} \sim 80M_\odot$. Recently, a continuous range of BH masses up to $10 - 15M_\odot$ was derived in calculations [75]. Since the present day calculations are still unable to reproduce self-consistently even the supernova explosion, we have parameterized the BH mass $M_{\rm BH}$ by specifying the fraction of the pre-supernova mass M_* that collapses into the BH: $k_{\rm BH} = M_{\rm BH}/M_*$. In fact, the pre-supernova mass M_* is directly related to $M_{\rm cr}$, but the form of this relationship is somewhat different in different scenarios of massive star evolution. According to our parameterization, the minimal BH mass can be $M_{\rm BH}^{\rm min} = k_{\rm BH}M_*$, where M_* itself depends on $M_{\rm cr}$. We have varied $k_{\rm BH}$ over a wide range, from 0.1 to 1.

The third parameter, as in the case of NS formation, is a possible kick velocity $\mathbf{w}_{\rm BH}$ attributed to a newly formed BH. In general, one expects that a BH should acquire a smaller

kick velocity than an NS, since black holes are more massive than neutron stars. In our calculations we adopted the relation

$$\frac{w_{\rm BH}}{w_{\rm NS}} = \frac{M_* - M_{\rm BH}}{M_* - M_{\rm OV}} = \frac{1 - k_{\rm BH}}{1 - M_{\rm OV}/M_*}\,, \qquad (7)$$

where $M_{\rm OV} = 2.5M_\odot$ is the maximum NS mass. When $M_{\rm BH}$ is close to $M_{\rm OV}$, the ratio $w_{\rm BH}/w_{\rm NS}$ approaches 1, and the low-mass black holes acquire kick velocities similar to those of neutron stars. When $M_{\rm BH}$ is significantly larger than $M_{\rm OV}$, the parameter $k_{\rm BH} = 1$, and the BH kick velocity becomes vanishingly small[4]. As we show below in Section 2.3, the allowance for a quite moderate $w_{\rm BH}$ strongly increases the coalescence rate of binary BH. A recent analysis of spatial velocities of some BH candidates did not reveal any need for a non-zero $w_{\rm BH}$ [76]. However, other studies show that some kick velocity can arise during the BH formation, and its presence does not contradict the observational data [77]. From a theoretical point of view, the presence of a moderate kick velocity imparted to a BH during its formation seems very plausible [75].

Binary black hole merging with $w_{\rm BH} = 0$: a typical example. We begin from the simplest assumption $w_{\rm BH} = 0$. More realistic cases $w_{\rm BH} \neq 0$ will be considered in Section 2.3. In contrast to NS + NS binaries, BH + BH and BH + NS binary systems have not been observed so far. There is no way of recovering from observations the range of progenitors for such binary systems. We can only apply the population synthesis method and derive theoretically the parameters of all binaries, including BH + BH and BH + NS pairs, that should be produced at the end of the evolution of very massive binary stars.

In addition to the evolutionary uncertainties existing for stars evolving to binary NS systems, new uncertainties arise for very massive stars with initial masses $M \gtrsim 40M_\odot$. First of all, a large mass loss via stellar wind is observed for such stars. According to current views, a massive single star can lose more than a half of its initial mass while on the main-sequence. Further rapid mass decrease is expected during the helium star stage. There is no general agreement as to how to describe the mass loss of a massive star. Yet, one can consider two extreme cases for the mass loss via stellar wind: a slow mass loss and a very fast one. Since an accurate description of the stellar wind mass loss is not known, we have considered both options in our numerical simulations (for more detail, see Ref. [64]). A typical evolutionary track that leads to the formation of a coalescing binary BH system according to a low mass-loss scenario is shown in Fig. 2.

Before proceeding further with evolutionary calculations, let us explain qualitatively how a merging BH binary can form even in the case of an extremely high mass loss. The high mass loss makes the binary system wider, according to Eqn (A.29), so it should be quite tight before the phase of active mass loss. Consider two already existing black holes, of mass $10M_\odot$ each, in a circular orbit. The orbital separation should be $< 20R_\odot$ for the binary to merge within the time $t_{\rm H}$ (see Fig. 19 in Appendix I). This means that the radius of a Wolf–Rayet (helium) star that collapses second in the course of a binary evolution, to form a BH, should have a mass less than $10R_\odot$. The pre-collapse mass of the star cannot be smaller than $10M_\odot$, since $k_{\rm BH} \leqslant 1$. Such massive helium stars have very small radii ($\sim 1R_\odot$) and do not expand too much before the collapse, so the requirement $R < 10R_\odot$ is fulfilled. The lifetime of a massive helium star is about 10^5 yr, since it loses mass at a high rate of $10^{-5}M_\odot$ per year. The star can lose a sizable fraction, for instance, a half, of its mass before the collapse. Thus, we will deal with a $20M_\odot$ helium star in a pair with the first, $10M_\odot$, BH in a circular orbit of radius $a \approx 13R_\odot$. (We applied Eqn (A.29) to calculate the radius of the resulting orbit.) Note that the $8R_\odot$ Roche lobe of a $20M_\odot$ helium star is still quite large. To form such a close WR + BH binary, a common envelope stage is needed. The $20M_\odot$ He core corresponds to at least a $55M_\odot$ main-sequence star, as follows from Eqn (A.44). According to the models [78], a massive star loses about a half of its initial mass on the main sequence; thus, to form a common envelope with a $10M_\odot$ BH, the star has to lose $\sim 25M_\odot$ while on the main sequence. This means that the common envelope stage should start with a $30M_\odot$ red super-giant filling its Roche lobe and having a $10M_\odot$ BH as a companion. The mass ratio in such a system is high enough for the common envelope to develop. The orbital separation at the common envelope stage should decrease by a factor of 6–12, according to Eqn (A.43), depending on the parameter $\alpha_{\rm CE}$ and the exact value of the red giant mass. So, before the common envelope stage, the orbital separation should be $\sim 130R_\odot$. The orbit should be somewhat smaller (i.e., about $120R_\odot$ or less) when the first BH forms, because of the strong wind from the red giant and the loss of mass. And in order to collapse first, the primary star must have a mass of

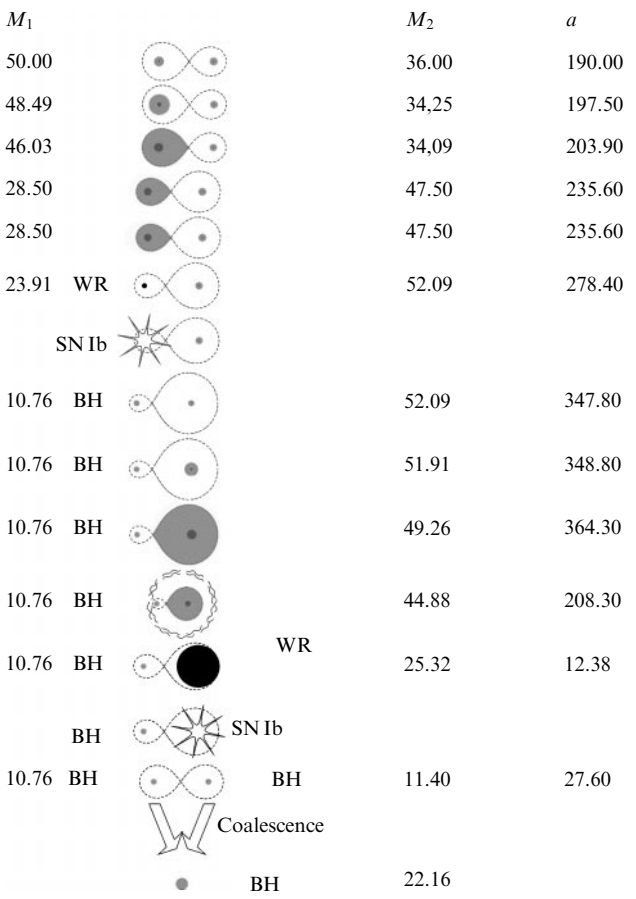

Figure 2. Evolutionary track of a massive binary star leading to the formation and coalescence of two BH. The scenario of a low stellar wind mass loss is used.

[4] Other possible relationships between $w_{\rm BH}$ have also been checked, but their different forms do not affect the results significantly.

at least $60M_\odot$. Assuming an isotropic stellar wind and using Eqn (A.29) again, we conclude that the system could become at most $(60+55)/(10+30)\approx 3$ times wider since the time of its formation, i.e. the initial separation of the progenitor binary should be larger than $40R_\odot$. The initial separation of $50R_\odot$ is sufficient to harbor two $60M_\odot$ stars, since their radii are less than $20R_\odot$ on the main sequence. Even though such initially close massive binaries are rare, they should exist. Thus, we see that some fraction of massive binary stars should end up as sufficiently close pairs of black holes.

2.3 Effects of the kick velocity

The picture outlined above changes if a non-zero kick velocity is present in the process of formation of an NS or a BH. This, in turn, has a significant effect on the expected rate of compact binary mergings, which is of primary interest for our study. In general, the formation of a compact object (NS or BH) is accompanied both by mass loss from the system and by a kick velocity. The effects of kick velocity during supernova explosions have been considered in many papers (see e.g. Refs [79, 80] and also Appendix II). General formulae for the conditions of system disruption and for the parameters of the resulting elliptical orbit, if the system remains bound, are derived in Appendix II; see Eqns (A.35), (A.36), (A.38). Here we will present qualitative arguments enabling the reader to see the main consequences of a non-zero kick velocity. We restrict our attention to circular orbits and assume equal probabilities for all possible orientations of the kick velocity vector **w**. We argue that a moderate (not too large) kick velocity increases the rate of binary mergings. This happens because a moderate kick velocity does not change the likelihood of the system's disruption too much, but, at the same time, always makes the periastron of the resulting elliptical orbit smaller than it would be without a kick. As a result, some of the binaries whose coalescence time without a kick would be longer than the Hubble time t_H now get a chance to merge in a time shorter than t_H. This increases the number of detectable gravitational wave sources.

Effect of the kick velocity on the disruption of a binary system. The collapse of a star to a BH, or its explosion leading to the formation of an NS, are normally considered as instantaneous. This assumption is well justified in binary systems, since typical orbital velocities before the explosion do not exceed a few hundred km s^{-1}, while most of the mass is expelled with velocities of several thousand km s^{-1}. If the exploding star M_1 leaves a remnant M_c, then the binary loses the portion $\Delta M = M_1 - M_c$ of its mass. The relative velocity of the stars before the event is

$$V_i = \sqrt{\frac{G(M_1+M_2)}{a_i}}\,. \tag{8}$$

Right after the event, the relative velocity is

$$\mathbf{V}_f = \mathbf{V}_i + \mathbf{w}\,. \tag{9}$$

Depending on the direction of the kick velocity vector **w**, the absolute magnitude of $\mathbf{V}_f$ varies from the smallest value $V_f = |V_i - w|$ to the largest value $V_f = V_i + w$. The system gets disrupted if (see Appendix II)

$$V_f \geqslant V_i\sqrt{\frac{2}{\chi}}\,, \tag{10}$$

where $\chi \equiv (M_1+M_2)/(M_c+M_2)$.

Let us start from the limiting case where the mass loss is virtually zero ($\Delta M = 0$, $\chi = 1$), while a non-zero kick velocity can still be present. This is a model for BH formation with $k_{BH} = 1$. It follows from Eqn (10) that, for relatively small kicks $w < (\sqrt{2}-1)V_i$, the system always (independently of the direction of **w**) remains bound, while for $w > (\sqrt{2}+1)V_i$ the system always disintegrates. By averaging over equally probable orientations of **w** with a fixed amplitude w, one can show that in the particular case $w = V_i$ the system disrupts or survives with equal probabilities. If $V_f < V_i$, the semi-major axis of the system becomes smaller than the original binary separation, $a_f < a_i$ [see Eqn (A.35)]. This means that the system becomes more bound than before, i.e. it has a greater negative total energy than the original binary. If $V_i < V_f < \sqrt{2}V_i$, the system remains bound, but $a_f > a_i$. For small and moderate kicks $w \lesssim V_i$, the probabilities for the system to become more or less bound are approximately equal.

In a more general case, the binary system loses some fraction of its mass ΔM. For BH formation this corresponds to $k_{BH} < 1$. In the absence of a kick velocity, the system remains bound if $\Delta M < M/2$ and gets disrupted if $\Delta M \geqslant M/2$ (see Appendix II). Clearly, a properly oriented kick velocity (directed against the vector $\mathbf{V}_i$) can keep the system bound, even if it would be disrupted without the kick. And, on the other hand, an unfortunate direction of **w** can disrupt the system, which otherwise would stay bound.

Consider, first, the case $\Delta M < M/2$. The parameter χ varies over the interval from 1 to 2, and the escape velocity V_e varies over the interval from $\sqrt{2}V_i$ to V_i (see Appendix II). It follows from Eqn (A.39) that the binary always remains bound if $w < V_e - V_i$, and always disintegrates if $w > V_e + V_i$. This is a generalization of the formulae derived above for the limiting case $\Delta M = 0$. Obviously, for a given w, the probability of the system disrupting or becoming less bound increases as ΔM becomes larger. Now turn to the case $\Delta M > M/2$. The escape velocity of the compact star becomes $V_e < V_i$. The binary is always disrupted if the kick velocity is too large or too small: $w > V_i + V_e$ or $w < V_i - V_e$. However, for all intermediate values of w, the system can remain bound, and sometimes even more bound than before, if the direction of **w** proves to be approximately opposite to $\mathbf{V}_i$. A detailed calculation of probabilities for the binary's survival or disruption requires integration over the kick velocity distribution function $f(\mathbf{w})$ (see e.g. Ref. [80]).

Effect of the kick velocity on the coalescence rate of compact binary systems. Here we consider binary systems that were not disrupted during the formation of a compact object. The parameters a_f and e of the resulting elliptical orbit are determined by Eqns (A.35) and (A.36). The distance of the closest approach between the stars is given by the orbit's periastron $a_p = a_f(1-e)$. It follows from Eqns (A.35), (A.36) that $a_p = a_i$ in the absence of a kick velocity. The importance of the kick velocity $w \neq 0$ lies in the fact that, although the semimajor axis can increase or decrease under the action of the kick, the periastron distance always becomes smaller: $a_p < a_i$. This relationship follows from the combination of Eqns (A.35) and (A.36) with the requirement that the system remain bound, i.e. the quantities that appear in (A.38) satisfy the opposite inequality (see Appendix II). The decrease of the periastron distance plays an important role in the subsequent evolution of the binary, which now consists of a newly born compact star and its companion.

Consider, first, a normal star as the companion. Since the kick diminishes the periastron distance, as compared with the no-kick case, the normal star, while passing through the periastron, will fill out its Roche lobe in a shorter time than in the absence of the kick. After the tidal circularization of the orbit, a tighter binary is formed. Accordingly, the subsequent common envelope stage makes the binary tighter than it would otherwise be [see Eqn (A.43)]. As a result, the final binary system, consisting of two compact objects, will coalesce in a shorter time due to GW radiation [see Eqn (A.22)]. In other words, some of the binaries, which would be too broad to coalesce in $t_{\rm H}$, become detectable sources of GW with the help of a moderate kick velocity. If the companion is already a compact star, the orbital evolution is driven exclusively by GW emission (Appendix I). Unless the kick velocity is so big that it makes the semimajor axis $a_{\rm f}$ very large, these binaries will also merge in a time interval shorter than that following from the evolution without a kick.

These qualitative considerations explain the outcomes of numerical simulations with many trial systems. We are interested in the results averaged over many systems with different input parameters. These results are presented below. As could be expected, a moderate kick velocity increases, on average, the rate of compact star mergings.

Coalescence rates of compact binaries. We can now present the results of our numerical calculations for the coalescence rate of compact binaries in a typical galaxy [81]. The total mass of a model galaxy is assumed to be $10^{11}M_{\odot}$. We adopt a constant star formation rate defined by Eqn (3). It is believed that this relation reflects well the situation in a galaxy like our own Milky Way.

In Fig. 3 we plot the NS + NS, BH + NS, and BH + BH merging rates as functions of the kick velocity parameter w_0 in the distribution (6). The calculations were performed for discrete values of w_0, but the resulting points were joined by smooth curves. The BH formation parameters were taken from the range $M_* = 15-50M_{\odot}$ with $k_{\rm BH} = 0.25$. Both the high-mass-loss and the low-mass-loss stellar winds were considered. The broad range of M_* and the uncertainty in specifying the stellar winds contributed to the spread of the results for BH + NS and BH + BH systems. The NS + NS systems arise from relatively low mass stars, so they are less sensitive to the uncertainty in the stellar wind. It is seen from Fig. 3 that the NS + NS rate lies in the range from $\sim 3\times10^{-4}$ to $\sim 3\times10^{-5}$ per year. The rates of BH + NS and NS + NS mergings are 10 – 100 times lower. For the limiting case of zero kick velocity ($w_0 = 0$) our rates agree with the independent estimates of Tutukov and Yungelson [82]. In the same limit ($w_0 = 0$), our rate for NS + BH binaries ($\sim 10^{-6}$ per year) is smaller than the estimate by Bethe and Brown [40], who obtained the rate $\sim 2\times10^{-5}$ per year. However, we believe that their estimate was derived from a somewhat simplified picture of binary evolution.

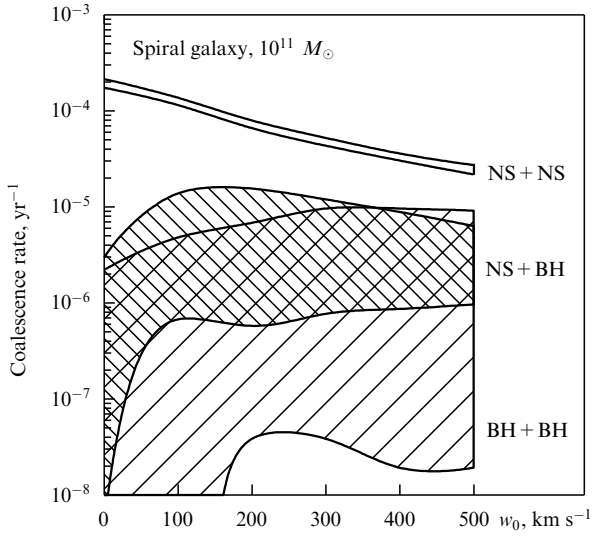

Figure 3. NS + NS, BH + NS, and BH + BH merging rates in a $10^{11}M_{\odot}$ galaxy as functions of the kick velocity parameter w_0 for a Lyne-Lorimer kick velocity distribution (6). The star formation rate in the galaxy is assumed constant. The BH formation parameters are $M_* = 15-50M_{\odot}$, $k_{\rm BH} = 0.25$.

As expected, the BH + NS and BH + BH rates have a tendency to grow with the increase of the kick velocity from zero. This is seen in the graph as the rise of the NS + BH and BH + BH curves for small and moderate w_0 (up to $w_0 \sim 100$ km s^{-1}). For much larger values of w_0, the kick velocity contributes mostly to the disruption of binary systems, and this is why the curves have a tendency to turn down. Generally speaking, the NS + NS rate should also grow with small increases in w_0 from zero. However, since the NS mass is smaller than the BH mass, the increase of the NS + NS rate takes place only for small values of w_0, not resolvable in the graph. For larger values of w_0, the kicks mostly disrupt the binaries, and the NS + NS curve goes down. The value of w_0 preferred by radio pulsar observations lies in the range 200 – 400 km s^{-1}.

For the broad range of parameters used and despite all the remaining uncertainties, the results of evolutionary calculations show that the number of coalescing BH + BH pairs is only a factor of 10 – 100 smaller than the number of coalescing NS + NS pairs. This relationship has a simple explanation and can be traced back to the initial conditions of star formation. The line of argumentation is as follows. Let us take the NS mass to be $1.4M_{\odot}$ (a typical mass well confirmed by existing observations) and the BH mass to be $8.5M_{\odot}$ (the mean value for BH candidates from Table 2). Assume that the lower bound for the initial mass of NS progenitors is $M_{\rm NS} \approx 10M_{\odot}$, while the threshold for BH formation is at the maximum of the estimates quoted above: $M_{\rm BH} > M_{\rm cr} = 80M_{\odot}$. Applying the Salpeter initial mass function for the formation rate of stars in the Galaxy [see Eqn (2)]

$$\frac{{\rm d}N}{{\rm d}t\,{\rm d}(M/M_{\odot})} \simeq \left(\frac{M}{M_{\odot}}\right)^{-2.35} {\rm yr}^{-1}$$

and using the lower limits of integration, one finds

$$\frac{N(M>80M_{\odot})}{N(M>10M_{\odot})} = \left(\frac{80M_{\odot}}{10M_{\odot}}\right)^{-1.35} \simeq 0.06\,.$$

This ratio should also be valid for binary stars. It is reasonable to expect that, despite the differences in and complexities of binary evolution, the ratio of coalescence rates will approximately be given by the same quantity,

$$\frac{\mathcal{R}_{\rm BH}}{\mathcal{R}_{\rm NS}} = \left(\frac{80M_{\odot}}{10M_{\odot}}\right)^{-1.35} \simeq 0.06\,. \qquad (11)$$

This expectation turns out to be in rough agreement with the results of detailed evolutionary calculations presented above.

The derived rates $\mathcal{R}_G$ for a single galaxy can be extrapolated to larger volumes. For the purposes of GW detection it is important to know the rate of events from distances accessible to LIGO, VIRGO, GEO-600. These are large distances, up to or even above 100 Mpc (see Section 7). In such a large volume one can regard galaxies to be distributed homogeneously, and at the same time, one can neglect the effects of cosmological evolution on star formation initial conditions, etc. To derive the average density of galactic events in a large volume one can use different approaches. One possibility is to use the luminosity of galaxies per Mpc^3 (as in Ref. [24]). Alternatively, one can rely on the estimate of the density of baryons bound in stars. The baryon density ρ_b is often expressed in terms of the dimensionless parameter $\Omega_b \equiv \rho_b/\rho_{cr}$, where $\rho_{cr} = 3H_0^2/8\pi G$ is the present-day value of the Hubble parameter and H_0 is the critical density. Then, one can relate the galactic rate $\mathcal{R}_G$ for a $10^{11}M_\odot$ galaxy with the volume rate $\mathcal{R}_V$ per Mpc^3,

$$\mathcal{R}_V = 3\times 10^{-3}\mathcal{R}_G\,\frac{\varepsilon}{0.5}\,\frac{\Omega_b\,h_{70}^2}{0.0045}\ \mathrm{Mpc}^{-3}\,, \tag{12}$$

where ε is the fraction of binary stars and $h_{70} = H_0/(70\ \mathrm{km\ s^{-1}\ Mpc^{-1}})$. This estimate agrees with that of Ref. [24] assuming $\varepsilon = 1$ (all stars are binaries). The available astronomical measurements of the total baryon budget give $\Omega_b \approx 0.0015h_{70}^{-1}$ in galactic disks and $\Omega_b \approx 0.003h_{70}^{-1}$ in bulges of spirals and ellipticals [83] (or somewhat larger values [84]). On the other hand, estimates of Ω_b based on the primordial nucleosynthesis considerations give as much as $\Omega_b h^2 = 0.016$, but this figure can also be a factor of two smaller [85]. Formula (12) can be rewritten for a sphere of radius r as

$$\mathcal{R}_V = 0.1\mathcal{R}_G\,\frac{\varepsilon}{0.5}\,\frac{\Omega_b\,h^2}{0.016}\left(\frac{r}{1\ \mathrm{Mpc}}\right)^3. \tag{13}$$

When comparing our numerical simulations described below, with qualitative estimates we rely on the relationship

$$\mathcal{R}_V = 0.1\mathcal{R}_G\left(\frac{r}{1\ \mathrm{Mpc}}\right)^3. \tag{14}$$

This result for $\mathcal{R}_V$ is based on $\mathcal{R}_G$ for spiral galaxies. For elliptical galaxies the star formation process is more likely an instantaneous event rather than the continuous process described by (3). The coalescence rates have also been calculated for elliptical galaxies. However, it was shown [86] that the contribution of elliptical galaxies to the coalescence rates from the discussed distances is only about 10–20%.

3. Detection rates

Having found the coalescence rates $\mathcal{R}_V$ for binaries of different nature, one can predict the detection rates of these binaries for a given GW detector. We note that binary black holes give a better SNR than binary neutron stars, and, despite their lower abundance, the BH + BH and BH + NS pairs should be seen more often than NS + NS pairs. In the first subsection, we will derive the detection rates based on the $\mathcal{R}_G$ described above. In the second subsection, we will discuss possible modifications to our conclusion in connection with the recently proposed scenario [87], which applies to very massive stars. Since the proposed scenario can affect only the BH + BH detection rates, we concentrate on these systems emphasizing the important role of kick velocities.

3.1 Detection rates in the usual picture

The rate of NS + NS coalescences is higher than the rate of NS + BH and BH + BH coalescences. However, the BH mass is significantly larger than the NS mass. A binary involving one or two black holes and situated at the same distance as an NS + NS binary produces gravitational waves of a significantly larger amplitude (see Appendix I). With a given sensitivity of the detector (fixed SNR), a BH + BH binary can be seen at a greater distance than an NS + NS binary. Hence, the registration volume for such bright binaries is significantly larger than the registration volume for relatively weak binaries. The detection rate of a given detector depends on the interplay between the coalescence rate (spatial density of sources) and the detector's response to sources of one kind or another.

Coalescing binaries emit gravitational wave signals with a well known time dependence (waveform). This allows one to use the technique of matched filtering [4]. The signal-to-noise ratio S/N depends mostly on the 'chirp' mass of the binary system $\mathcal{M} = (M_1+M_2)^{-1/5}(M_1M_2)^{3/5}$ and its distance r. An accurate formula for S/N is presented in Section 8 [formula (134)]. Here, we will use a simplified version, sufficient for our purposes ([4], see also Ref. [88]):

$$\frac{S}{N} \equiv \rho = 3^{-1/2}\pi^{-2/3}\,\frac{G^{5/6}}{c^{3/2}}\,\frac{\mathcal{M}^{5/6}}{r}\,\frac{f^{-1/6}}{h_{\mathrm{rms}}(f)}\,. \tag{15}$$

At a fixed level of S/N, the detection volume is proportional to r^3 and, therefore, proportional to $\mathcal{M}^{5/2}$. The detection rate $\mathcal{D}$ for binaries of a given class is the product of their coalescence rate $\mathcal{R}_V$ and the detector's registration volume $\mathcal{M}^{5/2}$ for these binaries.

Let us start from a qualitative discussion of the expected ratio

$$\frac{\mathcal{D}_{BH}}{\mathcal{D}_{NS}} = \frac{\mathcal{R}_{BH}}{\mathcal{R}_{NS}}\left(\frac{\mathcal{M}_{BH}}{\mathcal{M}_{NS}}\right)^{5/2}, \tag{16}$$

where $\mathcal{D}_{BH}$ and $\mathcal{D}_{NS}$ refer to BH + BH and NS + NS pairs, respectively. Here, we discuss the ratio of the detection rates, rather than their absolute values. The derivation of absolute values require detailed evolutionary calculations which will be discussed later. As a rough estimate for $\mathcal{R}_{BH}/\mathcal{R}_{NS}$ one can take Eqn (11). Then, Eqn (16) gives a remarkable result:

$$\frac{\mathcal{D}_{BH}}{\mathcal{D}_{NS}} = \left(\frac{80M_\odot}{10M_\odot}\right)^{-1.35}\left(\frac{8.5M_\odot}{1.40M_\odot}\right)^{15/6} \simeq 5.5\,. \tag{17}$$

This ratio becomes even larger than 5.5, if one takes $M_{cr} < 80M_\odot$, as is usually assumed. Thus, the registration rate of BH mergers is expected to be *higher* than that of NS mergers. This estimate is, of course, very rough, but it can serve as an indication of what one can expect from detailed calculations.

In Figure 4 we display the results of numerical calculations for the absolute registration rates of various binaries. The detector sensitivity is taken to be $h_{rms} = 10^{-21}$ at the frequency $f = 100$ Hz, as is expected for the first instruments in LIGO, VIRGO, GEO600. It is assumed that $S/N = 1$. Since the most interesting results refer to systems with black holes, we vary the black hole formation parameter k_{BH}. The

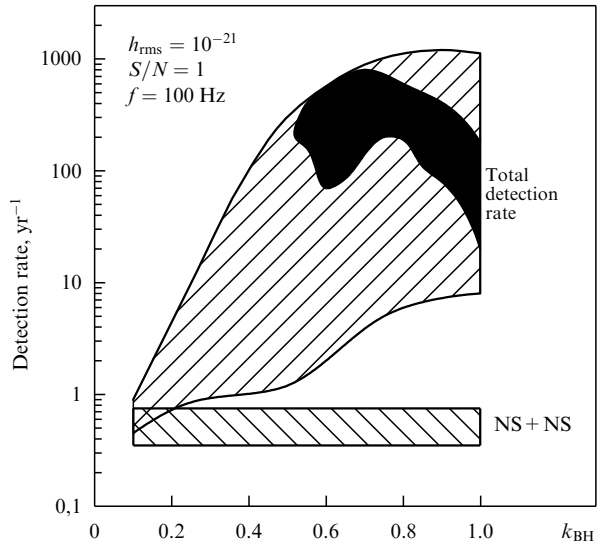

Figure 4. Detection rate $\mathcal{D}$ of GW events in a detector with sensitivity $h_{rms} = 10^{-21}$ at the frequency 100 Hz and the signal-to-noise level $S/N = 1$, as a function of the BH formation parameter k_{BH}. The calculations were performed for the Lyne-Lorimer kick velocity distribution (6) with $w_0 = 400$ km s^{-1}. The spread of $\mathcal{D}$ at fixed k_{BH} is due to the variation of the parameter M_{cr} from $15M_\odot$ to $50M_\odot$. The bottom rectangular area is drawn for binary NS coalescences. Their rate is independent of k_{BH} (as it should be) and predicts several events per 1–3 years at this level. The total detection rate may be 2–3 orders of magnitude higher then the NS + NS rate as a result of BH + BH and BH + NS coalescences. The hatched area shows the region of the most probable parameters for the low stellar wind mass loss scenario. Inside this region the outcomes of calculations are in agreement with the upper limit on the galactic number of binary BH with radiopulsars (less than 1 per 700 single pulsars) and with the galactic number of BH candidates similar to Cyg X-1 (from 1 to 10).

calculations were performed assuming the Lyne-Lorimer kick velocity distribution with $w_0 = 400$ km s^{-1}. The vertical dispersion of the results is due to the uncertainty in the parameter M_{cr}.

It is seen from the graph that under the formulated conditions one can expect several NS + NS coalescences in 1–3 years of observations, for $S/N = 1$. These systems are located, roughly, at a distance of 100 Mpc. The SNR is higher for closer systems, but the expected event rate is lower; for more distant systems, the event rate increases, but the SNR becomes smaller than 1. So, it is unlikely that NS + NS coalescences will be detected by the first-generation instruments.

The situation is significantly better for systems involving black holes. As is seen from Fig. 4, the total registration rate of all binaries, including BH + BH and BH + NS pairs, may be 2–3 orders of magnitude higher than the registration rate of NS + NS systems alone, mostly due to massive BH pairs. This is true unless the k_{BH} parameter is very small, $k_{BH} < 0.4$. The hatched area shows the results of calculations with the stellar wind parameters taken from the 'most probable' region. This means that for this choice of parameters, the outcomes of other evolutionary tracks are in agreement with observations, namely, in agreement with the upper limit for the number of binary pulsars with BH (less than 1 per 700 single radio pulsars) and with the number of Cyg X-1-like BH candidates (from 1 to 10 per Galaxy). Inside this region, one should count on 100 registrations (at the level $S/N = 1$), mostly from BH mergers. The mean total mass of the BH pairs in the hatched area is around $M = 30M_\odot$. The simplified formula (15), used in the construction of Fig. 4, overestimates the S/N for pairs heavier than $30M_\odot$, as shown in Section 8. However, a correction for more massive binaries is not expected to change significantly the derived total registration rate.

For a reliable detection, the S/N ratio should be at least 2 in each of four or more antennas of the network. Then, the calculated detection rates should be decreased by at least a factor $(S/N)^3 = 8$. This is because of the scaling $S/N \propto 1/r$ and $\mathcal{D} \propto r^3 \propto (S/N)^{-3}$. Then, the expected detection rate of merging BH + BH pairs is up to 10 events per year. As will be explained in Section 8, the S/N ratio is somewhat different for three different instruments — LIGO, VIRGO and GEO. For a coalescing pair with a total mass $M = 30M_\odot$ the SNR is roughly 4 for sources at a distance of 100 Mpc. (The performance of VIRGO is expected to be better than that of the other instruments, since VIRGO will be more sensitive at lower frequencies and will be able to track a binary for a larger number of cycles.) If one is satisfied with $S/N = 2$, the accessible radius increases to $r = 200$ Mpc. Then, the calculated detection rate (several events per year) is in agreement with formula (14) if one takes a reasonable galactic rate for coalescing black holes, $\mathcal{R}_G = 3 \times 10^{-6}$, and $r = 200$ Mpc. In turn, this value for $\mathcal{R}_G$ fits well the event rate derived from numerical simulations, as displayed in Fig. 3.

Thus, taking into account all the remaining uncertainties, we conclude that the original network is likely to see 2–3 coalescing black hole binaries each year with a total mass around $30M_\odot$, at an SNR level of about 2–3.

3.2 Non-standard scenarios and effects of kick velocities on the BH + BH detection rate

Some recent evolutionary calculations [87] assume that primary stars with initial masses $M_1 > 40M_\odot$ *never fill* their Roche lobes, so that the components of the binaries evolve like single stars. As a result, the binary BH systems should be too wide to merge in t_H. In principle, this effect can alter our conclusions on the binary BH merging rates. Although we think the scenario [87] will face observational difficulties, since it will lead to a too small number of binaries involving a BH and a massive blue star (Cyg X-1-like systems), we consider it necessary to follow in detail the possible fate of binary BH systems. We argue that the kick velocity accompanying the BH formation increases the eccentricity of the binary, decreases its coalescence time, and thus keeps the detection rate at almost the same level as discussed in Section 3.1. In addition, the kick velocity leads to interesting modifications in the relative orientations of the black hole spins with respect to each other and with respect to the orbital angular momentum.

We adopted the proposed scenario [87] and carried out population synthesis calculations by varying the kick velocity parameter. The binary BH merging rate was derived for a model galaxy of $10^{11}M_\odot$ (assuming that all stars are formed in binaries) with a constant star formation rate. For simplicity, the kick velocity distribution was taken to be a delta function. More complicated distributions do not change the results significantly and are not commented upon here. The results of these calculations are shown in Fig. 5. The left panel shows the merging rate, while the right panel shows the detection rate. The detection rate of binary BH coalescences is given for first-generation laser interferometers ($h_{rms} = 10^{-21}$ at

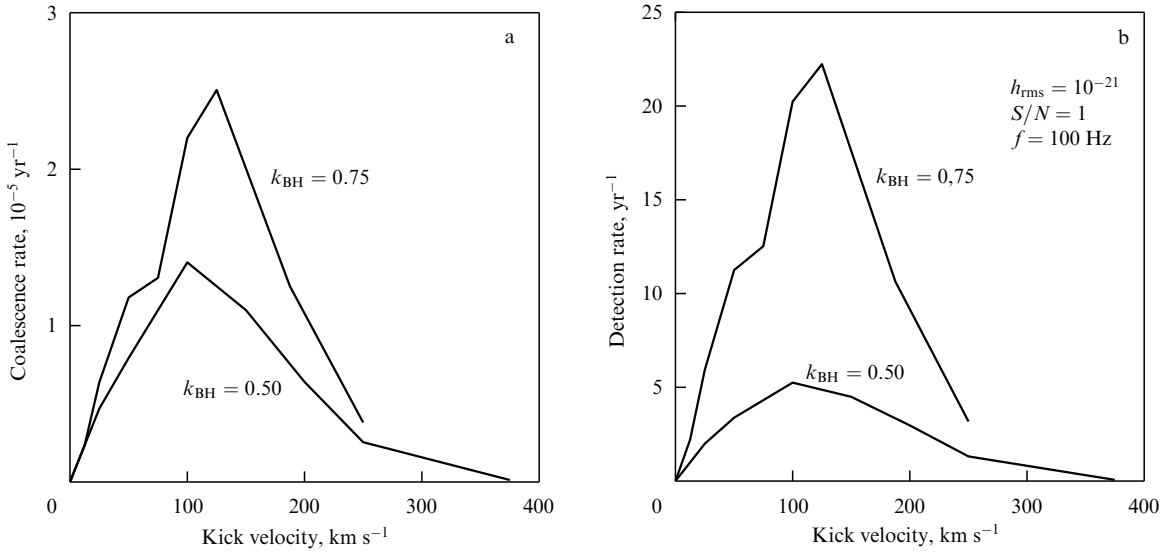

Figure 5. New scenario — superhigh wind. (a) BH + BH merging rate calculated for a $10^{11} M_\odot$ galaxy with a constant star formation rate, as a function of the kick velocity during BH formation with $M_{cr} = 35 M_\odot$, for $k_{BH} = 0.50$ and 0.75. (b) Detection rate of BH + BH mergings with the initial laser interferometers ($h_{rms} = 10^{-21}$ at $f = 100$ Hz), as a function of the kick velocity during BH formation.

