Lecture Notes 1

Fundamental Forces Relative Strength

Two out of the four fundamental forces are long range: gravity and electromagnetic force. Assume you have two *protons* of mass "m" and charge "e" who interact at a distance "r". Then (in cgs units):

$$F_g = \frac{Gm^2}{r^2} \qquad F_{em} = \frac{e^2}{r^2} \tag{1}$$

Therefore the ratio is:

$$\frac{F_{em}}{F_g} = \frac{e^2}{Gm^2} = \left(\frac{e^2}{\hbar\, c}\right) / \left(\frac{Gm^2}{\hbar\, c}\right) = \frac{\alpha}{\alpha_G} \approx 10^3 6 \tag{2}$$

Here $\alpha = \frac{e^2}{\hbar c} = \frac{1}{137}$ is the fine structure constant and $\alpha_G \approx 6 \times 10^{-39}$ is the gravitational coupling constant.
Conclusion: The e.m. force is dominant in atoms and nuclei, and gravity has basically no effect.
But when does gravity become important ?

Take now a body of mass M and radius R, composed by N atoms of mass A and atomic number Z:

$$M = N\, A\, m_u \approx N\, A\, m_p \tag{3}$$

$$R \simeq N^{1/3} \frac{a_0}{Z} \tag{4}$$

where a_0 is the Bohr radius of an atom (i.e., it is the "characteristic radius" of the electron cloud).
Therefore:

$$E_G = \frac{-GM^2}{R} = -N^2 \frac{GA^2 m^2}{R} \simeq -N^{5/3} A^2 Z \frac{Gm_p^2}{a_0} = -N^{5/3} \alpha_G \alpha\, m_e c^2 A^2 Z \tag{5}$$

The internal energy of an H atom is:

$$E_i = \frac{e^2}{a_0} = \alpha^2 m_e c^2 \tag{6}$$

And for N atoms with charge Z it becomes:

$$E_i = N\alpha^2 m_e c^2 Z^3 \tag{7}$$

The total energy is therefore:

$$E_{TOT} = E_G + E_i = \left(N^{-5/3} \alpha_G \alpha\, A^2 Z + N\alpha^2 Z^3\right) m_e c^2 \tag{8}$$

Taking $A = 2Z$ and $E_{TOT} = 0$ we get:

$$4N^{5/3} \alpha_G \alpha\, Z^3 = N\alpha^2 Z^3 \rightarrow N \propto \left(\frac{\alpha}{\alpha_G}\right)^{3/2} \simeq 10^{54} \tag{9}$$

This implies that the minimum mass required to have a predominance of gravity over the e.m. force is:

$$M_{min} \simeq 10^{54} \cdot m_p \approx 10^{30} g \sim 1\, M_{Jupiter} \sim 0.001 M_{\odot} \tag{10}$$

$$R_{min} \simeq 10^{54/3} \cdot a_0 \approx 10^{10} cm \sim 10^5 km \sim 1\, R_{Jupiter} \sim 0.1 R_{\odot} \tag{11}$$

Conclusion: objects with $M < M_{min}$ are not mainly stabilized by the gravitational force, but by solid state forces. Objects like White Dwarfs (henceforth referred to as WD) where the two most important forces are e.m. and gravitational, are stabilized by gravity.
Suppose now you have an object in which gravity competes with the *strong force* instead of the e.m. ones. What is the mass required for gravity to dominate ? We can use the identical argument outlined above, with the difference that we substitute the fine structure constant α with the strong force coupling constant α_s. Therefore:

$$N \propto \left(\frac{\alpha_s}{\alpha}\right)^{3/2} \left(\frac{\alpha}{\alpha_G}\right)^{3/2} \simeq 10^3 \cdot 10^{54} \simeq 10^{57} \tag{12}$$

from which it follows immediately that:

$$M_{min} \simeq 10^{33}\, g \simeq 1\, M_{\odot} \tag{13}$$

$$R_{min} \simeq 10^6\, cm \simeq 10\, Km \tag{14}$$

These values are those typical of Neutron Stars (henceforth referred to as NS), where the strong force is much more important than the e.m. force when competing with gravity.

Degeneracy Pressure of Fermions

In a non-degenerate star like the Sun, the gravitational force is balanced by thermal pressure and radiation pressure generated by an internal source of energy (the nuclear fusion). In WD and NS there is (almost) no internal source of energy. So how is the gravity balanced to avoid the collapse of the object ?
The gravitational force is balanced by the onset of degeneracy, which is a consequence of the Pauli exclusion principle. In WD, the degeneracy pressure is provided by (mainly relativistic) electrons, whereas in NS, the non-relativistic neutrons provide the degeneracy pressure.

Suppose now to have a fermion of momentum p_F, which is confined in a box of volume d^3. A consequence of the Heisenberg uncertainty principle is that:

$$p_f \cdot d \approx \hbar \tag{15}$$

If we take the non-relativistic Fermi energy, this can be written as:

$$\epsilon_F = \frac{p_F^2}{2m} \approx \frac{\hbar}{m \cdot d^2} \tag{16}$$

Conclusion: For a given distance "d", it is the lightest particle that has the largest Fermi energy, and that contributes the most to the pressure.
This is why for example in WDs the electrons provide most of the pressure, whereas the other

particles have basically no effect on the pressure. But then why are NS supported by the degeneracy pressure of neutrons ?
This is because NS are much more compact than WDs, and therefore the electrically neutral neutrons get much more closely packed than the electrons in a WD, and therefore their pressure overcomes by several orders of magnitude the electron degeneracy pressure.

Fermi-Dirac Distribution

In this section we will derive the Fermi-Dirac distribution (FD distribution), essential to understand the behaviour of degenerate matter as we find it in white dwarfs or neutron stars. At a fundamental level, all particles obey the law of quantum mechanics. When the number of (non-relativistic) fermions is N, the state of the system and its evolution in time is given by N Schroedinger equations (or Dirac equations in case of relativistic fermions). However, when the number N is very large, the description of particles with wave functions becomes unpractical and one introduces a so-called *distribution function* $f(x, p, t)$. Here t is time, x and p are the 3D positions and momenta of the particles. The distribution function $f(x, v, t)$ gives the particle number density in a 6D space (called the "phase space", i.e., the space of coordinate x and momentum p) and describes also the evolution in time of the system. In other words, if you have a very large number of particles (like for example those in a white dwarf), you do not follow the evolution of each single particle, but you define how many particles you find with a specific position and momentum. This quantity is given by $f(x, v, t)$.

If your system is an ideal gas (i.e., no interactions between particles) at equilibrium, the distribution function does not depend on time. For fermions (indistinguishable particles), this distribution function, called Fermi-Dirac distribution, has the form:

$$f(E) = \frac{1}{1 + exp(\alpha + E_i\beta)} \tag{17}$$

(note that E is related to the momentum p via the relativistic relation $E = \sqrt{p^2c^2 + m^2c^4}$). Here α and β are parameters related with the temperature and with the **Fermi energy** of the particles in the system (see the Syllabus for the correct definition of this energy):

$$\alpha = -\frac{E_F}{k_B T} \tag{18}$$

$$\beta = \frac{1}{k_B T} \tag{19}$$

Strictly speaking, the proportionality of α with the Fermi energy that appears above is correct **ONLY** when the temperature T of the system is zero. When $T > 0$ the Fermi energy has to be substituted with the *chemical potential* defined as:

$$\mu = \frac{\partial \epsilon}{\partial n} \tag{20}$$

where ϵ is the energy density of fermions (including rest mass energy) and n is the number density of fermions.When $T = 0$ it is immediately clear then why the chemical potential is

identical to the Fermi energy: to introduce a new particle in the system the particle must have at least an energy equal to the Fermi energy. However, even if WDs and NS are very hot objects (young NSs can reach temperatures of $\sim 10^9 K$), the Fermi energy of their particles is always $E_F \gg kT$, which means that the thermal energy is negligible when compared to the Fermi energy and therefore the system behaves *as if* it is a cold object close to absolute zero temperature. This justifies the presence of E_F in the FD distribution.

How do we derive this distribution function ? We need to make three simple hypotheses:

1. Fermions of the same type (e.g., electrons) are indistinguishable
2. Fermions obey the Pauli exclusion principle
3. the number of particles N and the total energy of the system E_T are fixed and all the many possible arrangements for the particles within the allowed state are equiprobable

Definition: f(E) is the probability that a state at energy level E is occupied

Since $f(E)$ is a probability that a state is occupied, we can start with a simple example. Let's assume that we have two balls of different colours (e.g., (R)ed and (Y)ellow) and two empty boxes. Let's assume that the probability of finding a ball in a box is uniform (hypothesis 3). The possible combinations of the balls in the two boxes are 4:
1. R+Y in the 1st box and the 2nd empty
2. 1st box empty and R+Y in the 2nd box
3. 1st box Y and 2nd box R
4. 1st box R and 2nd box Y

Now, let's assume our two balls have the same colour (hypothesis 1: fermions are indistinguishable): the total number of possible combinations is 3:
1. R+R in the 1st box, 2nd box empty
2. 1st box empty, R+R in the 2nd box
3. 1st box R, 2nd box R

Now, let's add the hypothesis 2 (Pauli exclusion principle: maximum one ball in each box): the total number of combinations is 1.
One can demonstrate that if we have a number M_i of boxes, and N_i of particles, the number of ways these particles can be distributed in the boxes under the hypotheses 1, 2 and 3 is:

$$W_i = \frac{M_i!}{N_i!(M_i - N_i)!} \tag{21}$$

for each energy level E_i.
Think at each box M_i as sub-levels of the energy E_i. For example, since fermions obey the Pauli principle, then each energy level has two sub-levels, one for a spin up fermion of energy E_i, and one for the spin down with the same energy.
Given $i = 1, 2..., S$ energy levels, the total probability is:

$$W = \Pi_i \, W_i \tag{22}$$

From basic thermodynamics, we now that at thermal equilibrium the entropy is maximized and it is related to the probability W via:

$$S = k_B \ln(W) \tag{23}$$

It can be shown (for example by using the Lagrange multipliers method) that when S is maximized we have:

$$\frac{N_i}{M_i} = f(E) = \frac{1}{1 + exp(\alpha + E_i\beta)} \tag{24}$$

which is our Fermi-Dirac distribution for an ideal gas of fermions at thermal equilibrium.

Lecture Notes 2

1 Stability of an Equilibrium Configuration

Solutions to the TOV equations correspond to stellar configurations that are in hydrostatic equilibrium. **However, equilibrium does not assure stability.** Equilibrium configurations may correspond either to a maximum or to a minimum in the energy with respect to radial compression or dilation. For a certain range of baryon number N that compose a star, there are at least two equilibrium solutions, one at higher density and energy than the other. How to determine therefore whether an equilibrium configuration is stable ?

For a given equation of state, there is a unique relationship between the stellar mass and the central energy density $\epsilon(0)$ (or, equivalently, the central density since $\epsilon = \rho \cdot c^2$), because the TOV equation is of first order. Thus, for each possible EoS, there is a unique *sequence* of stars parametrized by the central density or, equivalently, the central pressure. For a broad range of central densities the solution of the TOV equations gives a well established unique mass-to-density relation (see Figure 1).

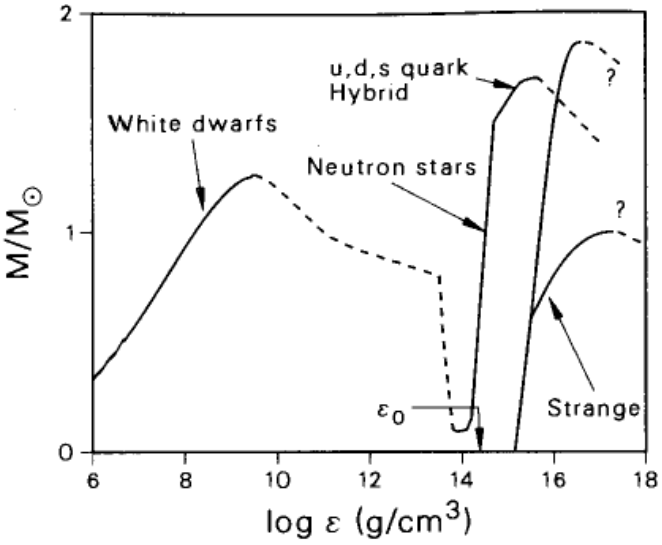

Figura 1: Mass of a fluid configuration as a function of central energy density (i.e., central density when $c = 1$). The solid lines show stable configurations, whereas dashed lines are unstable. The nuclear density is identified by the symbol ϵ_0. White dwarfs and neutron stars (henceforth referred to as WD and NS) are found for a certain EoS by integrating the TOV equations. The sequence labeled with "Strange" refer to a different EoS in which nucleons (i.e., protons and neutrons) are broken into quarks up and down and converted into strange quarks. Strange stars are an hypothetical class of compact objects, whose confirmation still needs observational validation.

WDs are the lowest density sequence of compact stable stars[1]. Their central densities range from about $5 \times 10^5\,\mathrm{g\,cm^{-3}}$ to $9 \times 10^9\,\mathrm{g\,cm^{-3}}$ (for comparison, the Earth average density

[1] At densities below WDs there is a long region of instability until the region populated by cold stable bodies like brown dwarfs and planets

is about $5\,\mathrm{g\,cm^{-3}}$). Stable WD that appear in Figure 1 are supported against gravitational collapse by the electron degeneracy pressure. However, with increasing WD mass, a point is reached in the equilibrium sequence where electrons are captured by protons via inverse β^- decay. This situation is more energetically favourable for high central densities. However, since electrons are captured, the degeneracy pressure falls off and gravity cannot be balanced anymore. This determines the unstable region (dashed line in Figure 1).