$f = 100$ Hz) as a function of the BH kick velocity. It is seen from the plot that the merging rate and the detection rates increase rapidly with the kick. The merging and detection rates reach maxima of $\mathcal{R} \sim 2.5 \times 10^{-5}$ yr^{-1} and $\mathcal{D} \sim 20$ detections per year for $w \simeq 120$ km s^{-1}. Since $\mathcal{D} \sim \mathcal{M}^{5/2}\mathcal{R}$, the $\mathcal{R}(w)$ and $\mathcal{D}(w)$ functions have similar shapes.

Obviously, the kick velocity imparted to newly born black holes makes the orbits of surviving systems highly eccentric. It is important to stress that some fraction of binary BH retain their large eccentricities until the late stages of their coalescence. This should be reflected by their emitted waveforms and should be modeled in templates.

The asymmetric explosions accompanied by a kick velocity change the spatial orientation of the orbital angular momentum. On the other hand, the star's spin axis remains fixed (unless the kick was non-central). As a result, some distribution of the angles between the BH spins and the orbital angular momentum (denoted by J) will be established [89]. It is interesting that even for small kicks of a few tens of km s^{-1} an appreciable fraction (30 – 50%) of merging binary BH should have $\cos J < 0$. This means that in these binaries the orbital angular momentum vector is oriented almost oppositely to the black hole spins. This is one more sign of imparted kicks that can be tested observationally. These effects were also discussed in a recent paper [90].

Thus, to conclude this analysis, we stress again that binary black hole coalescences remain the most likely sources to be detected first by the early network of laser interferometers.

4. Transient and continuous gravitational wave sources

In this section we will discuss two distinct types of signals: (1) transient events, which last a few to tens of milliseconds and, on astronomical grounds, are expected to occur but emit waves of unknown phase evolution, as in the case of supernovae and (2) continuous radiation, which lasts for several days or longer, from either newly born neutron stars or old, spun-up neutron stars. The strengths, duration and shapes of these signals is rather speculative and highly uncertain. On general physical grounds we should expect such sources to exist, and every effort should be made in searching for these sources by taking the best advantage of current knowledge. However, since the astrophysical uncertainties are so large, we shall keep the discussion of this topic only qualitative.

4.1 Transients

Supernovae and asymmetric explosions. Supernovae (of type II) are associated with violent mass ejections at velocities of order $0.01c$ and the formation of a compact remnant — a neutron star or black hole. The event has at its disposal the difference in the gravitational binding energy between the pre-collapse star and the newly formed compact star, which, neglecting the former, is:

$$|E| \sim 3 \times 10^{53} \left(\frac{M}{M_\odot}\right)^2 \left(\frac{R}{10 \text{ km}}\right)^{-1} \text{erg}\,. \tag{18}$$

As much as 99% of this energy is carried away by neutrinos, about 1% is transferred as the kinetic energy of the ejecta, and a fraction of 10^{-4} of the total energy is emitted in the form of electromagnetic radiation. Depending on how asymmetric the collapse is, some fraction of the total energy should be deposited into GWs; spherically symmetric collapse, of course, cannot emit any radiation. According to numerical simulations (see Ref. [91] for a review) one can expect up to 10^{-7} of the total energy to be emitted in GWs. Together with uncertainties in the event rate, this is not a very encouraging prognosis for the initial instruments [4, 5]. It appears that a star collapsing to form a black hole is also not particularly well suited for detection by the existing resonant detectors and forthcoming interferometers [92]. However, second-generation interferometers should be able to see a supernova event as far as the Virgo super-cluster, which contains about 200 bright galaxies and at least twice as many faint galaxies. In addition, there are a few other smaller clusters within that distance and a large number of field galaxies. Therefore, the supernova event rate for these instruments could be as large as tens per year. Such observations would undoubtedly be of great interest and would shed light on the hitherto little

understood processes that occur when a star collapses to form a compact object.

An asymmetric collapse generates GWs but is not necessarily accompanied by a change in the linear momentum of the exploding star. However, as we emphasized above, observations of single and binary radio pulsars require the presence of a significant kick velocity imparted to a newly born neutron star. The kick velocity can be as large as 200–400 km s^{-1}. This testifies to a non-axisymmetric explosion, leading to an additional linear momentum being acquired by the neutron star. A possible reason for this asymmetry may be related to the asymmetry of the neutrino emission during the collapse [93]. Using the observed kick velocities one can evaluate the asymmetry of the explosion and calculate the amplitude of emitted GWs. Nazin and Postnov [94] estimated that the mean energy carried away by a GW burst can reach the value $E_{\rm gw} = 5 \times 10^{-6} M_\odot c^2$. It is interesting to note that the radiated signal belongs to the category of bursts 'with memory' [95, 96]. However, the estimated strength of the asymmetric explosions is still too low to count on them as reliable sources for first detection.

Bar-mode and convective instabilities. Some quantity of GWs may be emitted during two special stages of the collapse: (1) rotation-induced bars and (2) convective instabilities set up in the core of the newly born neutron star.

If the core's rotation is high enough it may cause the core to flatten before it reaches nuclear density, leading to an instability that transforms the flattened core into a bar-like configuration which spins about its transverse axis. Some of these instabilities could also fragment the core into two or more pieces which then rotate about each other. Both are efficient ways of losing energy in the form of GWs. It was estimated [97] that the waves could carry away up to $10^{-3} M_\odot c^2$ in a few ms. Then, according to Eqn (20) (see Section 4.2), the LIGO and VIRGO detectors could see such an event at a distance up to 50 Mpc, or about 5–10 Mpc if the waves come off at 1 kHz. GEO will also be able to see these events provided the signal comes off around 200 Hz [98].

Convective instabilities in the core of a newly born neutron star, which last for about a second after the collapse, are likely to produce gravitational radiation due to anisotropic mass distribution and motion [99]. Müller and Janka [100] found that the gravitational wave amplitude of a source at 100 kpc would be $h \sim 10^{-23}$ and that the waves would come off at about 100 Hz. Since there should be about 100 cycles, we can enhance the amplitude to about 10^{-22} provided we know the shape of the signal; however, this is still far too weak to be detected beyond about 10 kpc with high confidence.

Merger waves. The physics of the merger phase that follows the adiabatic inspiral regime of a compact binary coalescence is not known. Presently, it is a subject of active research and contains a lot of uncertainties. Some authors expect that this phase could result in detectable amounts of radiation [88] while others conclude that the merger phase is just a continuation of the inspiral phase adding one more cycle or so to the inspiral [101]. There is a lot of effort to solve this important problem using semi-analytical and numerical techniques. It is likely that a solution will be in place by the time the first detectors begin to operate. The fact that the coalescence waves will be preceded by inspiral waves makes the search easier, though it is not inconceivable that, while the latter may be observable with the aid of accurate search templates, the former may not be. Flanagan and Hughes [88] estimate that for binary systems of total mass in excess of $25 M_\odot$, coalescence waves are likely to be significantly stronger than the inspiral waves. The span of a detector is larger for heavier binaries and, therefore, these authors conclude, it is likely that the first gravitational wave events will be the merger phases of massive binaries.

Sub-stellar mass black hole binaries. The results of recent micro-lensing experiments [102] have revealed massive compact halo objects (MACHOs) of mass $0.5^{+0.3}_{-0.2} M_\odot$. Nakamura et al. [103] argue that if MACHOs are black holes then they must have formed in the early Universe and they estimate that our Galaxy may contain about 10^8 black hole binaries with inspiral time scales less than the Hubble time. If this is the case then the rate of MACHO coalescences in our Galaxy is $\sim 5 \times 10^{-2}$ yr^{-1} implying an event rate of few per year within 15 Mpc. As we shall see in the data analysis sections, first-generation interferometers should be able to detect the final inspiral phase of these systems.

Quasi-normal modes. A compact binary coalescing as a result of gravitational radiation reaction should most likely give rise to the formation of a single black hole. The newly formed hole will initially be somewhat non-spherical, and this dynamical non-sphericity will be radiated away in the form of GWs. The late time behavior of this radiation is well studied in the black hole literature and there are detailed calculations of the (quasi) normal modes for both static, i.e. Schwarzschild, and stationary, i.e. Kerr, black holes. In all cases the time evolution of the emitted radiation is well modeled by a quasi-periodic signal of the form

$$h(t; \tau, \omega) = A \exp\left(-\frac{t}{\tau}\right) \cos \omega t \,, \tag{19}$$

where τ is the decay time scale of the mode in question and ω is the angular frequency of the mode, both of which depend on the black hole mass and angular momentum. In all but the extreme Kerr black holes (extreme Kerr black holes are those that are spinning at the maximum possible rate) the only dominant mode, i.e. the mode for which the decay time is the longest and the amplitude is the highest, is the fundamental mode whose frequency is related to the mass and spin of the black hole via $\omega = [1 - 0.63(1-a)^{0.3}]/M$, where M is the mass of the black hole in units $G = c = 1$ and $a = J/M$ is the spin angular momentum of the hole in units of black hole mass. The decay time τ is given by $\tau = 4/[\omega(1-a)^{0.45}]$ (see Ref. [104] and references therein for details.)

It was estimated [88] that during the quasi-normal mode ringing of a black hole the emitted energy may be as large as 3% of the system's total mass. By matched filtering (cf. Section 8.1) it should be possible to detect, with first-generation interferometers, quasi-normal modes from black holes of mass in the range $60 - 10^3 M_\odot$ and at a distance of 200 Mpc. Binary black hole mergers should result in the emission of such a ring down signal during the late stages. Thus, inspiral signals emitted before the merger might aid in identifying the quasi-normal modes.

4.2 Continuous waves

Our Galaxy is believed to have at least 10^8 spinning neutron stars that form roughly at a rate of one every 30 years. Some population of neutron stars is in binaries. There are a number of ways in which a single spinning neutron star could radiate away GWs (of course, if the neutron star is axisymmetric, then there will be no gravitational wave emission):

(1) Neutron stars normally spin at high rates (several to 500 Hz) and this must induce some equatorial bulge and flattening of the poles. The presence of a magnetic field may cause the star to spin about an axis other than the symmetry axis, leading to a time-varying quadrupole moment.

(2) The star may have some density inhomogeneities in the core/crust set up during its formation and/or subsequent convectively unstable motions of the core.

(3) The presence of an accretion disc, with its angular momentum not necessarily aligned with that of the neutron star, can potentially alter axisymmetry. This effect and electromagnetic radiation reaction torques can induce and sustain wobble.

(4) The normal modes of the neutron star fluid (radial and other oscillations) can extract rotational energy and re-emit it in the form of GWs.

(5) There are certain classical and relativistic instabilities in the neutron star fluid which may cause the star to radiate away energy in the form of gravitational radiation. In what follows we will only discuss a sample of recent work on the radiation from spinning neutron stars.

GW amplitude from spinning asymmetric neutron stars. If I_{zz} is the moment of inertia about the spin axis of a neutron star emitting GWs at a frequency f, then the gravitational amplitude at a distance r is

$$h = 3\times 10^{-27}\left(\frac{10\ \mathrm{kpc}}{r}\right)\left(\frac{I_{zz}}{10^{45}\ \mathrm{g\,cm^2}}\right)\left(\frac{f}{200\ \mathrm{Hz}}\right)^2\left(\frac{\epsilon}{10^{-6}}\right), \tag{20}$$

where ϵ is the ellipticity of the star. In a simple model of an elliptical equatorial cross section with a semi-major axis a_1 and a semi-minor axis a_2, the ellipticity is $\epsilon \equiv 1 - a_2/a_1$. The ellipticity is an unknown, but one can obtain an upper limit for it by attributing the observed spin-down of pulsars $\dot{P}$ to the gravitational radiation back reaction, namely, by assuming that the change in the rotational energy $E = I\Omega^2/2$ is equal to the gravitational wave luminosity. Then, the ellipticity is related to the spin-down rate of a pulsar via

$$\epsilon = 5.7\times 10^{-6}\left(\frac{P}{10^{-2}\ \mathrm{c}}\right)^{3/2}\left(\frac{\dot{P}}{10^{-15}}\right)^{1/2}. \tag{21}$$

If one knows the observed values of P and $\dot{P}$, one can obtain an upper limit for ϵ using the above equation. Following this method one can find that for the Crab pulsar $\epsilon \leqslant 7\times 10^{-4}$ and the gravitational amplitude is $h \leqslant 10^{-24}$. It can be concluded that the Crab will have a sufficiently large amplitude to be observable with the aid of the GEO detector [6] within a year of continuous observing, if all of its spin down can be attributed to the emission of GW. It is unlikely that the ellipticity is so large. Yet, the prospect of seeing the Crab event at a 10th or a 100th of this ellipticity is quite good with the first/second generation of interferometers.

In the next two subsections we will discuss some new developments in relativistic astrophysics that could lead to potential GW sources.

Relativistic instabilities in young neutrons stars. Chandrasekhar [105] and Friedman and Schutz [106] discovered an instability (now called the CFS instability) in the fundamental, or f, mode of a neutron star fluid, arising as a result of GW emission. The mode becomes unstable above a critical spin frequency of the star and progressively grows, instead of decaying, by emitting gravitational waves.

The physics behind this instability can be understood in the following manner: Imagine exciting a mass-quadrupole mode — a non-uniform distribution of mass — in a non-spinning star. The mass inhomogeneity will travel over the surface of the star, and this dynamical asymmetry will cause the star to radiate GWs. After a while, all the energy in the mode will be radiated away and the mode will decay. Now consider a spinning neutron star in which co-rotating and counter-rotating modes are excited. These fluid modes have a certain speed of the pattern on the surface of the star. For low spin rates both these modes will decay with time by emitting GWs. But if the neutron star is spun up above a critical rate, then, to an external inertial observer, both modes will appear to be traveling in the same sense as the rotation of the star. Therefore, the mode counter-rotating relative to the star will emit positive angular momentum, causing the angular momentum associated with this mode to be enhanced, or for the amplitude of the mode to increase. In other words, a mode counter-rotating relative to the star, but seen co-rotating relative to the inertial observer, can only emit negative angular momentum which causes its own angular momentum to increase. The energy for this enhancement is supplied by the spin angular momentum of the neutron star. Thus, while the mode co-rotating with the star will decay, the mode counter-rotating with the star will grow in amplitude and emit more and more radiation. This will go on until the mode has sucked out enough angular momentum from the star to make the counter-rotating mode also appear to be counter-rotating with respect to an inertial observer.

It is suspected that the CFS instability will not occur in the presence of viscosity and hence it may be unimportant in old neutron stars. However, newly born neutron stars will be very hot and viscous forces may be insignificant in them. Recently, Andersson [107] discovered another class of modes called r-modes, which — unlike CFS modes that are mass-quadrupole moments — are current-quadrupole moments and are unstable at all spin frequencies. The role of these modes may be significant in young neutron stars. It was suggested that r-modes are responsible for the limit on the spin frequencies of newly born neutron stars [108]. Owen et al. [109] computed the efficiency with which these modes extract energy out of the system and the expected gravitational wave amplitude from isolated neutron stars, as well as from the ensemble of all sources up to cosmological distances. They concluded that the second generation of interferometric antennas will be able to distinguish such a background by coincident observations with a nearby bar detector. Such observations should provide independent knowledge of the distribution of galaxies in the high-redshift Universe as well as on the star formation history, via the observation of r-modes associated with the formation of neutron stars and pulsars.

Neutron stars in X-ray binaries. In recent years, Rossi satellite observations of X-ray binaries revealed high-frequency quasi-periodic oscillations (QPO) in their X-ray power spectra. Some authors believe that these QPO could be a result of the beating of two frequencies, one of which is that of the neutron star. The neutron star spins inferred in this way seem to lie in a narrow range of 250 – 350 Hz and are all within 20% of 300 Hz. A neutron star may be born with a high spin rate (several 100 Hz) but quickly spins down to moderate rates (several 10 Hz). In a binary system, when the companion becomes a red giant the neutron star starts accreting mass and angular momentum. Though the mass accretion rate is very low ($\dot{M} \sim 10^{-10}\,M_\odot$ yr) the

accretion of angular momentum can spin up a neutron star and heat its crust substantially. Indeed, the millisecond pulsars are believed to be old pulsars in binaries spun up in this way. It is puzzling why the spin frequencies of neutron stars in X-ray binaries are all in a narrow range. Bildsten suggests [110] that the absence of efficient heat transport processes makes it possible to set up temperature gradients in accreting neutron stars. Provided that the temperature distribution has large-scale asymmetry, the temperature-sensitive electron captures in the deep crust can build up the mass quadrupole ($\sim 10^{-7} MR^2$) needed to radiate away the accreted angular momentum and to limit the spin frequency. This mechanism is present only during the accretion and decreases rapidly once the accretion halts. The frequency of GW radiation will be known in advance since one knows the spin frequency of the pulsar from X-ray observations. It is argued [110] that the gravitational wave strength will be $h \sim (0.5-3) \times 10^{-26}$ for many of these sources and that the LIGO/VIRGO and signal-recycled GEO600 could detect the strongest of these sources, Sco X-1, at an SNR of 5 within a few years of integration.

5. Astrophysical stochastic background of gravitational waves

Coalescing binaries of compact stars are at the center of our attention. A gravitational wave signal will be monitored by a detector as long as the changing frequency of the source sweeps through the detector's window of sensitivity. The detection of a useful gravitational wave source is limited by the instrumental noise and by possible gravitational wave noise produced by other sources. If a large population of astrophysical sources emit overlapping GWs, the resulting signal will be perceived by the detector as GW noise. In principle, this noise could be larger than the instrumental noise. The astrophysical backgrounds of gravitational waves are important in their own right, but we are discussing them mostly as possible noises preventing the detection of a signal of greater scientific importance. The aim of our discussion is the derivation of the gravitational mean square amplitude in a given frequency interval Δf and the comparison of this quantity with the instrumental noise in the same interval. If the background is below the instrumental noise, it will not prevent the detection of a useful signal. In Section 5.1 we derive general formulas for the performance of a large number of unresolved sources. In Section 5.1 we discuss the population of rotating neutron stars in our Galaxy. They could constitute a major danger for ground-based interferometers. The analysis shows, however, that the danger arises only for quite unrealistic assumptions about the parameters of this population. In Section 5.1 we discuss the stochastic GW background produced by binary white dwarfs in the Galaxy. This background dominates the LISA instrumental noise within the range from $\sim 10^{-4}$ Hz to $\sim 10^{-3}$ Hz, but leaves the GW sky transparent at lower and higher frequencies. Finally, Section 5.2 comments on astrophysical backgrounds of extragalactic origin. Generically, these backgrounds are an order of magnitude lower than those produced by the galactic sources.

5.1 Unresolved sources in our Galaxy

A large collection of independent sources produce signals whose intensities add. Consider N identical sources located at approximately the same distance r from the observer. The resulting GW field is characterized by the r.m.s. amplitude h_N:

$$h_N = h_1 \sqrt{N}\,, \tag{22}$$

where h_1 is the averaged amplitude of a single source. We are interested in the narrow frequency interval from f to $f + \Delta f$. The radiating sources gradually change their frequency and pass through the window of interest. The crossing of the window can occur either on the way from lower to higher frequencies, as in the case of binaries, or in the opposite direction, as in the case of rotating neutron stars. To find h_N, we need to know the number $N(f, \Delta f)$ of radiating systems in the discussed frequency interval. We denote by $\mathcal{R}$ the rate at which the sources appear in the window. A source is present in the window during the time Δt, where $\Delta t = \Delta f / \dot{f}$. Thus,

$$N(f, \Delta f) = \mathcal{R} \Delta t = \mathcal{R}\, \frac{\Delta f}{\dot{f}} \tag{23}$$

and

$$h_N = h_1 \sqrt{\mathcal{R}\, \frac{\Delta f}{\dot{f}}}\,. \tag{24}$$

If the frequency evolution is driven by GWs only, the quantity $\dot{f}$ is determined by the gravitational radiation damping. For example, in the case of a binary system, $\dot{f}$ is given by Eqn (A.19). Alternatively, the quantity $\dot{f}$ can be determined by the electromagnetic radiation damping. This takes place in the case of highly magnetized rotating neutron stars whose $\dot{f}$ is determined by electromagnetic, rather than gravitational, losses. The appearance rate $\mathcal{R}$ of sources of a given population in the discussed window Δf is equal to the birthrate or to the coalescence rate of the sources as such (see, for example, $\mathcal{R}_{\rm coal}$ in Section 4). This is true as long as the number of sources remains constant during the frequency evolution throughout the window, which we always assume.

For sources driven by gravitational radiation only, formula (24) can be expressed in terms of the energy of a single radiating system and the rate $\mathcal{R}$. One needs to use Eqn (A.16) in order to express h_1^2 in terms of $\mathrm{d}E/\mathrm{d}t$, and to take into account the relationship $\mathrm{d}E/\mathrm{d}t = (\mathrm{d}E/\mathrm{d}f)\dot{f}$. Then, Eqn (24) takes the form

$$h_N^2 = \frac{G}{c^3}\, \frac{\mathcal{R}}{r^2 (\pi f)^2}\, \frac{\mathrm{d}E}{\mathrm{d}f}\, \Delta f\,. \tag{25}$$

Usually, the energy of the radiating system is a power-law function of the frequency f: $E(f) \sim f^{\alpha}$. For example, in the case of a binary star in a circular orbit one derives from Eqn (A.6):

$$E(f) = \frac{G^{2/3}}{2}\, \mathcal{M}^{5/3} (\pi f)^{2/3}\,, \tag{26}$$

where $\mathcal{M}$ is the chirp mass. So, in the case of binaries, $\alpha = 2/3$. For a non-axisymmetric rotating star $E(f) \sim f^2$, so that $\alpha = 2$. Thus, Eqn (25) takes the universal form

$$h_N^2 = \frac{G}{c^3}\, \frac{\mathcal{R}}{r^2 (\pi f)^2}\, \alpha E(f)\, \frac{\Delta f}{f}\,. \tag{27}$$

The quantity $h_N^2 / \Delta f$ is the mean square noise amplitude h_f^2 (with dimensionality Hz^{-1}), which appears in the

expression

$$\langle h^2\rangle = \int_{f_{\min}}^{f_{\max}} h_f^2 \, \mathrm{d}f, \tag{28}$$

and can now be compared with the frequency-dependent instrumental noise. We will work with the dimensionless spectral amplitude

$$h_N(f) \equiv h_N\sqrt{\frac{f}{\Delta f}} = \frac{1}{r}\sqrt{\frac{G\alpha E(f)\mathcal{R}}{c^3(\pi f)^2}}\,. \tag{29}$$

Obviously, independent sources in a frequency bin form a kind of stochastic background if their number $N(f,\Delta f)$ is much larger than 1. A source becomes resolvable, if this number is of order 1. For a collection of evolving sources one can find the limiting frequency $f_{\lim}$ at which this happens. For fixed Δf and $\mathcal{R}$, one uses a particular function $\dot{f}(f)$ (arising due to the gravitational reaction force or some other cause) and finds $f_{\lim}$ from the requirement

$$\mathcal{R}\,\frac{\Delta f}{\dot{f}} = 1\,. \tag{30}$$

For a collection of binary stars one uses Eqn (A.19) and finds

$$f_{\lim} \approx (1.2\times 10^{-3}\ \mathrm{Hz})\mathcal{R}_{300}^{3/11}\left(\frac{\Delta f}{3\times 10^{-8}\ \mathrm{Hz}}\right)^{3/11} \times \left(\frac{\mathcal{M}}{0.52M_\odot}\right)^{-5/11}. \tag{31}$$

The appearance (coalescence) rate $\mathcal{R}_{300}$ is chosen for compact white dwarfs, which are expected to coalesce in our Galaxy once per 300 years. The chirp mass is normalised to $0.52M_\odot$ which is true for two CO white dwarfs with masses $0.6M_\odot$. This estimate will be needed in Section 5.1 which discusses the LISA noise.

Noise from old neutron stars at frequencies of ground-based interferometers. Rotating neutron stars as sources of gravitational radiation can be roughly divided into two populations. One consists of old neutron stars with relatively weak magnetic fields and small electromagnetic losses. Their rotational frequency slowly decreases due to GW damping. Another population consists of young, highly magnetized neutron stars. Their rotational frequency decreases much faster due to electromagnetic damping. In the course of their frequency evolution, members of both populations cross the window of sensitivity of ground-based interferometers, descending from 10^3 Hz to 10 Hz. The number of sources simultaneously radiating in a given frequency interval is proportional to the birthrate $\mathcal{R}$ of the population and inversely proportional to the velocity $\dot{f}$ of the population flow through the window. We start from old neutron stars and then discuss young neutron stars.

The frequency evolution of old, rotating, deformed neutron stars is governed by gravitational radiation damping. The formula for $\dot{f}$, analogous to Eqn (A.19) for double stars, is

$$\dot{f} = \frac{32\pi^4 G}{c^5}\, I\epsilon^2 f^5\,, \tag{32}$$

where I is the relevant moment of inertia and ϵ is the ellipticity (deformation) parameter. The birthrate $\mathcal{R}$ of neutron stars in this population can be estimated using the observed fraction of millisecond pulsars (which are thought to be old neutron stars with weak magnetic fields, which have been spun up by accretion in a binary system) among radio pulsars: $N_{\mathrm{ms}}/N_{\mathrm{PSR}} \approx 20/2000 = 1/100$. Remembering that the lifetime of a millisecond pulsar is $t_{\mathrm{ms}} = 10^8$ yr and that the lifetime of an ordinary radio pulsar is $t_{\mathrm{PSR}} = 10^6$ yr, and adopting the galactic birthrate of radio pulsars of 1 per 30 years, as for the core collapse supernovae, one can estimate the birthrate of millisecond pulsars:

$$\mathcal{R}_{\mathrm{ms}} = \mathcal{R}_{\mathrm{PSR}}\,\frac{N_{\mathrm{ms}}}{N_{\mathrm{PSR}}}\,\frac{t_{\mathrm{PSR}}}{t_{\mathrm{ms}}} \approx 3\times 10^{-6}\ \mathrm{yr}^{-1}\,.$$

This estimate is in agreement with that derived in Ref. [111] from the observed space density of millisecond pulsars.

From Eqn (30) one finds the limiting frequency

$$f_{\lim} \approx 53\ \mathrm{Hz}\left(\frac{\mathcal{R}}{3\times 10^5\ \mathrm{yr}}\right)^{1/5}\left(\frac{\Delta f}{3\times 10^{-8}\ \mathrm{Hz}}\right) \times \left(\frac{I}{10^{45}\ \mathrm{g\,cm^2}}\right)^{-1/5}\left(\frac{\epsilon}{10^{-9}}\right)^{-2/5}. \tag{33}$$

Thus, at frequencies below 53 Hz the population of old neutron stars is likely to produce a stochastic background. Taking $E(f) = \pi^2 I f^2/2$ one derives from Eqn (29):

$$h_N(f) = \frac{1}{r}\sqrt{\frac{G}{c^3}\,I\mathcal{R}} \approx 2\times 10^{-26}\left(\frac{10\ \mathrm{kpc}}{r}\right) \times \left(\frac{\mathcal{R}}{3\times 10^5\ \mathrm{yr}}\right)^{1/2}\left(\frac{I}{10^{45}\ \mathrm{g\,cm^2}}\right)^{1/2}. \tag{34}$$

It is interesting that this quantity does not depend on the deformation parameter ϵ, provided it is nonzero. The quantities h_1^2 and $\dot{f}$ are both proportional to ϵ^2, so that they mutually cancel out in the expression for h_N^2. The quantity $h_N(f)$ is also independent of frequency f. The numerical level of $h_N(f)$ is much lower than the instrumental noise of initial and advanced ground-based interferometers. The value of $h_N(f)$ can be increased by two orders of magnitude, and hence the GW noise from old neutron stars becomes marginally detectable only under the condition that one postulates a significantly larger (and, we believe, unrealistic) birthrate $\mathcal{R}$ for old neutron stars (compare with Ref. [112]).

At frequencies higher than the limiting frequency (33), the sources are resolvable during a 1-year interval of observations. However, to monitor a single neutron star, one needs to know its exact location on the sky and to take care of the Doppler frequency modulation due to the Earth's motion around the Sun (see Section 8).

Young neutron stars differ from old neutron stars in that their electromagnetic energy loss

$$\dot{E}_{\mathrm{em}} = \frac{2\pi^4}{3c^3}\,\mu^2 f^4\,, \tag{35}$$

where μ is the NS magnetic moment, is significantly larger than the GW loss

$$\dot{E}_{\mathrm{gw}} = \frac{32\pi^6 G}{c^5}\, I^2\epsilon^2 f^6\,. \tag{36}$$

The ratio $x = \dot{E}_{\rm em}/\dot{E}_{\rm gw}$ is

$$x \approx 4 \times 10^3 (\mu_{30})^2 (\epsilon_{-6})^{-2} (I_{45})^{-2} \left(\frac{100\ \text{Hz}}{f}\right)^2, \quad (37)$$

where $\mu_{30} = \mu/(10^{30}\ \text{G cm}^3)$, $\epsilon_{-6} = \epsilon/10^{-6}$ and $I_{45} = I/10^{45}\ \text{g cm}^2$. For typical parameters of young neutron stars one has $x \gg 1$. The ratio x becomes comparable with 1 only for relatively weak magnetic fields, such that the magnetic moment μ satisfies the condition $\mu < 1.5 \times 10^{26} (\text{G cm}^3) \epsilon_{-6} (f/100\ \text{Hz})$. The frequency change $\dot{f}$ is determined by the electromagnetic loss and reads

$$\dot{f} = \frac{2\pi^2}{3c^3} \frac{\mu^2 f^3}{I}. \quad (38)$$

For $f_{\rm lim}$ one derives

$$f_{\rm lim} = 0.5\ \text{Hz} (\mathcal{R}_{30})^{1/3} \left(\frac{\Delta f}{3 \times 10^{-8}\ \text{Hz}}\right)^{1/3} (\mu_{30})^{-2/3} (I_{45})^{1/3}. \quad (39)$$

The averaged amplitude h_1 of a single neutron star amounts to

$$h_1 = \frac{G\pi^2\sqrt{32}}{c^4} \frac{1}{r} I f^2 \epsilon. \quad (40)$$

Then, the $h_N(f)$ found from Eqn (24) with the help of Eqn (38) becomes

$$h_N(f) = \frac{4\sqrt{3}\,\pi G}{c^{5/2}} \frac{1}{r} I^{3/2} \mathcal{R}^{1/2} \epsilon \mu^{-1} f \approx 3 \times 10^{-26} \left(\frac{10\ \text{kpc}}{r}\right) \mathcal{R}_{30}^{1/2} I_{45}^{3/2} \epsilon_{-6}\, \mu_{30}^{-1} \frac{f}{100\ \text{Hz}}. \quad (41)$$

Thus, the GW noise from young neutron stars is at the same numerical level as the noise from old neutron stars. The appearance rate of young neutron stars is much higher than that of old neutron stars, but they flow through the window of sensitivity much faster, so that there is not a sufficiently large number of sources to produce a background of a high enough level. We conclude that the populations of neither old nor new neutron stars present any danger to the sensitivity curves of initial and advanced ground-based instruments. It is important that the calculated noise level is below the expected level of the relic GWs (see Section 6).

Noise from galactic binary white dwarfs in LISA. The LISA frequency range, $10^{-4} - 10^{-1}$ Hz, may be contaminated by GW noise from coalescing binary white dwarfs (WD) and binary neutron stars (NS). The appearance rates $\mathcal{R}$ of the two populations in the sensitivity window are numerically equal to their coalescence rates. Binary WD systems are much more numerous than binary NS systems, and the coalescence rate of the former population is significantly higher than that of the latter. The binary WD coalescence rate is about 1 per 300 yr, while the coalescence rate of binary NS is about 1 per several 10000 yr. At the same time, the chirp masses, $\mathcal{M} \approx 0.52 M_\odot$ for a WD binary and $\sim 1.2 M_\odot$ for an NS binary, are not significantly different. This is why the WD background is more important than the NS background, and we consider only the former.

For a collection of sources consisting of binary stars, one can use Eqns (A.13), (A.19) in Eqn (24), or, alternatively, Eqn (26) and $\alpha = 2/3$ in Eqn (27). Either way one obtains

$$h_N(f) = \frac{G^{5/6}}{\sqrt{3}\,\pi^{2/3} c^{3/2}} \frac{1}{r} \mathcal{R}^{1/2} \mathcal{M}^{5/6} f^{-2/3} = 10^{-20} \left(\frac{10\ \text{kpc}}{r}\right) \times \left(\frac{\mathcal{R}}{300\ \text{yr}}\right)^{1/2} \left(\frac{f}{10^{-3}\ \text{Hz}}\right)^{-2/3} \left(\frac{\mathcal{M}}{0.52 M_\odot}\right)^{5/6}. \quad (42)$$

Roughly, this is a result of the performance of $\approx 10^6$ binaries in the frequency bin $\Delta f = f = 10^{-3}$ Hz with the averaged amplitude [see Eqn (A.13)]

$$h_1 = \frac{\sqrt{32}\,\pi^2 G^{5/3}}{\sqrt{5}\,c^4} \frac{1}{r} \mathcal{M}^{5/3} f^{2/3} = 2.5 \times 10^{-23} \left(\frac{10\ \text{kpc}}{r}\right) \left(\frac{\mathcal{M}}{0.52 M_\odot}\right)^{5/3} \left(\frac{f}{10^{-3}\ \text{Hz}}\right)^{2/3}. \quad (43)$$

It is necessary to note that Eqn (42) gives an estimate for the amplitude averaged over the whole sky, while the real background is strongly concentrated near the galactic plane [36]. The response of a space-based interferometer should be modulated while the instrument turns in its orbit. In principle, this distinctive feature of the galactic background can be used in order to distinguish it from the backgrounds of cosmological origin [113].

In formula (42), the quantity $\mathcal{R}$ replaces all the astrophysical uncertainties in the binary WD evolution. At frequencies higher than 3×10^{-4} Hz the evolution of the vast majority of binary white dwarfs is totally controlled by GW emission. So, at frequencies of interest, the GW noise is fully determined by the galactic rate of binary WD mergers and is independent of complicated details of binary evolution at lower frequencies. (For examples of calculated spectra at all frequencies see Refs [33, 34, 114, 115].) The coalescence rate of close binary WD is known only to a factor of a few. One way to estimate $\mathcal{R}$ is based on the search for nearby WD binaries. A recent study [116] revealed a larger number of such systems than had been previously believed to exist. However, the statistics of such binaries in the Galaxy remains very poor. If coalescing binary WD are associated with SN Ia explosions, as suggested in Ref. [117] and further investigated by many authors (for a recent review of SN Ia progenitors see Ref. [118]), their coalescence rate can be constrained using the much more representative SN Ia statistics. The authors of Ref. [118] have concluded that coalescing CO–CO binary WD remain the most plausible candidates responsible for SN Ia events. The galactic rate of SN Ia events is estimated to be 4×10^{-3} per year [119, 120], which is close to the calculated rate of CO–CO coalescences [$\sim (1-3) \times 10^{-3}$]. The coalescence rate of He–CO and He–He WD pairs (other possible progenitors of SN Ia) falls ten times short of that for CO–CO WD [118]. As SN Ia explosions may well be triggered by other mechanisms as well, we conclude that the observed SN Ia rate provides an upper limit on the double WD merger rate, regardless of the evolutionary considerations.

In Figure 6 we plot the LISA sensitivity curve (thick solid line) calculated for the frequency bins $\Delta f = 3 \times 10^{-8}$ Hz (which correspond to one year of observation) as a function of frequency. The binary confusion limit is shown with the dash-dotted line. At frequencies below $\sim 4 \times 10^{-4}$ Hz the binary GW background is produced not only by coalescing

WD, but by other binaries as well. So, in this part of the graph we rely upon numerical calculations [33, 121]. At the limiting frequency $\sim 10^{-3}$ Hz, individual galactic WD binaries become resolvable in a 1 year observation time, and the binary WD noise drops below the LISA sensitivity. It then continues as a noise produced by an isotropic distribution of extragalactic binaries (see Section 5.2). Figure 6 also shows the expected background of relic GWs (see Section 6). Keeping in mind that the real noise caused by merging galactic WD may be lower than the plotted one and, in any case, is direction dependent, we conclude that in the frequency interval $\sim 10^{-3}-10^{-1}$ Hz none of GW backgrounds of galactic origin should be higher than the LISA sensitivity in the frequency bins $\Delta f = 3\times 10^{-8}$ Hz. If LISA detects a GW background in this frequency interval, it is expected to be of a primordial origin. Although the low-frequency part of the binary confusion limit is somewhat more model-dependent (it is determined by the actual number of binary stars in the Galaxy, their space distribution and details of binary evolution), the calculated noise drops below the LISA sensitivity curve at frequencies below $\sim 10^{-4}$ Hz. This leaves open the search for cosmological backgrounds of primordial origin in some low-frequency portion of the LISA sensitivity window, in addition to the already discussed interval $\sim 10^{-3}-10^{-1}$ Hz.

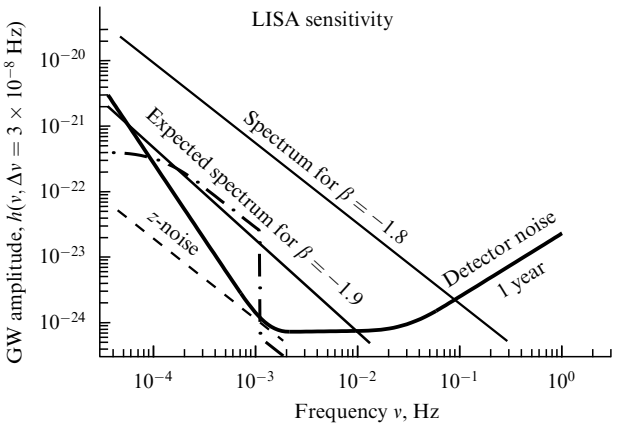

Figure 6. Expected spectrum $\beta = -1.9$ and other possible spectra in comparison with the LISA sensitivity.

5.2 Gravitational wave noise from extragalactic binaries

Simple estimates show that the isotropic extragalactic background is expected to be one and a half orders of magnitude smaller than the sky-averaged GW noise from galactic binaries (see Refs [33, 34, 36, 114]. Consider a volume of space with radius $r = 300$ Mpc. This is a large volume, but one can still neglect effects of curvature and the cosmological time dependence of the star formation rate. According to Eqn (14), the event rate in this volume $\mathcal{R}_{\rm V}$ is related to the galactic event rate $\mathcal{R}_{\rm G}$ as $\mathcal{R}_{\rm V} = 3\times 10^6 \mathcal{R}_{\rm G}$. Formula (29), written for extragalactic sources, should now contain $\mathcal{R}_{\rm V}$ instead of the galactic rate $\mathcal{R}$, and $r = 300$ Mpc instead of the characteristic galactic distance $r = 10$ kpc. Combining the numbers, one finds the relationship between the extragalactic amplitude $h_N(f)$ and the galactic amplitude $h_N^{\rm eg}(f)$:

$$h_N^{\rm eg}(f) \approx 5\times 10^{-2} h_N(f)\,. \tag{44}$$

Thus, the noise amplitude from extragalactic binaries is expected to be a factor 20 smaller than the noise amplitude from binaries in our Galaxy.

More sophisticated calculations take into account the somewhat larger star formation rate at larger red shifts [115, 122]. According to these studies, unresolved extragalactic binaries can contribute up to 10% of the mean galactic noise. This is still smaller than the projected LISA sensitivity and thus presents no danger of contamination. The contribution of unresolved extragalactic binaries is shown in Fig. 6 to the right of the frequency $\approx 10^{-3}$ Hz, where galactic binaries become resolvable and their noise contribution sharply drops down. An additional distinctive feature of GW backgrounds from sources in distant galaxies is a certain angular anisotropy of the background caused by inhomogeneities in the distribution of galaxies over the sky [123].

6. Relic gravitational waves and their detection

The existence of relic gravitational waves is a consequence of quite general assumptions. Essentially, we rely only on the validity of general relativity and the basic principles of quantum field theory. The strongly variable gravitational field of the early Universe amplifies the inevitable zero-point quantum oscillations of the GWs and produces a stochastic background of relic GWs measurable today [9, 124, 125]. The detection of relic GWs is the only way to learn about the evolution of the very early Universe, back to the limits of the Planck era and the Big Bang. It is important to appreciate the fundamental and unavoidable nature of this mechanism. Other physical processes can also generate stochastic backgrounds of GWs. But those processes either involve many additional hypotheses, which may turn out to be not true, or produce a gravitational wave background (like that from binary stars in the Galaxy) which should be treated as unwanted noise rather than a useful and interesting signal. The scientific importance of detecting relic GWs has been stressed on several occasions (see, for example, Refs [4–6]).

The central notion in the theory of relic gravitons is the phenomenon of super-adiabatic (parametric) amplification. The roots of this phenomenon are known in classical physics, and we will recall its basic features. As any wave-like process, GWs are amenable to the concept of a harmonic oscillator. The fundamental equation for a free harmonic oscillator is

$$\ddot{q} + \omega^2 q = 0\,, \tag{45}$$

where q may be the displacement of a mechanical pendulum or the time-dependent amplitude of a mode of a physical field. The energy of the oscillator can be changed by an external force or, alternatively, by a parametric influence, that is, when a parameter of the oscillator, for instance the length of the pendulum, is being changed. In the first case, the fundamental equation takes the form

$$\ddot{q} + \omega^2 q = f(t)\,, \tag{46}$$

whereas in the second case Eqn (45) takes the form

$$\ddot{q} + \omega^2(t) q = 0\,. \tag{47}$$

Equations (46) and (47) are profoundly different, both mathematically and physically.

Let us concentrate on the parametric influence. We consider a pendulum of length L oscillating in a constant

gravitational field g. The unperturbed pendulum oscillates with the constant frequency $\omega = \sqrt{g/L}$. Figure 7a illustrates the variation of the length of the pendulum $L(t)$ by an external agent, shown by alternating arrows. Since $L(t)$ varies, the frequency of the oscillator also varies: $\omega(t) = \sqrt{g/L(t)}$. The variation in $L(t)$ need not be periodic, but it cannot be too slow (i.e., adiabatic) for the result of the process to be significant. Otherwise, in the adiabatic regime of slow variations, the energy of the oscillator E and its frequency ω do change slowly, but E/ω remains constant, so one can say that the 'number of quanta' $E/\hbar\omega$ in the oscillator remains fixed. In other words, for the creation of new 'particle excitations', the characteristic time of the variation should be comparable with the period of the oscillator and the adiabatic behavior should be violated. After some duration of the appropriate parametric influence, the pendulum will oscillate at the original frequency, but will have a significantly larger amplitude and energy than before. This is shown in Fig. 7b. Obviously, the energy of the oscillator has been increased at the expense of the external agent (pump field). For simplicity, we have considered a familiar case, when the length of the pendulum varies, while the gravitational acceleration g remains constant. Variation in g represents a gravitational parametric influence that would be in even closer analogy with what we study below.

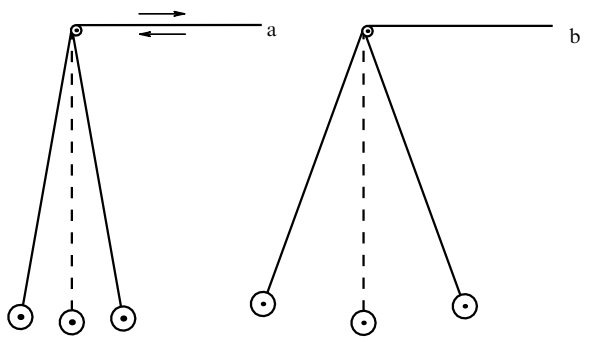

Figure 7. Parametric amplification: (a) variation of the length of the pendulum, (b) increased amplitude of oscillations.

A classical oscillator must have a non-zero initial amplitude for the amplification mechanism to work. Otherwise, if the initial amplitude is zero, the final amplitude will also be zero. Indeed, imagine the pendulum to be strictly at rest, hanging straight down. No variation in its length will cause the pendulum to oscillate and gain energy. In contrast, a quantum oscillator does not need to be excited from the very beginning. The oscillator can initially be in its quantum-mechanical vacuum state. The inevitable zero-point quantum oscillations are associated with the vacuum state energy $(1/2)\hbar\omega$. One can imagine a pendulum hanging straight down, but fluctuating with a tiny amplitude determined by the 'half quantum in the mode'. In the classical picture, it is this tiny amplitude of quantum-mechanical origin that is parametrically amplified.

The Schrödinger evolution of a quantum oscillator depends crucially on whether the oscillator is excited parametrically or by a force. Consider the phase diagram (q, p), where q is the displacement and p is the conjugate momentum. The vacuum state is described by a circle at the center of the phase space (Fig. 8). The mean values of q and p are each zero, but their variances (i.e. the zero-point quantum fluctuations)

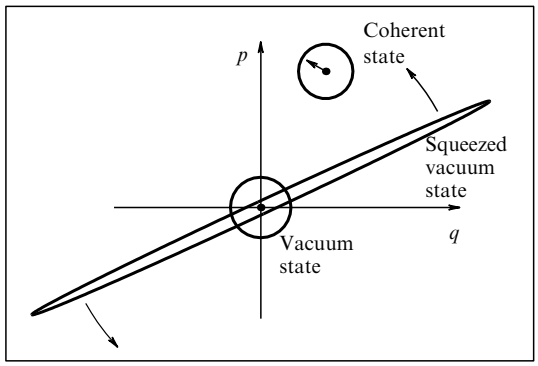

Figure 8. Some quantum states of a harmonic oscillator.

are not zeroes and are equal to each other. The magnitudes of the variances are represented by the radius of the circle at the center. Under the action of a force, the vacuum state evolves into a coherent state. The mean values of p and q have increased, but the variances are still equal and are described by a circle of the same size as for the vacuum state. On the other hand, under a parametric influence, the vacuum state evolves into a squeezed vacuum state. (For a recent review of squeezed states see, for example, [126] and references therein.) The variances of the conjugate variables q and p are now significantly unequal and are described by an ellipse. As a function of time, the ellipse rotates with respect to the origin of the (q, p) diagram, and the numerical values of the variances also oscillate. The mean numbers of quanta in the two states, one of which is coherent and the other a squeezed vacuum, may be equal (similar to the coherent and squeezed states shown in Fig. 8), but the statistical properties of these states are significantly different. Among other things, the variance of the phase of the oscillator in a squeezed vacuum state is very small (hence the name, squeezed). Graphically, this is reflected in the fact that the ellipse is very thin, so that the uncertainty in the angle between the horizontal axis and the orientation of the ellipse is very small. This highly elongated ellipse can be regarded as a portrait of the gravitational wave quantum state that is inevitably generated by parametric amplification, and which we will be dealing with below.

A wave field is not a single oscillator; it depends on spatial coordinates and time, and may have several independent components (polarization states). However, the field can be decomposed into a set of spatial Fourier harmonics. In this way we represent the gravitational wave field as a collection of many modes, many oscillators. Because of the nonlinear character of the Einstein equations, each of these oscillators is coupled to the variable gravitational field of the surrounding Universe. For sufficiently short GWs of experimental interest, this coupling was especially efficient in the early Universe, when the condition of adiabatic behavior of the oscillator was violated. It is this homogeneous and isotropic gravitational field of all the matter in the early Universe that played the role of an external agent — the pump field. The variable pump field acts parametrically on the gravity-wave oscillators and drives them into multi-particle states. Specifically, the initial vacuum state of each pair of waves with oppositely directed momenta evolves into a highly correlated state known as the two-mode squeezed vacuum state [127–130]. The strength and duration of the effective coupling

depends on the oscillator's frequency. They all start in the vacuum state but get excited by various amounts. As a result, a broad spectrum of relic GWs is formed. This spectrum is accessible to our observations today.

6.1 Cosmological gravitational waves

In the framework of general relativity, a homogeneous isotropic gravitational field is described by the line element

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - a^2(t)\delta_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j\,.$$

It is more convenient to introduce a new time coordinate η and to write $\mathrm{d}s^2$ in the form

$$\mathrm{d}s^2 = a^2(\eta)(\mathrm{d}\eta^2 - \delta_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j)\,. \tag{48}$$

In cosmology, the function $a(t)$ [or $a(\eta)$] is called the scale factor. In our discussion, it will represent the gravitational pump field.

Cosmological GWs are small corrections h_{ij} to the metric tensor. They are defined by the expression

$$\mathrm{d}s^2 = a^2(\eta)\left[\mathrm{d}\eta^2 - (\delta_{ij} + h_{ij})\,\mathrm{d}x^i\,\mathrm{d}x^j\right]. \tag{49}$$

The functions $h_{ij}(\eta, \mathbf{x})$ can be expanded over spatial Fourier harmonics $\exp(+\mathrm{i}\mathbf{n}\mathbf{x})$ and $\exp(-\mathrm{i}\mathbf{n}\mathbf{x})$, where $\mathbf{n}$ is a constant wave vector. In this way, we reduce the dynamical problem to the evolution of time-dependent amplitudes for each mode $\mathbf{n}$. Among six functions h_{ij} there are only two independent (polarization) components. This decomposition can be made for both a classical and a quantized field h_{ij}. In the quantum version, the functions h_{ij} are treated as quantum-mechanical operators. We will use the Heisenberg picture, in which the time evolution is carried by the operators while the quantum state is fixed. This picture is fully equivalent to the Schrödinger picture, discussed in the Introduction, in which the vacuum state evolves into a squeezed vacuum state while the operators are time-independent.

The Heisenberg operator for the quantized real field h_{ij} can be written as

$$h_{ij}(\eta, \mathbf{x}) = \frac{C}{(2\pi)^{3/2}}\int_{-\infty}^{\infty}\mathrm{d}^3\mathbf{n}\sum_{s=1}^{2}\overset{s}{p}_{ij}(\mathbf{n}) \times \frac{1}{\sqrt{2n}}\left[\overset{s}{h}_n(\eta)\exp(\mathrm{i}\mathbf{n}\mathbf{x})\,\overset{s}{c}_{\mathbf{n}} + \overset{s}{h}{}^*_n(\eta)\exp(-\mathrm{i}\mathbf{n}\mathbf{x})\,\overset{s}{c}{}^{\dagger}_{\mathbf{n}}\right], \tag{50}$$

where C is a constant which will be discussed later. The creation and annihilation operators satisfy the conditions

$$\left[\overset{s'}{c}_{\mathbf{n}}, \overset{s}{c}{}^{\dagger}_{\mathbf{m}}\right] = \delta_{s's}\delta^3(\mathbf{n} - \mathbf{m})\,, \quad \overset{s}{c}_{\mathbf{n}}|0\rangle = 0\,,$$

where $|0\rangle$ (for each $\mathbf{n}$ and s) is the fixed initial vacuum state discussed below. The wave number n is related to the wave vector $\mathbf{n}$ by $n = (\delta_{ij}n^i n^j)^{1/2}$. The two polarization tensors $\overset{s}{p}_{ij}(\mathbf{n})$ $(s = 1, 2)$ obey the conditions

$$\overset{s}{p}_{ij}n^j = 0\,, \quad \overset{s}{p}_{ij}\delta^{ij} = 0\,, \quad \overset{s'}{p}_{ij}\overset{s}{p}{}^{ij} = 2\delta_{ss'}\,, \quad \overset{s}{p}_{ij}(-\mathbf{n}) = \overset{s}{p}_{ij}(\mathbf{n})\,.$$

The time evolution, one and the same for all $\mathbf{n}$ belonging to a given n, is represented by the complex time-dependent function $\overset{s}{h}_n(\eta)$. This evolution is dictated by the Einstein equations. The nonlinear nature of the Einstein equations leads to the coupling of $\overset{s}{h}_n(\eta)$ with the pump field $a(\eta)$. For any wave number n and each polarization component s, the functions $\overset{s}{h}_n(\eta)$ have the form

$$\overset{s}{h}_n(\eta) = \frac{1}{a(\eta)}\left[\overset{s}{u}_n(\eta) + \overset{s}{v}{}^*_n(\eta)\right], \tag{51}$$

where $\overset{s}{u}_n(\eta)$ and $\overset{s}{v}_n(\eta)$ can be expressed in terms of three real functions (the polarization index s is omitted): r_n is the squeeze parameter, ϕ_n is the squeeze angle, and θ_n is the rotation angle,

$$u_n = \exp(\mathrm{i}\theta_n)\cosh r_n\,, \qquad v_n = \exp\left[-\mathrm{i}(\theta_n - 2\phi_n)\right]\sinh r_n\,. \tag{52}$$

The dynamical equations for $u_n(\eta)$ and $v_n(\eta)$

$$\mathrm{i}\,\frac{\mathrm{d}u_n}{\mathrm{d}\eta} = nu_n + \mathrm{i}\,\frac{a'}{a}\,v_n^*\,, \qquad \mathrm{i}\,\frac{\mathrm{d}v_n}{\mathrm{d}\eta} = nv_n + \mathrm{i}\,\frac{a'}{a}\,u_n^* \tag{53}$$

lead to dynamical equations governing the functions $r_n(\eta)$, $\phi_n(\eta)$, and $\theta_n(\eta)$ [129, 130]:

$$r_n' = \frac{a'}{a}\cos 2\phi_n\,, \qquad \phi_n' = -n - \frac{a'}{a}\sin 2\phi_n\coth 2r_n\,,$$
$$\theta_n' = -n - \frac{a'}{a}\sin 2\phi_n\tanh r_n\,, \tag{54}$$

where $' \equiv \mathrm{d}/\mathrm{d}\eta$, and the evolution begins from $r_n = 0$. This value of r_n characterizes the initial vacuum state $|0\rangle$ which is defined long before the interaction with the pump field becomes efficient, that is, long before the coupling term a'/a becomes comparable with n. The constant C should be taken as $C = \sqrt{16\pi}\,l_{\mathrm{Pl}}$, where $l_{\mathrm{Pl}} = (G\hbar/c^3)^{1/2}$ is the Planck length. This particular value of the constant C guarantees a correct quantum normalization of the field: an energy $(1/2)\hbar\omega$ per mode in the initial vacuum state. The dynamical equations and their solutions are identical for both polarization components s.

Equations (53) can be translated into the more familiar form of the second-order differential equation for the function

$$\overset{s}{\mu}_n(\eta) \equiv \overset{s}{u}_n(\eta) + \overset{s}{v}{}^*_n(\eta) \equiv a(\eta)\overset{s}{h}_n(\eta)$$

(see Refs [9, 124, 125]):

$$\mu_n'' + \mu_n\left(n^2 - \frac{a''}{a}\right) = 0\,. \tag{55}$$

Clearly, this is the equation for a parametrically disturbed oscillator [compare with Eqn (47)]. In the absence of the gravitational parametric influence represented by the term a''/a, the frequency n of the oscillator defined in terms of η-time would be a constant. Whenever the term a''/a can be neglected, the general solution to Eqn (55) has the usual oscillatory form

$$\mu_n(\eta) = A_n\exp(-\mathrm{i}n\eta) + B_n\exp(+\mathrm{i}n\eta)\,, \tag{56}$$

where the constants A_n and B_n are determined by the initial conditions. On the other hand, whenever the term a''/a is dominant, the general solution to Eqn (55) has the form

$$\mu_n(\eta) = C_n a + D_n a\int_{-\infty}^{\eta}\frac{\mathrm{d}\eta}{a^2}\,. \tag{57}$$

In fact, this approximate solution is valid as long as n is small in comparison with $|a'/a|$. This is more clearly seen from the equivalent form of Eqn (55) written in terms of the function $h_n(\eta)$ [2]:

$$h_n'' + 2\,\frac{a'}{a}\,h_n' + n^2 h_n = 0\,. \tag{58}$$

For growing functions $a(\eta)$, that is, in expanding universes, the second term in Eqn (57) is usually smaller than the first one (see below), so that, as long as $n \ll a'/a$, the dominant solution is the growing function $\mu_n(\eta) = C_n a(\eta)$, and

$$h_n = \text{const}\,. \tag{59}$$

Equation (55) can also be treated as the Schrödinger equation for a particle moving in the presence of an effective potential $U(\eta) = a''/a$. In the situations normally considered, the potential $U(\eta)$ has a bell-like shape and forms a barrier (Fig. 9). When a given mode n is outside the barrier, its amplitude h_n adiabatically decreases with time: $h_n \propto \exp(\pm in\eta)/a(\eta)$. This is shown in Fig. 9 by oscillating lines with decreasing amplitudes of oscillation. The modes with sufficiently high frequencies do not interact with and stay above the barrier. Their amplitudes h_n behave adiabatically all the time. For these high-frequency modes, the initial vacuum state (in the Schrödinger picture) remains the vacuum state forever. On the other hand, the modes that interact with the barrier are subject to super-adiabatic amplification. Under the barrier and as long as $n < a'/a$, the function h_n remains constant instead of decreasing adiabatically. For these modes, the initial vacuum state evolves into a squeezed vacuum state.

After having formulated the initial conditions, the present-day behavior of r_n, ϕ_n, θ_n (or, equivalently, the present-day behavior of h_n) is essentially all we need to find. The mean number of particles in a two-mode squeezed state is $2\sinh^2 r_n$ for each s. This number determines the mean square amplitude of the gravitational wave field. The time behavior of the squeeze angle ϕ_n determines the time dependence of the correlation functions of the field. The amplification (that is, the growth of r_n) governed by Eqn (54) is different for different wave numbers n. Therefore, the present-day results depend on the present-day frequency ν measured in Hz.

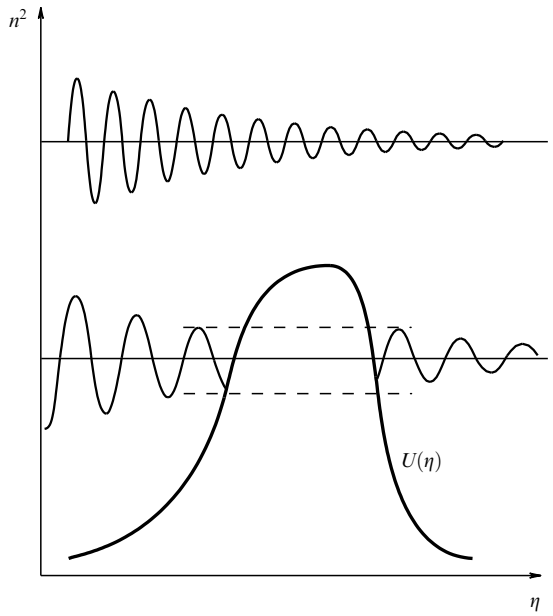

Figure 9. Effective potential $U(\eta)$.

In cosmology, the function $H \equiv \dot{a}/a \equiv ca'/a^2$ is the time-dependent Hubble parameter. The function $l \equiv c/H$ is the time-dependent Hubble radius. The time-dependent wavelength of the mode n is $\lambda = 2\pi a/n$. The wavelength λ has this universal definition in all regimes. In contrast, the quantity ν defined as $\nu = cn/2\pi a$ has the usual meaning of the frequency of an oscillating process only in the short-wavelength (high-frequency) regime of mode n, that is, in the regime where $\lambda \ll l$. As we have seen above, the qualitative behavior of the solutions to Eqns (55), (58) depends crucially on the comparative values of n and a'/a, or, in other words, on the comparative values of $\lambda(\eta)$ and $l(\eta)$. This relationship is also crucial for the solutions to Eqn (54), as we shall see now.

In the short-wavelength regime, that is, during intervals of time when the wavelength $\lambda(\eta)$ is shorter than the Hubble radius $l(\eta) = a^2/a'$, the term containing n in (54) is dominant. The functions $\phi_n(\eta)$ and $\theta_n(\eta)$ are $\phi_n = -n(\eta + \eta_n)$ and $\theta_n = \phi_n$, where η_n is a constant. The factor $\cos 2\phi_n$ is a rapidly oscillating function of time, so the squeeze parameter r_n remains virtually constant. This is the adiabatic regime for a given mode.

In the opposite, long-wavelength regime, the term n can be neglected. The function ϕ_n is $\tan\phi_n(\eta) \approx \text{const}/a^2(\eta)$, and the squeeze angle quickly approaches one of the two values, $\phi_n = 0$ or $\phi_n = \pi$ (analog of 'phase bifurcation' [131, 132]). When the long-wavelength regime for a given n begins at some $\eta = \eta_*$, the squeeze parameter $r_n(\eta)$ grows with time according to the equation

$$r_n(\eta) \approx nl\,\frac{a(\eta)}{a_*}\,, \tag{60}$$

where a_* is the value of $a(\eta)$ at η_*. The final value of r_n is

$$r_n \approx nl\,\frac{a_{**}}{a_*}\,, \tag{61}$$

where a_{**} is the value of $a(\eta)$ at η_{**}, the point of cessation of the long-wavelength regime and amplification. It is important to emphasize that it is not a 'sudden transition' from one cosmological era to another that is responsible for the amplification, but the entire interval of the long-wavelength (non-adiabatic) regime.

After the end of amplification, the accumulated (and typically large) squeeze parameter r_n remains approximately constant. The mode is again in the adiabatic regime. In the course of evolution, the complex functions $\overset{s}{u}_n(\eta) + \overset{s}{v}{}_n^*(\eta)$ become virtually real, and one has

$$\overset{s}{h}_n(\eta) \approx \overset{s}{h}{}_n^*(\eta) \approx \frac{1}{a}\exp(r_n)\cos\phi_n(\eta)\,.$$

Any amplified mode n of the field (50) takes the form of a product of a function of time and a (random, operator-valued) function of spatial coordinates; the mode acquires a standing-wave pattern. The periodic dependence $\cos\phi_n(\eta)$ will be further discussed below [see Eqn (75)].

It is clearly seen from the fundamental equations (54), (55), (58) that the final results depend only on $a(\eta)$. The equations do not ask us the names of our favorite cosmological prejudices; they ask us about the pump field $a(\eta)$. Conversely, from the measured relic GWs, we can deduce

the behavior of $a(\eta)$, which is essentially the purpose of detecting relic gravitons.

6.2 Cosmological pump field

With the chosen initial conditions, the final numerical results for relic GWs depend on the particular behavior of the pump field represented by the cosmological scale factor $a(\eta)$. We know a great deal about $a(\eta)$. We know that $a(\eta)$ behaves as $a(\eta) \propto \eta^2$ at the present matter-dominated stage. We know that this stage was preceded by a radiation-dominated stage in which $a(\eta) \propto \eta$. At these two stages of evolution the functions $a(\eta)$ are simple power-law functions of η. What we do not know is the function $a(\eta)$ describing the initial stage of expansion of the very early Universe, that is, before the era of primordial nucleosynthesis. It is convenient to parameterize $a(\eta)$ at this initial stage also by power-law functions of η. First, this is a fairly broad class of functions, which, in addition, allows us to find exact solutions to our fundamental equations. Second, it is known (see Refs [9, 124, 125]) that the pump fields $a(\eta)$ which have a power-law dependence in terms of η, produce GWs with simple power-law spectra in terms of ν. These spectra are easy to analyze and discuss in the context of detection.

We model cosmological expansion by several successive eras. Specifically, we take $a(\eta)$ at the initial stage of expansion (the i-stage) as

$$a(\eta) = l_0 |\eta|^{1+\beta} , \tag{62}$$

where η grows from $-\infty$, and $1 + \beta < 0$. We will show later how the available observational data constrain the parameters l_0 and β. The i-stage lasts until a certain $\eta = \eta_1$, $\eta_1 < 0$. To make our analysis more general, we assume that the i-stage was followed by some interval of a z-stage (the notation z is derived from the name Zeldovich). It is known that an interval of evolution governed by the 'stiffest' matter (with the effective equation of state $p = \epsilon$), advocated by Zeldovich, leads to a relative increase of gravitational wave amplitudes [9, 124, 125]. It is also known that the requirement of consistency of the graviton production with the observational restrictions does not allow the 'stiff' matter interval to be too long [9, 124, 125, 133]. However, we want to investigate any interval of cosmological evolution that can be consistently included. In fact, the z-stage of expansion that we include is quite general. It can be governed by matter 'stiffer than radiation' [134], as well as by matter 'softer than radiation'. It can also be simply a part of the radiation-dominated era. Specifically, we take $a(\eta)$ for the interval of time from η_1 to some η_s (z-stage) in the form

$$a(\eta) = l_0 \, a_z (\eta - \eta_p)^{1+\beta_s} , \tag{63}$$

where $1 + \beta_s > 0$. For this particular choice the z-stage reduces to an interval of expansion governed by the radiation-dominated matter. Starting from η_s and until η_2 the Universe was governed by radiation-dominated matter (e-stage). So, in this interval of evolution, we take the scale factor in the form

$$a(\eta) = l_0 \, a_e (\eta - \eta_e) . \tag{64}$$

And, finally, from $\eta = \eta_2$ the expansion switched to the matter-dominated era (the m-stage):

$$a(\eta) = l_0 \, a_m (\eta - \eta_m)^2 . \tag{65}$$

A link between the arbitrary constants that appear in Eqns (62) – (65) is provided by the conditions of continuous joining of the functions $a(\eta)$ and $a'(\eta)$ at the points of transition η_1, η_s, η_2.

We denote the present time by η_R (the notation R is derived from 'reception'). This time is defined by the observationally known value of the present-day Hubble parameter $H(\eta_R)$ and the Hubble radius $l_H = c/H(\eta_R)$. For numerical estimates we will use $l_H \approx 2 \times 10^{28}$ cm. It is convenient to choose $\eta_R - \eta_m = 1$, so that $a(\eta_R) = 2 l_H$. The ratio

$$\frac{a(\eta_R)}{a(\eta_2)} \equiv \zeta_2$$

is believed to be around $\zeta_2 = 10^4$. We also denote

$$\frac{a(\eta_2)}{a(\eta_s)} \equiv \zeta_s , \qquad \frac{a(\eta_s)}{a(\eta_1)} \equiv \zeta_1 .$$

With these definitions, all the constants present in Eqns (62) – (65) (except the parameters β and β_s which should be chosen from other considerations) can be expressed in terms of l_H, ζ_2, ζ_s, and ζ_1. For example,

$$|\eta_1| = \frac{|1 + \beta|}{2 \zeta_2^{1/2} \zeta_s \zeta_1^{1/(1+\beta_s)}} .$$

The important constant l_0 is expressed as

$$l_0 = b \, l_H \, \zeta_2^{(\beta-1)/2} \zeta_s^{\beta} \zeta_1^{(\beta-\beta_s)/(1+\beta_s)} , \tag{66}$$

where $b \equiv 2^{2+\beta}/|1 + \beta|^{1+\beta}$. Note that $b = 1$ for $\beta = -2$. (This expression for l_0 can help us to relate the formulas written here with the equivalent treatment [135] which was given in slightly different notation.) The sketch of the entire evolution $a(\eta)$ is given in Fig. 10.

We work with spatially flat models (48). At any instant of time, the energy density $\epsilon(\eta)$ of matter driving the evolution is related to the Hubble radius $l(\eta)$ by

$$\varkappa \epsilon(\eta) = \frac{3}{l^2(\eta)} , \tag{67}$$

where $\varkappa = 8\pi G/c^4$. For the case of power-law scale factors $a(\eta) \propto \eta^{1+\beta}$, the effective matter pressure $p(\eta)$ is related to the energy density $\epsilon(\eta)$ by the effective equation of state

$$p = \frac{1 - \beta}{3(1 + \beta)} \epsilon . \tag{68}$$

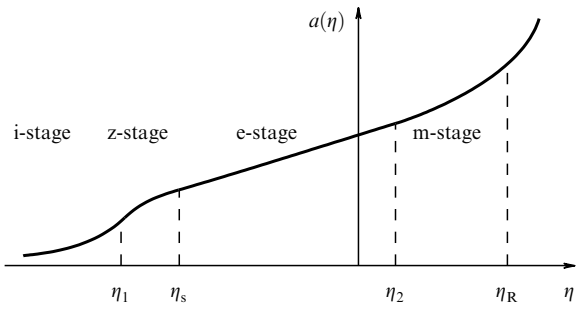

Figure 10. Scale factor $a(\eta)$.\

For instance, $p = 0$ for $\beta = 1$, $p = (1/3)\epsilon$ for $\beta = 0$, $p = -\epsilon$ for $\beta = -2$, and so on. Each interval of the evolution (62)–(65) is governed by one of these equations of state.

In principle, the function $a(\eta)$ could be even more complicated than the one that we consider. It could even include an interval of early contraction, instead of expansion, leading to a 'bounce' of the scale factor. In the case of a decreasing $a(\eta)$ the gravitational-wave equation can still be analyzed and the amplification is still efficient [9, 124, 125]. However, the Einstein equations for spatially flat models do not permit a regular 'bounce' of $a(\eta)$ (unless ϵ vanishes at the moment of 'bounce'). Possibly, a 'bounce' solution can be realized in alternative theories, such as string-motivated cosmologies [136–138]. For a recent discussion of spectral slopes of GWs produced in 'bounce' cosmologies, see Ref. [139].

6.3 Solving gravitational wave equations

The evolution of the scale factor $a(\eta)$ given by Eqns (62)–(65) and sketched in Fig. 10 allows us to calculate the function a'/a. This function is sketched in Fig. 11. In all theoretical generality, the left-hand side of the barrier in Fig. 11 could also consist of several pieces, but we do not consider this possibility here. The graph also shows the important wave numbers $n_{\rm H}$, n_2, $n_{\rm s}$, n_1: $n_{\rm H}$ marks the wave whose present-day wavelength $\lambda(\eta_{\rm R}) = 2\pi a(\eta_{\rm R})/n_{\rm H}$ is equal to the present-day Hubble radius $l_{\rm H}$. With our parameterization $a(\eta_{\rm R}) = 2l_{\rm H}$, this wave-number is $n_{\rm H} = 4\pi$. The number n_2 marks the wave whose wavelength $\lambda(\eta_2) = 2\pi a(\eta_2)/n_2$ at $\eta = \eta_2$ is equal to the Hubble radius $l(\eta_2)$ at $\eta = \eta_2$. Since

$$\frac{\lambda(\eta_{\rm R})}{\lambda(\eta_2)} = \frac{n_2}{n_{\rm H}}\,\frac{a(\eta_{\rm R})}{a(\eta_2)}$$

and

$$\frac{l(\eta_{\rm R})}{l(\eta_2)} = \frac{a(\eta_{\rm R})}{a(\eta_2)}\left[\frac{a(\eta_{\rm R})}{a(\eta_2)}\right]^{1/2},$$

this gives us $n_2/n_{\rm H} = [a(\eta_{\rm R})/a(\eta_2)]^{1/2} = \zeta_2^{1/2}$. Working out the other ratios in a similar fashion, we find

$$\frac{n_2}{n_{\rm H}} = \zeta_2^{1/2}\,, \qquad \frac{n_{\rm s}}{n_2} = \zeta_{\rm s}\,, \qquad \frac{n_1}{n_{\rm s}} = \zeta_1^{1/(1+\beta_{\rm s})}\,. \tag{69}$$

Solutions to the gravitational wave equations exist for any $a(\eta)$. Within the intervals of power-law dependence $a(\eta)$, solutions to Eqn (55) have the simple form of Bessel functions. We could find exact piece-wise solutions to Eqn (55) and join them at the transition points. However, we will use a much simpler treatment, which is sufficient for our purposes. We know that the squeeze parameter r_n remains constant in the short-wavelength regimes and grows according to Eqn (60) in the long-wavelength regime. All modes start in the vacuum state, that is, $r_n = 0$ initially. After the end of amplification, the accumulated value (61) remains constant up to now. To find today's value of $\exp r_n$, we need to calculate the ratio $a_{**}(n)/a_*(n)$. For any given n, the quantity a_* is determined by the condition $\lambda(\eta_*) = l(\eta_*)$, whereas a_{**} is determined by the condition $\lambda(\eta_{**}) = l(\eta_{**})$.