Stars in the unstable region between WDs and NS are subject to *vibrational modes* that will either cause their disassembly or else their collapse to NS. Indeed some NS are believed to have been formed by accretion induced collapse of a WD. In this process matter is transferred from a normal star to a WD until the WD mass enters the instability region.

The Fermi pressure of degenerate neutrons reestablishes stability against collapse for the lightest NS. Form more massive stars with higher densities, the repulsion of the nuclear fore at short range adds additional resistance to compression. NS have central densities in a range 5-10 times higher than the nuclear density ($\rho_0 \sim 2.5 \times 10^{14}\,\mathrm{g\,cm^{-3}}$). A similar situation then happens when moving along the sequence of stable NS configurations. At very high densities NS become unstable, and they collapse into black holes. However, recent calculations have shown that there might exist a third family of stable solutions called "strange stars" in which normal matter is converted into "strange matter" (i.e., strange quarks). Although fascinating, we will not consider this speculative hypothesis here any further.

2 Necessary condition for stability

It is possible to prove that along the sequence of equilibrium configurations of the TOV equations, perfect fluid[2] stars can pass from stability to instability with respect to any radial mode of oscillation only at a value of the central density at which the equilibrium mass is *stationary*:

$$\frac{\partial M(\epsilon_c)}{\partial \epsilon_c} = 0 \tag{1}$$

This is the key result in the discussion of stability. We will now give an heuristic proof of Eq. 1 which at the same time will provide a *necessary* but not sufficient condition for stability of stars along the sequence of equilibrium configurations. Refer to Figure 2 for a depiction of a sequence of stars in the vicinity of a maximum of the curve describing equilibrium configurations.

The figure above shows typical central densities and masses of a neutron star. Suppose an equilibrium solution S is perturbed so that the central (energy) density ϵ_c is increased. The star is displaced to C (i.e., it is compressed). The equilibrium star with this new central density lies in mass at C^*. The star at C therefore has a deficit of mass with respect to the equilibrium configuration, a deficit such that gravity under-balances the increased central pressure corresponding to the increased density at C, as compared to the equilibrium configuration of the star of this mass that lies at S. The gravitational force acting on the matter of the star will therefore act to return it towards S. A similar argument shows that, if the star at S undergoes a perturbation that decreases its central density, the force acting

[2] a fluid is *perfect* if it can be completely characterized by its rest frame energy density and pressure

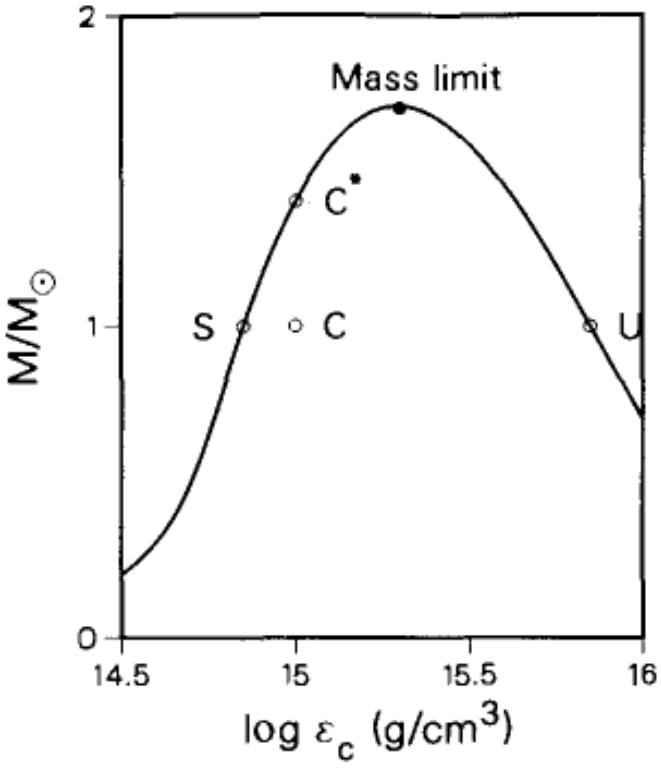

Figura 2: Schematic TOV solution in a region of increasing and decreasing mass as a function of central density. Hydrostatic equilibrium is stable along positive slopes of the curve until the mass limit is reached.

on the matter of the star will act to return it to S. Going through a similar analysis of the star U on the decreasing curve of M vs. ϵ_c, we arrive at the straightforward conclusion that, if U is either compressed or decompressed by a radial perturbation, the force acting on the perturbed star will act to drive it further from U, that is to say, equilibrium configurations must satisfy:

$$\frac{\partial M(\epsilon_c)}{\partial \epsilon_c} > 0 \tag{2}$$

This is the *necessary but not sufficient* condition to have stability of an equilibrium configuration. So we have proved that the passage from stability to instability occur only at the turning points defined by Eq. 1. If we now plot the mass of an equilibrium configuration vs. the number of baryons N (see Figure 3), then the number of baryons reaches its turning point at the maximum mass star, as prover above, and that stars with central density higher than that of the limiting mass star have higher mass than those of the same N on the lower branch.

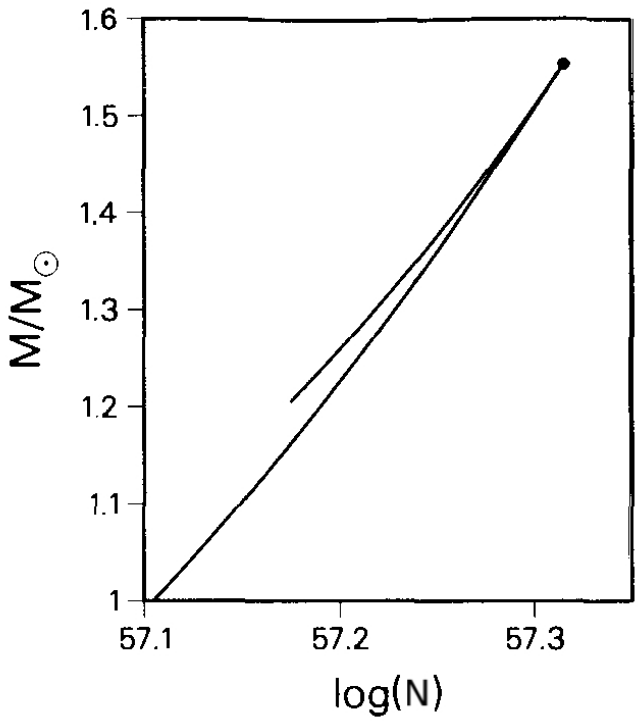

Figura 3: Mass of a fluid configuration as a function of baryon number N. Note that the x-axis is in Log scale. Central density increases from lower N up to the dot. The upper curve represents unstable fluid configurations, the dot the maximum mass of a stable equilibrium configuration and the lower curve stable fluid configurations.

We now can apply the necessary condition for stability to equilibrium configuration constructed with a polytropic EoS (i.e., $P = K\rho^{\Gamma}$, with K=constant). We saw in the previous lectures that the mass of a polytrope with $n = 3$ (i.e., relativistic case) is related with the central density via:

$$M \propto \epsilon_c^{(3\Gamma-4)/2} \tag{3}$$

which means that by using Eq. 1 we obtain the necessary condition for a n=3 polytrope: $\Gamma > 4/3$.

3 Sufficient condition for stability

We have now to determine whether an fluid configuration is stable against perturbations about the equilibrium. In other words we need to verify whether a fluid configuration (a WD or a NS) is not only in hydrostatic equilibrium, but it is also stable against radial oscillations. WDs and NS, like normal stars, can vibrate like most objects if perturbed. The vibrations may grow and destroy the object or be damped exponentially until equilibrium is restored. To determine which, we need to find the eigenfrequencies of the normal radial modes of vibration. The reason we use normal modes is because any type of vibration can be analyzed in terms of the normal modes. We start with the general relativistic expression of a spherically symmetric metric and we impose the condition that the perturbed star moves adiabatically (i.e., there are no energy losses associated with the compression/decompression process during the oscillation). Let's define the amplitude of the normal mode "n" of vibration with the function $u_n(r)$. The perturbation of the star at a radius r at time t can then be defined with the function:

$$\delta\, r(r,t) = e^{\phi} u_n(r) \frac{e^{i\omega_n t}}{r^2} \tag{4}$$

where ω_n are the star's oscillatory eigenfrequencies we want to find for each value of the central density, that is, for each equilibrium configuration of the stellar sequence under discussion. The fundamental mode corresponds to $n = 0$, which is also called node-less mode (see for example Figure 4). The equation that describes the n^{th} normal mode is the well known *Sturm-Liouville* eigenvalue equation:

$$\frac{d}{dr}\left(\Pi \frac{du_n}{dr}\right) + \left(Q + \omega_n^2 W\right) u_n = 0 \tag{5}$$

where the functions Π, Q and W are related with the spherically symmetric metric and with the adiabatic index $\Gamma = \frac{d\ln p}{d\ln \rho}$. We will not explicitly solve this equation here, but the interested reader can find a more detailed explanation of the problem in Ch. 6 of Shapiro & Teukolsky. What we need to know is just that this equation can be solved with suitable boundary conditions and gives the set of eigenfrequencies that we are looking for. The frequency spectrum has the form: $\omega_0^2 < \omega_1^2 < \omega_2^2 < ....$ If any of these is negative for a particular star, the frequency is purely imaginary and therefore any perturbation of the star will grow exponentially in amplitude of oscillation as $e^{|\omega|t}$. Such stars are unstable. As a consequence of the ordering of the eigenfrequencies and the fact that one mode of oscillation changes in stability at every stationary point (i.e., Eq. 1), we may infer the following. If the fundamental mode ($n = 0$) becomes unstable at the maximum neutron star mass, and it does, then at the next minimum in the sequence, either stability is restored to the fundamental mode or the next mode ($n = 1$) becomes unstable, and so on.

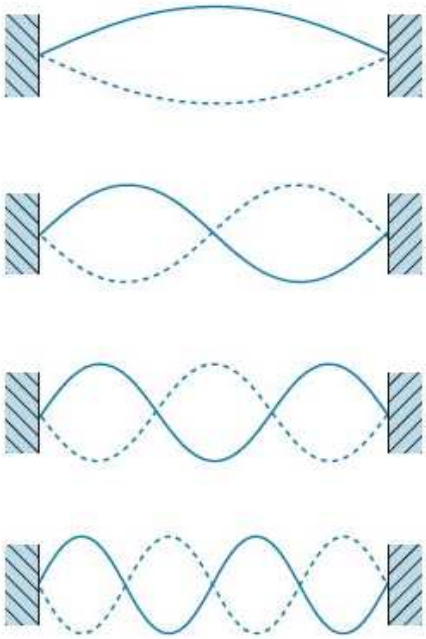

Figura 4: The first four normal modes of vibration of a string fixed at each end. The solid and dashed lines indicate the positions of the string at opposite phase positions in the cycle. From top to bottom: normal mode with n=0,1,2,3. Note that for a string the normal mode n=0 has two nodes (i.e., two stationary points) that correspond to the boundaries of the string. In a radial mode in the interior of a star instead there are no nodes for n=0.

3.1 Stability of the fundamental radial mode

As discussed earlier, we are primarily interested in the stability of the fundamental radial mode, which has no nodes in the interior of the star. Now the eigenfunction at a critical point of the M-ρ_c curve is simply the Lagrangian displacement $\delta\, r(r,t)$ that carries an equilibrium configuration on the low-density side of the critical point into an equilibrium configuration on the high-density side. In such a motion, ρ_c, increases and so $\delta\, r$ is negative near the center of the star. Since there are no nodes in the fundamental radial mode, $\delta\, r$ is also negative near the surface. Thus R decreases with increasing ρ_c, at a critical point where the fundamental mode changes stability. (More generally, $dR/d\rho_c < 0$ for the change in stability of a mode with an even number of nodes, $dR/d\rho_c > 0$ for an odd mode.) A schematic diagram of M vs. ρ_c, with multiple critical points is shown in Figure 5 and 6. At low density (large radius), all modes are stable. The first critical point that is reached is A, the maximum white dwarf mass. Since R is decreasing with increasing ρ_c, an even mode is changing stability. Since $\omega_0^2 < \omega_1^2 < \omega_2^2 < \ldots$, and since $\omega_0^2 > 0$ for low density, the only possibility is that ω_0^2 becomes negative. At B, an even mode is changing stability ($dR/d\rho_c < 0$). It cannot be that ω_2^2 is becoming negative, because ω_1^2 is still positive. Thus ω_0^2 must be turning positive again: B corresponds to the minimum neutron star mass. The critical point C is analogous to A and ω_0^2 becomes negative again; it corresponds to the maximum neutron star mass. At D, an odd mode changes stability. The only possibility is that ω_1^2 becomes negative. At E, an even mode changes stability. It is impossible for ω_0^2 to become positive while ω_1^2 remains negative, so it must be that ω_2^2 becomes negative.