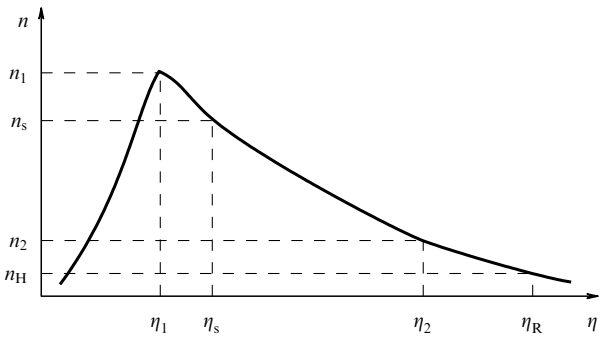

Figure 11. Function a'/a for the scale factor from Fig. 10.

Let us start from the mode $n = n_1$. For this wave number we have $a_* = a_{**} = a(\eta_1)$ and, therefore, $r_{n_1} = 0$. Higher frequency modes, i.e. $n > n_1$ (above the barrier in Fig. 11), have never been in the amplifying regime, so we can write

$$\exp r_n = 1\,, \qquad n \geqslant n_1\,. \tag{70}$$

Let us now consider the modes n in the interval $n_1 \geqslant n \geqslant n_{\rm s}$. For a given n we need to know $a_*(n)$ and $a_{**}(n)$. Using Eqn (62) one has $a_*(n)/a_*(n_1) = (n_1/n)^{1+\beta}$, and using Eqn (63) one finds $a_{**}(n)/a_{**}(n_{\rm s}) = (n_{\rm s}/n)^{1+\beta_{\rm s}}$. Therefore, one finds

$$\frac{a_{**}(n)}{a_*(n)} = \frac{a_{**}(n_{\rm s})}{a_*(n_1)}\left(\frac{n_{\rm s}}{n}\right)^{1+\beta_{\rm s}}\left(\frac{n}{n_1}\right)^{1+\beta}.$$

Since $a_{**}(n_{\rm s}) = a(\eta_{\rm s})$, $a_*(n_1) = a(\eta_1)$, and

$$\frac{a(\eta_{\rm s})}{a(\eta_1)} = \zeta_1 = \left(\frac{n_1}{n_{\rm s}}\right)^{1+\beta_{\rm s}},$$

we arrive at

$$\frac{a_{**}(n)}{a_*(n)} = \left(\frac{n}{n_1}\right)^{\beta-\beta_{\rm s}}.$$

Repeating this analysis for other intervals of decreasing n, we come to the conclusion that

$$\begin{aligned}
\exp r_n &= \left(\frac{n}{n_1}\right)^{\beta-\beta_{\rm s}}, \qquad n_1 \geqslant n \geqslant n_{\rm s}\,,\\
\exp r_n &= \left(\frac{n}{n_{\rm s}}\right)^{\beta}\left(\frac{n_{\rm s}}{n_1}\right)^{\beta-\beta_{\rm s}}, \qquad n_{\rm s} \geqslant n \geqslant n_2\,,\\
\exp r_n &= \left(\frac{n}{n_2}\right)^{\beta-1}\left(\frac{n_2}{n_1}\right)^{\beta}\left(\frac{n_{\rm s}}{n_1}\right)^{-\beta_{\rm s}}, \qquad n_2 \geqslant n \geqslant n_{\rm H}\,.
\end{aligned} \tag{71}$$

A mnemonic rule for constructing $\exp r_n$ at successive intervals of decreasing n is simple. If the interval begins at n_x, one takes $(n/n_x)^{\beta_*-\beta_{**}}$ and multiples by $\exp r_{n_x}$, that is, by the preceding interval's value of $\exp r_n$ calculated at the end of the interval n_x. For the function a'/a that we are working with, the quantity β_* is always equal to β, whereas β_{**} takes the values $\beta_{\rm s}$, 0, 1, in successive intervals.

The modes with $n < n_{\rm H}$ are still in the long-wavelength regime. For these modes, we should take $a(\eta_{\rm R})$ instead of $a_{**}(n)$. Combining with $a_*(n)$, we find

$$\exp r_n = \left(\frac{n}{n_{\rm H}}\right)^{\beta+1}\left(\frac{n_{\rm H}}{n_2}\right)^{\beta-1}\left(\frac{n_2}{n_1}\right)^{\beta}\left(\frac{n_{\rm s}}{n_1}\right)^{-\beta_{\rm s}}, \qquad n \leqslant n_{\rm H}\,. \tag{72}$$

Formulas (70)–(72) give approximate values of r_n for all n. The factor $\exp r_n$ obeys the following inequalities: $\exp r_n \geqslant 1$ for $n \leqslant n_1$, and $\exp r_n \gg 1$ for $n \ll n_1$, and determines the mean square amplitude of the GWs.

The mean value of the field h_{ij} is zero at any moment of time η and at any spatial point $\mathbf{x}$: $\langle 0|h_{ij}(\eta,\mathbf{x})|0\rangle = 0$. The variance

$$\langle 0|h_{ij}(\eta,\mathbf{x})h^{ij}(\eta,\mathbf{x})|0\rangle \equiv \langle h^2\rangle$$

is not zero and determines the mean square amplitude of the generated field — the quantity of interest for the experiment. Taking the product of two expressions (50), one can show that

$$\langle h^2\rangle = \frac{C^2}{2\pi^2}\int_0^\infty n\sum_{s=1}^{2}\left|{}^{s}h_n(\eta)\right|^2 \mathrm{d}n \equiv \int_0^\infty h^2(n,\eta)\,\frac{\mathrm{d}n}{n}\,. \qquad (73)$$

Using the representation (51), (52) in Eqn (73) one can also write

$$\langle h^2\rangle = \frac{C^2}{\pi^2 a^2}\int_0^\infty n\,\mathrm{d}n\,(\cosh 2r_n + \cos 2\phi_n \sinh 2r_n)\,. \qquad (74)$$

We can now consider the present era and use the fact that $\exp r_n$ are large numbers for all n in the interval of interest $n_1 \geqslant n \geqslant n_\mathrm{H}$. Then, we can derive

$$h(n,\eta) \approx \frac{C}{\pi}\,\frac{1}{a(\eta_\mathrm{R})}\,n\exp r_n\,\cos\phi_n(\eta)$$

$$= 8\sqrt{\pi}\,\frac{l_\mathrm{Pl}}{l_\mathrm{H}}\,\frac{n}{n_\mathrm{H}}\,\exp r_n\,\cos\phi_n(\eta)\,. \qquad (75)$$

The quantity $h(n,\eta)$ is the dimensionless spectral amplitude of the field whose numerical value is determined by the calculated squeeze parameter r_n. The oscillatory factor $\cos\phi_n(\eta)$ reflects the squeezing (standing wave pattern) acquired by modes with $n_1 > n > n_\mathrm{H}$. For modes with $n < n_\mathrm{H}$ this factor is approximately 1. For high-frequency modes $n \gg n_\mathrm{H}$ one has $\phi_n(\eta) \sim n(\eta - \eta_n) \gg 1$, so that $h(n,\eta)$ makes many oscillations while the scale factor $a(\eta)$ is practically fixed at $a(\eta_\mathrm{R})$.

The integral (74) extends formally from 0 to ∞. Since $r_n \approx 0$ for $n \geqslant n_1$, the integral diverges at the upper limit. This is a typical ultra-violet divergence. It should be removed (renormalized to zero) because it comes from the modes which are always in their vacuum state. At the lower limit, the integral diverges, if $\beta \leqslant -2$. This is an infra-red divergence which comes from the assumption that the amplification process started from an infinitely remote time in the past. One can deal with this divergence either by introducing a lower frequency cut-off (equivalent to a finite duration of amplification) or by considering only the parameters $\beta > -2$, in which case the integral is convergent at the lower limit. It turns out that the available observational data (see Section 6.4) favor this second option. The particular case $\beta = -2$ corresponds to the de Sitter evolution $a(\eta) \propto |\eta|^{-1}$. In this case, the $h(n)$ found in Eqns (72), (75) does not depend on n. This is known as the Harrison–Zeldovich, or scale-invariant, spectrum.

The spectral amplitudes $h(n)$ can also be derived using the approximate solutions (56), (57) to the wave equation (55). This method gives exactly the same numerical values of $h(n)$ as in Eqns (70)–(72), (75), but does not reproduce the oscillatory factor $\cos\phi_n(\eta)$.

One begins with the initial spectral amplitude $h_\mathrm{i}(n)$ determined by quantum normalization: $h_\mathrm{i}(n) = 8\sqrt{\pi}(l_\mathrm{Pl}/\lambda_\mathrm{i})$. This is the amplitude of mode n at the moment η_* of entering the long wavelength regime, i.e. when the mode's wavelength λ_i is equal to the Hubble radius $l(\eta_*)$. For λ_i one derives

$$\lambda_\mathrm{i} = \frac{1}{b}\,l_0\left(\frac{n_\mathrm{H}}{n}\right)^{2+\beta}. \qquad (76)$$

Thus, we have

$$h_\mathrm{i}(n) = A\left(\frac{n}{n_\mathrm{H}}\right)^{2+\beta}, \qquad (77)$$

where A denotes the constant

$$A = b8\sqrt{\pi}\,\frac{l_\mathrm{Pl}}{l_0}\,. \qquad (78)$$

The numbers $h_\mathrm{i}(n)$ are specified at the beginning of the long-wavelength regime. In other words, they are given along the left-hand-side slope of the barrier in Fig. 11. We want to know the final numbers (spectral amplitudes) $h(n)$ which describe the field today, at η_R.

According to the dominant solution $h_n(\eta) = \mathrm{const}$ for the long-wavelength regime [see Eqn (59)], the initial amplitude $h_\mathrm{i}(n)$ remains virtually constant until the end of the long-wavelength regime at η_{**}, that is, until the right-hand-side slope of the barrier. [The second term in Eqn (57) could be important only at the z-stage and only for parameters $\beta_\mathrm{s} \leqslant -1/2$, which correspond to the effective equations of state $p \geqslant \epsilon$. In order to keep the analysis simple, we do not consider those cases.] After the completion of the long-wavelength regime, the amplitudes decrease adiabatically in proportion to $1/a(\eta)$, up to the present time. Thus, we have

$$h(n) = A\left(\frac{n}{n_\mathrm{H}}\right)^{2+\beta}\frac{a_{**}(n)}{a(\eta_\mathrm{R})}\,. \qquad (79)$$

Let us start from the lower end of the spectrum, $n \leqslant n_\mathrm{H}$, and go upward in n. The modes $n \leqslant n_\mathrm{H}$ have not yet started the adiabatic decrease of amplitudes, so we have

$$h(n) = A\left(\frac{n}{n_\mathrm{H}}\right)^{2+\beta}, \qquad n \leqslant n_\mathrm{H}\,. \qquad (80)$$

Now consider the interval $n_2 \geqslant n \geqslant n_\mathrm{H}$. In this interval, $a_{**}(n)/a(\eta_\mathrm{R})$ scales as $(n_\mathrm{H}/n)^2$, so we have

$$h(n) = A\left(\frac{n}{n_\mathrm{H}}\right)^{\beta}, \qquad n_2 \geqslant n \geqslant n_\mathrm{H}\,. \qquad (81)$$

In the interval $n_\mathrm{s} \geqslant n \geqslant n_2$ the ratio

$$\frac{a_{**}(n)}{a(\eta_\mathrm{R})} = \frac{a_{**}(n)}{a(\eta_2)}\,\frac{a(\eta_2)}{a(\eta_\mathrm{R})}$$

scales as $(n_2/n)(n_\mathrm{H}/n_2)^2$, so we have

$$h(n) = A\left(\frac{n}{n_\mathrm{H}}\right)^{1+\beta}\frac{n_\mathrm{H}}{n_2}\,, \qquad n_\mathrm{s} \geqslant n \geqslant n_2\,. \qquad (82)$$

Repeating the same analysis for the interval $n_1 \geqslant n \geqslant n_\mathrm{s}$ we find

$$h(n) = A\left(\frac{n}{n_\mathrm{H}}\right)^{1+\beta-\beta_\mathrm{s}}\left(\frac{n_\mathrm{s}}{n_\mathrm{H}}\right)^{\beta_\mathrm{s}}\frac{n_\mathrm{H}}{n_2}\,, \qquad n_1 \geqslant n \geqslant n_\mathrm{s}\,. \qquad (83)$$

It is seen from Eqn (83) that the interval of the z-stage with $\beta_s < 0$ (the already imposed restrictions also require $-1/2 < \beta_s$) bends the spectrum $h(n)$ upwards, as compared with Eqn (82), for larger n. If one recalls the relationship (66) between l_0 and l_H and uses (71), (72) in Eqn (75), one arrives exactly at Eqns (80)–(83) up to the oscillating factor $\cos\phi_n(\eta)$.

Different parts of the barrier in Fig. 11 are responsible for the amplitudes and spectral slopes in different intervals of n. The sketch of the generated spectrum $h(n)$ in conjunction with the form of the barrier a'/a is shown in Fig. 12. The present-day frequency of the oscillating modes, measured in Hz, is defined as $\nu = cn/2\pi a(\eta_R)$. The lowest (Hubble) frequency is $\nu_H = c/l_H$. For numerical estimates we will use $\nu_H \approx 10^{-18}$ Hz. The ratios of n are equal to the ratios of ν, so that, for example, $n/n_H = \nu/\nu_H$. For high-frequency modes we will now often use the ratios of ν instead of ratios of n.

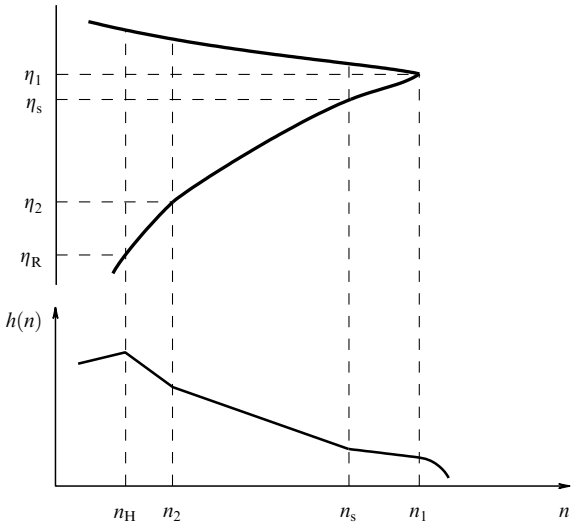

Figure 12. Amplitudes and spectral slopes of $h(n)$ are determined by different parts of the barrier a'/a.

In addition to the spectral amplitudes $h(n)$, the generated field can also be characterized by the spectral energy density parameter $\Omega_{gw}(n)$. The energy density ϵ_{gw} of the gravitational wave field is

$$\varkappa\epsilon_{gw} = \frac{1}{4}\, h^{ij}{}_{,0} h_{ij,0} = \frac{1}{4a^2}\, h'^{ij} h'_{ij}\,.$$

The mean value $\langle 0|\epsilon_{gw}(\eta,\mathbf{x})|0\rangle$ is given by

$$\varkappa\langle\epsilon_{gw}\rangle = \frac{1}{4a^2}\,\frac{C^2}{2\pi^2}\int_0^\infty n \sum_{s=1}^{2}\left|\overset{s}{h}{}'_n(\eta)\right|^2 \mathrm{d}n\,. \tag{84}$$

For high-frequency modes, it is only the factor $\exp(\pm in\eta)$ that needs to be differentiated by η. After averaging out the oscillating factors, one gets

$$\left|\overset{s}{h}{}'_n\right|^2 = n^2\left|\overset{s}{h}_n\right|^2,$$

so that

$$\varkappa\langle\epsilon_{gw}\rangle = \frac{1}{4a^2}\int_0^\infty n^2 h^2(n)\,\frac{\mathrm{d}n}{n}\,. \tag{85}$$

In fact, the high-frequency approximation that has been used permits integration over lower n only up to n_H. And the upper limit, as was discussed above, is in practice n_1, not infinity. The parameter Ω_{gw} is defined as $\Omega_{gw} = \langle\epsilon_{gw}\rangle/\epsilon$, where ϵ is given by Eqn (67) (the critical density). So, we derive

$$\Omega_{gw} = \int_{n_H}^{n_1} \Omega_{gw}(n)\,\frac{\mathrm{d}n}{n} = \int_{\nu_H}^{\nu_1} \Omega_{gw}(\nu)\,\frac{\mathrm{d}\nu}{\nu}$$

and

$$\Omega_{gw}(\nu) = \frac{\pi^2}{3}\, h^2(\nu)\left(\frac{\nu}{\nu_H}\right)^2. \tag{86}$$

The dimensionless quantity $\Omega_{gw}(\nu)$ is useful because it allows us to quickly evaluate the cosmological importance of the generated field in a given frequency interval. However, the primary and more universal quantity is $h(\nu)$ rather than $\Omega_{gw}(\nu)$. It is the field, not its energy density, that is directly measured by the gravity-wave detector. One should also note that some authors use the misleading definition

$$\Omega_{gw}(f) = \frac{1}{\rho_c}\,\frac{\mathrm{d}\rho_{gw}}{\mathrm{d}\ln f}\,,$$

which implies frequency differentiation of the gravity-wave energy density. This would be incorrect and could cause disagreements in numerical values of Ω_{gw}. Wherever we use $\Omega_{gw}(\nu)$, we mean relationship (86); and for order-of-magnitude estimates one can use [9, 124, 125]

$$\Omega_{gw}(\nu) \approx h^2(\nu)\left(\frac{\nu}{\nu_H}\right)^2. \tag{87}$$

6.4 Theoretical and observational constraints

The theoretical approach is entirely based on the assumption that a weak quantized gravity-wave field interacts with a classical pump field. We should follow the validity of this approximation throughout the analysis. The pump field can be treated as a classical gravitational field as long as the driving energy density ϵ is smaller than the Planck energy density, or, in other words, as long as the Hubble radius $l(\eta)$ is greater than the Planck length l_{Pl}. This is a restriction on the pump field, but it can also be used as a restriction on the wavelength λ_i of the gravity-wave mode n at the time when it enters the long-wavelength regime. If $l(\eta_*) > l_{Pl}$, then $\lambda_i > l_{Pl}$. The quantity λ_i is given by Eqn (76). So, we need to ensure that

$$b\,\frac{l_{Pl}}{l_0}\left(\frac{\nu}{\nu_H}\right)^{2+\beta} < 1\,.$$

At the lowest-frequency end $\nu = \nu_H$ this inequality gives $b(l_{Pl}/l_0) < 1$. In fact, the observational constraints (see below) are stronger:

$$b\,\frac{l_{Pl}}{l_0} \approx 10^{-6}\,, \tag{88}$$

which we accept. Then, at the highest-frequency end $\nu = \nu_1$ we need to satisfy

$$\left(\frac{\nu_1}{\nu_H}\right)^{2+\beta} < 10^6\,. \tag{89}$$

Let us now turn to the generated spectral amplitudes $h(\nu)$. According to Eqn (80) we have $h(\nu_H) \approx b8\sqrt{\pi}(l_{Pl}/l_0)$. The measured microwave background anisotropies, which we discuss below, require this number to be at a level of 10^{-5}, which gives the already mentioned Eqn (88). The quantity $h(\nu_1)$ at the highest frequency ν_1 is given by Eqn (83):

$$h(\nu_1) = b8\sqrt{\pi}\,\frac{l_{Pl}}{l_0}\left(\frac{\nu_1}{\nu_H}\right)^{1+\beta-\beta_s}\left(\frac{\nu_s}{\nu_H}\right)^{\beta_s}\frac{\nu_H}{\nu_2}\,.$$

Using Eqn (66), this expression for $h(\nu_1)$ can be rewritten as

$$h(\nu_1) = 8\sqrt{\pi}\,\frac{l_{Pl}}{l_H}\,\frac{\nu_1}{\nu_H} = 8\sqrt{\pi}\,\frac{l_{Pl}}{\lambda_1}\,, \tag{90}$$

where $\lambda_1 = c/\nu_1$. This last expression for $h(\nu_1)$ is not surprising: the modes with $\nu \geqslant \nu_1$ are still in a vacuum state, so the numerical value of $h(\nu_1)$ is determined by quantum normalization.

All the amplified modes started with small initial amplitudes h_i at the level of zero-point quantum fluctuations. These amplitudes are also small today, since h_i could only remain constant or decrease. However, even these relatively small amplitudes should obey observational constraints. We do not want Ω_{gw} in the high-frequency modes (which could affect the rate of primordial nucleosynthesis) to exceed the level of 10^{-5}. This means that $\Omega_{gw}(\nu_1)$ cannot exceed the level of 10^{-6} or so. The use of Eqn (86) in combination with $\Omega_{gw}(\nu_1) \approx 10^{-6}$ and $h(\nu_1)$ from Eqn (90), gives us the highest allowed frequency $\nu_1 \approx 3 \times 10^{10}$ Hz. We will use this value of ν_1 in our numerical estimates. Returning with this value of ν_1 to Eqn (89) we find that the parameter β can only be $\beta \leqslant -1.8$. We will treat $\beta = -1.8$ as an upper limit for the allowed values of β.

We can now check whether the accepted parameters leave room for the postulated z-stage with $\beta_s < 0$. Using Eqn (66) we can rewrite Eqn (88) in the form

$$10^{-6}\,\frac{l_H}{l_{Pl}} = \left(\frac{\nu_1}{\nu_H}\right)^{-\beta}\left(\frac{\nu_1}{\nu_s}\right)^{\beta_s}\frac{\nu_2}{\nu_H}\,. \tag{91}$$

We know that $\nu_2/\nu_H = 10^2$ and ν_1/ν_s is not smaller than 1. Substituting all the numbers into Eqn (91) one can find that this equation cannot be satisfied for the largest possible $\beta = -1.8$. In the case $\beta = -1.9$, Eqn (91) is only marginally satisfied, in the sense that a significant deviation from $\beta_s = 0$ toward negative β_s can only last for a relatively short time. For instance, one can bring $\beta_s = -0.4$ into conformity with $\nu_s = 10^8$ Hz. On the other hand, if one takes $\beta = -2$, a somewhat longer interval of the z-stage with $\beta_s < 0$ can be included. For instance, Eqn (91) is satisfied if one assumes $\nu_s = 10^{-4}$ Hz and $\beta_s = -0.3$. This allows us to slightly increase $h(\nu)$ in the interval $\nu_s < \nu < \nu_1$, as compared with the values of $h(\nu)$ reached in the more traditional case $\beta = -2$, $\beta_s = 0$. In what follows, we will consider the consequences of this assumption for the prospects of detection of the gravitational wave signal produced. Finally, we will see what can be inferred from the available information on the microwave background anisotropies [140, 141] about the parameters β and l_0.

Usually, cosmologists operate with the spectral index n (not to be confused with the wave number n) of primordial cosmological perturbations. Taking into account the way in which the spectral index n is defined, one can relate n to the spectral index $\beta + 2$ that shows up in Eqn (80). The relationship between them is $n = 2\beta + 5$. This relationship is valid independently of the nature of cosmological perturbations. In particular, it is valid for density perturbations, in which case $h(n)$ of Eqn (80) is the dimensionless spectral amplitude of metric perturbations associated with density perturbations. If primordial GWs and density perturbations are generated by the mechanism that we discuss here (an assumption that is likely to be true) then the parameter β that appears in the spectral index is the same as the one that appears in the scale factor of Eqn (62). Primordial GWs and primordial density perturbations with the same spectral index $\beta + 2$ produce approximately the same lower-order multipole distributions of large-scale anisotropies.

Evaluation of the spectral index n of primordial perturbations results in $n = 1.2 \pm 0.3$ [141] or even in a somewhat higher value. A recent analysis [142] of all available data favors $n = 1.2$ and a quadrupole contribution of GWs twice as large as that of density perturbations. One can interpret these evaluations as an indication that the true value of n lies somewhere near $n = 1.2$ (hopefully, the planned new observational missions will determine this index more accurately). This gives us the parameter β somewhere near $\beta = -1.9$. We will use $\beta = -1.9$ in our estimates below, as the observationally preferred value. The parameter β may be somewhat larger than $\beta = -1.9$. However, as we already discussed, the value $\beta = -1.8$ ($n = 1.4$) is the largest one for which the entire approach is well posed. The Harrison – Zeldovich spectral index $n = 1$ corresponds to $\beta = -2$.

The observed quadrupole anisotropy of the microwave background radiation is at the level $\delta T/T \sim 10^{-5}$. The quadrupole anisotropy that would be produced by the spectrum (80) – (83) is mainly accounted for by the wave numbers near n_H. Thus, the numerical value of the quadrupole anisotropy produced by relic GWs is approximately equal to A [see Eqn (78)]. According to general physical considerations and detailed calculations [143 – 145], the metric amplitudes of long-wavelength GWs and density perturbations generated by the discussed amplification mechanism are of the same order of magnitude. Therefore, they contribute roughly equally to the anisotropy at lower multipoles. This gives us the estimate $A \sim 10^{-5}$, which we already used in Eqn (88). It has not yet been proven observationally that a significant part of the observed anisotropies at lower multipoles is indeed provided by relic GWs, but we can at least assume this with some degree of confidence. It is likely that the future measurements of the microwave background radiation, and especially its polarization, will help us to verify this theoretical conclusion.

Upon combining all the evaluated parameters together, we show the expected spectrum of $h(\nu)$ for the case $\beta = -1.9$ in Fig. 13. A small allowed interval of the z-stage is also included. The intervals of the spectrum accessible to space- and ground-based interferometers are indicated by vertical lines.

It is necessary to note [143 – 146] that a confirmation of any $n > 1$ ($\beta > -2$) will mean that the very early Universe was not driven by a scalar field — the cornerstone of inflationary considerations. Indeed, $n > 1$ ($\beta > -2$) requires the effective equation of state at the initial stage of expansion to be $\epsilon + p < 0$ [see Eqn (68)]. But this requirement cannot be accommodated by any scalar field $\varphi(t)$ whatever the scalar field potential $V(\varphi)$ may be. The energy density of the scalar field is $\epsilon = \dot{\varphi}^2/2 + V(\varphi)$, whereas its pressure is

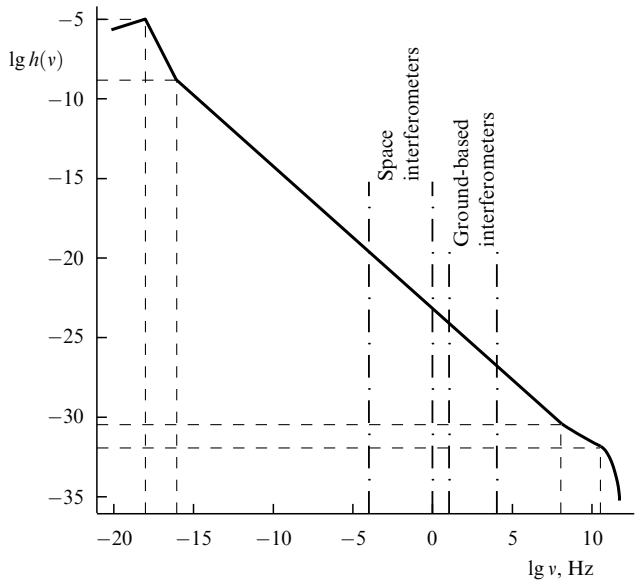

Figure 13. Expected spectrum $h(\nu)$ for the case $\beta = -1.9$.

$p = \dot{\varphi}^2/2 - V(\varphi)$, so that $\epsilon + p$ cannot be negative. The available data do not yet prove that $n > 1$, but this possibility seems likely.

It is also necessary to comment on the damage to gravitational wave research that was inflicted by the so-called 'standard inflationary result'. The 'standard inflationary result' predicts infinitely large amplitudes of density perturbations $\delta\rho/\rho$ in the interval of spectrum with the Harrison–Zeldovich slope $n = 1$ $(\beta = -2)$: $\delta\rho/\rho \propto V^{3/2}/V' \propto 1/\sqrt{1-n}$. The metric (curvature) perturbations h_S accompanying $\delta\rho/\rho$ are also predicted to be infinitely large, in the same proportion to $1/\sqrt{1-n}$. The inflationary literature conceals this predicted infinity of density perturbations by writing the ratio of the gravitational wave amplitude h_T to the scalar metric amplitude h_S: $h_T/h_S \sim 7\sqrt{1-n}$, and declaring that the contribution of GWs to the cosmic microwave background (CMB) anisotropies should be zero, or almost zero, for $n \approx 1$. Thus, the 'standard' inflationary theory shifts the spectrum of relic GWs, similar in shape to that shown in Fig. 13, down by many orders of magnitude. This claim has incorrectly led to discarding the GW contribution in the analysis of CMB data. Although one of the recent best fits to the available data indicates the presence of a GW contribution [147], this fit has been ignored. For many years, inflationary theorists claimed that their arbitrarily large density perturbations were caused by the 'big amplification during reheating'. It is now universally accepted that this explanation is false. The scalar metric perturbation, similarly to the GW perturbation [see Eqn (59)], remains constant during the long wavelength regime, that is, its numerical value does not change on the way from the first 'Hubble-radius-crossing' to the second 'Hubble-radius-crossing'. This fact is also reflected by the constancy of the so-called 'conserved' gauge-invariant quantity ζ, which is at the center of inflationary analysis (see, for example, Ref. [148]). Since the 'standard' inflationary theory predicts an arbitrarily large numerical value of ζ, and this quantity is 'conserved' during the evolution, this arbitrarily large number must be postulated from the very beginning, as quantum normalization. This happens if one incorrectly assigns the quantum normalization to the scalar field perturbations alone, as if they were a free scalar test field in the De-Sitter space-time. Then, the quantity ζ calculated from the perturbed Einstein equations is arbitrarily large at the first 'Hubble-radius-crossing', and this figure is transmitted to the second 'Hubble-radius-crossing'. In reality, however, scalar field perturbations are always coupled to metric (gravitational field) perturbations. A correct quantization of the combined degree of freedom renders the scalar metric (curvature) perturbation h_S finite and small, and of the same order of magnitude as the GW perturbation h_T. The inflationary formula for $\delta\rho/\rho$ is incorrect in that it misses the dimensionless factor $(-\dot{H}/H^2)^{1/2}$, which cancels out the zero in the denominator and makes the generated density perturbations finite and small, even in the interval of spectrum with the Harrison–Zeldovich slope. The 'standard inflationary result' is in severe conflict not only with theory but also with observations: when observers marginalize their data to $n = 1$ (enforce this value of n in data analysis) they find finite and small density perturbations, instead of the infinitely large perturbations predicted by inflationary theorists. (For analytical expressions of the 'standard inflationary result' see inflationary articles, including recent reviews. For a graphical illustration of the divergent density perturbations and quadrupole anisotropies, predicted by inflationary theorists, see, for example, Ref. [149]. For a critical analysis and disagreement with the 'standard inflationary result' see Refs [143–145].) The most recent articles dealing with perturbations in quasi-De-Sitter models rightly emphasize the expected substantial contribution of GWs to the large scale CMB anisotropies [150]. In short, general relativity and quantum field theory do not yield the 'standard inflationary result'.

6.5 Detectability of relic gravitational waves

We switch now from cosmology to the prospects of detecting the predicted relic GWs. The ground-based [14, 15, 151] and space-based [16, 152] laser interferometers (see also Refs [153–155]) will be the focus of our attention. We use laboratory frequencies ν and intervals of laboratory time t $(c\,dt = a(\eta_R)\,d\eta)$. Formulas (82) and (83), with $A = 10^{-5}$, $\nu_2/\nu_H = 10^2$, and the oscillating factor restored, can be written as

$$h(\nu, t) \sim 10^{-7} \cos\left[2\pi\nu(t - t_\nu)\right] \left(\frac{\nu}{\nu_H}\right)^{\beta+1}, \qquad \nu_2 \leqslant \nu \leqslant \nu_s \tag{92}$$

and

$$h(\nu, t) \sim 10^{-7} \cos\left[2\pi\nu(t - t_\nu)\right] \left(\frac{\nu}{\nu_H}\right)^{1+\beta-\beta_s} \left(\frac{\nu_s}{\nu_H}\right)^{\beta_s}, \quad \nu_s \leqslant \nu \leqslant \nu_1\,, \tag{93}$$

where the deterministic (not random) constant t_ν does not vary significantly from one frequency to another over the intervals $\Delta\nu \approx \nu$. The explicit time dependence of the spectral variance $h^2(\nu, t)$ of the field, or, in other words, the explicit time dependence of the (zero-lag) temporal correlation function of the field at any given frequency, demonstrates that we are dealing with a non-stationary process (a consequence of squeezing and severe reduction of the phase uncertainty). We will first ignore the oscillating factor and compare the predicted amplitudes with the sensitivity curves of advanced detectors. The potential reserve of improving the

SNR by exploiting squeezing will be discussed at the end of Section 6.5.

Let us start from the Laser Interferometer Space Antenna (LISA) [16]. The instrument will be most sensitive in the interval, roughly, from 10^{-3} Hz to 10^{-1} Hz, and will be reasonably sensitive over a broader range, up to frequencies 10^{-4} Hz and 1 Hz. The graph of LISA sensitivity to a stochastic background is usually plotted under the assumption of a 1-year observation time, that is, the root-mean-square (r.m.s.) instrumental noise is evaluated in frequency bins $\Delta\nu = 3 \times 10^{-8}$ Hz around each frequency ν. We need to rescale our predicted amplitude $h(\nu)$ to these bins.

The mean square amplitude of the gravitational wave field is given by the integral (73). Thus, the r.m.s. amplitude in the band $\Delta\nu$ centered at a given frequency ν is represented by the expression

$$h(\nu, \Delta\nu) = h(\nu)\sqrt{\frac{\Delta\nu}{\nu}}\,. \tag{94}$$

We use Eqns (92), (93) and calculate expression (94) assuming $\Delta\nu = 3 \times 10^{-8}$ Hz. The results are plotted in Fig. 6. Formula (92) was used throughout the covered frequency interval for the realistic case $\beta = -1.9$ and for the extreme case $\beta = -1.8$. The line marked 'z-model' describes the signal produced in the composite model with $\beta = -2$ up to $\nu_s = 10^{-4}$ Hz [formula (92)] and then followed by formula (93) with $\beta_s = -0.3$. This model gives a signal a factor of 3 higher, at $\nu = 10^{-3}$ Hz, than the model $\beta = -2$ extrapolated down to this frequency.

There is no doubt that the signal $\beta = -1.8$ would be easily detectable even with a single instrument. The signal $\beta = -1.9$ is marginally detectable, with an SNR of 3 or so, in a quite narrow frequency interval near and above the frequency 3×10^{-3} Hz. However, at lower frequencies one would need to be concerned with the possible GW noise from unresolved binary stars in our Galaxy (see Section 5). Further improvement of the expected LISA sensitivity by a factor of 3 may prove to be crucial for a confident detection of the predicted signal with $\beta = -1.9$.

Let us now turn to the ground-based interferometers operating in the interval from 10 Hz to 10^4 Hz. The best sensitivity is reached in the band around $\nu = 10^2$ Hz. We take this frequency as the representative frequency for comparison with the predicted signal. We will work directly in terms of the dimensionless quantity $h(\nu)$. If necessary, the r.m.s. amplitude per $\mathrm{Hz}^{1/2}$ at a given ν can be found simply as $h(\nu)/\sqrt{\nu}$. The instrumental noise will also be quoted in terms of the dimensionless quantity $h_{\mathrm{ex}}(\nu)$.

The expected sensitivity of the first-generation instruments at $\nu = 10^2$ Hz is $h_{\mathrm{ex}} = 10^{-21}$ or better. The theoretical prediction at this frequency, following from (92) and (93) with $\beta_s = 0$, is $h_{\mathrm{th}} = 10^{-23}$ for $\beta = -1.8$, and $h_{\mathrm{th}} = 10^{-25}$ for $\beta = -1.9$. Therefore, the gap between the signal and noise levels is from 2 to 4 orders of magnitude. The expected sensitivity of the advanced interferometers, such as LIGO-II [17], can be as high as $h_{\mathrm{ex}} = 10^{-23}$. In this case, the gap vanishes for the $\beta = -1.8$ signal and reduces to 2 orders of magnitude for the $\beta = -1.9$ signal. Figure 14 illustrates the expected signal in comparison with the LIGO-II sensitivity. Since the signal lines are plotted in terms of $h(\nu)$, the LISA sensitivity curve (shown for periodic sources) should be raised and adjusted in accordance with Fig. 6.

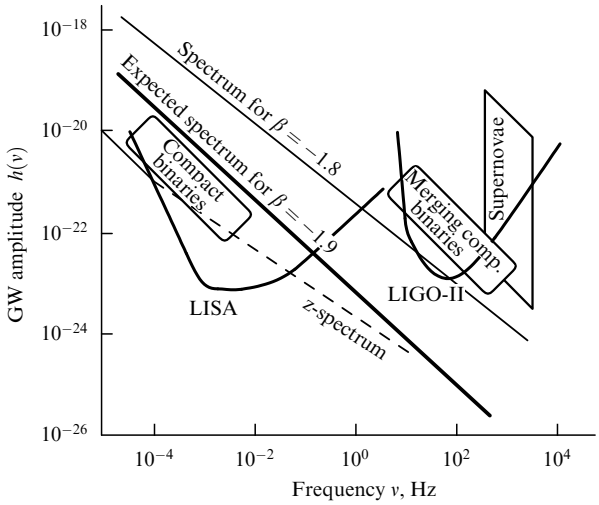

Figure 14. Full spectrum $h(\nu)$ and the intervals accessible to ground-based and cosmic interferometers.

A signal below noise can be detected if the outputs of two or more detectors can be cross correlated. (For the early estimates of detectability of relic GWs see Ref. [156].) This cross correlation will be possible for ground-based interferometers, several of which are currently under construction. The gap between the signal and the noise level should be covered by a sufficiently long observation time τ. The duration τ depends on whether the signal has any temporal characterization known in advance. We start from the assumption that no temporal parameters are known in advance. In other words, we first ignore the squeezed nature of the relic background and work under the assumption that the squeezing cannot be exploited to our advantage.

The response of an instrument to the incoming radiation is $s(t) = F_{ij}h^{ij}$ where F_{ij} depends on the position and orientation of the instrument. Since h^{ij} is a quantum-mechanical operator [see Eqn (50)], we need to calculate the mean value of a quadratic quantity. The mean value of the cross correlation of responses from two instruments $\langle 0|s_1(t)s_2(t)|0\rangle$ will involve an overlap reduction function [157 – 160], which we assume to be not much smaller than 1 [159]. The SNR in the measurement of the amplitude of a signal with no specific features known increases as $(\tau\nu)^{1/4}$, where ν is some characteristic central frequency (for more detail see Section 8).

We apply the guaranteed law $(\tau\nu)^{1/4}$ to early and advanced instruments at the representative frequency $\nu = 10^2$ Hz. This law requires a reasonably short time $\tau = 10^6$ s in order to improve the S/N in the first instruments by two orders of magnitude and to reach the level of the signal with the extreme spectral index $\beta = -1.8$. A longer integration time or a better sensitivity will make the S/N larger than 1. In the case of a realistic spectral index $\beta = -1.9$ the remaining gap of 4 orders of magnitude can be covered by the combination of a significantly better sensitivity and a longer observation time (not necessarily in one non-interrupted run). The sensitivity of the advanced laser interferometers, such as LIGO II, at the level $h_{\mathrm{ex}} = 10^{-23}$ and the same observation time $\tau = 10^6$ s would be sufficient for reaching the level of the predicted signal with $\beta = -1.9$.

An additional increase of S/N can be achieved if the statistical properties of the signal can be properly exploited. Squeezing is automatically present at all frequencies from ν_{H}

to ν_1. The squeezing parameter r is larger in GWs of cosmological scales, and possibly the periodic structure in Eqn (75) can be better revealed at those scales. However, we are interested here in frequencies accessible to ground-based interferometers, say, in the interval 30–100 Hz. If our intention were to monitor one given frequency ν from the beginning of its oscillating regime until now, then, in order to avoid the destructive interference from neighboring modes during all that time, the frequency resolution of the instrument should be made incredibly high, of the order of 10^{-18} Hz. Certainly, this is not something we can, or intend to do. Although the amplitudes of the waves have adiabatically decreased and their frequencies have redshifted since the beginning of their oscillating regime, the general statistical properties of the discussed signal are essentially the same now as they were 10 years after the Big Bang or will be within 1 million years from now.

The periodic structure (92) may survive at some level in the instrumental window of sensitivity from $\nu_{\min}$ (minimal frequency) to $\nu_{\max}$ (maximal frequency). The mean square value of the field in this window is

$$\int_{\nu_{\min}}^{\nu_{\max}} h^2(\nu, t)\,\frac{\mathrm{d}\nu}{\nu} = 10^{-14}\,\frac{1}{\nu_{\mathrm{H}}^{2\beta+2}}\int_{\nu_{\min}}^{\nu_{\max}} \cos^2\left[2\pi\nu(t-t_\nu)\right]\nu^{2\beta+1}\,\mathrm{d}\nu\,. \quad (95)$$

Because of the strong dependence of the integrand on frequency, $\nu^{-2.6}$ or $\nu^{-2.8}$, the value of the integral (95) is determined by its lower limit. Apparently, the search through the data should be based on a periodic structure that could survive at $\nu = \nu_{\min}$. As an illustration, we can consider a narrow interval $\Delta\nu = \nu_{\max} - \nu_{\min}$ such that the integral (95) can be approximated by the formula

$$\int_{\nu_{\min}}^{\nu_{\max}} h^2(\nu, t)\,\frac{\mathrm{d}\nu}{\nu} \sim 10^{-14}\left(\frac{\nu_{\min}}{\nu_{\mathrm{H}}}\right)^{2\beta+2}\frac{\Delta\nu}{\nu_{\min}}\cos^2\left[2\pi\nu_{\min}(t-t_{\min})\right].$$

Clearly, the correlation function is strictly periodic and its structure is known in advance, in contrast to other possible signals. This is a typical example of using *a priori* information. Ideally, the gain in S/N can grow as $(\tau\nu_{\min})^{1/2}$. This would significantly reduce the required observation time τ. For a larger $\Delta\nu$, even an intermediate gain between the guaranteed law $(\tau\nu)^{1/4}$ and the law $(\tau\nu)^{1/2}$, from the matched filtering technique, would help. This could potentially make the signal with $\beta = -1.9$ measurable even by the early laser interferometers. A straightforward application of (95) exploiting squeezing may not be possible, as argued in a recent study [161], but more sophisticated methods are not excluded.

For frequency intervals covered by bar detectors and electromagnetic detectors, the expected results follow from the same formulas (92), (93) and have been briefly discussed elsewhere [135, 156].

6.6 Short summary

It will be strange for the predicted signal at the level corresponding to $\beta = -1.9$ not to be seen by the instruments capable of its detection. There are not so many cosmological assumptions involved in the derivation that could prove wrong, thus invalidating our predictions. On the other hand, it would be even more strange (and even more interesting) if the relic GWs should be detected at a level above the $\beta = -1.8$ line. This would mean that there is something fundamentally wrong in our basic cosmological premises. In the most favorable case, the detection of relic GWs could be achieved by the cross correlation of the outputs of the first-generation laser interferometers in LIGO, VIRGO, GEO600. In a more realistic case, the sensitivity of advanced ground-based and space-based laser interferometers will be needed. The specific statistical properties of relic GWs, associated with the phenomenon of squeezing, is a potential reserve for further improvement of the SNR.

7. Gravitational wave detectors and their sensitivity

7.1 Current status of gravitational wave antennas

Currently, there are a number of bar detectors in operation: some of these operate at room temperature and some others at cryogenic temperatures. Bar detectors are resonant, narrow-band detectors. They can detect signals of amplitudes $h \sim 10^{-20}$ in a band width of 10–20 Hz around a central frequency of ~ 1 kHz. Asymmetric supernovae in our Galaxy are the best candidate sources for these detectors. They can also see the continuous radiation emitted by a neutron star if the frequency falls in their sensitivity band.

Interferometric detectors currently under construction will increase our ability to directly observe GWs. A 300 m detector has already been built in Tokyo, Japan [162]. Several other projects are now nearing completion: a British-German collaboration is constructing a 600 m interferometer (GEO) in Hannover, Germany [13, 163], a French-Italian collaboration is building a 3 km detector (VIRGO) near Pisa, Italy [14, 164] and Americans are building two 4 km antennas (LIGO), one in Livingston and the other in Hanford [15], U.S.A. These detectors will start collecting data between 2001 and 2003. The largest of these detectors, LIGO and VIRGO, are likely to be upgraded in sensitivity by an order of magnitude with a better low-frequency performance in 2005. These ground-based interferometers will eventually be sensitive to sources in the frequency range from 10 Hz to several kHz. In addition to these ground-based antennas, it is planned to place an interferometer in space by the end of this decade [16]. The Laser Interferometer Space Antenna (LISA) consists of three drag-free satellites, forming an equilateral triangle of side 5 million km, in a heliocentric orbit, lagging behind the Earth by 20°. LISA will be sensitive to waves in the low-frequency band of 10^{-4}–10^{-1} Hz.

7.2 Sensitivity of a gravitational wave antenna

The performance of a GW detector is characterized by the *power spectral density* (henceforth denoted PSD) of its noise background. One can construct the noise PSD as follows: A GW detector outputs a dimensionless data train, say $x(t)$, which in the case of an interferometer is the relative strain in the two arms. In the absence of any GW signal the detector output is just an instance of noise $n(t)$, that is, $x(t) = n(t)$. The noise auto-correlation function $\varkappa$ is defined as [5]

$$\varkappa(t_1, t_2) \equiv \overline{n(t_1)n(t_2)}\,, \quad (96)$$

[5] Note that in earlier sections we used angular brackets to denote the ensemble average. In this section, however, we shall use an overbar to denote ensemble average while angular brackets will be reserved for denoting the scalar product of functions.

where the overbar indicates averaging over an ensemble of noise realizations. In general, $\varkappa$ depends on both t_1 and t_2 However, if the detector output is a stationary noise process, i.e. its performance is, statistically speaking, independent of time, then $\varkappa$ depends only on $\tau \equiv t_2 - t_1$. We shall, furthermore, assume that $\varkappa(\tau) = \varkappa(-\tau)$. For data from real detectors the above average can be replaced by a time average under the assumption of ergodicity:

$$\varkappa(\tau) = \frac{1}{T}\int_{-T/2}^{T/2} n(t)n(t-\tau)\,\mathrm{d}t\,. \tag{97}$$

It should be noted that in real detectors there are noises other than thermal. In particular, there may be extra (overthermal) noises (related to the properties of suspensions, mirrors, etc.; see, e.g., Refs [165, 166]), to which these considerations cannot be applied.

The assumption of stationarity is not strictly valid in the case of real GW detectors; however, if their performance does not vary greatly over time scales much larger than the typical observation time scales, stationarity could be used as a working rule. While this may be good enough in the case of binary inspiral and coalescence searches, it is a matter of concern for the observation of continuous and stochastic GWs. In this review, for simplicity, we shall assume that the detector noise is stationary. In this case the *one-sided* noise PSD, defined only at positive frequencies, is the Fourier transform of the noise auto-correlation function:

$$S_\mathrm{n}(f) \equiv \begin{cases} \dfrac{1}{2}\displaystyle\int_{-\infty}^{\infty} \varkappa(\tau)\exp(2\pi\mathrm{i}f\tau)\,\mathrm{d}\tau\,, & f \geqslant 0\,, \\ 0\,, & f < 0\,, \end{cases} \tag{98}$$

where the factor of $1/2$ is included by convention. Since we have assumed that $\varkappa(\tau)$ is an even function, the above equation immediately implies that $S_\mathrm{n}(f)$ is real. It is quite straightforward to show that

$$\overline{\tilde{n}(f)\tilde{n}^*(f')} = S_\mathrm{n}(f)\delta(f-f')\,, \tag{99}$$

where $\tilde{n}(f)$ is the Fourier transform of $n(t)$ and $\tilde{n}^*(f)$ denotes the complex conjugate of $\tilde{n}(f)$. The above identity (99) implies that $S_\mathrm{n}(f)$ is positively definite. The above identity can be obtained by expressing the Fourier transforms on the left-hand side by their respective time-domain functions, i.e., $\tilde{n}(f) \equiv \int_{-\infty}^{\infty} n(t)\exp(2\pi\mathrm{i}ft)\,\mathrm{d}t$ and using Eqns (96) and (99). The autocorrelation function $\varkappa(\tau)$ at $\tau = 0$ can be expressed as an integral of $S_\mathrm{n}(f)$. Indeed, it is easy to see that

$$\overline{n^2(t)} = 2\int_0^\infty S_\mathrm{n}(f)\,\mathrm{d}f. \tag{100}$$

The above equation justifies the name *power spectral density* given to $S_\mathrm{n}(f)$.

It is obvious that $S_\mathrm{n}(f)$ has dimensions of time but it is conventional to use the dimensions of Hz^{-1} since $S_\mathrm{n}(f)$ is a quantity defined in a frequency domain. The square root of $S_\mathrm{n}(f)$ is the noise amplitude, $\sqrt{S_\mathrm{n}(f)}$, and has dimensions of $\mathrm{Hz}^{-1/2}$. It is often useful to define the dimensionless quantity $h_\mathrm{n}^2(f) \equiv fS_\mathrm{n}(f)$ called the *effective noise*. In GW interferometer literature one also comes across 'displacement noise' or 'strain noise' defined as $h_\ell(f) \equiv \ell h_\mathrm{n}(f)$ and the corresponding noise spectrum $S_\ell(f) \equiv \ell^2 S_\mathrm{n}(f)$, where ℓ is the arm length of the interferometer. The displacement noise gives the smallest strain $\delta\ell/\ell$ in the arms of an interferometer that can be measured at a given frequency.

7.3 Source amplitudes vs sensitivity

One compares the GW amplitudes of astronomical sources with the instrumental sensitivity, and assesses what sort of sources will be observable, in the following way. Firstly, as comparisons are almost always made in a frequency domain, it is important to note that the Fourier component $\tilde{h}(f) \equiv \int_{-\infty}^{\infty} \mathrm{d}t\,h(t)\exp(2\pi\mathrm{i}ft)$ of a deterministic signal $h(t)$ has dimensions of Hz^{-1} and the quantity $f|\tilde{h}(f)|$ is dimensionless. It is this last quantity that should be compared with $h_n(f)$ to deduce the strength of a source relative to the detector noise. Secondly, it is also quite common to compare the amplitude spectrum per logarithmic bin for a source, $\sqrt{f}\,|\tilde{h}(f)|$, with the amplitude spectrum of noise $\sqrt{S_\mathrm{n}(f)}$, both of which have dimensions of $\mathrm{Hz}^{-1/2}$. A justification for these comparisons is given in Section 8. Finally, for monochromatic sources, one compares the effective noise in a long integration period with the expected 'instantaneous' amplitudes in the following way: a monotonic wave of frequency f_0 observed for a time T is simply a narrow line in a frequency bin of width $\Delta f \equiv 1/T$ around f_0. The noise in this bin is $S_\mathrm{n}(f)\Delta f = S_\mathrm{n}(f)/T$. Thus, the SNR ρ after the period of observation T is

$$\rho = \frac{h_0}{\sqrt{S_\mathrm{n}(f_0)/T}}\,. \tag{101}$$

One therefore computes this dimensionless noise spectrum for a given duration of observation, $S_\mathrm{n}(f)/T$, to assess the detectability of a continuous GW.

Surely, if the observation time is T then the total *energy* (that is, the integrated power spectrum) of both the signal and noise must increase in proportion to T^2. Then how does the SNR for a continuous wave improve with the duration of observation? The point is that, while the signal energy is all concentrated in one bin, the noise is distributed over the entire frequency band. As T increases, the frequency resolution improves as $1/T$ and the number of frequency bins increases in proportion to T. Consequently, the noise intensity *per frequency bin* decreases only as $1/T$. Now, the signal intensity is concentrated in just one bin since the signal is assumed to be monochromatic. Therefore, the power SNR increases as T and the amplitude SNR increases as $\sqrt{T}$.

7.4 Noise power spectral density in the first interferometers

As mentioned in Section 7.2, the performance of a GW detector is characterized by the one-sided noise power spectral density (PSD). The noise PSD plays an important role in signal analysis. We shall only discuss the PSDs of interferometric gravitational wave detectors, since the two prime candidate sources discussed in this review are both detectable conclusively only with a broadband detector. Interferometers have a very broad band sensitivity and are, therefore, ideal for the detection of these sources. Since our aim is to foresee the first possible detections, we shall mainly concentrate on the first-generation interferometers, GEO, LIGO-I, TAMA and VIRGO, mentioning future ground- and space-based detectors where appropriate. The sensitivity of ground-based detectors is limited to frequencies less than 1 Hz by the time-varying local gravitational field produced by a variety of different noise sources, e.g. low frequency seismic

Table 5. Noise power spectral densities of initial interferometers $S_n(f)$. For each detector the noise PSD is given in terms of a dimensionless frequency $x = f/f_0$ and raises steeply above a lower cutoff f_s (also see Figs 15 and 16).

Detector	f_s, Hz	f_0, Hz	$10^{46} \times S_n(x)$, Hz^{-1}
GEO	40	150	$(3.4x)^{-30} + 34x^{-1}$ $+20(1 - x^2 + x^4/2)/(1 + x^2/2)$
LIGO-I	40	150	$9.00\left[(4.49x)^{-56} + 0.16x^{-4.52}\right.$ $\left.+0.52 + 0.32x^2\right]$
TAMA	75	400	$75\left[x^{-5} + 13x^{-1} + 9(1 + x^2)\right]$
VIRGO	20	500	$3.24\left[(6.23x)^{-5} + 2x^{-1} + 1 + x^2\right]$
LISA	10^{-5}	10^{-3}	$420\left[(x/5.62)^{-14/3} + 10^3 + x^2\right]$

vibrations, density variation in the atmosphere due to winds, etc. Thus, for data analysis purposes, the noise PSD is assumed to be essentially infinite below a certain lower cutoff f_s. Table 5 lists the noise PSD $S_n(f)$ above this cutoff, i.e. for $f \geqslant f_s$, for various interferometric detectors. The square root of the noise PSD, $\sqrt{S_n(f)}$, expected for these detectors is summarized in Fig. 15. The GEO noise curve is somewhat different from the others as it uses a signal enhancement technique called 'signal recycling' [167]. It is also for this reason that GEO is close to LIGO and VIRGO in sensitivity though it is only a sixth/fifth in size compared to the larger interferometers.

For LISA Table 5 gives the internal instrumental noise only. However, in the frequency range $10^{-4} - 3 \times 10^{-3}$ Hz, LISA will be limited in its sensitivity by the background produced by several populations of galactic binary systems, such as closed white-dwarf binaries, cataclysmic variables, etc. This binary *confusion noise* has been well modeled (see Ref. [16] and Section 5) and gives the following 'external' LISA noise $S_{ex}(x)$:

$$10^{46} S_{ex}(x) = \begin{cases} \left(\dfrac{x}{5.56 \times 10^4}\right)^{-1.9}, & x < 10^{-0.15}, \\ \left(\dfrac{x}{12.3}\right)^{-7.5}, & 10^{-0.15} < x < 10^{0.25}, \\ \left(\dfrac{x}{471}\right)^{-2.6}, & 10^{0.25} < x, \end{cases} \tag{102}$$

where $x = f/f_0$ and $f_0 = 10^{-3}$ Hz.

7.5 False alarms and detection threshold

Gravitational wave event rates in the first interferometers are expected to be rather low, a few per year. Therefore, one has to set a high threshold so that noise-generated false alarms mimicking an event be negligible.

For a detector output sampled at 1 kHz and processed through a large number (say, 10^3) of filters one has $\sim 3 \times 10^{13}$ instants of noise a year. If the noise is Gaussian then demanding that no more than one false alarm occur a year of observation sets the threshold to be about 7.5 times the standard deviation of noise. Therefore, a source is detectable only if its amplitude is significantly larger than the effective noise amplitude, i.e. $f\tilde{h}(f) \gg h_n(f)$. The reason for selecting only such high-sigma events is because the event rate of a transient source, i.e. a source lasting for a few seconds to minutes, such as a binary inspiral, may be as low as a few per year, while noise-generated false alarms at low SNRs

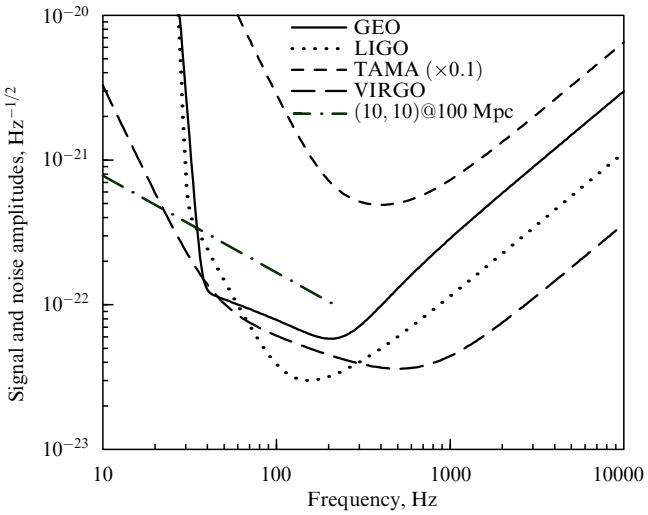

Figure 15. Amplitude noise spectral density $\sqrt{S_n(f)}$ in the first-generation interferometers. On the same graph we also plot the signal amplitude, $\sqrt{f}|h(f)|$, of a binary black hole inspiral occurring at a distance of 100 Mpc. Each black hole is taken to be of mass $10M_\odot$ (see text in the Section 8.2 for a discussion). The TAMA dependency is multiplied by 0.1.

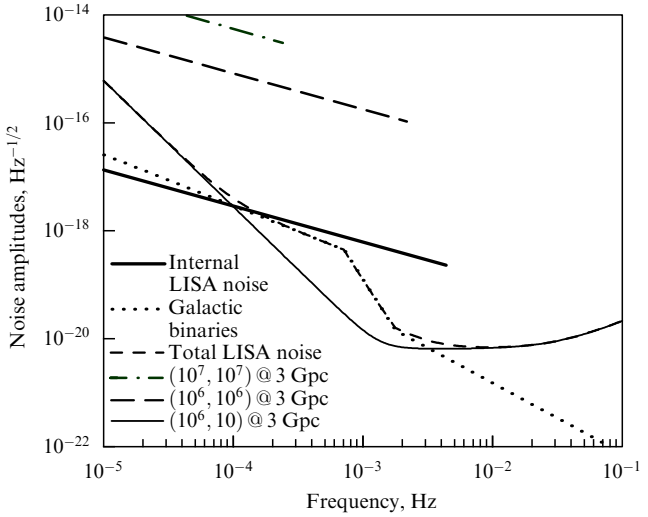

Figure 16. Same as Fig. 15 but for the LISA detector. Note that this figure, in contrast to Fig. 6, uses amplitudes per Hz$^{-1/2}$. We also plot signals from supermassive black holes. The supermassive BH sources are assumed to lie at a redshift of $z = 1$ but LISA can detect these sources with a good SNR almost anywhere in the Universe.

$S/N \sim 3-4$ tend to be quite frequent over a period of a year. Setting higher thresholds for detection helps in removing spurious, noise-generated events. However, signal enhancement techniques (cf. Section 8) make it possible to detect a signal of a relatively low amplitude provided there are many wave cycles and the shape of the wave is known accurately.

7.6 Beam pattern functions

Gravitational wave detectors are sensitive to waves coming from almost any direction in the sky although the degree of sensitivity depends on the actual direction. The sensitivity of a detector to the direction of the wave is described by what is called the *beam pattern function*, also referred to as the

antenna pattern [4]. A GW antenna responds best if the waves are incident at right angles to the principal direction, as in the case of a cylindrical bar antenna, or to the principal plane, as in the case of an interferometric detector. For waves incident in any other direction the response will alter by trigonometric factors that will be different for the two polarizations, h_+ and $h_\times$.

Denoting the two beam pattern functions by F_+ and $F_\times$, the response C of an antenna to a GW of polarization amplitudes h_+ and $h_\times$ is given by

$$C = \left[F_+^2(\theta,\varphi,\psi)h_+^2 + F_\times^2(\theta,\varphi,\psi)h_\times^2\right]^{1/2}, \tag{103}$$

where θ and φ denote the direction to the source and ψ is the polarization angle. For a resonant bar detector with a longitudinal axis aligned with the z-axis, the response is

$$F_+ = \sin^2\theta\cos 2\psi\,, \qquad F_\times = \sin^2\theta\sin 2\psi\,, \tag{104}$$

and for an interferometer with its arms in the (x,y) plane and at right angles to each other and the x-axis bisecting the two arms [4, 168]

$$\begin{aligned} F_+(\theta,\varphi,\psi) &= \frac{1}{2}(1+\cos^2\theta)\cos 2\varphi\cos 2\psi \\ &\quad - \cos\theta\sin 2\varphi\sin 2\psi\,, \\ F_\times(\theta,\varphi,\psi) &= \frac{1}{2}(1+\cos^2\theta)\cos 2\varphi\sin 2\psi \\ &\quad + \cos\theta\sin 2\varphi\cos 2\psi\,. \end{aligned} \tag{105}$$

If a GW lasts long enough, then the detector's motion relative to the source will cause the detector to see different polarizations at different times (i.e. C is a function of time). Thus, long duration observation will help in resolving the two polarizations. However, this will not be possible for inspirals observed by ground-based antennas as the waves will only last for a few minutes, during which C remains essentially constant. In this case an interferometric antenna will observe only a certain combination of the two polarizations and information on the source direction and wave polarization can be extracted only if several widely separated antennas observe the same signal.

8. Data analysis

Observing GWs requires a data analysis strategy which is in many ways different from conventional astronomical data analysis. There are several reasons why this is so:

— Gravitational wave antennas are essentially omni-directional with a response better than 50% on average over 75% of the sky. Hence our data analysis systems should carry out all-sky searches for sources.

— Interferometers are typically broad-band, covering 3 to 4 orders of magnitude in frequency. Although this is obviously to our advantage as it helps to track sources whose frequency may change rapidly, it calls for searches to be carried out over a wide band of frequencies.

— In contrast to EM radiation, most astrophysical GWs are tracked in phase and the SNR is built up by the coherent superposition of many wave cycles emitted by a source. Consequently, the SNR is proportional to the amplitude and only falls off with the distance to the source r, as $1/r$. Therefore, the number of sources of a limiting SNR increases as r^3 for a homogeneous distribution of sources in a flat Universe, as opposed to EM sources that increase only as $r^{3/2}$.

— Finally, GW antennas acquire data continuously for many years at a rate of several megabytes per second. It is expected that about a hundredth of this data will pass through our search analysis systems. Unless real time processing can be done we cannot hope to make our searches. This places huge demands on the speed of our data analysis hardware and requires a careful study of the search algorithms with a view to making them as optimal (maximum SNR) and efficient (least search time) as possible.

Let us first clarify our notation in this section and recall two important theorems. We shall use $x(t)$ to denote the detector output, which is assumed to consist of background noise $n(t)$ and a useful GW signal $h(t)$. The Fourier transform of a function $x(t)$ will be denoted as $\tilde{x}(f)$ and defined as

$$\tilde{x}(f) \equiv \int_{-\infty}^{\infty} x(t)\exp(2\pi\mathrm{i}ft)\,\mathrm{d}t\,. \tag{106}$$

With this definition the inverse Fourier transform is $x(t) \equiv \int_{-\infty}^{\infty}\tilde{x}(f)\exp(-2\pi\mathrm{i}ft)\,\mathrm{d}f$. The Fourier transform of a real function $x(t)$ obeys $\tilde{x}(-f) = \tilde{x}^*(f)$. A shift in the time domain simply appears as a constant phase shift in the Fourier domain, i.e. if $\tilde{x}(f)$ is the Fourier transform of $x(t)$ then the Fourier transform of $x(t-t_\mathrm{a})$ is $\exp(2\pi\mathrm{i}ft_\mathrm{a})\tilde{x}(f)$.

In Section 8 we use a system of units in which $G = c = 1$. Thus, for instance, $1M_\odot = 4.925\times10^{-5}$ s and 1 Mpc $= 1.08\times10^{14}$ s. We would like to caution the reader that $h(f)$ is the dimensionless gravitational wave amplitude, while in this section $\tilde{h}(f)$ is the Fourier transform of $h(t)$ and has physical dimensions of time.

8.1 Matched filtering and optimal signal-to-noise ratio

Matched filtering is a data analysis technique that efficiently searches for a signal of known shape buried in noisy data [169]. The technique consists in correlating the output of a detector with a waveform, variously known as a template or a filter. Given a signal $h(t)$ buried in noise $n(t)$, the task is to find an 'optimal' template $q(t)$ that would produce, on average, the best possible SNR. In this review we shall treat the problem of matched filtering as an operational exercise. However, this intuitive picture has a solid basis in the theory of hypothesis testing. The interested reader may consult any standard text book on signal analysis, for example Helstrom [169], for details.

Optimal filter. Let $x(t)$ denote the detector output. If no signal is present then $x(t)$ is just a realization of noise $n(t)$, i.e. $x(t) = n(t)$, while in the presence of a deterministic signal $h(t)$ it takes the form

$$x(t) = h(t-t_\mathrm{a}) + n(t)\,, \tag{107}$$

where t_a is the delay of the signal relative to the origin of time. In the case of signals $h(t)$ whose frequency changes with time, different values of t_a correspond to different frequencies at $t = 0$. Since a wave detector can record signals only in a certain frequency band, t_a marks the time at which the signal enters the detector's sensitivity band. For this reason t_a is called the 'time of arrival'. Given the time of arrival, the moment when a signal leaves the sensitivity band of the detector is determined by the parameters characterizing the signal, such as the masses of the component stars in the case of a binary inspiral signal. The time of arrival is an important

parameter in data analysis as it is usually unknown and must be determined by observation. Measuring arrival times in a network of detectors also helps in determining the direction to a source.

The correlation c of a template $q(t)$ with the detector output is defined as

$$c \equiv \int_{-\infty}^{\infty} x(t)q(t)\,\mathrm{d}t\,. \tag{108}$$

The above correlation integral is needed to improve the visibility of the signal. The following analysis reveals how this can be achieved: we shall work out the *optimal* filter $q(t)$ that maximizes the statistical average of the correlation c when a signal $h(t)$ is present in the detector output. To do this, let us first write the correlation integral in the Fourier domain by substituting the Fourier transforms $\tilde{x}(f)$ and $\tilde{q}(f)$ for $x(t)$ and $q(t)$ in the above integral, i.e.,

$$x(t) \equiv \int_{-\infty}^{\infty} \tilde{x}(f)\exp(-2\pi\mathrm{i}ft)\,\mathrm{d}f$$

and

$$q(t) \equiv \int_{-\infty}^{\infty} \tilde{q}(f)\exp(-2\pi\mathrm{i}ft)\,\mathrm{d}f,$$

respectively. After some straightforward algebra one obtains

$$c = \int_{-\infty}^{\infty} \tilde{x}(f)\tilde{q}^{*}(f)\,\mathrm{d}f, \tag{109}$$

where $\tilde{q}^{*}(f)$ denotes the complex conjugate of $\tilde{q}(f)$. In general, c consists of the sum of two terms — a filtered signal S and filtered noise N:

$$c = S + N, \tag{110}$$

where

$$S \equiv \int_{-\infty}^{\infty} \tilde{h}(f)q^{*}(f)\exp(2\pi\mathrm{i}ft_{\mathrm{a}})\,\mathrm{d}f,$$

$$N \equiv \int_{-\infty}^{\infty} \tilde{n}(f)q^{*}(f)\,\mathrm{d}f.$$

Since n is a real random process, c is also a real random process. If n is specified by a Gaussian random process with zero mean then c will also be described by a Gaussian distribution function, although its mean and variance will, in general, differ from those for n. The mean value of c is, clearly, S — the correlation of the template q with the signal h, since the mean value of n is zero:

$$\overline{c} = S = \int_{-\infty}^{\infty} \tilde{h}(f)\tilde{q}^{*}(f)\exp(2\pi\mathrm{i}ft_{\mathrm{a}})\,\mathrm{d}f. \tag{111}$$

The variance of c, that is $\overline{(c-\overline{c})^2}$, turns out to be

$$\overline{(c-\overline{c})^2} = \overline{N^2} = \int_{0}^{\infty} S_{\mathrm{n}}(f)\left|\tilde{q}(f)\right|^2\,\mathrm{d}f, \tag{112}$$

where $S_{\mathrm{n}}(f)$ is the power noise spectral density defined in Eqn (98). Now the power SNR is defined as $\rho^2 \equiv S^2/\overline{N^2}$ and the amplitude SNR is ρ. The form of the integrals in Eqns (111) and (112) resembles the definition of the scalar product of functions, which may be either templates or waveforms. Given two functions $a(t)$ and $b(t)$, we define their scalar product $\langle a,b\rangle$ to be

$$\langle a,b\rangle \equiv 2\int_{0}^{\infty} \frac{\mathrm{d}f}{S_{\mathrm{n}}(f)}\left[\tilde{a}(f)\tilde{b}^{*}(f) + \tilde{a}^{*}(f)\tilde{b}(f)\right]. \tag{113}$$

Recall that $S_{\mathrm{n}}(f)$ is real and positively definite, [cf. Eqn (99)]; consequently, the above scalar product determines a positively definite norm. The norm of a, denoted as $\|a\|$, is given by

$$\|a\| = 2\left[\int_{0}^{\infty} \frac{\mathrm{d}f}{S_{\mathrm{n}}(f)}\left|\tilde{a}(f)\right|^2\right]^{1/2}. \tag{114}$$

Using the reality of the time-domain function $h(t)$ we can write down the SNR in terms of the above scalar product as

$$\rho^2 = \frac{\left\langle h\exp(2\pi\mathrm{i}ft_a), S_{\mathrm{n}}q\right\rangle^2}{\left\langle S_{\mathrm{n}}q, S_{\mathrm{n}}q\right\rangle}\,. \tag{115}$$

Now, the scalar product of two functions $\langle a,b\rangle$ acquires its maximum value when $a = b$. Applying this to the above equation one finds that the template q maximizing ρ and called the *optimal template*, denoted as $\tilde{q}_{\mathrm{opt}}(f)$, is simply

$$\tilde{q}_{\mathrm{opt}}(f) = \gamma\,\frac{\tilde{h}(f)\exp(2\pi\mathrm{i}ft_{\mathrm{a}})}{S_{\mathrm{n}}(f)}\,, \tag{116}$$

where γ is an arbitrary constant. The inverse Fourier transform of Eqn (116) gives the optimal template $q_{\mathrm{opt}}(t)$ in the time domain. One can see that $q_{\mathrm{opt}}(t)$ is the convolution of the *time-translated* signal $h(t)$ with the inverse Fourier transform of $1/S_{\mathrm{n}}(f)$. Note that, to achieve the maximum of the SNR, the optimal template should match not only the shape of the signal but also its time of arrival t_{a} [cf. the factor $\exp(2\pi\mathrm{i}ft_{\mathrm{a}})$ in the expression for the optimal template]. Since one does not know the time of arrival of the signal beforehand, one has to construct the correlation of the detector output for several different relative lags of the template with respect to the detector output. In other words, one constructs the correlation function

$$\begin{aligned} c(t') &\equiv \int_{-\infty}^{\infty} x(t)q(t-t')\,\mathrm{d}t \\ &= \int_{-\infty}^{\infty} \tilde{x}(f)\tilde{q}^{*}(f)\exp(-2\pi\mathrm{i}ft')\,\mathrm{d}f, \end{aligned} \tag{117}$$

where t' is called the lag parameter.