Note that the criterion $dM/d\rho_c > 0$ (< 0) for identifying stability (instability) has limited applicability: thus, the range D-E is unstable, even thuogh $dM/d\rho_c > 0$. For typical cold equations of state, however, the lowest density at which $dM/d\rho_c = 0$ and $d^2M/d\rho_c^2 < 0$ corresponds to the maximum mass and density of a stable white dwarf (e.g., point A in Figure 5). The next point at which $dM/d\rho_c = 0$ and $d^2M/d\rho_c^2 < 0$ gives the maximum mass and density of a stable neutron star (e.g., point C in Figure 5 and 6).

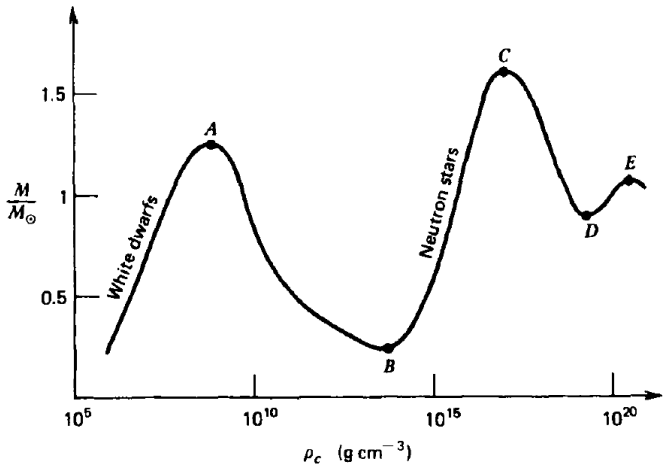

Figura 5: Schematic diagram showing the turning points in the mass vs. central density diagram for equilibrium configurations of cold matter (cf. Shapiro &Teukolsky, pp. 150, Fig. 6.2)

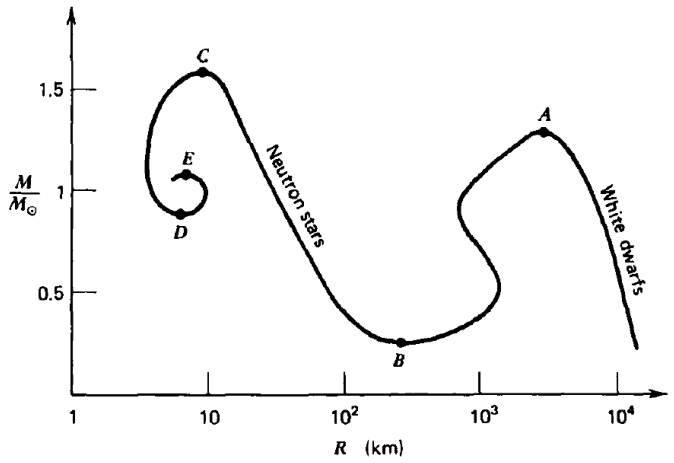

Figura 6: Schematic diagram showing the turning points in the mass vs. radius diagram for equilibrium configurations of cold matter (cf. Shapiro &Teukolsky, pp. 150, Fig. 6.3)

4 Physical Significance of a M-R relation for Neutron Stars

Once a sequence of stable equilibrium configurations has been established for a certain EoS, it is possible to build a mass-radius relationship for neutron stars. In Figure 7 several M-R relations are shown for NS with different EoS. The M-R relation is just analogous to the mass-density relation plotted in Figure 1. However, while the central density of a NS is not an observable, the mass and radius are measurable (at least in principle), and therefore with a suitable measurement of these quantities it may be possible to identify the EoS of ultra-dense matter. Once the EoS is identified, we can understand how particles and fundamental forces behave at densities not reproducible in Earth-based labs. However, while measuring the mass of a neutron star is possible in some situations, the measurement of the NS radius is a formidable challenge and no reliable measurements exist so far. The determination of the mass and radius *for the same NS* and therefore the identification of the EoS of ultra-dense matter is one of the most exciting areas of compact object astrophysics.

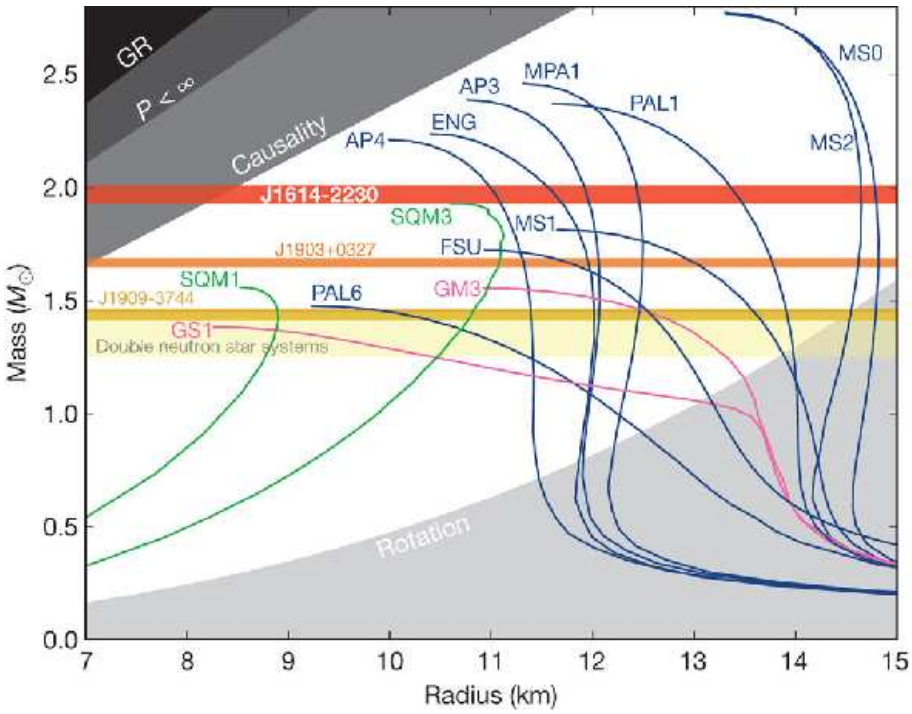

Figura 7: Mass-Radius relation for neutron stars with different equations of state. The gray areas are regions that can be excluded because of GR effects, pressure limit, causality and rotational centrifugal instability of a NS. The horizontal colored bands refer to precise radio measurements done on radio pulsars that allow a precise determination of the neutron star mass. The EoS labeled with "SQM" refer to "Strange Quark Matter" models, whereas all other EoS are calculated with different assumptions on the fundamental interactions of nucleons and on the particle composition of the inner core. By using the maximum mass observed for the pulsar J1614-2230 it is possible to immediately rule out equations of state like SQM1 (with pure strange quark matter), GS1 (with a Bose-Einstein kaon condensate) and many others. From P. B. Demorest,T. Pennucci, S. M. Ransom M. S. E. Roberts & J. W. T. Hessels 2010, Nature, Vol. 467, pp. 1081-1083

Lecture Notes 3

Equation of state above Neutron Drip

1 Neutron Drip Point

1.1 Structure of the inner crust

With increasing density, the ground-state value of Z/A decreases and neutrons become less and less bound. Let us define the "net neutron chemical potential" of a neutron in a nucleus

$$\mu_n' \equiv \mu_n - m_n c^2 = \left(\frac{\partial E\{A,Z\}}{\partial N}\right)_Z - m_n c^2 \,. \tag{1}$$

As long as $\mu_n' < 0$, all neutrons are bound within nuclei. The neutron drip point corresponds to $\mu_n' = 0$; beyond this point neutrons "drip out of nuclei", i.e. they begin to fill states in the continuous part of the energy spectrum. We can roughly localize the neutron drip point using the approximate mass formula for $E'\{A,Z\} = E\{A,Z\} - Amc^2$, where, for simplicity, we neglect the neutron-proton mass difference, putting $m_n \approx m_p \approx m = 939\ \mathrm{MeV}/c^2$. Neglecting surface and Coulomb terms, we have

$$E'\{A,Z\} \approx A\ \left(E_{\rm vol} + E_{\rm sym}\delta^2\right)\,, \tag{2}$$

where $\delta \equiv (N-Z)/A$, and $E_{\rm vol}$ and $E_{\rm sym}$ are nuclear volume and symmetry energies, respectively. Experimentally, $E_{\rm vol} \simeq -16$ MeV and $E_{\rm sym} \simeq 32$ MeV. Using Equation (2), we can easily show that the value of δ at which $\mu_n' = 0$ is

$$\delta_{\rm ND} = \sqrt{1 - E_{\rm vol}/E_{\rm sym}} - 1. \tag{3}$$

Using experimental values of $E_{\rm vol}$ and $E_{\rm sym}$, we find $\delta_{\rm ND} = 0.225$.

Neglecting neutron-proton mass difference, the beta equilibrium condition

$$p + e \rightarrow n + \nu_e \tag{4}$$

reads

$$\mu_e = \mu_n - \mu_p \approx 4E_{\rm sym}\delta\,. \tag{5}$$

At the density under consideration, electrons are ultra-relativistic, so that the chemical potential of the electrons $\mu_e \simeq 5.16\ (\rho_9 Z/A)^{1/3}$ MeV, where $\rho_9 = \rho/10^9$ g cm^{-3}. It is now easy to show that the neutron drip density is roughly $\rho_{\rm ND} \approx 2\times10^{11}$ g cm^{-3}, which is about half of the value obtained in complete calculations (cf. Sections 3.6 and 3.7 of Shapiro & Teukolsky). For $\rho > \rho_{\rm ND} \approx 4\times10^{11}$ g cm^{-3} a fraction of neutrons thus forms a gas outside the nuclei.

The inner crust of a neutron star is a unique system, which is not accessible in the laboratory due to the presence of this neutron gas. Sometimes the "nuclei" in the inner crust are referred to as "clusters" in order to emphasize these peculiarities. The description of

the crust beyond neutron drip therefore relies on theoretical models only. Many-body calculations starting from the realistic nucleon-nucleon interaction are out of reach at present due to the presence of spatial inhomogeneities of nuclear matter. Even in the simpler case of homogeneous nuclear matter, these calculations are complicated by the fact that nucleons are strongly interacting via two-body, as well as three-body, forces, which contain about twenty different operators. As a result, the inner crust of a neutron star has been studied with phenomenological models. Most of the calculations carried out in the inner crust rely on purely classical (compressible liquid drop) and semi-classical models (Thomas–Fermi approximation and its extensions). The state-of-the-art calculations performed so far are based on self-consistent mean field methods, which have been very successful in predicting the properties of heavy laboratory nuclei. Today we can say that our knowledge of the EoS in the inner crust of neutron stars is known relatively well. The outer core instead, defined as the region where $\rho_{nucl} < \rho < 2\rho_{nucl}$, has a less understood EoS, which however can still be reasonably described with an n-p-e gas (see Shapiro & Teukolsky, Section 2.5) with appropriate corrections for the nucleon-nucleon interaction (remember that the n-p-e gas we have seen so far was an *ideal* gas, i.e., non-interacting). When the density crosses the threshold of $2\rho_{nucl}$, the EoS becomes extremely uncertain, with many different possibilities that depend on the poorly understood nucleon-nucleon interaction. Degenerate nucleons become relativistic ($p_F^N > m_N c^2$) with new particles species appearing. The new particles might be light mesons (like pions π^-, that is a bound states of down and anti-up quarks), hyperons ($\Lambda, \Sigma^{0,\pm}$) or even a deconfined state of free quarks[1]. Our knowledge of the EoS of dense matter above the neutron drip point has to be considered more and more poorly understood as the density increases, with the EoS becoming basically completely uncertain when $\rho > 2\rho_{nucl}$. This matter is sometimes referred to as *ultra-dense matter*.

1.2 Nucleon-Nucleon interaction

When the density of matter reaches about the nuclear density $\rho_{nucl} \approx 2.8 \times 10^{14}\,\mathrm{g\,cm^{-3}}$, the nucleons start to "touch" each other. In this case the nucleon-nucleon interaction plays a dominant role in determining the EoS. However, present day nucleon-nucleon interaction is still poorly understood, and this is reflected on the EoS. In this section we will describe a rough sketch of the main properties that the nuclear force has. First of all, the nuclear force is not one of the four fundamental forces, and it exists as a consequence of the strong force that acts on quarks that compose nucleons. Nuclear forces are sometimes called "residual strong forces", because they emerge as a residual strong interaction between different quarks in different nucleons (see Fig. 1).

The force that holds protons and neutrons together is extremely strong. It has to be strong to overcome the electric repulsion between the positively charged protons. It is also of very short range, acting only when two particles are within 1 or 2 fm (1 fm = $10^{-15}m$) of each other.

In Fig. 2 an empirical description of the nucleon-nucleon potential is shown.

- There is an extremely strong short-range repulsion that pushes protons and neutrons apart before they can get close enough to touch. (This is shown in light blue.) This repulsion can be understood to arise because the quarks in individual nucleons are forbidden to be in the same area by the Pauli exclusion principle.

[1] quarks can exist as free particles only when the densities reach the extreme values at the center of a NS. In this case asymptotic freedom guarantees that the strong interaction is sufficiently weak to allow quarks in nucleons to behave as "free" particles

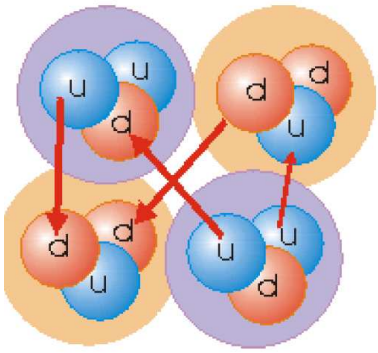

Figura 1: The residual strong force can be understood as a residual strong interaction between quarks in different nucleons. In the picture neutrons (up+down+down quarks) and protons (up+up+down quarks) are bound together via the residual interaction of up and down quarks.