Optimal signal-to-noise ratio. We can now work out the maximum, or optimal, SNR by substituting Eqn (116) for the optimal template in Eqn (115),

$$\rho_{\mathrm{opt}} = \langle h,h\rangle^{1/2} = 2\left[\int_{0}^{\infty} \mathrm{d}f\,\frac{\left|\tilde{h}(f)\right|^2}{S_{\mathrm{n}}(f)}\right]^{1/2}. \tag{118}$$

We note that the optimal SNR is *not* proportional to the total energy of the signal, which is $4\int_0^\infty \mathrm{d}f|\tilde{h}(f)|^2$, but rather to the integrated signal power weighted with the noise PSD. This is in accordance with what we could guess intuitively: the contribution to the SNR from a frequency bin where the noise PSD is high and hence less reliable should be smaller

than that from a bin where the noise PSD is low. Thus, an optimal filter automatically takes into account the nature of the noise PSD. In this final expression for the optimal SNR the parameter t_a does not appear because the time-of-arrival optimal template matches that of the signal and hence t_a cancels out in the scalar product.

The expression for the optimal SNR (118) suggests how one should compare signal strengths with the noise performance of the detector. Note that one cannot directly compare $|\tilde{h}(f)|^2$ with $S_n(f)$ as they have different physical dimensions. In the GW literature one writes the optimal SNR in one of the following equivalent ways

$$\rho_{\rm opt} = 2\left[\int_0^\infty \frac{{\rm d}f}{f}\,\frac{\left|\sqrt{f}\,\tilde{h}(f)\right|^2}{S_n(f)}\right]^{1/2} = 2\left[\int_0^\infty \frac{{\rm d}f}{f}\,\frac{\left|f\tilde{h}(f)\right|^2}{fS_n(f)}\right]^{1/2}, \tag{119}$$

which facilitates the comparison of signal strengths with the noise performance. As seen from (119), one can compare between the dimensionless quantities $f|\tilde{h}(f)|$ and $[fS_n(f)]^{1/2}$ or between the dimensioned quantities $\sqrt{f}|\tilde{h}(f)|$ and $[S_n(f)]^{1/2}$.

Signals of interest to us are characterized by several (*a priori* unknown) parameters, such as the masses of the component stars in the binary, their intrinsic spins, etc., and an optimal filter must agree with both the signal shape and its parameters. A filter whose parameters are slightly mismatched with that of a signal can greatly degrade the SNR. For example, even a mismatch of one cycle out of 10^4 cycles can degrade the SNR by a factor of 2.

What is the improvement brought about if the parameters of a filter and its shape are precisely matched with that of a signal, as opposed to the case where no knowledge of the signal is available? Matched filtering helps in enhancing the SNR in proportion to the square root of the number of signal cycles covered by the detector band, as opposed to the case where the signal shape is not known and all that can be done is to Fourier-transform the detector output and compare the signal energy in a frequency bin to the noise energy in the same bin (see, e.g., Ref. [170] for a proof).

Matched filtering of continuous GWs. We will now apply the matched filtering theorem to observations of sources emitting continuous gravitational waves (CW) at a single frequency. If we observe a monochromatic source of frequency f_0, that is $h(t) = h_0\cos(2\pi f_0 t)$, for a duration T, then its Fourier transform is a sinc function ($\mathrm{sinc}\,x \equiv \sin x/x$):

$$\tilde{h}(f) = \frac{h_0 T}{2}\,\mathrm{sinc}\left[2\pi(f-f_0)T\right]. \tag{120}$$

In the above expression we have ignored the sinc function for $f = -f_0$; it does not contribute to the SNR since the integral in Eqn (118) runs only over positive frequencies. Since the sinc function is strongly peaked near $f = f_0$ and also because the noise PSD varies slowly in the frequency range 100 – 1000 Hz, one can treat the power spectrum appearing in the SNR integrand of Eqn (118) as a constant and write the optimal SNR as

$$\rho_{\rm opt}^{\rm CW} = \frac{2}{\sqrt{S_n(f_0)}}\left[\int_0^\infty \left|\tilde{h}(f)\right|^2 {\rm d}f\right]^{1/2} = \frac{h_0}{\sqrt{S_n(f_0)/T}}, \tag{121}$$

where we have used the identity $\int_0^\infty \mathrm{sinc}^2(2\pi f T)\,{\rm d}f = 1/4T$. Equation (121) justifies Eqn (101), which was derived heuristically.

8.2 Matched filtering inspiral waves from compact binaries

As pointed out in earlier sections, the last few minutes in the evolution of a compact binary (NS + NS, NS + BH, BH + BH) are a promising source for interferometers that are currently being built. The waveform from these sources is known very accurately and therefore matched filtering is the best choice for detecting these sources. Matched filtering is very sensitive to the phasing of the waves. It is important, therefore, to keep accurate phase information in our search templates; their amplitudes can be taken from the lowest order post-Newtonian theory. Such an approximation, which works only with phase corrections and neglects amplitude corrections, is called the *restricted post-Newtonian approximation* [171].

Accurate templates for inspiral search. To compute the waveform, one must know the evolution of its phase and amplitude, which involves the computation of the relative velocity $v_A(t)$ of the two stars and the phase evolution $\varphi_A(t)$ of the binary. The subscript A (which will hereafter be dropped, for convenience, from v_A) denotes the fact that we know these quantities only in a certain approximation. In the restricted post-Newtonian approximation the binary inspiral waveform is given by (see, e.g., Ref. [172])

$$h^A(t) = h_0 v_A^2(t)\cos\phi_A\left[v_A(t)\right]. \tag{122}$$

The amplitude h_0 depends on the masses M_1 and M_2 of the component stars, the distance r to the binary, and the orientation of the source relative to the antenna. More precisely,

$$h_0 = \frac{4\eta M}{r}\,C(i,\theta,\varphi,\psi). \tag{123}$$

Here, $M = M_1 + M_2$ is the total mass of the binary. $\eta = M_1M_2/M^2$ is the dimensionless (symmetric) mass ratio, which takes a maximum value of $1/4$ when $M_1 = M_2$. The quantity η can also be regarded as a measure of how strongly the geometry is different from the Schwarzschild geometry of a single body and is sometimes referred to as the *deformation parameter*. C is a function of various angles describing the polarization of the wave and the orientation of the source relative to the antenna. For a binary whose orbit is inclined at an angle i to the line of sight, C is given by

$$C(i,\theta,\varphi,\psi) = \sqrt{A^2 + B^2}\,,$$

$$A = \frac{1}{2}F_+(\theta,\varphi,\psi)(1+\cos^2 i), \quad B = F_\times(\theta,\varphi,\psi)\cos i. \tag{124}$$

In this equation the angles θ, φ, and ψ parameterize both the propagation direction and the polarization of the GW with respect to the detector, and F_+ and $F_\times$ are the detector beam pattern functions given in Eqn (105).

For the purpose of computing the SNR from candidate binaries we can either assume the source to be *ideally* oriented, that is

$$i = \theta = \varphi = \psi = 0\,, \quad F_+ = 1\,, \quad F_\times = 0\,, \quad C = 1 \tag{125}$$

(which gives us the best possible SNR $\rho_{\rm ideal}$), or compute the rms (root-mean-square) SNR, $\rho_{\rm rms}$, by averaging over all angles. In the latter case we have the following rms values:

$$\langle F_+^2\rangle^{1/2}_{\theta,\varphi,\psi} = \langle F_\times^2\rangle^{1/2}_{\theta,\varphi,\psi} = \frac{1}{\sqrt{5}}\,, \qquad \langle C^2\rangle^{1/2}_{i,\theta,\varphi,\psi} = \frac{2}{5}\,, \quad (126)$$

which will be used in the next section to compute various SNRs.

The relative velocity $v_A(t)$ and the GW phase $\phi_A(t) \equiv 2\varphi_A(t)$ in Eqn (122) can be found from the following ordinary differential equations [173]:

$$\frac{{\rm d}v}{{\rm d}t} + \frac{\mathcal{F}_A(v)}{ME'_A(v)} = 0\,, \qquad \frac{{\rm d}\phi_A}{{\rm d}t} - \frac{2v^3}{M} = 0\,, \quad (127)$$

where $E_A(v)$ is the relativistic total energy per unit mass, i.e., $E = (E_{\rm total} - M)/M$, while $E'_A(v) \equiv {\rm d}E_A/{\rm d}v$, and $\mathcal{F}_A \equiv -{\rm d}E_A/{\rm d}t$ is the GW flux escaping the system, at the given approximation [6]. The above differential equations can be formally solved to obtain

$$t_A = t_0 - M\int_{v_0}^{v} {\rm d}v\,\frac{E'_A(v)}{\mathcal{F}_A(v)}\,, \qquad \phi_A = \phi_0 - 2\int_{v_0}^{v} {\rm d}v\,v^3\,\frac{E'_A(v)}{\mathcal{F}_A(v)}\,. \quad (128)$$

In general, there is no closed-form representation of the above integrals. However, while working with the post-Newtonian theory, one can use Taylor series expansions of GW energy and flux; therefore, the above integrals can be calculated to obtain a solution of the form $t_A = t_A(v)$ and $\phi_A = \phi_A(v)$, which can in turn be inverted to obtain $v_A = v_A(t)$ and $\phi_A = \phi_A(t)$. However, there is much debate in the literature on what is the best representation of the exact inspiral wave emitted by a binary system. Some authors work with post-Newtonian expansions for flux and energy and the closed form solution mentioned above [174]; others, with the ODEs (127) or, equivalently, with the integrals (128) [173, 175]; and yet others, with a more accurate representation of the energy and flux functions, called P-approximants [172, 176]. Such representations have been shown to be extremely accurate in the test mass limit (i.e. $\eta \to 0$) and are also expected to be well-behaved when $\eta \neq 0$ [172, 173, 176].

To the lowest post-Newtonian order, that is, in the quadrupole approximation, and for circular orbits we have

$$E_{\rm N}(v) = -\frac{\eta v^2}{2}\,, \qquad \mathcal{F}_{\rm N}(v) = -\frac{32\eta^2 v^{10}}{5}\,, \quad (129)$$

where the subscript N is used to denote that the quantity is given to the lowest, i.e. Newtonian, order in the post-Newtonian theory. Substituting these expressions into Eqn (128) and inverting the resulting equations gives the following waveform:

$$t(v) = t_0 + \frac{5M}{256\eta}\left(v_0^{-8} - v^{-8}\right), \quad (130)$$

$$\phi(t) = \phi_0 + \frac{1}{16\eta}\left(v_0^{-5} - v^{-5}\right), \quad (131)$$

$$h(t) = h_0 v^2(t)\cos\phi(t)\,, \quad (132)$$

[6] See Appendix III for the post-Newtonian expansions of the energy and flux functions.

where we have dropped the subscript N for brevity; v_0 and ϕ_0 are the relative velocity of the two stars and the phase of the wave at time t_0; and v in the last two equations is obtained by inverting the first equation.

Signal-to-noise ratios for binary inspiral signals. The optimal SNR is written in Eqn (118) in terms of the Fourier transform of the signal. For compact binary inspiral events an estimate of the optimal SNR can be obtained by using the stationary phase approximation to the Fourier transform of the signal [177]. In this approximation, the modulus of the Fourier transform of the signal in Eqn (122) is found to be [176]

$$|\tilde{h}(f)| = h_0\left(\frac{5}{384\eta}\right)^{1/2}\frac{f^{-7/6}}{\pi^{2/3}M^{1/6}}\,. \quad (133)$$

Substituting C from (123) for h_0 and using the resulting expression in Eqn (118) we find the following optimal SNR:

$$\rho_{\rm opt} = C\,\frac{M^{5/6}}{r\pi^{2/3}}\left(\frac{5\eta}{6}\right)^{1/2}\left[\int_{f_{\rm s}}^{f_{\rm LSO}} {\rm d}f\,\frac{f^{-7/3}}{S_{\rm n}(f)}\right]^{1/2}, \quad (134)$$

where we have now introduced specific lower and upper limits in the integral. The lower cutoff is dictated by the noise PSD, which rises steeply below a certain frequency $f_{\rm s}$. SNR can hardly be increased by extending our templates to frequencies below this lower cutoff. For ground-based interferometers $f_{\rm s}$ is in the range 1 – 50 Hz, while for the space-based LISA $f_{\rm s}$ is slightly smaller than 0.1 mHz. The upper cutoff $f_{\rm LSO}$ is determined by the location of the last stable orbit of the binary. In the test mass limit (i.e., $\eta \to 0$) the inspiral signal will terminate when the test mass is at the distance $R_{\rm LSO} = 6M$ from the central body. This corresponds to the GW frequency of $f_{\rm LSO}$ given by

$$f_{\rm LSO}^2 = \frac{M}{\pi^2 R_{\rm LSO}^3} = \frac{1}{6^3\pi^2 M^2}\,, \quad (135)$$

where the first equality simply represents Kepler's third law and in the second equality we have used $R_{\rm LSO} = 6M$. This upper limit [7] is important since it limits the inspiral signals that an interferometer can observe. For instance, a ground-based interferometer could not observe the inspiral of massive binary black holes of mass $\sim 10^6 - 10^9 M_\odot$ since this signal terminates at frequencies in the milli-Hertz region, where ground-based interferometers are too noisy. However, a space-based detector, such as LISA, will be sensitive to the inspiral of massive black holes occurring anywhere in the Universe.

We are finally in a position to write down the rms and the ideal SNRs. For a binary at a distance r from the Earth, consisting of stars of individual masses M_1 and M_2, total mass M, and symmetric mass ratio η, the rms and ideal SNRs are obtained by using the rms and ideal values of C from Eqn (126) and Eqn (125), respectively, in Eqn (134):

$$\rho_{\rm rms} = \frac{M^{5/6}}{r\pi^{2/3}}\left(\frac{2\eta}{15}\right)^{1/2}\left[\int_{f_{\rm s}}^{f_{\rm LSO}} {\rm d}f\,\frac{f^{-7/3}}{S_{\rm n}(f)}\right]^{1/2}, \qquad \rho_{\rm ideal} = \frac{5}{2}\,\rho_{\rm rms}\,. \quad (136)$$

[7] The justification for using an upper limit in the frequency domain, based on the upper limit for the time domain, is somewhat technical; the interested reader is referred to Ref. [171] and references therein for a discussion.

Note that the SNR depends only on the combination $\mathcal{M} = M\eta^{3/5}$ of the two masses, called the chirp mass [178], and on the integral whose value crucially depends on the upper limit. In ground-based detectors, for $f_{\rm LSO}$ larger than about 300 Hz, the upper limit is not too important while, for lower values of the LSO frequency, the integral begins to degrade.

Though the SNR depends only on the chirp mass, this does not mean that we cannot measure the two masses by GW observations. Recall that the optimal SNRs are obtained only when the phases of the signal and the template are finely matched. Matching the phase resolves the degeneracy in the two masses. Indeed, post-Newtonian theory offers an opportunity to measure more than two parameters, although, in general relativity, each of those parameters in the point particle limit (i.e., with intrinsic spins neglected) depends only on the two component masses. This can potentially be used to test general relativity in a strongly non-linear regime (see Refs [179, 180] for a discussion.)

A detector's sensitivity is sometimes measured by the distance $r_5^{(1.4,\,1.4)}$ to an ideally oriented NS + NS source that produces an SNR of 5. We find that $r_5^{(1.4,\,1.4)} = 30$ Mpc for GEO, 45 Mpc for LIGO, 1.8 Mpc for TAMA and 50 Mpc for VIRGO.

The integral in the expression (136) for the SNR depends on the choice of the detector via the noise PSD $S_{\rm n}(f)$ and can be (numerically) computed using the expression for noise PSDs given in Table 5. The SNRs thus obtained are plotted in Fig. 17 as a function of the total mass of a binary consisting of equal masses (i.e. $\eta = 1/4$) and listed in Table 6 for three types of binaries at 100 Mpc and hypothetical MACHO binaries at 1 Mpc. We note that an inspiral event from a BH + BH binary at 100 Mpc will have a reasonably good SNR in all the three larger first-generation interferometers, namely GEO, LIGO and VIRGO. A coincidence search for these sources can, therefore, unambiguously pick up these signals and hence we believe that the first detected signals will be binary black hole inspirals.

For intermediate mass black hole binaries, of a mass of 100 solar masses or so, VIRGO and GEO can achieve excellent SNRs. There is no convincing evidence for the existence of BHs of such masses, or their binaries, but future GW observations should shed light on such systems.

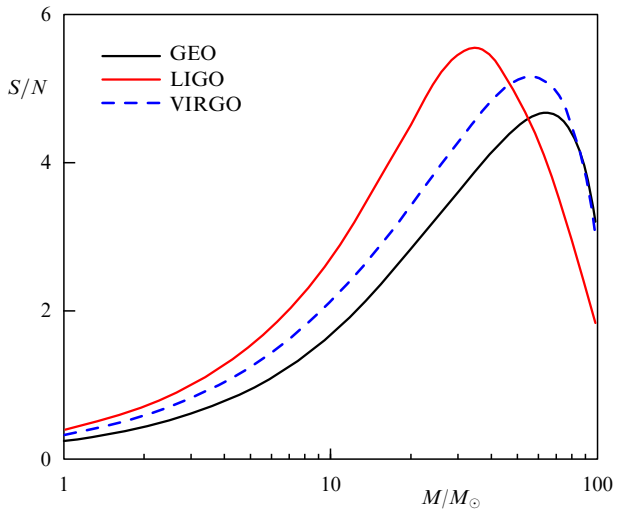

Figure 17. Signal-to-noise ratio for early ground-based interferometers as a function of total mass, for inspiral signals from binaries of equal masses at 100 Mpc, averaged over source inclination and location. (TAMA is left, since the SNRs in that case are too low for these sources.)

Table 6. Signal-to-noise ratios for some typical binaries consisting of neutron stars of $1.4M_\odot$ and/or black holes of $10M_\odot$ and $15M_\odot$ at 100 Mpc, as well as for hypothetical primordial MACHO BH binaries of $0.5M_\odot$ at 10 Mpc.

Detector\Binary	(0.5, 0.5)	(1.4, 1.4)	(1.4, 10)	(10, 10)	(15, 15)
GEO	2.6	0.58	1.2	2.8	3.6
LIGO	3.8	0.95	2.0	4.6	5.5
VIRGO	3.5	0.78	1.5	3.4	4.3

As mentioned before, improved interferometers will have ten times better sensitivity and bandwidth and will therefore be able to achieve similar SNRs at 1 Gpc. As we have seen in Section 3, at such distances the event rate builds up to several per week. The construction of such advanced interferometers will, therefore, make it possible to routinely detect compact binary inspirals and coalescences and thereby open up a new gravitational window for observation.

The SNRs achievable for the LISA detector are phenomenal and are listed in Table 7. LISA should be able to observe a million solar mass BH inspiral almost anywhere in the Universe. The time spent by a binary in the LISA band width could be larger than the lifetime of the detector. For this reason we have expressed the SNR in the above equation for the frequency range (f_1, f_2) in which the signal is extracted during a given observational period. For a given duration of observation the strongest binary signal will be recorded if the observation starts closest to the instant of coalescence; therefore, we shall take $f_2 = f_{\rm LSO}$.

Table 7. Signal-to-noise ratios in LISA for supermassive binary black holes and stellar mass compact objects falling into them, at a cosmological distance of 3 Gpc.

$M_1/M_\odot$	$M_2/M_\odot$	f_1, Hz	$f_{\rm LSO}$, Hz	S/N
10^7	10^7	1.08×10^{-5}	2.2×10^{-4}	1900
10^6	10^6	4.54×10^{-5}	2.2×10^{-3}	4900
10^5	10^5	1.92×10^{-4}	2.2×10^{-2}	1700
10^6	10	2.75×10^{-3}	4.4×10^{-3}	21
10^5	10	5.55×10^{-3}	4.4×10^{-2}	11
10^6	1,4	3.93×10^{-3}	4.4×10^{-3}	3.8
10^5	1,4	1.15×10^{-2}	4.4×10^{-2}	2.2

8.3 Sensitivity to stochastic gravitational waves

In Sections 5 and 6 we have seen how astrophysical processes can generate a stochastic background of GWs and what physical interactions can produce a background in the early Universe, respectively. A stochastic background can easily be detected if the instrumental and environmental background noise levels in the detector are much lower than the GW background and especially if the statistical properties of the GW background or its spectral characteristics are different from those expected for the background noise.

Even when the GW background is far too weak compared to the instrumental noise backgrounds, the theory of matched filtering developed in Section 8.2 for the detection of deterministic signals can be generalized to the case of a stochastic background of gravitational radiation. As opposed to deterministic signals, we will not have the advantage of using a template computed beforehand.

Rather, the idea is to use the (noisy) data in one detector, in which a stochastic background may be present, as a template to detect the background in (noisy) data from another detector. In other words, we can use the 'noisy template' from one detector for the matched filtering of the background in another. Because our template is noisy, the enhancement in SNR will not be as good as in the case of matched filtering of a deterministic signal. However, as the background will always be present, if we can correlate sufficiently long outputs of two detectors, we will be able, in principle, to single out the background from other noise sources.

The idea that one detector output can be used as a template to enhance the visibility of the background in another essentially assumes that (1) the two detectors record identical GW signals and (2) the instrumental noises in the two detectors are uncorrelated. In reality neither of these assumptions will hold perfectly. If two detectors are close together on the Earth, so that they record identical stochastic signals, then they will also be affected by similar environmental disturbances such as seismic activity, wind, storm, etc., and the background noise may have a large non-GW correlation. If the two detectors are far apart then their environmental disturbances are unlikely to be correlated but they may record different polarizations and phases of the background GW so that the cross-correlation may be insignificant. Clearly, a compromise is in order: two widely separated detectors are good for cancellation of the noise background, while two nearby detectors are favored for enhancing the signal background. In what follows we will derive the SNR enhancement achieved on cross-correlating the data from two identical detectors placed at the same location. The following derivation is heuristic and indicates how the cross-correlation works; see Refs [159, 160] for a more rigorous derivation for two detectors of different orientations.

Let $x_1(t)$ and $x_2(t)$ denote the outputs of two interferometric antennas located near each other and having the same orientation. They will both receive the same stochastic signal $h(t)$ and we will assume that the internal noises $n_1(t)$ and $n_2(t)$ are uncorrelated. We will also assume that the rms value of the internal noise is much larger than the rms value of the stochastic signal we wish to detect. Thus, we can write

$$x_1(t) = n_1(t) + h(t)\,, \quad x_2(t) = n_2(t) + h(t)\,, \quad \overline{n_1^2} = \overline{n_2^2} \gg \overline{h^2}\,, \tag{137}$$

where the overbar indicates the ensemble averaging and we have assumed that the ensemble averages of the background noise and stochastic signal are both zero. In order to extract the stochastic background we construct a cross-correlation integral C. The cross-correlation of two detector outputs lasting a time T each is given by

$$C^2 \equiv \int_{-T/2}^{T/2} x_1(t)x_2(t)\,\mathrm{d}t \equiv (x_1, x_2)\,, \tag{138}$$

where we have introduced a bracket notation for the cross-correlation of two data sets, x_1 and x_2 The cross-correlation is denoted as C^2 since it is quadratic in the useful signal h. In view of Eqn (137), C^2 can be written as

$$C^2 = (n_1, n_2) + (n_1, h) + (h, n_2) + (h, h)\,. \tag{139}$$

C^2 is a random process whose average, under the assumption that the internal detector noises are uncorrelated, gives the useful signal

$$S^2 = \overline{C^2} = \overline{(h, h)}\,. \tag{140}$$

From this equation we conclude that the signal component of C^2 grows in proportion to the integration time T, i.e., $S^2 \propto T$. This, of course, is only the statistical average of C^2 and it must be compared with the fluctuation in C^2 to deduce its significance. Let us define the noise component of C^2 by

$$N^2 \equiv C^2 - \overline{C^2} = (n_1, n_2)\,. \tag{141}$$

N^2 is also a random variable and its rms value will involve the noise spectral densities [cf. Eqns (98) and (100)] $S_{\mathrm{n}_1}(f)$ and $S_{\mathrm{n}_2}(f)$ of the two detectors. Assuming that the two detectors have identical noise spectra, say $S_{\mathrm{n}}(f)$, the rms value of N^2 turns out to be

$$\overline{N^4} \propto T \int_{-\infty}^{\infty} \mathrm{d}f\, S_{\mathrm{n}}^2(f)\,. \tag{142}$$

We see from Eqns (140) and (142) that the SNR for the stochastic background grows as $T^{1/4}$. Therefore, in principle, a stochastic background that is below the internal noise can be extracted by integrating for a sufficiently long time. Conversely, if we have data sets of a certain duration T with internal noise spectrum $S_{\mathrm{n}}(f)$ then we can estimate the minimum level of the stochastic background that can be extracted from observations. These estimates show that early interferometers will be able to detect a stochastic background at a level of $h(f) \approx 2 \times 10^{-23}$, or $\Omega_{\mathrm{gw}} \sim 5 \times 10^{-6}$ (cf. Section 6), with a 90% confidence level, after three months of integration [160]. Advanced ground-based detectors should be able to reach a level that is five orders of magnitude lower (in terms of Ω_{gw}). The space-based LISA will have a sensitivity to the primordial background similar to that of advanced ground-based detectors, but in the frequency range $\sim 10^{-4} - 10^{-1}$ Hz, where the primordial background has larger spectral amplitudes $h(f)$. Since LISA is likely to be in place at about the same time as the third generation advanced ground-based detectors will be, there is an exciting possibility of detecting the primordial background, of the kind discussed in Section 6, within the next decade. They will also help us to understand the nature of astrophysical sources that generate a stochastic GW background of the sort discussed in Section 5.

8.4 Computational costs

Matched filtering places stringent demands on the knowledge of the signal's phase evolution which depends on (1) our modeling of the signal and (2) the parameters characterizing the signal. If our model signal is inaccurate or if the signal's parameters are unknown, there may be a loss in the SNR extracted. For instance, in the case of inspiral signals, a mismatch of one cycle in 10^4 cycles leads to a drop in the SNR by more than a factor of two, which reduces the number of potentially detectable events by a factor of eight. (Recall that the SNR is inversely proportional to the distance to the source; thus, an SNR loss by a factor a will reduce the range of the detector by the same factor, resulting in a decrease in the volume of observation, and hence in the number of events, by the factor a^3.) Moreover, since the parameters of the signal will not be known in advance, it is necessary to filter the data with a family of templates located at various points in the

parameter space — e.g., at the sites of a lattice, so that any signal will lie close enough to at least one of the templates to have a good cross-correlation with that template. The number of such templates is typically very large. This places a great demand on the computational resources needed to make an in-real-time search, that is to search for signals in the detector output at the same rate at which data are acquired. We shall discuss below the method of finding the number of templates to filter any known signal and estimating the computational resources required to analyse the data in real time. In Section 8.4 we will discuss the tools needed for parameter estimation.

Ambiguity function. The ambiguity function, well known in the statistical theory of signal detection [169], is a very powerful tool in signal analysis. It helps one in making estimates of variances and covariances involved in the measurement of various parameters, in computing biases introduced by a family of templates whose shape is not the same as that of the family of signals intended to be detected, in assessing the number of templates required to span the parameter space of the signal, etc. We discuss parameter estimation in this section and computational costs in the next one.

We begin by defining a *normalised* waveform. A waveform a is said to be normalised if it has a unit norm: $\|a\| \equiv \langle a, a\rangle^{1/2} = 1$. The norm h_0 of the signal h is also its optimal SNR: $\rho_{\rm opt} = \langle h, h\rangle^{1/2} = \|h\| = h_0$. For this reason the norm of a signal is also referred to as the signal strength.

The *ambiguity function* is defined as the scalar product of two normalised waveforms maximized at the initial phase of one of the waveforms. More precisely, it is the absolute value of the scalar product of two normalised waveforms [8].

Let $a(t;\alpha)$ be a normalised waveform. Note that we use the symbol a to denote a family of waveforms all having the same functional form but differing from one another in the parameter values. Indeed, $\alpha = \{\alpha^A | A = 0, \ldots, p\}$ denotes the parameter vector comprising $p+1$ parameters. It is convenient to choose the parameter α^0 to be the lag t' that simply corresponds to the time when the event occurs; therefore, α^0 is called an *extrinsic* parameter, while the other p parameters α^k, $k = 1, \ldots, p$ are called *intrinsic* parameters and characterize the GW source. Given two normalised waveforms, $a(t;\alpha_1)$ and $a(t;\alpha_2)$, whose parameter vectors are not necessarily the same, the ambiguity $\mathcal{A}$ is defined as

$$\mathcal{A}(\alpha_1, \alpha_2) \equiv \max_{\alpha^0} \left| \langle a(\alpha_1), a(\alpha_2)\rangle \right| . \tag{143}$$

Since the waveforms are normalised, $\mathcal{A}(\alpha_1, \alpha_1) = 1$ and, if α_1 is not equal to α_2, $\mathcal{A}(\alpha_1, \alpha_2) \leqslant 1$.

It is important to note that in the definition of the ambiguity function there is no need for the functional forms of the template and signal to be the same; the definition holds for any signal – template pair of waveforms. Moreover, the number of template parameter does not need to be (and usually is not) equal to the number of parameters characterizing the signal. For instance, a binary can be characterized by a large number of parameters, such as the masses, spins, eccentricity of the orbit, etc., while, as a model waveform, we may take one involving only the masses. In the context of inspiral waves, $a(t;\alpha_2)$ is the exact general-relativistic waveform emitted by a binary, whose form we do not know, while the template family is a post-Newtonian, or some other, approximation to it, which will be used in detecting the true waveform. Another example could be signals emitted by spinning neutron stars, isolated or in binaries, whose time evolution is unknown, either because we cannot anticipate all physical effects that affect their spin, or because the parameter space is so large that we cannot take into account all of them in a realistic search. Of course, in such cases we cannot compute the ambiguity function since one of the arguments to the ambiguity function is unknown. These are indeed issues where substantial work is called for. For this review it suffices to assume that the signal and template waveforms are of identical shape and the number of parameters in the two cases is the same.

In the definition of the ambiguity function Eqn (143) α_1 can be thought of as the parameters of the (normalised) template, while α_2 are those of the signal. With the template parameters α_1 fixed, the ambiguity function is a function of p signal parameters α_2^k, $k = 1, \ldots, p$, giving the SNR obtained by the template for different signals. However, one can interpret equally well the ambiguity function as the SNR obtained for a given signal by filters of different parameter values. Now, the region in the signal parameter space for which a template obtains SNRs larger than a chosen value, called the *minimal match* [181], is the *span* of that template. Templates should be chosen so that they together span the entire signal parameter space of interest with the least overlap of one another's spans.

Metric on the space of waveforms. The computational costs of our search for and estimation of the parameters of a signal can be assessed based on a lucid geometrical picture [181, 182]. To form such a picture, we begin with the space of signal waveforms. A waveform $e(t;\alpha)$ with a given set of values of its parameters can be thought of as a unit *vector*. (In much of the discussion below we will deal with unit signal and template vectors.) The set of sample values $h = \{h_k | k = 0, \ldots, N-1\}$ of a waveform for times t_k, $k = 0, \ldots, N-1$ can be thought of as an N-dimensional signal vector but not necessarily of unit norm. Indeed, the output of a detector sampled in the same way can also be regarded as an N-dimensional vector. The set of all detector output vectors forms a vector space. Return to the signal vector. As its parameters are varied, the signal vector spans a sub-space in the underlying N-dimensional vector space of detector outputs. The dimension of this sub-space is equal to the number of parameters of the signal and is called the *signal space*.

The signal space, which is a sub-space of the full vector space, has a manifold structure, the parameters of the wave constituting a coordinate system, and the dimensionality of the space being equal to the number of parameters. Having defined a manifold we can ask if it is possible to define a meaningful metric on this manifold. Indeed, we already have a necessary tool to define the metric, namely the ambiguity function.

Let us fix the parameters α_1 of the template; then the ambiguity function is a function of $p+1$ signal parameters α_2^k, $k = 0, \ldots, p$. Expanding $\mathcal{A}(\alpha_1, \alpha_2)$ about its maximum $\alpha_2 = \alpha_1$ and retaining only quadratic terms yields

$$\mathcal{A}(\alpha_1, \alpha_2) = 1 - \gamma_{AB}\, \Delta\alpha^A \Delta\alpha^B + O\left[(\Delta\alpha)^3\right] , \tag{144}$$

[8] Working with analytic signals $h(t) = a(t)\exp\{{\rm i}[\phi(t) + \phi_0]\}$, where $a(t)$ and $\phi(t)$ are the time-varying amplitude and phase of the signal, respectively, we observe that the initial phase ϕ_0 of the signal simply factors out as a constant phase in the Fourier domain and the maximization at this initial phase can be done by taking the absolute value of the scalar product of the template and the signal.

where $\Delta\alpha \equiv \alpha_2 - \alpha_1$ and γ_{AB} is the metric on the space of waveforms:

$$\gamma_{AB}(\alpha_1) \equiv -\frac{1}{2}\frac{\partial^2 \mathcal{A}(\alpha_1,\alpha_2)}{\partial\alpha_2^A\,\partial\alpha_2^B}\bigg|_{\alpha_2=\alpha_1}. \tag{145}$$

This is the metric at the point α_1 on the manifold. From now on we shall drop the subscript 1 of the parameter α since it really represents an arbitrary point on the manifold where a template resides. Since one can easily maximize $\mathcal{A}$ over the lag parameter α^0, it is desirable to work with the metric g_{ij} that is projected orthogonal to α^0, namely

$$g_{ij} \equiv \gamma_{ij} - \frac{\gamma_{i0}\gamma_{j0}}{\gamma^{ij}}, \qquad i,j = 1,\ldots,p. \tag{146}$$

Let us suppose that it is required to choose a template bank. By a template bank we mean a discrete family of signal waveforms chosen in a given region of the signal parameter space. In such a template bank no template waveform will perfectly match an incoming signal but if the density of templates is large enough then it may be possible to extract any signal with an SNR larger than a certain fraction of the optimal SNR. Most of the early GW data analyses concerned studying efficient algorithms to set up a template bank so as to minimize the computational costs of the search. The density of templates in the bank depends on what is the largest fraction of the optimal SNR one is prepared to lose in the search: the smaller this fraction, the greater the density of templates. One is normally interested in setting up the bank so that each possible signal have a *maximized overlap* larger than a certain *minimal match* $\mathfrak{M}$ with at least one template in the bank. By overlap we mean the scalar product of a (normalised) signal waveform with a (normalised) template waveform, and maximization is carried out over all the template parameters. The maximized overlap, sometimes referred to as the *match*, is always smaller than or equal to one. Demanding that the proper distance between templates, namely $g_{ij}\Delta\alpha^i\Delta\alpha^j$, be as large as possible, for a given minimal match $\mathfrak{M}$, we can obtain the following formula for the spacing between templates using Eqn (144):

$$\Delta\alpha^k = \sqrt{\frac{2(1-\mathfrak{M})}{g_{kk}}}, \tag{147}$$

where the factor 2 in the numerator arises because the proper distance $(1-\mathfrak{M})$ between the least matched signal and template implies that the proper distance between templates is twice that value, assuming that the templates are placed on a square lattice.

The distance between templates can also be computed more accurately by employing numerical methods, as was done in Ref. [177] (see also Ref. [183]). A comparison of the metric-based method discussed above and the numerical method shows that the quadratic approximation (147) is typically good for $\mathcal{A} \geqslant 0.95$. Thus, in the limit of close template spacing, Eqn (147) can be used to make a choice of templates.

We now turn our attention to the use of the metric in calculating the number of templates needed for a search. If the number $\mathcal{N}$ of templates needed to cover a region of interest is large, $\mathcal{N}$ is well approximated by dividing the proper volume of the region of interest on the signal manifold by the proper volume per template. The proper volume per template, ΔV, depends on the packing algorithm used, which in turn depends on the dimension p of the parameter space. For instance, we used a square lattice above. For $p = 2$, the optimal packing is a hexagonal lattice, and thus

$$\Delta V = \frac{3\sqrt{3}}{2}(1-\mathfrak{M}). \tag{148}$$

There is no packing scheme optimal for all p, but it is always possible (though inefficient) to use a hypercubic lattice, for which

$$\Delta V = \left(2\sqrt{\frac{1-\mathfrak{M}}{p}}\right)^p, \tag{149}$$

as a starting point. Once the span of a template is known, the total number of templates is straightforward to compute using

$$\mathcal{N} = \frac{\int \mathrm{d}^p\alpha\,\sqrt{\det\|g_{ij}\|}}{\Delta V}. \tag{150}$$

where p is, as before, the dimension of the parameter space.

Computational costs for binary inspiral search. The number of templates for binary inspiral searches has been computed for post-Newtonian signals and ground-based interferometers discussed earlier [183] (see Section 7.4). The number required at a high minimal match $\mathfrak{M} = 0.97$ is several hundred thousand (cf. Table 8), implying huge computational costs. For instance, to search for inspiral waves in real time (i.e. to search at the same rate as the rate at which the data are recorded) we would need a dedicated computer that could carry out 10^{10} floating point operations per second (i.e., 10 GFLOPS). The development of economical computers of such a speed is the prime concern to various data analysis groups at the moment.

Table 8. Number of templates required to search for inspiral waves from a binary consisting of stars of masses $m_1, m_2 \geqslant 0.2M_\odot$, with a minimal match of 0.97. The numbers are given for different post-Newtonian families of waveforms.

Interferometer	1PN	1.5PN	2PN
LIGO-I	2.5×10^5	5.3×10^5	4.7×10^5
VIRGO	1.4×10^7	1.4×10^7	1.3×10^7
GEO600	4.3×10^5	8.5×10^5	7.5×10^5

Computational costs of search for continuous waves. The search problem for continuous waves from spinning neutron stars is computationally the most expensive job in gravitational wave data analysis. Today, there is little hope that all-sky searches lasting for a year or more can be made. It is easy to see why this is the case. Firstly, the data should be collected continuously for months at a good sensitivity. No one has run interferometers for such long periods, and we do not yet know if this will be possible in the future. Secondly, though a neutron star emits a periodic signal in its rest frame (if we ignore the neutron star spin-down, which actually induces some modulation in the waveform), the detector does not see a periodic wave because of the Earth's acceleration relative to the source. The wave is both frequency- and amplitude-modulated. One can, fortunately, de-modulate these effects since the Earth's motion is known quite precisely, and hence recover the original periodic signal. But de-modulation requires the knowledge of the source's direction and frequency, which are unknown in a blind search. The angular

resolution one obtains in a year's integration is $\Delta\theta = \lambda/D$, where λ is the wave length of radiation and D is the baseline of the detector in a year's integration, namely 1 A.U. Thus, for $f = 100$ Hz we have $\Delta\theta = 10^{-5}$ rad, or about two arcsec. Now, assuming that the source may be in any of 4 arcsec^2 patches on the sky, we get the number of patches on the sky for which we will try out a de-modulation correction to be $4\pi/(\Delta\theta)^2 = 4\pi \times 10^{10}$. It is quite impossible to apply Doppler de-modulation to the detector output for each of these $\sim 10^{11}$ patches and compute as many Fourier transforms.

One, therefore, asks the question: what is, for a given computational power, the best possible search one can do? Is there any advantage in going from a one-step search to a two- or multi-step hierarchical search? What about directional searches? These are problems for which we have some answers; but a great deal of work is needed and is currently in progress, to improve and optimize search algorithms. In the following we will provide a summary of the current status.

The differential geometric formalism discussed above was used in Ref. [184] to compute the number of days of data acquisition during which a TFLOPS-class computer could analyse the data in real time (that is, analyse T-hour data in T hours) and carry out a blind (that is, unknown direction, frequency and spin-down rate) search. Unfortunately, the longest data series we can integrate in real time, for neutron stars with spin frequencies $f \leqslant 100$ Hz and spin-down rates less than 1000 years, is about 18 days. This yields a SNR lower by a factor of 5 as compared to a year's period of observing. On-line searches for neutron stars with $f \leqslant 500$ Hz (largest observed frequencies of millisecond pulsars) and spin-down rates of 40 years (shortest observed spin-down rates) can only be made for a data set lasting for 20 hours or less. If the source's position, but not its frequency, is known in advance, then one can carry out an in-real-time search, again with a TFLOPS-class computer, for the frequency of the source in a data set that is 3 months long. This is good news since there are many known pulsars and X-ray binary systems that are potential sources of radiation. In addition, obvious targeted-search locations are the center of the Galaxy and globular clusters.

There have been efforts [185, 186] to study the efficiency of a two-step hierarchical method for a blind search. Here the basic idea is to construct Fourier transforms of data sets of duration smaller than the period in which Doppler modulations will be important, to stack spectral densities obtained in this way, and to add them all up. This is an incoherent way of constructing the signal, since one adds spectral densities that have no phase information. Therefore, one gains in SNR less than what an optimal matched filtering method is able to achieve. However, this does not make matter since (i) the targeted SNR's are quite high ($\sim$10) and (ii) candidate events can always be followed-up using coherent integration methods. These methods afford an in-real-time all-sky blind search (i.e., a search in which no assumptions are made about the parameters of the source) for continuous GWs during a period of 4 months or less using a 20 GFLOPS computer. Detector groups are planning to build computers of this kind to aid in search for continuous GWs.

8.5 Covariance matrix and parameter estimation

After a detection is made, say, because a high SNR has been recorded that cannot be neglected, the next step in data analysis is the estimation of the parameters characterizing the event and the determination of the error bounds for the measured values. The first thing to note is that one can never be absolutely certain that a signal is present in a data train; one can only give confidence levels for its presence, which may be close to 100% at high values of the SNR. The confidence level is a measure of probability for the observation of an event (such as a large peak in the correlated output) to be generated by a GW signal as opposed to a random, non-GW process. The next thing to note is that, however high the SNR is, one cannot be absolutely certain about the true parameters of the signal: at best one can compute a range of values in which the true parameters of the signal are most likely to lie. The width of the range depends on the confidence level demanded, being larger for higher confidence levels.

In our search for a signal in the output of a detector we use a discrete, rather than a continuous, family of templates. Each template has a particular set of values of the parameters and the templates together span an interesting region of the parameter space. The spacing between templates in the parameter space should, in general, be quite small. A common estimate of the signal parameters is given by the parameters of the template that provides a maximum SNR. Such an estimate is called the 'maximum likelihood estimate', so named because the parameters of this template maximize what is called the *likelihood ratio* [169]. Maximum likelihood estimates are not always minimum uncertainty estimates, as particularly demonstrated for the case of binary inspiral signals [182]. *Bayesian* estimates, which take into account any prior knowledge that may be available about the distribution of the source parameters as well as the information from the output of the whole bank of templates rather than a single template, often give a much better estimate [187]. The reason for the better reliability of Bayesian estimates is that they make quantitative use of the information at hand.

In a measurement process any estimation of parameters, however efficient, robust and accurate, is unlikely to give the actual parameters of the signal since, at any finite SNR, the presence of noise alters the input signal. In geometric terms, the signal vector is altered by the noise vector, resulting in a vector that lies outside the signal manifold. Techniques such as matched filtering aim at computing the best projection of this altered vector onto the signal space. The true parameters of the signal are expected to lie within an ellipsoid of p dimensions at a certain level of confidence, the volume of the ellipsoid increasing with the level of confidence. The axes of the ellipsoid are the 1σ uncertainties in the estimation of parameters; the confidence level corresponding to a 1σ uncertainty is 0.67^p, the confidence level corresponding to a 2σ uncertainty is 0.95^p, and so on.

The topic of parameter estimation deserves a much wider discussion than given here. However, our goal here is only to assemble the necessary tools of the estimation theory for ready use. An interested reader can consult the ever-growing literature for further details. (See Ref. [169] for the estimation theory and Refs [182, 188, 189] for applications in GW observations.)

Covariance matrix. The scalar product (113) induces the following metric on the signal manifold [182]:

$$G_{ij} = \left\langle \frac{\partial a(\alpha)}{\partial \alpha^i}\,,\, \frac{\partial a(\alpha)}{\partial \alpha^j} \right\rangle, \tag{151}$$

where $a(\alpha)$ is a signal vector of unit norm and a^i are the coordinates. The metric G_{ij} is defined on the manifold of all

the parameters, including the lag t' and the initial phase ϕ_0 of the waveform. It is easy to show that G projected orthogonal to the initial phase ϕ_0 yields the metric γ in Eqn (145).

The metric (151) on the space of waveforms, introduced above, is also called the Fisher information matrix [169]. Indeed, it contains the 'information' about how similar or dissimilar are the waveforms in a small neighborhood of the parameter α of the signal. Large values of the metric imply that even small changes in the signal and template parameters can greatly affect their overlap, while the opposite is true when the metric coefficients are small.

The inverse of the information matrix is the *covariance matrix* C_{ij}, whose diagonal and off-diagonal elements are, in the limit of large SNR, the variances in the measured values of the parameters and coefficients of correlation between different parameters, respectively [169]:

$$C_{ij} = (G)^{-1}_{ij} . \tag{152}$$

Covariance-matrix-based errors in the estimates of the total mass, reduced mass, and the instant of coalescence $t_{\rm coal}$ have been computed by Poisson and Will [189] for the second-approximation post-Newtonian inspiral waveforms. They are listed in Table 9 for three typical binaries. The relative errors are smaller, for a given SNR, in the case of lighter binaries. This is because the signals of lighter binaries last longer and have larger numbers of cycles in the detector, making it relatively easier, as compared to higher mass binaries, to discriminate between waveforms of different parameters.

Table 9. Errors in the estimation of the instant of coalescence phase $\Delta t_{\rm coal}$, coalescence $\Delta\phi_{\rm coal}$, chirp mass $\Delta\mathcal{M}/\mathcal{M}$ and symmetric mass ratio $\Delta\eta/\eta$, and percentage biases in the estimation of the total mass $\mathcal{B}_m = 100(1 - m^T/m^X)$ and mass ratio $\mathcal{B}_\eta = 100(1 - \eta^T/\eta^X)$ (where T stands for the parameter of the template and X for the parameter of the exact waveform), for the second-post-Newtonian corrected inspiral waveform neglecting the effect of spins. Values are quoted for three typical binaries consisting of two $1.4M_\odot$ neutron stars (system NS + NS), a $1.4M_\odot$ neutron star and a $10M_\odot$ black hole (system NS + BH), and two $10M_\odot$ black holes (system BH + BH).

System	$\Delta t_{\rm coal}$, ms	$\Delta\phi_{\rm coal}$, rad	$\Delta\mathcal{M}/\mathcal{M} \times 10^{-3}$	$\Delta\eta/\eta$	$\mathcal{B}_m$	$\mathcal{B}_\eta$
NS+NS	1.07	2.94	0.36	0.28	0.214	0.211
NS+BH	1.72	2.27	2.20	0.50	−6.96	12.2
BH+BH	1.50	2.19	5.40	1.50	1.40	0.282

Biases in estimation. There are two ways in which an error can arise in the estimation of signal parameters. Firstly, an error in the measurement of a parameter occurs because of internal noise, which alters the input signal and hence, in the process of maximizing the SNR, we err in the estimation of the parameter by an amount that depends on the SNR. This type of error is a *random error* and normally goes down in inverse proportion to the SNR. In the limit of an infinitely large ensemble of measurements the estimated values converge to the true values of the signal parameters. Secondly, our estimator may be *biased* in the sense that the value averaged over the ensemble of measurements may converge to values different from the true values of the parameters. This can happen for one or both of the following two reasons. The search templates we use in our detection algorithms may be only an approximation to the true signal, as in the case of inspiral wave searches, where we use post-Newtonian search templates to look for a fully general-relativistic signal. As we will show below, such a search should induce a bias in the estimation of parameters, which we will refer to as a bias of the *first kind*. Alternatively, the estimator may be inherently biased and may give erroneous values even if the search templates are not faulty, which we will refer to as a bias of the *second kind*.

An example of a bias of the second kind is the following. In the maximum likelihood method we always aim at the largest value of the likelihood ratio. This results in overestimating the amplitude parameter of the signal, as shown in Ref. [182]. For inspiral signals, this means that the distance to the source is underestimated.

Biases in parameter estimates for inspiral signals have been discussed in Ref. [190]. Biases in the estimates of the total mass and the symmetric mass ratio in using the standard post-Newtonian approximation are given in Table 9 and discussed in detail in Ref. [172]. These studies demonstrated that, while using the standard post-Newtonian approximation to the waveform, the bias may be quite large, whereas P-approximants to inspiral waves greatly reduce the bias in the estimation of parameters.

9. Conclusion

With the aid of matched filtering, a network of first-generation interferometers, consisting of GEO, LIGO and VIRGO, should be able to survey a volume of 3×10^6 Mpc^3 for binary black holes at a minimum SNR of 3. For this volume conservative estimates of coalescence rates predict a few per year. Thus, binary black holes may turn out to be the first events recorded with our detectors.

The first-generation interferometers should also be able to detect primordial GW background at the level of $\Omega_{\rm gw} \sim 5 \times 10^{-6}$ by cross-correlating data from close interferometer pairs (LIGO – LIGO or GEO – VIRGO). However, this is a level insufficient from a theoretical point of view. New data analysis algorithms that exploit the specific properties of the primordial background could aid in detecting the background at a much lower level but, at the moment, we do not know how to achieve this.

The first interferometers will also be able to detect continuous gravitational waves from newly born or rapidly spinning non-spherical neutron stars provided that the amplitude is $h \geqslant 10^{-26}$ and the signal lasts for a few months. This, of course, assumes that we know the evolution of the phase of the emitted signals, which, for old neutron stars, is a simple sinusoid. However, in the case of waves emitted due to an instability we do not have right templates. This is also true of transient waves emitted during a supernova explosion or a merger of black holes. Thus, much work is needed to understand the sources of GWs. Indeed, it may very well be that the new generation of gravitational wave antennas will open up a new window for observing the Universe, which will aid in our understanding of strong and non-linear gravity.

We appreciate useful discussions with K Thorne, V Braginskiĭ, and B Schutz. This work was supported in part by a joint research grant from the Royal Society. Support from the Russian Foundation for Basic Research (grants 00-02-17884 and 00-02-17164) is also acknowledged.

10. Appendices

Here we summarize some basic formulae relevant to the description of massive binary evolution.

I. Keplerian binary system and radiation back reaction

Binary stars are among the main topics of the present paper, so it is necessary to remind the reader of some basic facts about Keplerian motion in a binary system. The stars are highly concentrated objects, so their treatment as point masses is usually adequate for the description of their interaction in the binary. Further, the Newtonian gravitation theory is sufficient for this purpose as long as the orbital velocities are small in comparison with the speed of light c. The systematic change of the orbit caused by the emission of GWs will be considered in a separate subsection below.

Keplerian motion. Let us consider two point masses M_1 and M_2 orbiting each other under the force of gravity. It is well known (see Ref. [191]) that this problem is equivalent to the problem of a single body with mass μ moving in an external gravitational potential. The value of the external potential is determined by the total mass M. The total mass M of the system is

$$M = M_1 + M_2 \,, \tag{A.1}$$

and the reduced mass μ is

$$\mu = \frac{M_1 M_2}{M} \,. \tag{A.2}$$

The body μ moves along an elliptic orbit with eccentricity e and major semiaxis a. The orbital period P and orbital frequency $\Omega = 2\pi/P$ are related to M and a by Kepler's third law

$$\Omega^2 = \left(\frac{2\pi}{P}\right)^2 = \frac{GM}{a^3} \,. \tag{A.3}$$

This relationship is true for any eccentricity e.

Individual bodies M_1 and M_2 move around the barycenter of the system in elliptic orbits with the same eccentricity e. The semimajor axes a_1, a_2 of the two ellipses are inversely proportional to the masses:

$$\frac{a_1}{a_2} = \frac{M_2}{M_1} \tag{A.4}$$

and satisfy the relationship $a = a_1 + a_2$. The position vectors of the bodies with respect to the system's barycenter are $\mathbf{r}_1 = M_2\,\mathbf{r}/(M_1 + M_2)$ and $\mathbf{r}_2 = -M_1\mathbf{r}/(M_1 + M_2)$, where $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ is the relative position vector. Therefore, the velocities of the bodies with respect to the system's barycenter are related by

$$-\frac{\mathbf{V}_1}{\mathbf{V}_2} = \frac{M_2}{M_1} \,, \tag{A.5}$$

and the relative velocity is $\mathbf{V} = \mathbf{V}_1 - \mathbf{V}_2$.

The total conserved energy of the binary system is

$$E = \frac{M_1\mathbf{V}_1^2}{2} + \frac{M_2\mathbf{V}_2^2}{2} - \frac{GM_1M_2}{r} = \frac{\mu\mathbf{V}^2}{2} - \frac{GM_1M_2}{r} = -\frac{GM_1M_2}{2a} \,, \tag{A.6}$$

where r is the distance between the bodies. The orbital angular momentum vector is perpendicular to the orbital plane and can be written as

$$\mathbf{J}_{\rm orb} = M_1\mathbf{V}_1 \times \mathbf{r}_1 + M_2\mathbf{V}_2 \times \mathbf{r}_2 = \mu\mathbf{V} \times \mathbf{r} \,. \tag{A.7}$$

The absolute value of the orbital angular momentum is

$$|\mathbf{J}_{\rm orb}| = \mu\sqrt{GMa(1-e^2)} \,. \tag{A.8}$$

For circular binaries (with $e = 0$) the distance between orbiting bodies does not depend on time,

$$r(t, e = 0) = a \,,$$

and is usually referred to as the orbital separation. In this case, the velocities of the bodies, as well as their relative velocity, are also time-independent:

$$V \equiv |\mathbf{V}| = \Omega a = \sqrt{\frac{GM}{a}} \,, \tag{A.9}$$

and the orbital angular momentum is

$$|\mathbf{J}_{\rm orb}| = \mu V a = \mu\Omega a^2 \,. \tag{A.10}$$

Gravitational radiation from a binary. The plane of the orbit is determined by the orbital angular momentum vector $\mathbf{J}_{\rm orb}$. The line of sight is specified by a unit vector $\mathbf{n}$. The binary inclination angle i is defined by the relation $\cos i = (\mathbf{n}, \mathbf{J}_{\rm orb}/J_{\rm orb})$ such that $i = 90^\circ$ corresponds to a system visible edge-on.

Let us start from two point masses M_1 and M_2 in a circular orbit. In a quadrupole approximation [1], two polarization amplitudes of GW at a distance r from the source are given by

$$h_+ = \frac{G^{5/3}}{c^4}\,\frac{1}{r}\,2(1+\cos^2 i)(\pi f M)^{2/3}\mu\cos(2\pi f t) \,, \tag{A.11}$$

$$h_\times = \pm\frac{G^{5/3}}{c^4}\,\frac{1}{r}\,4\cos i\,(\pi f M)^{2/3}\mu\sin(2\pi f t) \,. \tag{A.12}$$

Here M is the total mass, μ is the reduced mass, and $f = \Omega/\pi$ is the frequency of the emitted GW (twice the orbital frequency). Note that, for a fixed distance r and a given frequency f, the GW amplitudes are fully determined by $\mu M^{2/3} = \mathcal{M}^{5/3}$, where the combination

$$\mathcal{M} \equiv \mu^{3/5} M^{2/5}$$

is called the 'chirp mass' of the binary. After averaging over the orbital period (so that the squares of periodic functions are replaced by $1/2$) and over the orientations of the binary orbital plane, one arrives at the averaged (characteristic) GW amplitude

$$h(f, \mathcal{M}, r) = \left(\langle h_+^2\rangle + \langle h_\times^2\rangle\right)^{1/2} = \left(\frac{32}{5}\right)^{1/2}\frac{G^{5/3}}{c^4}\,\frac{\mathcal{M}^{5/3}}{r}\,(\pi f)^{2/3} \,. \tag{A.13}$$

Energy and angular momentum loss. In the approximation and under the choice of coordinates that we are working with, it is sufficient to use the Landau – Lifshitz gravitational pseudotensor [2] when calculating the gravitational wave's energy and flux. (This calculation technique can be justified with the help of a fully satisfactory gravitational energy – momentum tensor, which can be derived in the field-theoretical formulation of general relativity [192]). The energy $\mathrm{d}E$ carried by a GW along its direction of propagation

per area dA per time dt is given by

$$\frac{dE}{dA\,dt} \equiv F = \frac{c^3}{16\pi G}\left[\left(\frac{\partial h_+}{\partial t}\right)^2 + \left(\frac{\partial h_\times}{\partial t}\right)^2\right]. \quad (A.14)$$

The energy output dE/dt from a localized source in all directions is given by the integral

$$\frac{dE}{dt} = \int F(\theta,\phi)\, r^2\, d\Omega\,. \quad (A.15)$$

Replacing

$$\left(\frac{\partial h_+}{\partial t}\right)^2 + \left(\frac{\partial h_\times}{\partial t}\right)^2 = 4\pi^2 f^2 h^2(\theta,\phi)$$

and introducing

$$h^2 = \frac{1}{4\pi}\int h^2(\theta,\phi)\, d\Omega\,,$$

we write Eqn (A.15) in the form

$$\frac{dE}{dt} = \frac{c^3}{G}(\pi f)^2 h^2 r^2\,. \quad (A.16)$$

Specifically for a binary system in a circular orbit, one finds the energy *loss* from the system (as indicated by the minus sign) with the help of Eqns (A.16) and (A.13):

$$\frac{dE}{dt} = -\frac{32}{5}\frac{G^{7/3}}{c^5}(\mathcal{M}\pi f)^{10/3}\,. \quad (A.17)$$

This expression is exactly the same as can be obtained directly from the quadrupole formula [2]:

$$\frac{dE}{dt} = -\frac{32}{5}\frac{G^4}{c^5}\frac{M_1^2 M_2^2 M}{a^5} \quad (A.18)$$

and rewritten using the definition of the chirp mass and Kepler's law. Since energy and angular momentum are continuously carried away by gravitational radiation, the two masses in the orbit spiral toward each other, thus increasing their orbital frequency Ω. The GW frequency $f = \Omega/\pi$ and the GW amplitude h are also increasing functions of time. The rate of the frequency change [9] is

$$\dot{f} = \frac{96}{5}\frac{G^{5/3}}{c^5}\pi^{8/3}\mathcal{M}^{5/3} f^{11/3}\,. \quad (A.19)$$

In a spectral representation, the flux of energy per unit area per unit frequency interval is given by the right-hand side of the expression

$$\frac{dE}{dA\,df} = \frac{c^3}{G}\frac{\pi f^2}{2}\left(|\tilde{h}(f)_+|^2 + |\tilde{h}(f)_\times|^2\right) \equiv \frac{c^3}{G}\frac{\pi f^2}{2}\,S_h^2(f)\,, \quad (A.20)$$

where we have introduced the spectral density $S_h^2(f)$ of the gravitational wave field h. In the case of a binary system, the quantity S_h^2 is calculable from Eqns (A.11) and (A.12):

$$S_h^2 = \frac{G^{5/3}}{c^3}\frac{\pi}{12}\frac{\mathcal{M}^{5/3}}{r^2}\frac{1}{(\pi f)^{7/3}}\,. \quad (A.21)$$

Binary coalescence time. A binary system in a circular orbit loses energy according to Eqn (A.17). For orbits with a non-zero eccentricity e, the right-hand side of this formula should be multiplied by the factor

$$f(e) = \left(1 + \frac{73}{24}e^2 + \frac{37}{96}e^4\right)(1-e^2)^{-7/2}$$

(see Ref. [193]). The initial binary separation a_0 decreases and, assuming Eqn (A.18) to be always valid, it should vanish in the time

$$t_0 = \frac{c^5}{G^3}\frac{5a_0^4}{256M^2\mu} = \frac{5c^5}{256}\frac{(P_0/2\pi)^{8/3}}{(G\mathcal{M})^{5/3}} \approx (9.8\times 10^6\ \mathrm{yr})\left(\frac{P_0}{1\ \mathrm{h}}\right)^{8/3}\left(\frac{\mathcal{M}}{M_\odot}\right)^{-5/3}. \quad (A.22)$$

As we noted above, gravitational radiation from the binary depends on the chirp mass $\mathcal{M}$, which can also be written as $\mathcal{M} \equiv M\eta^{3/5}$, where η is the dimensionless ratio $\eta = \mu/M$. Since $\eta \leqslant 1/4$, one has $\mathcal{M} \lesssim 0.435M$. For example, for two NS with equal masses $M_1 = M_2 = 1.4M_\odot$ the chirp mass is $\mathcal{M} \approx 1.22M_\odot$. This explains the choice of normalization in Eqn (A.22).

The coalescence time for an initially eccentric orbit with $e_0 \neq 0$ and separation a_0 is shorter than the coalescence time for a circular orbit with the same initial separation a_0 [193]:

$$t_{\mathrm{coal}}(e_0) = t_0 f(e_0)\,, \quad (A.23)$$

where the correction factor $f(e_0)$ is

$$f(e_0) = \frac{48}{19}\frac{(1-e_0^2)^4}{e_0^{48/19}\left[1+(121/304)e_0^2\right]^{3480/2299}} \times \int_0^{e_0}\frac{\left[1+(121/304)e^2\right]^{1181/2299}}{(1-e^2)^{3/2}}\,e^{29/19}\,de\,. \quad (A.24)$$

To merge in a time interval shorter than the Hubble time ($t_{\mathrm{H}} \approx 10^{10}$ yr), the binary should have a small enough initial orbital period $P_0 \leqslant P_{\mathrm{cr}}(e_0,\mathcal{M})$ and, accordingly, a small enough initial semimajor axis $a_0 \leqslant a_{\mathrm{cr}}(e_0,\mathcal{M})$. These critical orbital periods and semimajor axes are shown as functions of the initial eccentricity e_0 in Figs 18 and 19, respectively. The lines are plotted for three typical sets of masses: two neutron stars with equal masses ($1.4M_\odot + 1.4M_\odot$), a black hole and a neutron star ($10M_\odot + 1.4M_\odot$), and two black holes with equal masses ($10M_\odot + 10M_\odot$). Note that, in order to get a significantly shorter coalescence time, the initial binary eccentricity should be $e_0 \gtrsim 0.6$.

II. Mass transfer modes and mass loss in binary systems

Gravitational wave emission is the sole factor responsible for the change of orbital parameters of a pair of compact (degenerate) stars. However, at the early stages of binary evolution, it is the mass transfer between the components and the loss of matter and its orbital momentum that play the dominant dynamical role. Strictly speaking, these processes

[9] A signal with such an increasing frequency is reminiscent of the chirp of a bird. This explains the origin of the term 'chirp mass' for the parameter $\mathcal{M}$, which fully determines the GW frequency and amplitude behavior.

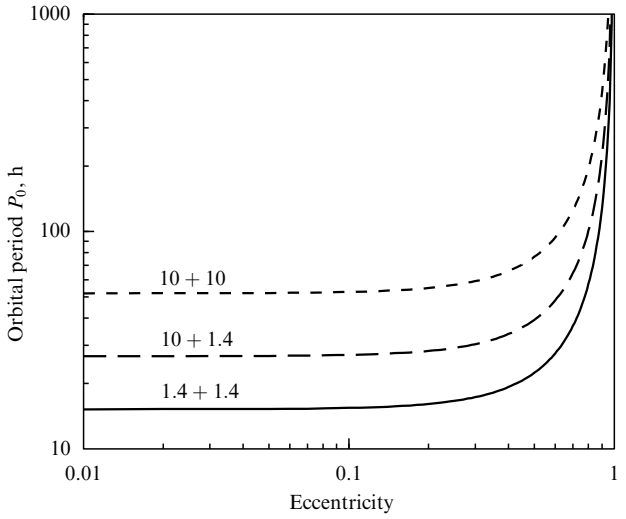

Figure 18. Initial orbital period P_0 (in hours) of two point masses which will coalesce due to GW emission in a time interval 10^{10} yr, as a function of the initial eccentricity e_0. The lines correspond to binaries with $10M_\odot + 10M_\odot$ (BH + BH), $10M_\odot + 1.4M_\odot$ (BH + NS), and $1.4M_\odot + 1.4M_\odot$ (NS + NS). A binary with a shorter (longer) orbital period will coalesce in a shorter (longer) time.

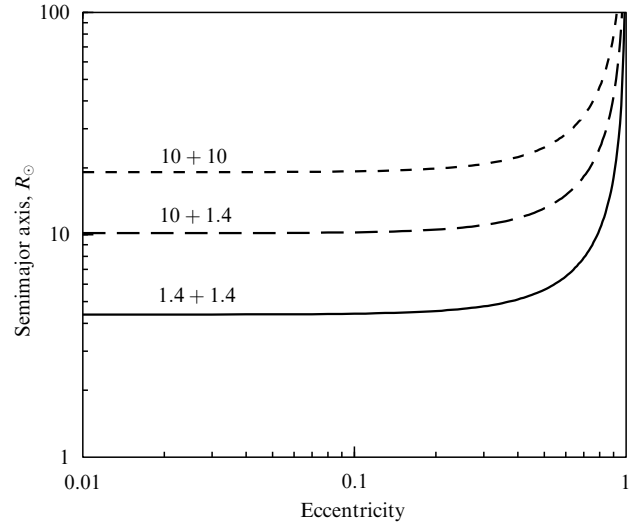

Figure 19. Initial semimajor axis a_0 (in $R_\odot$) of two point masses which will coalesce due to GW emission in a time interval 10^{10} yr, as a function of the initial eccentricity e_0. The lines correspond to binaries with $10M_\odot + 10M_\odot$ (BH + BH), $10M_\odot + 1.4M_\odot$ (BH + NS), and $1.4M_\odot + 1.4M_\odot$ (NS + NS). A binary with a smaller (larger) semimajor axis will coalesce in a shorter (longer) time.

should be treated hydrodynamically and require complicated numerical calculations. However, binary evolution can also be described semi-qualitatively, using a simplified description in terms of point-like bodies. The change of their integrated physical quantities, such as masses, orbital angular momentum, etc. governs the evolution of the orbit. This description turns out to be successful in reproducing the results of more rigorous numerical calculations (see e.g. Ref. [41] for a review). In this approach, the key role is attributed to the total orbital angular momentum $J_{\rm orb}$ of the binary.

Let star 2 lose matter at a rate $M_2 < 0$ and let β be the fraction of the ejected matter that leaves the system ($0 \leqslant \beta \leqslant 1$; the rest falls onto the first star), i.e. $\dot{M}_1 = -(1-\beta)\dot{M}_2 \geqslant 0$. Consider circular orbits with an orbital angular momentum given by (A.10). Differentiate both parts of Eqn (A.10) by time t and exclude $\mathrm{d}\Omega/\mathrm{d}t$ with the help of Kepler's third law (A.3). This gives us the rate of change of the orbital separation:

$$\frac{\dot{a}}{a} = -2\left[1 + (\beta - 1)\frac{M_2}{M_1} - \frac{\beta}{2}\frac{M_2}{M}\right]\frac{\dot{M}_2}{M_2} + 2\,\frac{\dot{J}_{\rm orb}}{J_{\rm orb}}\,. \qquad \text{(A.25)}$$

One defines the mass transfer as conservative if both $\beta = 0$ and $\dot{J}_{\rm orb} = 0$. The mass transfer is called non-conservative if at least one of these conditions is violated.

For massive binaries, which we are mostly interested in, it is important to distinguish some specific cases (modes) of mass transfer. They are: (1) the conservative accretion mode, (2) the non-conservative Jeans' mode (or fast wind mode), (3) non-conservative isotropic re-emission, (4) sudden mass loss from one of the components during supernova explosion, and (5) the common-envelope stage. Separately we consider orbit evolution due to (6) GW emission, which becomes the main factor for short-period compact binaries. For non-conservative modes, one can also introduce some subcases, such as a ring-like mode in which a circumbinary ring of expelled matter is formed (e.g. Ref. [194]). Here, we will not go into the details of such subcases.

1. Conservative accretion. In the case of conservative accretion, matter from M_2 is fully deposited onto M_1. The transfer process preserves the total mass $M = \text{const}$ ($\beta = 0$) and the orbital angular momentum $J_{\rm orb} = \text{const}$ of the system. It follows from Eqn (A.25) that

$$M_1 M_2 \sqrt{a} = \text{const}\,,$$

so that the initial and final binary separations are related as

$$\frac{a_{\rm f}}{a_{\rm i}} = \left(\frac{M_{1\rm i} M_{2\rm i}}{M_{1\rm f} M_{2\rm f}}\right)^2 . \qquad \text{(A.26)}$$

The well-known 'rule of thumb' for this case says that the orbit shrinks when the more massive component loses matter, and the orbit widens in the opposite situation. During such a mass exchange, the orbital separation passes through a minimum, if the masses become equal in the course of the mass transfer.

2. The Jeans (fast wind) mode. In this mode the ejected matter completely escapes from the system, that is, $\beta = 1$. The escape of matter can take place either in a spherically symmetric way or in the form of bipolar jets moving away from the system at high velocity. In both cases, matter carries away some amount of the total orbital momentum proportional to the orbital angular momentum $J_2 = (M_1/M)J_{\rm orb}$ of the star losing mass (we neglect the possible proper rotation of the star, see Ref. [195]). For the loss of orbital momentum it is reasonable to take

$$\dot{J}_{\rm orb} = \frac{\dot{M}_2}{M_2}\,J_2\,. \qquad \text{(A.27)}$$

In the case $\beta = 1$, Eqn (A.25) can be written as

$$\frac{(\Omega \dot{a^2})}{\Omega a^2} = \frac{\dot{J}_{\rm orb}}{J_{\rm orb}} - \frac{M_1 \dot{M}_2}{M M_2}\,. \qquad \text{(A.28)}$$

Then Eqn (A.28) in conjunction with Eqn (A.27) give $\Omega a^2 =$ const, that is, $\sqrt{GaM} =$ const. Thus, as a result of such a mass loss, the change in orbital separation is

$$\frac{a_\mathrm{f}}{a_\mathrm{i}} = \frac{M_\mathrm{i}}{M_\mathrm{f}} \,. \tag{A.29}$$

Since the total mass decreases, the orbit always widens.

3. Isotropic re-emission. The matter lost by star 2 can first accrete to star 1, and then, a fraction β of the accreted matter, can be expelled from the system. This happens when a massive star transfers matter to a compact star on a thermal timescale ($< 10^6$ yr). The accretion luminosity can exceed the Eddington luminosity limit, and the radiation pressure will then push the infalling matter away from the system, in a manner similar to the spectacular example of the SS 433 binary system. In this mode of mass transfer, the binary orbital momentum carried away by the expelled matter is determined by the orbital momentum of the accreting star M_1, rather than by the orbital momentum of the mass-losing star M_2. The orbital momentum loss can be written as

$$\dot{J}_\mathrm{orb} = \beta \, \frac{\dot{M}_2}{M_1} \, J_1 \,, \tag{A.30}$$

where $J_1 = (M_2/M) J_\mathrm{orb}$ is the orbital momentum of the star M_1. In the limiting case where all the mass attracted by M_1 is fully pushed away ($\beta = 1$) Eqn (A.30) simplifies to

$$\frac{\dot{J}_\mathrm{orb}}{J_\mathrm{orb}} = \frac{\dot{M}_2 M_2}{M_1 M} \,. \tag{A.31}$$

After the substitution of this formula into Eqn (A.25) and integration over time, one arrives at

$$\frac{a_\mathrm{f}}{a_\mathrm{i}} = \frac{M_\mathrm{i}}{M_\mathrm{f}} \left(\frac{M_{2\mathrm{i}}}{M_{2\mathrm{f}}}\right)^2 \exp\left(-2\,\frac{M_{2\mathrm{i}} - M_{2\mathrm{f}}}{M_1}\right) . \tag{A.32}$$

The exponential term makes this mode of mass transfer very sensitive to the components' mass ratio. If $M_1/M_2 \ll 1$, the separation a between the stars can decrease so greatly that the approximation of point masses becomes invalid. Tidal orbital (Darwin) instability can set in, and the compact star can start spiralling toward the companion star's center (the common envelope stage; see section 5 below).

4. Supernova explosion. Supernova explosion in a binary system occurs on a timescale much shorter than the orbital period, so the loss of mass is virtually instantaneous. This case can be treated analytically (e.g. Refs [79, 196, 197]). In general, the loss of matter and radiation is aspherical, so that the remnant of the supernova explosion (neutron star or black hole) acquires some recoil velocity called the kick velocity $\mathbf{w}$. In a binary, the kick velocity should be added to the orbital velocity of the pre-supernova star.