- There is a medium-range attraction (pulling the neutrons and protons together) that is strongest for separations of about 1 fm. (This is shown in light green). This attraction can be understood to arise from the exchange of quarks between the nucleons.
- At long ranges (separation $> 2 - 3$ fm) nuclear forces become ineffective.

The density of nuclei is limited by the short range repulsion. The maximum size of nuclei is limited by the fact that the attractive force dies away extremely quickly (exponentially) when nucleons are more than a few fm apart.

The real nuclear force is however more complex than suggested above. For example, if a proton and a neutron form a bound state (deuterium, or ${}^{2}_{1}H$), why can't two neutrons form a bound state? If the nuclear force were a simple central force (like gravity) that only depended on distance like shown above, we would expect the di-neutron to be stable just like ${}^{2}_{1}H$ because neither has any Coulomb repulsion to deal with. (The two protons that would make up ${}^{2}_{2}He$ would be pushed apart because like charges repel.)

The answer is that the measured properties of the nucleon-nucleon interaction depend on many things, including the kind of particle and spin orientation, and is not strongly attractive under all possible situations. We will not go into the details of realistic nucleon-nucleon interaction here, but we will use only the rough sketch outlined above.

2 Nuclear properties of Ultra-dense Matter

We have seen the semi-empirical mass formula (Shapiro & Teukolsky, Section 2.6) which, in one of its simplest formulation, can be written as:

$$E_b = [a_1 A - a_2 A^{2/3} - a_3 Z^2 A^{1/3} - a_4 (A - 2Z)^2 A^{-1}]c^2 \tag{6}$$

According to this formula, when ignoring the Coulomb and the surface term, the binding energy of nuclei decreases (i.e., it becomes stronger) in direct proportion to the atomic mass number A. We have already discussed in Lecture 3 that when the nucleon-nucleon interaction between A nucleons is considered, the "volume" term in the semi-empirical mass formula is proportional to A instead of $A(A-1)$, as one would expect if A nucleons interact with the remaining (A-1). The heuristic explanation we gave back then was that the nucleons interact

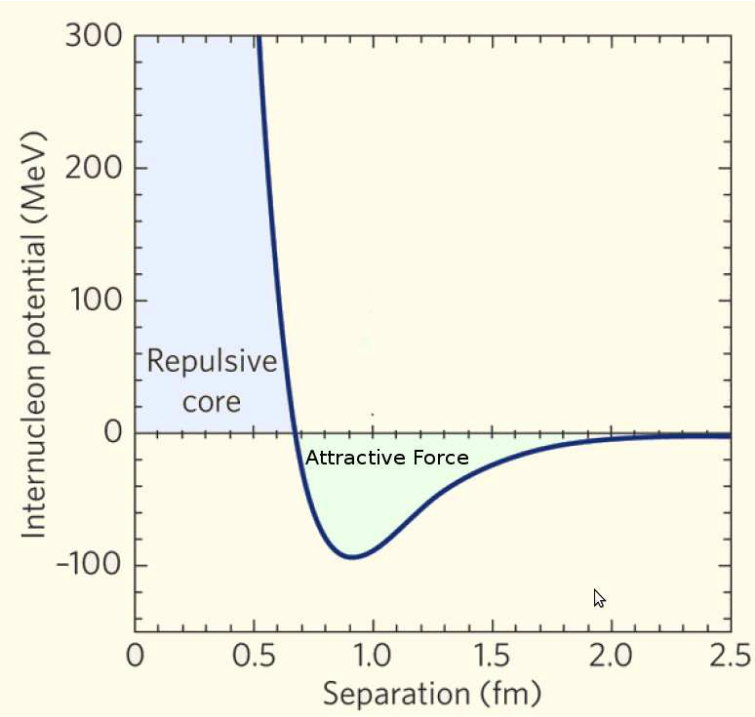

Figura 2: Empirical description of nucleon-nucleon potential. At very short distances ($d \lesssim 0.1\,\mathrm{fm}$) the nucleons feel a repulsive core, whereas at longer distances the potential is negative and generates a weakly attractive force (the nuclear force)

only with their closest neighbours. This property of nuclear matter is called "saturation". Nuclear matter represents saturated systems. The reason for this is in the short range of nuclear force, its strong repulsion at short distances and the Pauli principle. Saturation means that, as more nucleons are added to nuclei, the density of the central region remains constant. The Saturation occurs in nuclei as well as in assemblies of free nucleons at sufficient densities. The force between nucleons must be attractive for a small number of nucleons, but become repulsive for a larger number. Chemical forces show saturation: two hydrogen atoms combine to form a hydrogen molecule, but a third atom cannot be bound. In the hydrogen molecule the two electrons can occupy the same orbital with opposite spins, but a third electron cannot because of the Pauli principle. A purely attractive nuclear potential leads to nuclear collapse as A increases, while the presence of a repulsive core helps bring about saturation. Nuclear saturation corresponds to a baryon density $n_s = 0.16 \pm 0.01\,\mathrm{fm}^{-3}$. The central densities of neutron stars may even range up to (3–10) n_s. At densities around n_s and below, matter may be regarded as a mixture of neutrons, protons, electrons, positrons, neutrinos and anti-neutrinos, and photons. At higher densities, additional constituents may be present, such as muons, hyperons, kaons, pions and quarks. Up to now, there is no consensus on the composition of ultra-dense matter.

Whatever form is chosen for the nucleon-nucleon interaction in calculating the EoS of ultra-dense matter, a key experimental constraint is that the chosen potential reproduce the observed properties of nuclear matter at saturation. Four important parameters need to be reproduced:

- The saturation density $n_s \simeq 0.16 \pm 0.01\,\mathrm{fm}^{-3}$.

- The energy W_V of symmetric nuclear matter.
- The compressibility K of nuclear matter.
- The volume symmetry coefficient S_V.

The energy and compressibility of nuclear matter are defined in terms of the quantity $W(k, x)$ (see Shapiro & Teukolsky, Section 8.2) for the bulk energy of nuclear matter per nucleon. Symmetric nuclear matter (i.e., Z=A/2) has $x = \frac{1}{2}$. In the vicinity of the saturation density, we can write:

$$W\left(k, \frac{1}{2}\right) = -W_V + \frac{1}{2}K\left(1 - \frac{k}{k_0}\right)^2 \tag{7}$$

The most recent nuclear experiments provide the values:

$$W_V = 16 \pm 1.0\,\mathrm{MeV} \tag{8}$$

$$K = k^2 \frac{\partial^2 W\left(k, \frac{1}{2}\right)}{\partial\, k^2}\bigg|_{k=k_0} \approx 231 \pm 5\,\mathrm{MeV} \tag{9}$$

Here $k_0 = 1.33\,\mathrm{fm}^{-3}$ follows from the value of the saturation density n_s and the nucleon number density $n = \frac{2k^3}{3\pi^2}$ (see also Eq. (8.2.27) on Shapiro & Teukolsky). An interpretation of K is that it gives the associated increase of the energy per nucleon of the symmetric nuclear matter due to a small compression of rarefaction.
The volume symmetry coefficient S_V measures the curvature of the bulk nuclear energy $W(k, x)$ (see also §8.2 on Shapiro & Teukosly) due to changes in x:

$$S_V = \frac{1}{8}\frac{\partial^2 W(k_0, x)}{\partial\, x^2}\bigg|_{x=1/2} = 32 \pm 6\,MeV \tag{10}$$

Notice how the uncertainty on S_V is rather large, which is a consequence of the ambiguity in separating the bulk and surface symmetry term in binding energies.

Accordingly, we can write for k near k_0 and x near $\frac{1}{2}$:

$$W(k, x) = -W_V + \frac{1}{2}K\left(1 - \frac{k}{k_0}\right)^2 + 4S_V\left(x - \frac{1}{2}\right)^2 \tag{11}$$

3 Minimum Neutron Star Mass

If you look at the M-R or M-ρ_c relation for equilibrium configurations of cold matter (e.g., Figure 6.2 and 6.3 on the Shapiro & Teukolsky, or see Lecture Notes 2), you might have noticed that the white dwarf sequence has no minimum mass, whereas there is a clear minimum for neutron stars.

To understand this minimum, let's start by calculating the gravitational energy of a neutron star. As an example, we can roughly estimate the gravitational energy for a uniform density neutron star with a nominal two solar mass and a 10 km radius.

Using Eq. (5.7.8) from the Shapiro & Teukolsky:

$$M = \int_0^R 4\pi\, r^2 \rho\, dr \tag{12}$$

we find:

$$E_G \approx \int_0^R \frac{M(r)dM(r)}{r} = \frac{3}{5}\frac{M^2}{R} \approx 7 \times 10^{53}\,\mathrm{erg} = 4 \times 10^{59}\,\mathrm{MeV} \tag{13}$$

Thus the gravitational energy is a significant fraction of the total gravitational mass of the star, namely, 18% for a two solar mass star. To find the gravitational energy per nucleon in the star, refer above to the estimated number of nucleons A to find 157 MeV per nucleon. By contrast, the binding of a nucleon at saturation is about 16 MeV (see Eq. 8) and 8 MeV in a finite nucleus. The classical estimate of the gravitational energy is not the binding energy of the star; it does not take account of the compression energy resisting the assembly of nucleons in a star arising from the Fermi pressure and the short-range repulsive nuclear force. Nonetheless, the binding energy per nucleon in a limiting mass star is about 100 MeV per nucleon. Thus, the weakest force binds nucleons in a neutron star ten times more strongly than the strong force binds them in nuclei. How can this be so? The answer lies in their ranges. The nuclear force acts only on neighbors that lie within its range, typically the several next nearest neighbors, but each particle in the star feels the gravitational force of all others. From the absolute limit on M/R of a stable star derived in Problem Set 6 $\left(\frac{2M}{R} < \frac{8}{9}\right)$, we can find an upper limit on the gravitational energy of a star relative to its mass. It is $(3/5)(M/R) < 4/15$ or 27%.

We can now compute the mass of the equivalent numbers of baryons dispersed at rest at infinity.

$$M_B = A\,m_B \tag{14}$$

where the suffix "B" refers to the baryon species present in the neutron star, and A is the baryon number. The mass M_B is sometimes called the baryon mass of the star. In Fig. 3 we compare the gravitational masses with the baryon masses. By reading the mass difference from the graph, one can find that the binding per neutron is about 100 MeV. This compares with our previous estimate of 160 MeV, which actually is the classical gravitational energy. The latter refers only to the gravitational attraction and does not include the work required to overcome the Fermi pressure and the resistance to compression of high-density matter. The difference $M - M_B$, which is the binding energy of the neutron star, contains all such effects and is the net energy gained by assembling the A nucleons from infinity to form an equilibrium neutron star. The binding energy is the energy released when the core of an evolved star of $\sim 8 - 10$ or more solar masses collapses. About one percent actually appears as the kinetic energy of the supernova explosion in which the neutron star is formed. Ninety-nine percent is carried off by neutrinos in the first few seconds following birth of the neutron star, and a small fraction of one percent appears as light in the supernova remnant. The baryon mass is usually larger than the gravitational mass and always larger for configurations near the limiting mass. However, less massive neutron stars have a larger gravitational mass than that of all the individual baryons dispersed to infinity. This fact is shown in Fig. 4 where the binding energy per nucleon, $B/A = (M_B - M)/A$, is plotted. These stars have even higher gravitational mass as compared to the mass of the equivalent baryon number disassembled in the form of iron nuclei at infinity. Nevertheless, even in the case in which a star is not bound with respect to a simultaneous removal of all constituents to infinity, it is bound with respect to removal of single nucleons (or groups of nucleons, but not all nucleons). The reason for this is the long-range attraction of all the remaining nucleons in the star. Such neutron stars are metastable with an essentially infinite barrier for disintegration. It may be relevant to observe that there may exist no creation mechanism for low-mass neutron stars having small or negative binding, even though they are stable in the sense discussed

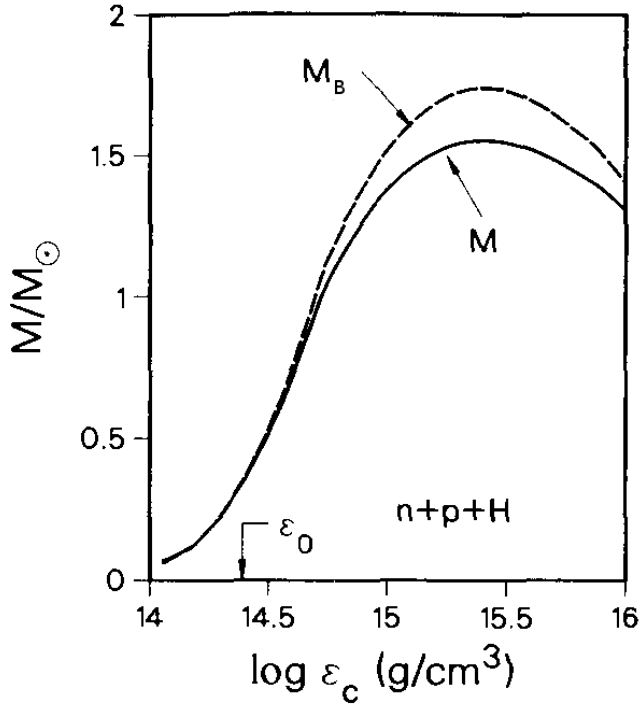

Figura 3: Gravitational Mass (M) and Baryon Mass (M_B) vs. central density. The difference $M_B - M$ provides the binding energy of the star