The usual treatment proceeds as follows. Let us consider a pre-SN binary with initial masses M_1 and M_2. The stars move in a circular orbit with orbital separation a_i and relative velocity $\mathbf{V}_\mathrm{i}$. The star M_1 explodes leaving a compact remnant of mass M_c. The total mass of the binary decreases by the amount $\Delta M = M_1 - M_\mathrm{c}$. The compact star acquires some kick velocity $\mathbf{w}$. Unless the binary is disrupted, it will end up in a new orbit with eccentricity e, major semiaxis a_f, and an angle θ between the orbital planes before and after the explosion. In general, the new barycenter will also receive some velocity, but we neglect this motion. Our goal is to evaluate the parameters a_f, e, and θ.

It is convenient to work in an instantaneous reference frame centered at M_2 right at the time of explosion. The x-axis is the line from M_2 to M_1, the y-axis points in the direction of $\mathbf{V}_\mathrm{i}$, and the z-axis is perpendicular to the orbital plane. In this frame, the pre-SN relative velocity is $\mathbf{V}_\mathrm{i} = (0, V_\mathrm{i}, 0)$, where $V_\mathrm{i} = [G(M_1 + M_2)/a_\mathrm{i}]^{1/2}$ [see Eqn (A.9)]. The initial total orbital momentum is $\mathbf{J}_\mathrm{i} = \mu_\mathrm{i} a_\mathrm{i} (0, 0, -V_\mathrm{i})$. The explosion is considered to be instantaneous. Right after the explosion, the position vector of the exploded star M_1 has not changed: $\mathbf{r} = (a_\mathrm{i}, 0, 0)$. However, other quantities have changed:

$$\mathbf{V}_\mathrm{f} = (w_x, V_\mathrm{i} + w_y, w_z)$$

and

$$\mathbf{J}_\mathrm{f} = \mu_\mathrm{f} a_\mathrm{i} (0, w_z, -(V_\mathrm{i} + w_y)) \,,$$

where $\mathbf{w} = (w_x, w_y, w_z)$ is the kick velocity and $\mu_\mathrm{f} = M_\mathrm{c} M_2/(M_\mathrm{c} + M_2)$ is the reduced mass of the system after the explosion. The parameters a_f and e are found from equating the total energy and the absolute value of orbital momentum at the initial circular orbit to their expressions in the resulting elliptical orbit [see Eqns (A.6), (A.8), and (A.10)]:

$$\mu_\mathrm{f} \frac{V_\mathrm{f}^2}{2} - \frac{G M_\mathrm{c} M_2}{a_\mathrm{i}} = -\frac{G M_\mathrm{c} M_2}{2 a_\mathrm{f}} \,, \tag{A.33}$$

$$\mu_\mathrm{f} a_\mathrm{i} \sqrt{w_z^2 + (V_\mathrm{i} + w_y)^2} = \mu_\mathrm{f} \sqrt{G(M_\mathrm{c} + M_2) a_\mathrm{f} (1 - e^2)} \,. \tag{A.34}$$

For the resulting a_f and e one finds

$$\frac{a_\mathrm{f}}{a_\mathrm{i}} = \left\{ 2 - \chi \left[\frac{w_x^2 + w_z^2 + (V_\mathrm{i} + w_y)^2}{V_\mathrm{i}^2} \right] \right\}^{-1} \tag{A.35}$$

and

$$1 - e^2 = \chi \, \frac{a_\mathrm{i}}{a_\mathrm{f}} \left[\frac{w_z^2 + (V_\mathrm{i} + w_y)^2}{V_\mathrm{i}^2} \right] , \tag{A.36}$$

where $\chi \equiv (M_1 + M_2)/(M_\mathrm{c} + M_2) \geqslant 1$. The angle θ is defined by

$$\cos\theta = \frac{\mathbf{J}_\mathrm{f} \cdot \mathbf{J}_\mathrm{i}}{|\mathbf{J}_\mathrm{f}| \, |\mathbf{J}_\mathrm{i}|} \,,$$

which results in

$$\cos\theta = \frac{V_\mathrm{i} + w_y}{\sqrt{w_z^2 + (V_\mathrm{i} + w_y)^2}} \,. \tag{A.37}$$

The condition of disruption of the binary system depends on the absolute value V_f of the final velocity, and on the parameter χ. The binary disrupts if its total energy determined by the left-hand side of Eqn (A.33) becomes non-negative or, equivalently, if its eccentricity determined by Eqn (A.36) becomes $\geqslant 1$. From either of these requirements one derives the condition of disruption:

$$\frac{V_\mathrm{f}}{V_\mathrm{i}} \geqslant \sqrt{\frac{2}{\chi}} \,. \tag{A.38}$$

The system remains bound if the opposite inequality is satisfied. Condition (A.38) can also be written in terms of

the escape (parabolic) velocity V_e determined by the requirement

$$\mu_f \frac{V_e^2}{2} - \frac{GM_cM_2}{a_i} = 0\,.$$

Since $\chi = M/(M-\Delta M)$ and $V_e^2 = 2G(M-\Delta M)/a_i = 2V_i^2/\chi$, one can write inequality (A.38) in the form

$$V_f \geqslant V_e\,. \tag{A.39}$$

The condition of disruption simplifies in the case of a spherically symmetric SN explosion, that is, when there is no kick velocity, $\mathbf{w} = 0$, and, therefore, $V_f = V_i$. In this case, condition (A.38) reads as $\chi \geqslant 2$, which is equivalent to $\Delta M \geqslant M/2$. Thus, the system unbinds if more than a half of mass of the binary is lost. In other words, the resulting eccentricity

$$e = \frac{M_1 - M_c}{M_c + M_2}\,, \tag{A.40}$$

following from (A.35), (A.36), and the condition $\mathbf{w} = 0$, becomes larger than 1, if $\Delta M > M/2$. So far, we have considered an originally circular orbit. If the pre-SN star moves in an originally eccentric orbit, the condition of disruption of the system under symmetric explosion reads:

$$\Delta M = M_1 - M_c > \frac{1}{2}\frac{r}{a_i}\,,$$

where r is the distance between the components at the moment of explosion.

5. Common envelope stage. This is a very important stage in binary evolution. A possibility of this stage was first suggested in Ref. [198]. Generally, it occurs in binary systems where the mass transfer from the mass-losing star is high, and the companion cannot accrete all the matter. The common envelope stage appears unavoidable on observational grounds. Evidence for a dramatic orbital angular momentum decrease at some preceding evolutionary stage follows from observations of certain types of close binary stars. They include cataclysmic variables, in which a white dwarf accretes matter from a small red dwarf main-sequence companion, planetary nebulae with double cores, low-mass X-ray binaries and X-ray transients (neutron stars and black holes accreting matter from low-mass main-sequence dwarfs). The radii of progenitors of compact stars in these binaries should typically be of 100 – 1000 solar radii, that is, much larger than the observed binary separations. This testifies to some dramatic reduction of the orbital momentum at earlier stages of evolution and eventual removal of the common envelope.

There is no exact criterion for the formation of a common envelope. However, a high mass overflow onto the compact star from the normal star is always expected when the normal star leaves the main sequence and develops a convective envelope. The critical mass ratio for an unstable Roche lobe overflow depends on the specifics of the stars, but is always close to 1. Another way for the formation of a common envelope is the direct penetration of the compact star into the dense outer layers of the companion. This can happen as a result of the Darwin tidal orbital instability in binaries [199, 200], or when a compact remnant of a supernova explosion with an appropriately directed kick velocity finds itself in an elliptic orbit whose minimum periastron distance $a_f(1-e)$ is smaller than the stellar radius of the companion.

A simplified treatment of the common envelope stage is usually done as follows [201]. The orbital evolution of the compact star M_c inside the envelope of the normal star M_1 is driven by the dynamical friction drag. This leads to a gradual spiral-in process of the compact star. The released orbital energy ΔE_{orb}, or a fraction of it, can become numerically equal to the binding energy E_{bind} of the envelope with the rest of the binary system. It is usually assumed that this equality provides a condition for expulsion of the common envelope. What remains of the normal star M_1 is its stellar core M_{sc}. The final orbital parameters can be derived from the condition $E_{bind} = \alpha_{CE}\Delta E_{orb}$, where α_{CE} is the efficiency parameters less or equal to one. This condition reads:

$$\frac{GM_1(M_1 - M_{sc})}{\lambda R_L} = \alpha_{CE}\left(\frac{GM_cM_{sc}}{2a_f} - \frac{GM_1M_c}{2a_i}\right), \tag{A.41}$$

where a_i and a_f are the initial and the final orbital separation, λ is a numerical coefficient of order 1, R_L is the Roche lobe radius of the normal star, approximated by [202]

$$\frac{R_L}{a_i} = \frac{0.49}{0.6 + q^{2/3}\ln(1+q^{-1/3})}\,, \tag{A.42}$$

and $q \equiv M_1/M_2$. From Eqn (A.41) one derives

$$\frac{a_f}{a_i} = \frac{M_{sc}}{M_1}\left(1 + \frac{2a_i}{\lambda\alpha_{CE}R_L}\frac{M_1 - M_{sc}}{M_c}\right)^{-1} \lesssim \frac{M_{sc}}{M_1}\frac{M_c}{\Delta M}\,, \tag{A.43}$$

where $\Delta M = M_1 - M_{sc}$. Recent studies (e.g. Refs [70, 203]) show that $\alpha_{CE}\lambda$ falls within the range from 0.5 to 2. The mass M_{sc} of the helium core of a massive star is (see Ref. [204])

$$M_{He} \approx 0.073\left(\frac{M_1}{M_\odot}\right)^{1.42}, \tag{A.44}$$

so the orbital separation during the common envelope stage may decrease by as much as a factor of 30 – 60.

III. Post-Newtonian expansions of gravitational wave flux and energy

The gravitational wave flux escaping a system of two compact stars in a quasi-circular orbit when their orbital frequency is f_{orb} (the dominant GW frequency being $f = 2f_{orb}$), is given by [205 – 208]

$$\mathcal{F}(v) = \frac{32\eta^2 v^{10}}{5}\left[1 - \left(\frac{1247}{336} + \frac{35\eta}{12}\right)v^2 + 4\pi v^3 - \left(\frac{44711}{9072} + \frac{9271\eta}{504} + \frac{65\eta^2}{18}\right)v^4 - \left(\frac{8191}{672} + \frac{535\eta}{24}\right)\pi v^5\right], \tag{A.45}$$

where $v = (\pi M f)^{1/3}$ is the relative velocity of the two stars and $\eta = M_1M_2/M^2$ is the symmetric mass ratio. The η-parameter reaches a maximum value of 1/4 when the two masses are equal. It characterizes the extent to which two-body effects in the system are important. The relativistic energy $E(v)$ of the system is given by

$$E(v) = -\frac{\eta v^2}{2}\left[1 - \left(\frac{9+\eta}{12}\right)v^2 - \left(\frac{81 - 57\eta + \eta^2}{24}\right)v^4\right]. \tag{A.46}$$

Solving the differential equations in Eqn (127) for time and phase in terms of v gives the following equations:

$$t(v) = -\frac{5m}{256\eta v^8}\left[1 + \left(\frac{743}{252} + \frac{11\eta}{3}\right)v^2 - \frac{32\pi}{5}v^3 + \left(\frac{3058673}{508032} + \frac{5429\eta}{504} + \frac{617\eta^2}{72}\right)v^4 - \left(\frac{7729}{252} + \eta\right)\pi v^5\right], \tag{A.47}$$

$$\phi(v) = -\frac{1}{16\eta v^5}\left[1 + \left(\frac{3715}{1008} + \frac{55\eta}{12}\right)v^2 - 10\pi v^3 + \left(\frac{15293365}{1016064} + \frac{27145\eta}{1008} + \frac{3085\eta^2}{144}\right)v^4 + \left(\frac{38645}{672} + \frac{15\eta}{8}\right)\pi \ln\left(\frac{v}{v_{\rm LSO}}\right)v^5\right]. \tag{A.48}$$

One can invert Eqn (A.47) to express v in terms of a post-Newtonian expansion in t and then use the resulting expression in Eqn (A.48) to arrive at an explicit phasing formula. Introducing the new time parameter θ defined by $\theta = [\eta(t_{\rm LSO} - t)/(5m)]^{-1/8}$, where $t_{\rm LSO}$ is the reference time taken to be the time at which the GW frequency is equal to twice the orbital frequency in the last stable circular orbit, we find

$$\phi(\theta) = -\frac{2}{\eta\theta^5}\left[1 + \left(\frac{3715}{8064} + \frac{55\eta}{96}\right)\theta^2 - \frac{3\pi}{4}\theta^3 + \left(\frac{9275495}{14450688} + \frac{284875\eta}{258048} + \frac{1855\eta^2}{2048}\right)\theta^4 + \left(\frac{38645}{21504} + \frac{15\eta}{256}\right)\pi \ln\left(\frac{\theta}{\theta_{\rm LSO}}\right)\theta^5\right]. \tag{A.49}$$

References

1. Weber J *General Relativity and Gravitational Waves* (New York: Interscience Publ., 1961)
2. Landau L D, Lifshitz E M *Teoriya Polya* (The Classical Theory of Fields) (Moscow: Nauka, 1973) [Translated into English (Oxford: Pergamon Press, 1975)]
3. Misner C W, Thorne K S, Wheeler J A *Gravitation* (San Francisco: W.H. Freeman and Co., 1973)
4. Thorne K S, in *Three Hundred Years of Gravitation* (Eds S W Hawking, W Israel) (Cambridge: Cambridge Univ. Press, 1987) p. 330
5. Thorne K S, in *Particle and Nuclear Astrophysics and Cosmology in the Next Millennium* (Eds E W Kolb, R D Peccei) (Singapore: World Scientific, 1995) p. 160
6. Schutz B F *Classical Quant. Grav.* **16** A131 (1999)
7. Braginskiĭ V B, Rudenko V N *Usp. Fiz. Nauk* **100** 395 (1970) [*Sov. Phys. Usp* **13** 165 (1970)]
8. Grishchuk L P *Usp. Fiz. Nauk* **121** 629 (1977) [*Sov. Phys. Usp.* **20** 319 (1977)]
9. Grishchuk L P *Usp. Fiz. Nauk* **156** 297 (1988) [*Sov. Phys. Usp.* **31** 940 (1988)]
10. Will C M *Usp. Fiz. Nauk* **164** 765 (1994) [*Phys. Usp.* **37** 697 (1994)]
11. King A R, in *Relativistic Gravitation and Gravitational Radiation*: *Proc. of the Les Houches School of Physics, 1995* (Eds J-A Marck, J-P Lasota) (Cambridge: Cambridge Univ. Press, 1997) p. 105
12. Taylor J H *Rev. Mod. Phys.* **66** 711 (1994)
13. Danzmann K et al., in *First Edoardo Amaldi Conf. on Gravitational Wave Experiments* (Edoardo Amaldi Foundation Ser., Vol. 1, Eds E Coccia, G Pizzella, F Ronga) (Singapore: World Scientific, 1995) p. 100
14. Bradaschia C et al. *Nucl. Instrum. Meth. A* **289** 518 (1990)
15. Abramovici A et al. *Science* **256** 325 (1992)
16. Bender P et al. *LISA*: *Pre-Phase a Report* 2nd ed. (Garching, 1998)
17. *LIGO II Conceptual Project Book* (LIGO-M990288-00-M, 1999)
18. Taylor J H, Manchester R N, Lyne A G *Astrophys. J. Suppl.* **88** 529 (1993)
19. D'Amico N et al., astro-ph/9911482
20. Camilo F et al., in *Pulsar Astronomy — 2000 and Beyond*: *IAU Colloq. 177* (ASP Conf. Ser., Vol. 202, Eds M Kramer, N Wex, R Wielebinski) (San Francisco, Calif.: Astron. Soc. Pacific, 2000) p. 3
21. Cherepashchuk A M *Usp. Fiz. Nauk* **166** 809 (1996) [*Phys. Usp.* **39** 759 (1996)]
22. Thorsett S E, Chakrabarty D *Astrophys. J.* **512** 288 (1999)
23. Nice D J, Sayer R W, Taylor J H *Astrophys. J. Lett.* **466** L87 (1996)
24. Phinney E S *Astrophys. J. Lett.* **380** L17 (1991)
25. Narayan R, Piran T, Shemi A *Astrophys. J. Lett.* **379** L17 (1991)
26. Curran S J, Lorimer D R *Mon. Not. R. Astron. Soc.* **276** 347 (1995)
27. Van den Heuvel E P J, Lorimer D R *Mon. Not. R. Astron. Soc.* **283** L37 (1996)
28. Bailes M, in *Compact Stars in Binaries*: *Proc. of the 165th Symp. of Inst. Astron. Union, The Netherlands, 1994* (Eds J Van Paradijs, E P J Van den Heuvel, E Kuulkers) (Dordrecht: Kluwer Acad. Publ., 1996) p. 213
29. Arzoumanian Z, Cordes J M, Wasserman I *Astrophys. J.* **520** 696 (1999)
30. Stairs I H et al. *Astrophys. J.* **505** 352 (1998)
31. Kalogera V, Lorimer D R *Astrophys. J.* **530** 890 (2000)
32. Clark J P A, Van den Heuvel E P J, Sutantyo W *Astron. Astrophys.* **72** 120 (1979)
33. Lipunov V M, Postnov K A, Prokhorov M E *Astron. Astrophys.* **176** L1 (1987)
34. Hils D, Bender P L, Webbink R F *Astrophys. J.* **360** 75 (1990)
35. Tutukov A V, Yungel'son L R *Astron. Zh.* **70** 812 (1993) [*Astron. Rep.* **37** 411 (1993)]
36. Lipunov V M et al. *Astron. Astrophys.* **298** 677 (1995)
37. Portegies Zwart S F, Spreeuw H N *Astron. Astrophys.* **312** 670 (1996)
38. Lipunov V M, Postnov K A, Prokhorov M E *Astrophys. Scape Phys. Rev.* **9** 1 (1996)
39. Portegies Zwart S F, Yungelson L R *Astron. Astrophys.* **332** 173 (1998)
40. Bethe H A, Brown G E *Astrophys. J.* **517** 318 (1999)
41. Van den Heuvel E P J, in *Interacting Binaries* (Saas-Fee Advanced Course 22 Lecture Notes, 1992, Eds H Nussbaumer, A Orr) (Berlin: Springer-Verlag, 1994) p. 263
42. Abt H A *Annu. Rev. Astron. Astrophys.* **21** 343 (1983)
43. Salpeter E E *Astrophys. J.* **121** 161 (1955)
44. Popova E I, Tutukov A V, Yungelson L R *Astrophys. Space Sci.* **88** 55 (1982)
45. Trimble V *Nature* **303** 137 (1983)
46. Lipunov V M, Postnov K A, Prokhorov M E *Astron. Astrophys.* **310** 489 (1996)
47. Kaspi V M et al. *Nature* **381** 584 (1996)
48. Prokhorov M E, Postnov K A *Pis'ma Astron. Zh.* **23** 503 (1997) [*Astron. Lett.* **23** 439 (1997)]
49. Karastergiou A et al., in *Pulsar Astronomy — 2000 and Beyond*: *IAU Colloq. 177* (ASP Conf. Ser., Vol. 202, Eds M Kramer, N Wex, R Wielebinski) (San Francisco, Calif.: Astron. Soc. Pacific, 2000) p. 125
50. Karastergiou A et al., astro-ph/0002099
51. Iben I, Jr., Tutukov A V *Astrophys. J.* **456** 738 (1996)
52. Portegies Zwart S F, Van den Heuvel E P J *New Astron.* **4** 355 (1999)
53. Van den Heuvel E P J, Van Paradijs J *Astrophys. J.* **483** 399 (1997)
54. Tauris T M, Van den Heuvel E P J, in *Pulsar Astronomy — 2000 and Beyond*: *IAU Colloq. 177* (ASP Conf. Ser., Vol. 202, Eds M Kramer, N Wex, R Wielebinski) (San Francisco, Calif.: Astron. Soc. Pacific, 2000) p. 595
55. Tauris T M, Van den Heuvel E P J, astro-ph/0001015
56. Lai D, astro-ph/9912522
57. Dorofeev O F, Rodionov V N, Ternov I M *Pis'ma Astron. Zh.* **11** 302 (1985) [*Sov. Astron. Lett.* **11** 123 (1985)]
58. Bisnovatyi-Kogan G S *Astron. Astrophys. Trans.* **3** 287 (1993)
59. Lyne A G, Lorimer D R *Nature* **369** 127 (1994)
60. Blaauw A, Ramachandran R J *Astron. Astrophys.* **19** 19 (1998)
61. Frail D A, Goss W M, Whiteoak J B Z *Astrophys. J.* **437** 781 (1994)
62. Lorimer D R (to be published)

63. Hansen B M S, Phinney E S *Mon. Not. R. Astron. Soc.* **291** 569 (1997)
64. Lipunov V M, Postnov K A, Prokhorov M E *Mon. Not. R. Astron. Soc.* **288** 245 (1997)
65. Belczynski K, Bulik T *Astron. Astrophys.* **346** 91 (1999)
66. Bagot P *Astron. Astrophys.* **322** 533 (1997)
67. Wettig T, Brown G E *New Astron.* **1** 17 (1996)
68. Terman J L, Taam R E, Savage C O *Mon. Not. R. Astron. Soc.* **281** 552 (1996)
69. Rasio F A, Livio M *Astrophys. J.* **471** 366 (1996)
70. Van den Heuvel E P J, Habets G M H J *Nature* **309** 598 (1984)
71. Woosley S E, Langer N, Weaver T A *Astrophys. J.* **448** 315 (1995)
72. Portegies Zwart S F, Verbunt F, Ergma E *Astron. Astrophys.* **321** 207 (1997)
73. Timmes F X, Woosley S E, Weaver T A *Astrophys. J.* **457** 834 (1996)
74. Fryer C L *Astrophys. J.* **522** 413 (1999)
75. Fryer C L, Kalogera V, astro-ph/9911312; submitted to *Astrophys. J.*
76. Nelemans G, Tauris T M, Van den Heuvel E P J *Astron. Astrophys.* **352** L87 (1999)
77. Kuranov A G, Postnov K A, Prokhorov M E *Astron. Rep.* (2001) (in press)
78. Schaller G et al. *Astron. Astrophys. Suppl.* **96** 269 (1992)
79. Yamaoka H, Shigeyama T, Nomoto K *Astron. Astrophys.* **267** 433 (1993)
80. Brandt N, Podsiadlowski P *Mon. Not. R. Astron. Soc.* **274** 461 (1995)
81. Lipunov V M, Postnov K A, Prokhorov M E *Pis'ma Astron. Zh.* **23** 563 (1997) [*Astron. Lett.* **23** 492 (1997)]
82. Tutukov A V, Yungelson L R *Mon. Not. R. Astron. Soc.* **260** 675 (1993)
83. Fukugita M, Hogan C J, Peebles P J E *Astrophys. J.* **503** 518 (1998)
84. Edmunds M G, in *The Low Surface Brightness Universe*: *IAU Colloquium 171* (ASP Conf. Ser., Vol. 170, Eds J I Davies, C Impey, S Phillipps) (San Francisco, Calif.: Astron. Soc. Pacific, 1999) p. 383
85. Olive K A *Nucl. Phys. B Proc. Suppl.* **80** 79 (2000)
86. Lipunov V M et al. *Astrophys. J.* **454** 593 (1995)
87. Vanbeveren D et al. *New Astron.* **3** 443 (1998)
88. Flanagan É É, Hughes S A *Phys. Rev. D* **57** 4535 (1998)
89. Postnov K A, Prokhorov M E, in *Gravitational Waves and Experimental Gravity* (Eds J Tran Thanh Van et al.) (Hanoi, Vietnam: World Publishers, 2000) p. 113; astro-ph/9903193
90. Kalogera V, astro-ph/9911417; *Astrophys. J.* **541** 319 (2000)
91. Müller E *Classical Quant. Grav.* **14** 1455 (1997)
92. Palomba C, astro-ph/0003321; submitted to the *Int. J. Mod. Phys. D*
93. Burrows A, Hayes J C *Am. Astron. Soc. Meeting* **187** 1704 (1995)
94. Nazin S N, Postnov K A *Astron. Astrophys.* **317** L79 (1997)
95. Braginskiĭ V B, Grishchuk L P *Zh. Eksp. Teor. Fiz.* **89** 744 (1985) [*Sov. Phys. JETP* **62** 427 (1985)]
96. Braginskii V B, Thorne K S *Nature* **327** 123 (1987)
97. Bonnell I A, Pringle J E *Mon. Not. R. Astron. Soc.* **273** L12 (1995)
98. Schutz B F, gr-qc/9802020
99. Burrows A, Hayes J *Phys. Rev. Lett.* **76** 352 (1996)
100. Müller E, Janka H-T *Astron. Astrophys.* **317** 140 (1997)
101. Buonanno A, Damour T, gr-qc/0001013; *Phys. Rev. D* **62** 064015 (2000)
102. Alcock C et al. *Astrophys. J.* **486** 697 (1997)
103. Nakamura T et al. *Astrophys. J. Lett.* **487** L139 (1997)
104. Echeverria F *Phys. Rev. D* **40** 3194 (1989)
105. Chandrasekhar S *Phys. Rev. Lett.* **24** 611 (1970)
106. Friedman J L, Schutz B F *Astrophys. J.* **222** 281 (1978)
107. Andersson N *Astrophys. J.* **502** 708 (1998)
108. Andersson N, Kokkotas K, Schutz B F *Astrophys. J.* **510** 846 (1999)
109. Owen B J et al. *Phys. Rev. D* **58** 14834 (1998)
110. Bildsten L *Astrophys. J. Lett.* **501** L89 (1998)
111. Kulkarni S R et al., in *X-ray Binaries and Recycled Pulsars* (NATO ASI Series, Ser. C, Vol. 377, Eds E P J van den Heuvel, S A Rappaport) (Dordrecht: Kluwer Acad. Publ., 1992) p. 99
112. Caron B et al. *Nucl. Phys. B. Proc. Suppl.* **54** 167 (1997)
113. Giampieri G, Polnarev A G *Mon. Not. R. Astron. Soc.* **291** 149 (1997)
114. Lipunov V M, Postnov K A *Astron. Zh.* **64** 438 (1987) [*Sov. Astron.* **31** 228 (1987)]
115. Schneider R et al., astro-ph/0002055; submitted to *Mon. Not. R. Astron. Soc.*
116. Marsh T R, Dhillon V S, Duck S R *Mon. Not. R. Astron. Soc.* **275** 828 (1995)
117. Iben J, Tutukov A V *Astrophys. J. Suppl.* **54** 335 (1984)
118. Branch D et al. *Publ. Astron. Soc. Pac.* **107** 1019 (1995)
119. Tammann G A, Löffler W, Schröder A *Astrophys. J. Suppl.* **92** 487 (1994)
120. Van den Bergh S, McClure R D *Astrophys. J.* **425** 205 (1994)
121. Postnov K A, Prokhorov M E *Astrophys. J.* **494** 674 (1998)
122. Kosenko D I, Postnov K A *Astron. Astrophys.* **336** 786 (1998)
123. Kosenko D I, Postnov K A *Astron. Astrophys.* **355** 1209 (2000)
124. Grishchuk L P *Zh. Eksp. Teor. Fiz.* **67** 825 (1974) [*Sov. Phys. JETP* **40** 413 (1975)]
125. Grishchuk L P *Ann. N.Y. Acad. Sci.* **302** 439 (1977)
126. Knight P L, in *Quantum Fluctuations*: *Les Houches, Session LXIII, 1995* (Eds S Reynaud, E Giacobino, J Zinn-Justin) (Amsterdam: Elsevier, 1997) p. 5
127. Grishchuk L P, Sidorov Yu V *Classical Quant. Grav.* **6** L161 (1989)
128. Grishchuk L P, Sidorov Y V *Phys. Rev. D* **42** 3413 (1990)
129. Grishchuk L P, in *Workshop on Squeezed States and Uncertainty Relations* (NASA Conf. Ser., No. 3135) (New York: NASA Conf. Publ., 1992) p. 329
130. Grishchuk L P, in *Quantum Fluctuations*: *Les Houches, Session LXIII, 1995* (Eds S Reynaud, E Giacobino, J Zinn-Justin) (Amsterdam: Elsevier, 1997) p. 541
131. Schleich W, Wheeler J A *J. Opt. Soc. Am. B* **4** 1715 (1987)
132. Schleich W, Horowicz R J, Varro S *Phys. Rev. A* **40** 7405 (1989)
133. Zeldovich Y B, Novikov I D *Relativistic Astrophysics. The Structure and Evolution of the Universe* Vol. 2 (Chicago: Univ. of Chicago Press, 1983)
134. Giovannini M *Phys. Rev. D* **58** 14630 (1998)
135. Grishchuk L P, gr-qc/9810055
136. Veneziano G *Phys. Lett. B* **265** 287 (1991)
137. Gasperini M, Veneziano G *Astropart. Phys.* **1** 317 (1993)
138. Gasperini M, Giovannini M *Phys. Rev. D* **47** 1519 (1993)
139. Creighton T, gr-qc/9907045
140. Smoot G F et al. *Astrophys. J. Lett.* **396** L1 (1992)
141. Bennett C L et al. *Astrophys. J. Lett.* **464** L1 (1996)
142. Melchiorri A et al. *Astrophys. J.* **518** 562 (1999)
143. Grishchuk L P *Phys. Rev. D* **50** 7154 (1994)
144. Grishchuk L P, in *Current Topics in Astrofundamental Physics*: *Primordial Cosmology* (NATO ASI Series, Ser. C, Vol. 511, Eds N Sanchez, A Zichichi) (Dordrecht: Kluwer Acad. Publ., 1998) p. 539
145. Grishchuk L P, gr-qc/9801011
146. Grishchuk L P *Classical Quant. Grav.* **14** 1445 (1997)
147. Tegmark M, Zaldarriaga M *Astrophys. J.* **544** 30 (2000)
148. Kolb E W, Turner M *The Early Universe* (Reading, Mass.: Addison-Wesley, 1990)
149. Martin J, Schwarz D, astro-ph/9911225; *Phys. Rev. D* **62** 103520 (2000)
150. Hawking S W, Hertog T, Reall H S *Phys. Rev. D* **62** 901 (2000)
151. Hough J et al. *GEO600 Proposal*, 1994
152. Larson S L, Hiscock W A, Hellings R W *Phys. Rev. D* **62** 231 (2000)
153. *First Edoardo Amaldi Conf. on Gravitational Wave Experiments* (Edoardo Amaldi Foundation Ser., Vol. 1, Eds E Coccia, G Pizzella, F Ronga) (Singapore: World Scientific, 1995)
154. *Gravitational Waves*: *Sources and Detection* (Eds I Giufolini, F Fidecaro) (Singapore: World Scientific, 1997)
155. *Laser Interferometer Space Antenna* (AIP Conf. Proc., Vol. 456, Ed. W M Folkner) (Woodbury, N.Y.: AIP, 1998)
156. Grishchuk L P *Pis'ma Zh. Eksp. Teor. Fiz.* **23** 326 (1976) [*JETP Lett.* **23** 293 (1976)]
157. Michelson P F *Mon. Not. R. Astron. Soc.* **227** 933 (1987)
158. Christensen N *Phys. Rev. D* **46** 5250 (1992)
159. Flanagan É É *Phys. Rev. D* **48** 2389 (1993)
160. Allen B, in *Relativistic Gravitation and Gravitational Radiation*: *Proc. of the Les Houches School of Physics, 1995* (Eds J-A Marck, J-P Lasota) (Cambridge: Cambridge Univ. Press, 1997) p. 373
161. Allen B, Flanagan É É, Papa M A *Phys. Rev. D* **61** 024024 (2000)

162. Tsubono K, in *First Edoardo Amaldi Conf. on Gravitational Wave Experiments* (Edoardo Amaldi Foundation Ser., Vol. 1, Eds E Coccia, G Pizzella, F Ronga) (Singapore: World Scientific, 1995) p. 112
163. Lück H et al. *Classical Quant. Grav.* **14** 1471 (1997)
164. Caron B et al. *Classical Quant. Grav.* **14** 1461 (1997)
165. Ageev A Yu et al. *Phys. Lett. A* **227** 159 (1997)
166. Ageev A Yu, Bilenko I A, Braginsky V B *Phys. Lett. A* **246** 479 (1998)
167. Meers B, MacDonald N *Phys. Rev. A* **40** 3754 (1989)
168. Dhurandhar S V, Tinto M *Mon. Not. R. Astron. Soc.* **234** 663 (1988)
169. Helstrom C W *Statistical Theory of Signal Detection* 2nd ed. (Oxford: Pergamon Press, 1968)
170. Schutz B F, in *The Detection of Gravitational Waves* (Ed. D G Blair) (Cambridge: Cambridge Univ. Press, 1991) p. 406
171. Cutler C et al. *Phys. Rev. Lett.* **70** 2984 (1993)
172. Damour T, Iyer B R, Sathyaprakash B S *Phys. Rev. D* **57** 885 (1998)
173. Damour T, Iyer B R, Sathyaprakash B S (unpublished)
174. Allen B et al. ''GRASP software package'' (http://www.lsc-group.phys.uwm.edu/~ballen/grasp-distribution/)
175. Poisson E *Phys. Rev. D* **52** 5719 (1995)
176. Damour T, Iyer B R, Sathyaprakash B S (to be published)
177. Sathyaprakash B S, Dhurandhar S V *Phys. Rev. D* **44** 3819 (1991)
178. Schutz B F *Nature* **323** 310 (1986)
179. Blanchet L, Sathyaprakash B S *Classical Quant. Grav.* **11** 2807 (1994)
180. Blanchet L, Sathyaprakash B S *Phys. Rev. Lett.* **74** 1067 (1995)
181. Owen B J *Phys. Rev. D* **53** 6749 (1996)
182. Balasubramanian R, Sathyaprakash B S, Dhurandhar S V *Phys. Rev. D* **53** 3033 (1996)
183. Owen B J, Sathyaprakash B S *Phys. Rev. D* **60** 022002 (1999)
184. Brady P R et al. *Phys. Rev. D* **57** 2101 (1998)
185. Brady P R, Creighton T (to be published)
186. Schutz B F, Papa M A, in *Gravitational Waves and Experimental Gravity* (Eds J Tran Thanh Van et al.) (Hanoi, Vietnam: World Publishers, 2000) p. 199; gr-qc/9905018
187. Nicholson D, Vecchio A *Phys. Rev. D* **57** 4588 (1998)
188. Cutler C, Flanagan É É *Phys. Rev. D* **49** 2658 (1994)
189. Poisson E, Will C M *Phys. Rev. D* **52** 848 (1995)
190. Sathyaprakash B S, in *Relativistic Gravitation and Gravitational Radiation*: *Proc. of the Les Houches School of Physics, 1995* (Eds J-A Marck, J-P Lasota) (Cambridge: Cambridge Univ. Press, 1997) p. 361
191. Landau L D, Lifshitz E M *Mekhanika* (Mechanics) (Moscow: Nauka, 1972) [Translated into English (Oxford: Pergamon Press, 1969)]
192. Babak S V, Grishchuk L P *Phys. Rev. D* **61** 024038 (1999)
193. Peters P C *Phys. Rev. B* **136** 1224 (1964)
194. Soberman G E, Phinney E S, Van den Heuvel E P J *Astron. Astrophys.* **327** 620 (1997)
195. Van den Heuvel E P J, in *Formation and Evolution of X-ray Binaries* (Cambridge: Cambridge Univ. Press, 1983) p. 303
196. Blaauw A *Bull. Astron. Inst. Netherlands* **15** 265 (1961)
197. Flannery B P, Van den Heuvel E P J *Astron. Astrophys.* **39** 61 (1975)
198. Paczynski B *IAU Symp.* **73** 75 (1976)
199. Counselman C C *Astrophys. J.* **180** 307 (1973)
200. Bagot P *Astron. Astrophys.* **314** 576 (1996)
201. Webbink R F *Astrophys. J.* **277** 355 (1984)
202. Eggleton P P *Astrophys. J.* **268** 368 (1983)
203. Van den Heuvel E P J *Astron. Astrophys.* **291** L39 (1994)
204. Iben J, Tutukov A V *Astrophys. J. Suppl.* **58** 661 (1985)
205. Blanchet L et al. *Phys. Rev. Lett.* **74** 3515 (1995)
206. Blanchet L, Damour T, Iyer B R *Phys. Rev. D* **51** 5360 (1995)
207. Will C M, Wiseman A G *Phys. Rev. D* **54** 4813 (1996)
208. Blanchet L et al. *Classical Quant. Grav.* **13** 575 (1996)