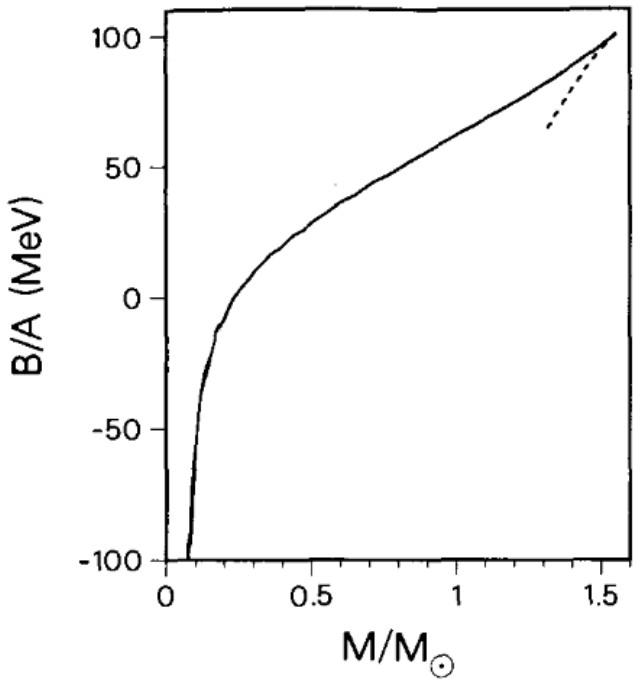

Figura 4: Binding energy per baryon of a neutron star sequence. The dashed line represents the unstable configurations with central densities higher than that of the limiting mass star. Note the negative binding for stars with $M < 0.2\, M_\odot$

above. They must evade collapse to a black hole. However, the neutrinos carry almost all the binding energy, and because of their weak interaction with nucleons, transmit very little to the in-falling material of the defunct star. Thus a low binding energy may fail to eject the ten or more solar masses of the original star, and there is no chance to do so at all if B/A is negative. A large release of binding energy, only a small part of which can be transmitted to the in-falling material because of the weak interaction of neutrinos that carry the energy, is required to eject most of the progenitor star and evade the formation of a black hole. This observation may account for the apparent absence of low-mass neutron stars (say less than $1M_{\odot}$).

Finally, we must observe that neutron stars with mass $M \rightarrow 0$ cannot exist also for another fundamental reason related with the instability due to neutron decay. We have seen that the beta decay reaction:

$$n \rightarrow p^{+} + e^{-} + \bar{\nu_e} \tag{15}$$

is suppressed in neutron stars because the electron Fermi sea is completely filled. The neutrons (whose mean lifetime is $\sim$ 15 min) are therefore stabilized against beta decay in neutron stars. However, when the density drops towards zero the electrons stop being degenerate and the beta decay reaction is unlocked. Neutrons start to decay and the neutron star is quickly disrupted on a dynamical timescale (i.e., between a few ms to a few seconds). Matter experiences abrupt acceleration and disperses to infinity, releasing a copious amount of anti-neutrinos.

4 Maximum Neutron Star Mass

The existence of M_{max} can be explained using two kinds of physical arguments, related to the behavior of ultra-dense cold matter and to General Relativity.

The **ultra-dense cold matter argument** was presented for the Newtonian case by Landau in 1932. We have seen how he was able to determine the Chandrasekhar limit with very good approximation simply using the Fermi energy and the gravitational energy for a number N of particles (see Shapiro & Teukolsky §3.4). When he used the electron mass as the source of degeneracy pressure, Landau obtained a maximum mass and a radius for white dwarfs very close to the results of Chandrasekhar when using a relativistic polytrope (n=3). If you substitute the electron mass with the neutron mass, i.e., if you assume that the degeneracy pressure is provided by free neutrons, then the maximum mass will be the same as in the WD case, but the radius will scale down to a few km (i.e., the radius of a neutron star). The great beauty of this argument was that the maximum mass of a neutron star was not depending on any quantity other than fundamental constant of nature and on the baryon mass that provides most of the neutron star mass (same identical argument was true for WDs).

Beautiful by its simplicity and astonishing by the realistic value of M_{max}, the Landau-type derivation is based on two bold assumptions. First, one assumes that the matter becomes a free Fermi gas with increasing density (i.e., interactions between matter constituents become negligible). Second, one uses the Newtonian gravity and the baryon mass of the star. The crucial effect of space-time curvature will be discussed in the end of the present section, while now we focus on the behavior of the EoS at higher densities.

For a certain model of the neutron-neutron interaction involving nucleons whose details we will skip here, it is possible to demonstrate that the interaction contribution dominates over the kinetic energy at high densities, so that:

$$\epsilon \propto n_b^2 \propto \frac{N^2}{R^6} \tag{16}$$

where n_b is the baryon number density, N is the baryon number and R is the radius where the baryons are enclosed. This is different from the value that Landau obtained with an internal energy for the body corresponding to $E_{int} = N \cdot E_F \simeq \frac{\hbar c N^{4/3}}{R}$ (note that in Eq. 3.4.1 of the Shapiro & Teukolsky, E_F is the Fermi energy *per particle*). The total gravitational energy (Newtonian case) is: $N \cdot E_G = -GN^2 m_B^2/R$ Therefore it is always possible to find a value N that minimizes the total energy $E_{int} + N \cdot E_G$ and therefore there is no upper bound on M or N for Newtonian stars. However, we know that at sufficiently high densities quarks become nearly deconfined (albeit we do not know whether the required densities are really reached in the inner core of neutron stars). When and if this happens, the internal energy of the star cannot be calculated anymore as above, because that depends on the nucleon-nucleon interaction, which is just not working for deconfined quarks. It is possible to demonstrate that for free quarks $E_{int} \propto N^{4/3}/R$, i.e., this is exactly the same scaling as in the Landau argument.

The whole point above has been given just to demonstrate the following: *until we do not know what is the EoS of ultra-dense matter, i.e., until we understand which fundamental interactions are playing in the inner core of neutron stars, the maximum mass of a neutron star will depend on the* ***assumed*** *fundamental interaction that dominates in the inner core.*

Moreover, irrespectively of the EoS, the upper bound on M is a consequence of a **General Relativistic argument**. If you look at the TOV hydrostatic equilibrium equation, the gravitational pull acting on a matter element of unit volume is:

$$GravitationalPull = -\frac{Gm\rho}{r^2}\left(1 + \frac{4\pi P}{m r^3}\right)\left(1 + \frac{P}{\rho c^2}\right)\left(1 - \frac{2Gm}{2c^2}\right)^{-1} \tag{17}$$

This pull is given by a Newtonian-like term, i.e., $-Gm\rho/r^2$ multiplied by three relativistic factors. With increasing M, all these factors amplify the pull compared to the Newtonian case, as you can verify by noticing that all of them are always positive. This means that there is a demand for higher pressure P to keep hydrostatic equilibrium. Mathematically, the derivative dM/dP_c decreases with growing M and makes stellar configurations less stable.

Let's illustrate this property with an *unphysical* case of incompressible fluid of constant density ρ_{inc}. The total gravitational mass is then $M = (4\pi/3)\rho_{inc}R^3$, and the pressure $P(r)$ in the star is determined analytically (the interested reader can find the details of this calculation in the book "Gravitation", Misner, Thorne & Wheeler (1971), Box 23.2). The central pressure P_c can be determined as a function of r_s/R, where $r_s = GM/c^2$ is the Schwarzschild radius of the star. As the mass increases, the pressure P_c and r_s/R grow up. In other words, both r_s/R and M increase with increasing P_c. For $P_c \to \infty$, the radius tends to a finite value $R_{max} = \frac{9}{8} r_s$ (cf. Exercise 5.5), and the mass tends so:

$$M_{max}^{inc} = \frac{4\pi}{9}\frac{R_{max}c^2}{G} \approx 5.09\left(\frac{5 \times 10^{14}\,\mathrm{g\,cm^{-3}}}{\rho_{inc}}\right)^{1/2} \tag{18}$$

This limit is the effect of General Relativity; there is no limit on the mass of the incompressible fluid, the it should a fortiori exist for any EoS of dense matter with *finite compressibility.* Calculations show that the numerical value of M_{max} is mainly determined by the EoS at $\rho \gtrsim 2\rho_{nucl}$, which is largely unknown.

Lecture Notes 4

1 Superconductivity

Superconductors are defined as materials with a definite phase transition and critical temperature T_c below which the resistivity is zero, and who *exhibit the Meissner effect*. This latter property will be explained in §1.2, and marks one of the *fundamental* differences between perfect conductors and superconductors.

In superconductors the resistivity, ρ, becomes zero, and so the conductivity σ appears to become infinite below T_c. To be consistent with Ohm's law ($\vec{\epsilon} = \rho\,\vec{j}$ we must always have zero electric field, $\vec{\epsilon} = 0$ at all points inside a superconductor. In this way the current, $\vec{j}$, can be finite. So we have current flow without electric field. Notice that the change from finite to zero resistivity at the superconducting critical temperature T_c is very sudden. This represents a thermodynamic phase transition from one state to another. As for other phase transitions, such as from liquid to gas, the properties of the phases on either side of the transition can be completely different. The change from one to the other occurs sharply at a fixed temperature rather than being a smooth cross-over from one type of behaviour to another. Here the two different phases are referred to as the *normal state* and the *superconducting state*. In the normal state the resistivity and other properties behaves similarly to a normal metal, while in the superconducting state many physical properties, including resistivity, are quite different. In some cases, notably the high temperature superconductors, there is a small range of temperatures where the resistance starts to decrease before becoming truly zero. This is due to thermodynamic critical fluctuations associated with the phase transition. The precise thermodynamic phase transition temperature T_c can be defined as the temperature where the resistivity first becomes exactly zero. The key characteristic of the superconducting state is that the resistivity is exactly zero:

$$\rho = 0 \tag{1}$$

or the conductivity, σ, is infinite[1]

1.1 Perfect Conductor

The most convincing evidence that perfect conductors really have $\rho = 0$ but they are not superconductors comes from the observation of persistent currents and the observation of the Meissner effect (see next section). If we have a closed loop of perfectly conducting wire, such as the ring shown in Fig. 1, then it is possible to set up a current, I, circulating in the loop. Because there is no dissipation of energy due to finite resistance, the energy stored in the magnetic field of the ring will remain constant and the current never decays. To see how this persistent current can be set up, consider the flux of magnetic field through the centre of the superconducting ring. The flux is defined by the surface integral

$$\phi = \int \vec{B} \cdot d\vec{S} \tag{2}$$

[1]It is important to stress once more that that the zero resistivity is not a **sufficient condition** to be a superconductor. Perfect conductors are **not** superconductors and still show $\rho = 0$.

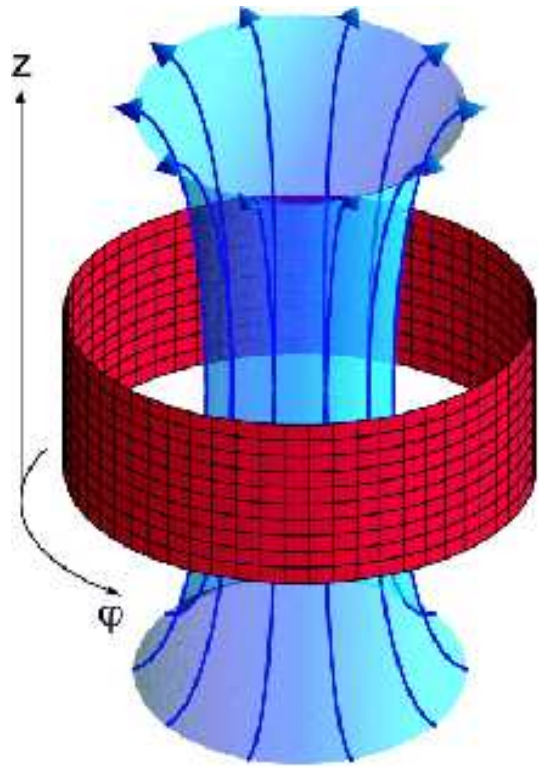

Figure 1: A persistent current can circulate in a ring of superconductor (in red). Since the resistivity is zero, there is no dissipation and the ring can sustain the current forever, and the energy stored in the magnetic field of the ring (blue lines) will remain constant in time.

where $d\vec{S}$ is a vector perpendicular to the plane of the ring. Its length dS, is an infinitesimal element of the area enclosed by the ring. But, by using the Maxwell equation

$$\Delta \times \vec{\epsilon} = -\frac{\partial \vec{B}}{\partial t} \tag{3}$$

and using Stokes's theorem we can see that

$$-\frac{d\phi}{dt} = \oint \vec{\epsilon} \cdot d\vec{r} \tag{4}$$

where the line integral is taken around the closed path around the inside of the ring. This path can be taken to be just inside the superconductor, and so $\vec{\epsilon} = 0$ everywhere along the path. Therefore

$$\frac{d\phi}{dt} = 0 \tag{5}$$

and hence the magnetic flux through the ring stays constant as a function of time. We can use this property to set up a persistent current in a perfect conductor. First we start with the perfect conductor at a temperature above T_c, so that it is in its normal state, with a non-zero resistance (normal conductor). Then apply an external magnetic field, $\vec{B}_{ext}$. This passes easily through the normal conductor. Now cool the system to below T_c, so that the resistance becomes zero and the conductor is now perfect. The flux in the ring is given by $\phi = \oint \vec{B}_{ext} \cdot d\vec{S}$. But we know that the flux remains constant, no matter what. Even if we turn off the source of external magnetic field, so that now $\vec{B}_{ext} = 0$, the flux must remain constant. The only way the perfect conductor can keep ϕ constant is to generate its own magnetic field $\vec{B}$ through the centre of the ring, which it must achieve by having a circulating current, I, around the ring. The value of I will be exactly the one required to induce a magnetic flux equal to ϕ inside the ring. Further, because ϕ is constant, the current I must also be constant. We therefore have a set up circulating persistent current in our perfect conducting ring. Furthermore if there were any electrical resistance at all in the ring there would be energy dissipation and hence the current I would decay gradually over time.

1.2 The Meissner Effect

As we have said earlier, the fact the the resistivity is zero, $\rho = 0$, is a common property between perfect conductors and superconductors. However, the latter cannot be defined merely with the condition $\rho = 0$. The fundamental proof that superconductivity occurs in a given material is the demonstration of the Meissner effect. This effect is the fact the a superconductor expels a weak external magnetic field. First, consider the situation illustrated in Fig. 2 in which a small spherical sample of material is held at temperature T and placed in a small external magnetic field, B_{ext}. Suppose initially we have the sample in its normal state, $T > T_c$, and the external field is zero, as illustrated in the Figure. Imagine that we first cool to a temperature below T_c while keeping the field zero. Then later as we gradually turn on the external field the field inside the sample must remain zero. This is because, by the Faraday's law (Maxwell equation written above), combined with $\vec{\epsilon} = 0$ we must have

$$\frac{\partial \vec{B}}{\partial t} = 0 \tag{6}$$

at all points inside the superconductor. Thus by applying the external field to the sample after it is already superconducting we must arrive at the state where the magnetic field

$\vec{B} = 0$ is zero everywhere inside the sample. But now consider doing things in the other order. Suppose we take the sample above T_c and first turn on the external field, B_{ext}. In this case the magnetic field will easily penetrate into the sample, $\vec{B} = \vec{B}_{ext}$, as shown in Fig. 2. What happens then we now cool the sample? The Meissner effect is the observation that upon cooling the system to below T_c the magnetic field is expelled. This fact cannot be deduced from the simple fact of zero resistivity and so this is a new and separate physical phenomenon associated with superconductors. There are several reasons why the existence of the Meissner effect affecting a sample is taken as definitive proof of superconductivity. At a practical level it is perhaps clearer to experimentally demonstrate the flux expulsion than zero resistivity, because, for example, it is not necessary to attach any electrical leads to the sample. A more fundamental reason is that the Meissner effect is a property of thermal equilibrium, while resistivity is a non-equilibrium transport effect. In fact one can see in Fig. 2 that we reach the same final state of the system whether we first cool and then apply the field, or the other way around. Therefore the final state of the system does not depend on the history of the sample, which is a necessary condition for thermal equilibrium. We shall always define a superconductor as a system which exhibits the Meissner effect.

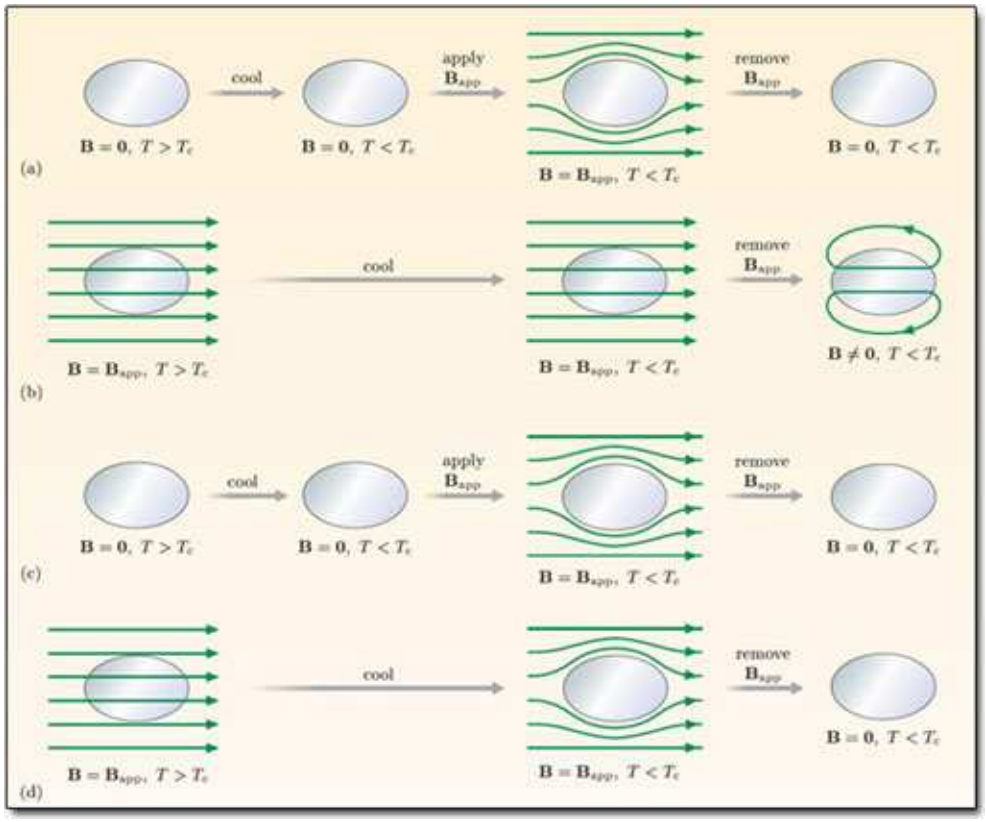

Figure 2: Panel a)-b): Perfect conductor. Panel c)-d): Superconductor

2 Pulsar Glitches

This section has to be considered an integration to §10.10 in Shapiro &Teukolsky. This is necessary because the *two-component* model presented in S&T can be considered outdated

in some aspects. Here we present a brief historical perspective on the evolution of pulsar glitch models from the '80s up to the present days (consider that Shapiro & Teukolsky has been published in 1983, so it is not updated in the content of some sections, especially on the observations). We will show that in the last decades many other pulsar glitches have been observed both for Crab and Vela pulsars, but also for other pulsars not mentioned in the S&T. New observations led to a reinterpretation of the two-component model, which however to date remain a valid base to start studying pulsar glitches and hadron superfluidity in neutron stars.

In 1981, the first real opposition to the two-component model appears, in two separate papers. Firstly, Boynton (1981, The IAU, 279) obtained a power spectrum of timing fluctuations from pulsar timing data (power spectra are obtained by squaring the Fourier amplitudes of the time series Fourier transform. More details will be given in Lecture 10, on the Observational techniques). He showed that for the calculated values of Q and τ for the Crab pulsar, according to the two-component model, there should be a prominent , characteristic, two-component signature in the power spectrum. however, in his analysis, Boynton found none. Instead, he found that the signature pointed towards rigid-body rotation. He then went on to say that the two-component model works well for macroscopic glitches, but not for minor perturbations, which give results resembling rigid-body behaviour.

The second objection came the same year from Downs 91981, ApJ 249, 687). In this paper, he thoroughly analyzed twelve years of observational data for the Vela pulsar. He mentioned that only 20% of each Vela period jump decayed exponentially towards the pre-glitch spin-down rate, having recovery time of around one year. He also showed that the recovery time τ is not the same from glitch to glitch for Vela, in contradiction with what was proposed by Pines et al. (1974). Another important observation is that the values of Q obtained in this analysis were much lower than expected, for large glitches ($\Delta\Omega/\Omega \simeq 10^{-6}$. These Q values were used to calculate neutron star masses which are roughly half those obtained using more direct methods. As a result, Downs stated that the two-component model is not dominant in case of large spin-up. He concluded that the interior structure of the Vela pulsar is too complex to be explained by the simplistic two-component model.

Later that year, Alpar et al. (1091, ApJ 249, L29), commented on the merit of Downs' conclusion that the Vela pulsar's relaxation behaviour is too complex to be described by the simplistic two-component model presented by Baym et al. (see S&T, §10.10). They used this result as the basis for their model in which spin-ups are caused by a change of vorticity in the superfluid component, and the post-glitch relaxation is caused by vortex creep mechanisms. This newer model relies on the processes of *vortex pinning and unpinning* as the main drivers of the glitch phenomenon. The main differences between this model and the two-component model are as follows:

- The vortex unpinning model involves a much smaller fraction (a few percent) of the neutron superfluid, i.e., the pinned portion of the superfluid, as compared to the entire superfluid in the two-component model.

- The vortex unpinning model is a non-linear theory, in that the vortex creep rate (i.e., the rate at which vortices migrate radially outwards) depends exponentially on the size of the glitch, as well as on the glitch-induced change in superfluid angular velocity.

- The vortex unpinning model produces relaxation timescales which are proportional to the internal temperature of the star. As a result, analyzes of timing data using

this model would enable the determination of the internal temperature, and hence an estimate of the surface temperature of the star.

A few months after Alpar et al. first proposed the vortex unpinning model, is early 1982, Downs produced a second detailed paper on pulsar timing analysis (Downs 1982, ApJ 257, L67), this time for PSR 0525+21. Once again, twelve years of data were analyzed, and in this period, two small glitches were observed. Downs contended that the original two-component model was not a good fit to these glitches. The reason for this is that after the spin-up, the period derivative $\dot{P}$ did not return to the pre-jump value, as is required by the model. Instead, the post-glitch behaviour seemed to mimic that displayed by the Vela pulsar, as discussed in Downs' previous investigation.

The following year, in 1983, Demiansky & Prozynsky (1983, MNRAS 202, 437) published their results on the 1975 Crab pulsar glitch. They too found a permanent change in $\dot{P}$ after the glitch event, citing a change in the external torque on the star as the cause. This further added to the evidence supporting the claim that the two component model is too simplistic to describe the processes occurring during a glitch. Link et al. (1992, ApJ 390, L21) proposed that this observed behaviour is a result of an increased *external torque* caused by the rearrangement of the stellar magnetic field.

In 1985, Cordes and Downs (1985, ApJ Supplements 59, 343) aptly commented on the then recent findings regarding the two-component model. They stated that the model was not to be scrapped, but that a redefinition of the two components was required. They emphasized the fact that the most successful mechanisms were those that involved a transfer of angular momentum from the more rapidly rotating superfluid to the slower crust.

In 1987, the two0component model received some observational support from the analysis of the 1986 glitch in the Crab pulsar, performed by Lyne & Pitchard (1987, MNRAS 229, 223). They found that the data for this glitch could be easily fitted using a simple exponential, of the form of that obtained using the two-component model. They did, however, find that the data are best fitted by two separate exponential. However, the two exponential factors ($\tau_1 = 2.5 \pm 0.5$ days and $\tau_2 = 5.0 \pm 0.5$ days) are multiples of each other within the errors. Hence, they could not conclude whether these do in fact represent two separate exponential, or a single more complicated one. They went on to say that the two-component model still represents an applicable model; one that can be best understood in terms of the vortex unpinning model of Alpar et al. It is interesting to note that the magnitude of this glitch is rather small, comparable in size to the 1969 Crab glitch, which was also fitted using a two-component model. The Crab glitch of 1975 was instead a lot larger, by about three orders of magnitude, and its data could not be fitted using the simple two-component model. This adds weight to the comments made by Downs that the two-component model fails in the large glitch regime.

In 1990, Flanagan (1990, Nature 345, 416) produced results from the analysis of the 1988 Vela pulsar glitch. She found that there was an additional coupling timescale when the data were fitted using exponential functions. This timescale was much shorter that those previously discovered ($\tau = 0.4 \pm 1$ days), most probably due to the fact that observations had not been made so soon after a glitch prior to this one. This extra coupling constant provided evidence for the presence of two superfluid components, both linearly coupled to the crust and its accompanying constituents. This finding was supported by Ali et al. (1990, Nature 348, 707).

Michel et al. (1990, MNRAS 246, 624) showed that the glitches occurring in PSR 1737-30 were very well fitted with a simple two-component model. This showed that in this pulsar

at least, there is no change in moment of inertia (i.e., $Q = 1$), and supported the case for continued validity of the two-component model.

In 1996, Lyne et al. (1996, MNRAS 281, L14) performed an analysis on a glitch in PSR B1757-24. It was one of the largest recorded glitches, having magnitude $\Delta\Omega/\Omega \approx 2 \times 10^{-6}$. The decay was fitted using an exponential function, as given by the two-component model, having coupling constant of $\tau = 42 \pm 14$ days, and a post glitch spin down rate equal to the pre-glitch value (i.e., $Q = 1$ once again), further strengthening the case of the two-component model as a contributor of pulsar glitches.

Also in that year, Shemar & Lyne (1996, MNRAS 282, 677)presented a collection of results on the analysis of 25 glitches occurring in ten pulsars. They found that the older pulsars did not show substantial recovery towards pre-glitch frequencies (i.e., they displayed small Q values), while the younger pulsars displayed substantial recoveries towards pre-glitch rotation frequencies. However, almost all the glitches showed some form of recovery, which was easily modeled using a single exponential having a time constant of the order of 100 days, followed by a long term relaxation, of at least 1000 days (in most cases, this value was greater than the inter-glitch spacing). The few glitches for which this was not found to be the case were those for which the timing data was either too sparse or too noisy for accurate analysis. These findings were later confirmed by Lyne & Shemar (2000, MNRAS 315, 534).

In 2005, Shabanova (2005, MNRAS 356, 1435), published results on three glitches in PSR B1822-09. These glitches exhibited recovery, similar to that described by the two-component model, with timescales of $\tau \sim 100$ days, 235 days and 80 days respectively.

In summary: the two-component model is the oldest model still in use for explaining the glitch phenomenon. It has been scrutinized and revised numerous times, taking on various guises, with a more complex model being created each time. However, the basics of the model have remained the same; there exist two separate rotating components in a neutron star, which are coupled to one another by some coupling parameter. Theoretically, consensus on the composition and proportion of the two components has not been reached yet. Neither has there been agreement on the method of coupling between the two components, although vortex pinning and unpinning is the prime candidate. Observationally, there has been debate over the agreement of the model with the observational data of some glitches. This has, however, been the exception rather than the rule. For most glitches the data could be explained using an exponential fit, based on the two-component model.

A final remark: in 1980, Tsakadze & Tsakadze (1980, Journal of Low Temp. Phys 39m 649) published the results of various experiments performed over a number of years to determine whether the glitching behaviour of pulsars could be reproduced in labs using ^4He superfluidity. The experiments did in fact agree with expectations based on the two-component model of Baym et al. (1969). This further strengthens the simplicity of the two-component model adding weight to its validity.

Lecture Notes 5

1 Formation of an Accretion Disc

1.1 Circularization Radius

When the accreting gas has some angular momentum associated with it, then it forms a so called *accretion disc*. The accreting gas forms a disc around the central attracting object and a particular parcel of matter gradually spirals inward, thereby decreasing its negative gravitational potential energy. There is strong circumstantial evidence that this is the process by which energy is released in many astronomical systems. The first astronomical observations in the X-ray band (Giacconi et al. 1962) established the existence of compact X-ray sources. After the launch in 1970 of the satellite *Uhuru*, devoted completely to X-ray astronomy, it was possible to identify several compact X-ray sources with binary stellar systems. The most plausible model for such system is that the binary contains a neutron star (or black hole) and another "normal" gaseous star (called "donor star") of mass M_d, as sketched in Figure 1.

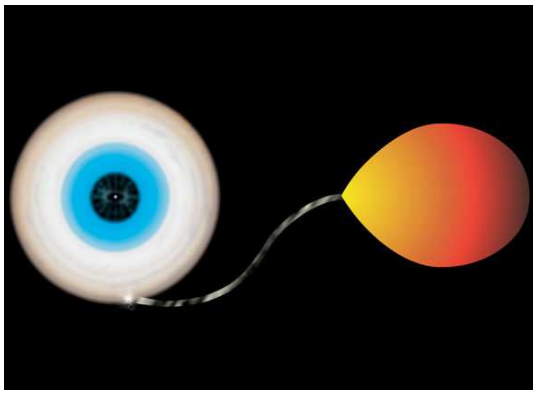

Figure 1: An X-ray binary, with a black hole accretor, an accretion disc and a low mass donor star ($M_d \lesssim 1\,M_\odot$). The donor star is filling the Roche lobe and transfers gas towards the accretor via the accretion disc.

Since the gas from the donor star has angular momentum due to the orbital motion of the donor around the center of mass of the binary, the gas flows towards the compact star by first forming an accretion disc. The gravitational energy lost is radiated in the form of X-rays.

Actually, the "accretor" can also be a white dwarf, although in that case the radiation emitted will not peak in X-rays but in the Optical/UV (for reason connected with the gravitational potential of the gas in the WD gravitational field, which is much weaker than in NS or BHs in the innermost regions of the accretion disc). Also the donor can indeed be different than a normal star, and be for example a brown or a white dwarf.

If the disc has a semi-thickness $h(r)$ and a radial coordinate r, then the thin disc model applies when $h(r)/r \ll 1$. This is true in many (but not all) X-ray binaries, and we will work

under this hypothesis from now on. The thin accretion disc model was studied in detail in a famous paper[1] by Shakura & Sunyaev (1973, Astron. Astrophys., Vol. 24, p. 337 - 355).

When the gas leaves the donor star via the Lagrangian point L_1 (see Fig. 13.12 in S&T), it has a substantial angular momentum which is equal to the specific orbital angular momentum. To a good approximation, we can take the stream trajectory as the orbit of a test particle released from rest at L_1, and thus with a given angular momentum, falling in the gravitational field of the accretor alone. This would give an elliptical orbit lying in the binary plane: the presence of the secondary causes this to precess slowly. (The presence of the donor star changes indeed the effective gravitational potential from the exact $1/r$ dependence which is required for closed periodic orbits, as it would be if only the accretor were present). A continuous stream trying to follow this orbit will therefore intersect itself, resulting in dissipation of energy via shocks. On the other hand, the gas has little opportunity to rid itself of the angular momentum it has on leaving L_1, so it will tend to the orbit of lowest energy for a given angular momentum, i.e., a circular orbit. We thus expect the gas initially to orbit the primary in the binary plane at a radius R_{circ} such that the Kepler orbit at R_{circ} has the same specific angular momentum as the transferring gas has on passing through L_1. Thus the gas will have a circular velocity:

$$v_\phi = \left(\frac{GM_d}{R_{circ}}\right)^{1/2} \tag{1}$$

where we have used cylindrical coordinates (r, ϕ, z), with ϕ the azimuthal coordinate. At this point the gas flows in a thin ring and has to expand in the radial direction (both inwards and outwards) to create an accretion disc. The inertial forces of the gas will be given by $R_{circ} \cdot v_\phi$, which for a typical accretion disc are equal to approximately $10^{18}\,\mathrm{cm^2\,s^{-1}}$. If the gas was collisionless, then the flow will proceed along the circularization ring and an accretion disc cannot form in any circumstance. The real gas flow instead will be an *hydrodynamic flow*, in the sense that collisions, and thus viscosity, has an important role. Now, to form a disc structure starting from a ring, part of the gas in the ring needs to move inward. But this is not sufficient, since the total angular momentum in the disc must be conserved because there is no external torque acting on the gas. Thus, to an inward motion there must be a corresponding outward gas motion. The gas spiraling inward/outward has less/more angular momentum than the original gas at the circularization radius. The net effect is that the total angular momentum is conserved and the original gas ring has now expanded into an extended accretion disc. To have this outward angular momentum transport, we need viscosity.

1.2 The Role of Viscosity

Viscosity is the fluid internal resistance to flow and can be though as the equivalent of a "fluid friction". All fluids have a viscosity, except ideal fluids and superfluids. The accreting gas behaves as a viscous fluid, and viscosity has a fundamental role for at least two reasons:

- for the formation of the accretion disc itself (the outward/inward gas motion)
- for conversion of gravitational potential energy into radiation

However, viscosity is perhaps the most poorly understood property of accreting discs, despite several decades of intensive research in this field. The reason, which might appear surprising

[1] this paper is one of the most cited in all fields of astronomy, with more than 5000 citations at the time of writing

at first, is that we do not understand completely the origin of viscosity in the accretion disc gas. Why ? Can't we treat viscosity just as in normal fluids we know in everyday life (like water) ? The short answer is: no, we cannot.

The accreting gas flowing at the circularization radius will be laminar or turbulent according to the Reynold number:

$$\Re = \frac{inertial\ forces}{viscous\ forces} = \frac{R\, v_\phi}{\nu} \tag{2}$$

where ν is the viscosity of the gas. In lab experiments, the $\Re$ number gives turbulence when it is larger than about $\sim$ 10–1000. Now, suppose the viscosity is given by the molecular scattering of particles, so that $\nu_{mol} = \lambda_{mfp}\, c_s$, where λ_{mfp} is the mean free path of gas particles and c_s is the sound speed. The molecular scattering is for example responsible for the viscosity of air or water, just to give two familiar examples. It might appear natural therefore to extend this also to accretion discs.

If the accreting gas is sufficiently hot, then it will constitute a plasma with each plasma particle having a mean free path $\lambda_{mfp} \simeq (n\sigma)^{-1} = \left(4\pi\, b_c^2\, n\right)^{-1}$. Here n is the number density of molecules (or particles in this case) and b_c is the "impact parameter" for plasma, which can be determined by considering the electrostatic interaction between the charged component of the plasma (for typical accretion discs the temperature is $\sim\ 10^4\,\mathrm{K}$ so a plasma of electrons and protons is formed, since the gas is mainly composed by Hydrogen which is ionized above $\approx$ 6500K). The impact parameter can be calculated by equating the electrostatic potential between Z charges of charge e and the thermal temperature of the plasma. Therefore $\left(Ze^2\right)/b_c \simeq (3/2)k_bT$. Numerically this gives $\lambda_{mfp} \approx 6 \times 10^4 \left(\frac{T^2}{n}\right)$ cm and $c_s \approx 10^4 T^{1/2}\,\mathrm{cm\,s^{-1}}$. Therefore the molecular viscosity can be calculated as $\nu_{mol} \simeq 6 \times 10^8\, T^{5/2} n^{-1}\,\mathrm{cm^2\,s^{-1}}$. For typical astrophysical accretion disc values ($n \approx 10^{15}\, cm^{-3}$, $T \approx 10^4 K$), the value of $\Re$ will be extremely high (typical values are $\Re > 10^{14}$). Therefore the molecular viscosity is completely irrelevant when compared to inertial forces, and therefore cannot be that important to determine the formation of an accretion disc (the situation will be almost identical to the collisionless gas case). To fix ideas, note that disturbances are propagated by viscous diffusion over a distance l on a time scale of order:

$$t_{visc} \sim\ l^2/\nu \tag{3}$$

where ν is the molecular viscosity. This timescale is about $\sim\ 10^8$ years for typical accretion disc sizes $l \sim\ 10^{10}$ cm and $\nu = 10^4\,\mathrm{cm^2\,s^1}$. This is orders of magnitude too long for the time variability seen in compact object accretion discs.

Since $\Re$ is very large for molecular viscosity, we might conjecture that the gas flow in accretion discs is also turbulent, although there is, as yet, no proof that this is so. If this is the case, the flow will be characterized by the size λ_{turb} and turnover velocity v_{turb} of the largest turbulent eddies. Since the turbulent motion is completely chaotic about the mean gas velocity, our simple viscosity calculations apply: there is a turbulent viscosity $\nu_{turb} = \lambda_{turb} v_{turb}$. Although this result has a neat appearance, it is here that our troubles with viscosity really begin. Turbulence is one of the major uncharted areas of classical physics and we do not understand the onset of turbulence, still less the physical mechanisms involved and how they determine the length-scale, λ_{turn}, and turnover velocity v_{turn}. The most we can do with present knowledge is to place plausible limits on these two parameters. First, the typical size of the largest turbulent eddies cannot exceed the disc thickness, so $\lambda_{turn} < H$, where we have used the disc scale height H. Second, it is unlikely that the turnover velocity is

supersonic, for, in this case, the turbulent motions would probably be thermalized by shocks. Thus, we can write:

$$\nu = \alpha\, c_s\, H \tag{4}$$

and expect $\alpha < 1$. This is the so-called α-prescription of think accretion discs, and all our ignorance on the viscosity ν has just been isolated in α. However, apart from the rather obvious expectation $\alpha < 1$ we have gained nothing so far. Nonetheless, the α-prescription has proved a useful parametrization of our ignorance and has encouraged a semi-empirical approach to the viscosity problem, which seeks to estimate the magnitude of α by a comparison of theory and observation. There is, for example, some reason to believe that $\alpha \sim 0.1$ in accretion discs in accreting white dwarfs, at least some of the time.

1.3 Rayleigh Stability

There is however still a fundamental problem in this whole picture, in the sense that even if we have quite reasonably assumed that turbulence plays a role in determining the viscosity of the accretion disc, we still have not identified any energy source to *sustain* this turbulence. The most obvious solution to this problem is to invoke a hydrodynamic instability. However, the angular velocity Ω of the gas in the accretion disc scales as $r^{-3/2}$ (i.e., Keplerian motion), and therefore the gas will obey the *Rayleigh stability criterion*:

$$\frac{d\left(r^2\Omega\right)^2}{dr} > 0 \tag{5}$$

The above simple criterion is derived and explained in the Appendix of these notes.

Any realistic accretion disc (in the thin accreton disc case for example $\Omega \propto r^{-3/2}$) is therefore linearly stable to pure hydrodynamic perturbations: they satisfy the classical Rayleigh criterion for axisymmetric perturbation by a safe margin and they are also stable to non-axisymmetric disturbances. Therefore even if the shear in the mean flow generates turbulence with the desired transport of angular momentum outward, the Reynolds turbulent stress will inevitably decay (because it is Rayleigh stable) unless some other effect feeds energy back into the turbulence.

1.4 Magneto-Rotational Instability (MRI)

A mechanism that currently looks very promising in that it seems to satisfy all the consistency requirements mentioned above and predicts a viscosity of the right magnitude and sign is magneto-hydrodynamic (MHD) turbulence. MHD turbulence differs from the hydrodynamic turbulence described above in the sense that magnetic fields have now an effect on the plasma flow and the Rayleigh stability criterion outlined above does not apply anymore. Now, it can be demonstrated that with a magnetic field in the plasma, the stability is guaranteed only if:

$$\frac{d\Omega^2}{dr} > 0 \tag{6}$$

Therefore for a Keplerian disc with $\Omega \propto r^{-3/2}$, the stability criterion is not satisfied and the disc becomes unstable to axisymmetric perturbations. This idea of MHD turbulence is not new, but has received fresh impetus in the context of accretion discs through the recent rediscovery by Balbus and Hawley (1991) of a weak field instability originally discussed by Velikhov (1959) and Chandrasekhar (1960, 1961). The addition of a magnetic field implies that even though the perturbations are still axisymmetric, the angular momentum of each

fluid element is no longer conserved because the magnetic field can apply stresses which move it about. Under these circumstances, instability to axisymmetric perturbations occurs when the angular velocity decreases outwards.

In an ideal plasma (i.e. a magnetized, perfectly conducting fluid), the action of the magnetic field is to link neighboring fluid parcels that lie along a common field line. One property of ideal plasmas is that magnetic field lines are frozen in the fluid parcels, so that the motion of the fluid carries the magnetic field along with it. This lends itself to a beautifully simple physical picture of the magnetic forces; the fluid parcels can be thought of a beads tied together on a string. If for some reason fluid parcels start to diverge, the tension in the magnetic string acts to bring the connected fluid elements back together (see Figure 2).

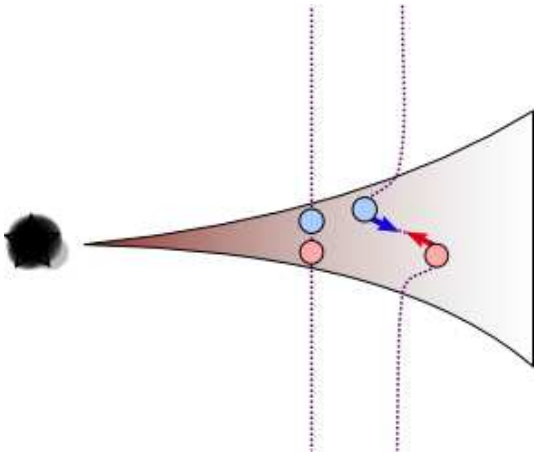

Figure 2: Motion of the fluid parcels (red and blue circles) carries the magnetic field lines (purple line) along with them. The resultant magnetic forces are shown (red and blue arrows) which arise due to tension in the field lines.

Intuitively, we can just think of the magnetic force acting on the fluid as springs tying neighboring fluid elements together! Normally, we think of a spring as a restoring force, which tends to preserve a systems stability. However, if the fluid motion occurs in a differentially rotating frame, it turns out that this restoring force can actually lead to the system destabilizing. Imagine that the red particle in Figure 2 is at a radius $r + \epsilon$ with respect to the blue particle, which is at a radius r. Here ϵ represents an infinitesimal displacement along the radial accretion disc coordinate. The red particle has therefore a slightly larger angular momentum than the blue particle and they start to rotate differentially on different Keplerian orbits. The blue particle will rotate faster than the red particle ($\Omega \propto r^{-3/2}$) and therefore the "magnetic spring", that connects the two particles, will pull the red particle. The pulled red particle will therefore increase its angular momentum, and move further outward. The blue particle will instead be slowed down by the red particle, and will move further inward. As the red and blue particles move outward and inward, the situation becomes worse and worse and the process keeps going on (see Figure 3). This is the origin of the MRI. Now it is clear how an accretion disc can develop from the simple circular ring of material we started with.

A schematic summary of the MRI goes as follows:

- a) Magnetically connected fluid elements have some initial displacement
- b) Differential rotation increases the displacement, and magnetic tension causes the

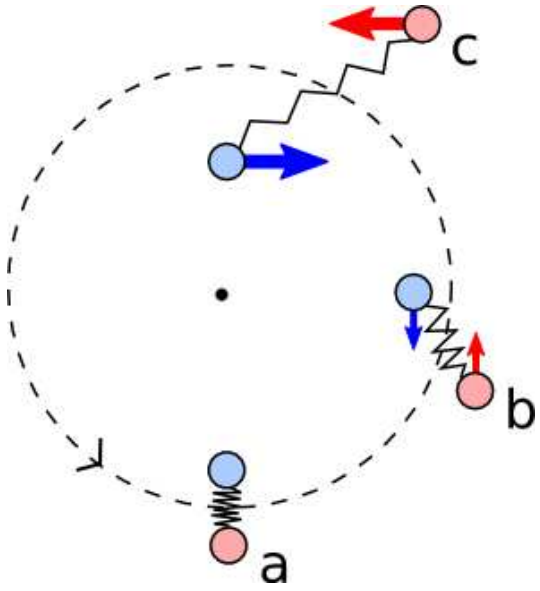

Figure 3: Growth of the MRI due to differential rotation

inner parcel to slow down, and the outer parcel to speed up

- c) This transfer of angular momentum causes the inner parcel to migrate inwards and the outer parcel to be pushed outwards.
- d) Repeat from step a, but now with a larger displacement.

However, if the magnetic field is too strong (i.e. the tension in the spring is too strong), the feedback cycle will not run. The tension will instead cause the displacement between fluid parcels to oscillate rather than grow in step b).

2 Accretion disc structure

An accretion disc can conveniently be divided into three distinct regions, depending on r. These regions are:

- 1. An outer region, at large r, in which gas pressure dominates radiation pressure and in which the opacity is controlled by free-free absorption;
- 2. A middle region, at smaller r, in which gas pressure dominates radiation pressure but the opacity is mainly due to electron scattering; and
- 3. An inner region, at very small r, in which radiation pressure dominates gas pressure and again, scattering dominates absorption in the opacity. (For some choices of h, the inner and middle regions may not exist at all.)

Typically, the transition from the outer to middle region occurs at a radius r_{om} when $\bar{k}_{ff} \simeq \bar{k}_{es}$,which occurs at

$$\frac{2r_{om}}{r_s} = 4 \times 10^3 \left(\frac{\dot{M}_{17}}{M/M_\odot} \right)^{2/3} \tag{7}$$

Here $r_s = 2GM/c^2$ is the Schwarzschild radius, $\dot{M}_{17}$ is in units of $10^{17}\,\mathrm{g\,s^{-1}}$. The transition from the middle to the inner region (point r_{mi}) instead can be found when $P_{gas} \simeq P_{rad}$:

$$\frac{2r_{mi}}{r_s} = 80\alpha^{2/21}\left(\frac{M}{M_\odot}\right)^{-2/3}\dot{M}_{17}^{16/21} \tag{8}$$

Now, the gas flows in the accretion disc because of some form of turbulent viscosity (like the MRI sketched above), and this process heats up the disc from the outer regions towards the innermost part. The viscosity generates entropy (heat) which is released as thermal radiation. The integrated flux emitted from the top and bottom faces of the disk at r is:

$$F(r) = h(r)\dot{Q} = \frac{3\dot{M}}{8\pi\, r^2}\frac{GM}{r}\left[1 - \left(\frac{r_I}{r}\right)^{1/2}\right] \tag{9}$$

with $\dot{Q}$ the heat production rate, r_I the innermost disc radius and h the semi-thickness of the disc at radius r. The curious reader can find the derivation of this equation in §14.5 of the textbook S&T. Here we need just to know that this equation is found when assuming that no heat is stored in the disc but it is totally radiated away, that the disc has a shear flow and that mass and angular momentum are conserved.

To summarize: the particle leaves the donor at the lagrangian point L_1 and enters the gravitational potential well of the accretor. The gravitational potential energy is converted in kinetic energy of the particle that rotates around the circularization radius. The kinetic energy is then converted into heat by viscosity, which is also responsible for the formation of the accretion disc. As the disc forms, an inward matter flow starts, with the gas sinking in the potential well of the accretor. More gravitational potential energy is converted into kinetic energy, and more kinetic energy is converted into heat. The heat is finally released as thermal radiation, with the inner regions which are hotter than the outer regions, because of the energy release of gravitational potential energy. The thermal radiation is then detected from the distant observer.

2.1 Vertical disc structure

So far we have discussed the horizontal structure of a thin accretion disc. The vertical structure can be calculated very easily by assuming that the gas is in hydrostatic equilibrium along the vertical coordinate z. This means that the vertical component of the gravitational force of the compact object has to be equal to the vertical pressure gradient. Therefore:

$$\frac{1}{\rho}\frac{dP}{dz} = -\frac{GM}{r^2}\frac{z}{r} \tag{10}$$

with $z \ll r$. Replacing the differentials with finite differences (that is, setting $\Delta P \approx P$, where P is the pressure at $z = 0$, and setting $\Delta z = h$) yields:

$$h = \left(\frac{P}{\rho}\right)^{1/2}\left(\frac{r^3}{GM}\right)^{1/2} \approx \frac{c_s}{\Omega} \tag{11}$$

where c_s is the sound speed in the disc midplane, and where we have used the definition of the sound speed $c_s = \sqrt{P/\rho}$.

Appendix: The Rayleigh Stability Criterion

Consider a fluid of uniform density rotating differentially around its axis of symmetry (like an accretion disc). To find the condition for its stability, we suppose that a fluid ring at a distance r_0 from the axis moving with velocity v_0 is interchanged with a fluid ring at a greater distance r_1 (i.e., $r_1 > r_0$) moving with velocity v_1. The system is stable if the displaced fluid rings tend to return to their initial positions, whereas it is unstable if the displaced rings move further away. Assuming conservation of angular momentum, we conclude that the fluid ring displaced from r_0 to r_1 acquires a velocity $(r_0/r_1)v_0$. The ring previously at r_1 had a centripetal acceleration v_1^2/r_1, which must have been provided by the various forces there such as the part of the pressure gradient left after balancing gravity. The ring brought to r_1 now acquires a centripetal acceleration $r_0^2 v_0^2/r_1^3$ to remain in its new position. If this is less that v_1^2/r_1, then we expect the forces present there to push the ring inward towards its initial position. The condition for stability is then

$$\frac{r_0^2 v_0^2}{r_1^3} < \frac{v_1^2}{r_1} \tag{12}$$

so that:

$$\left(r_0^2 \Omega_0\right)^2 < \left(r_1^2 \Omega_1\right)^2 \tag{13}$$

where $\Omega_n = v_n r_n$ is the angular velocity at radius r_n. This stability condition can be rewritten as:

$$\frac{d\left(r^2\Omega\right)^2}{dr} > 0 \tag{14}$$

Lecture Notes 6

1 Radio Pulsar Timing

1.1

Pulsars are excellent celestial clocks. The period of the first pulsar was found to be stable to one part in 10^7 over a few months. Following the discovery of the millisecond pulsar B1937+21 it was demonstrated that its period could be measured to one part in 10^{13} or better. This unrivaled stability leads to a host of applications including uses as time keepers, probes of relativistic gravity and natural gravitational wave detectors.

Each pulsar is typically observed at least once or twice per month over the course of a year to establish its basic properties. Figure 1 summarise the essential steps involved in a pulse time-of-arrival (TOA) measurement.

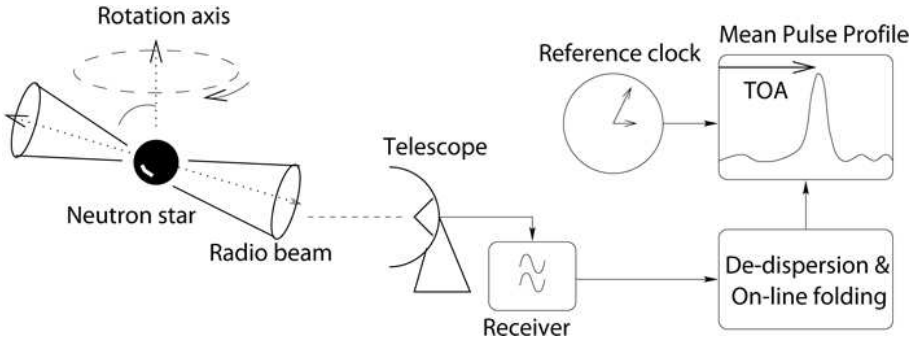

Figure 1: Schematic showing the main stages involved in pulsar timing observations.

Pulses from the neutron star traverse the interstellar medium before being received at the radio telescope where they are dedispersed and added to form a mean pulse profile. During the observation, the data regularly receive a time stamp, usually based on a caesium time standard or hydrogen maser at the observatory plus a signal from the Global Positioning System of satellites (GPS). The TOA is defined as the arrival time of some fiducial point on the integrated profile with respect to either the start or the midpoint of the observation. It is observed that a pulse profile has a stable form at any given observing frequency, so that the TOA can be accurately determined by cross-correlation of the observed profile with a high S/N template profile obtained from the addition of many observations at the particular observing frequency.

To model the rotational behaviour of the neutron star we ideally require TOAs measured by an inertial observer. An observatory located on Earth experiences accelerations with respect to the neutron star due to the Earths rotation and orbital motion around the Sun and is therefore not in an inertial frame. To a very good approximation, the solar system centre-of-mass (barycentre) can be regarded as an inertial frame. It is now standard practice to transform the observed TOAs to this frame using a planetary ephemeris such as the JPL DE405 (JPL DE stands for "Jet Propulsion Laboratory Development Ephemeris"). The

ephemerides DE405 are expressed in coordinates referred to the ICRF (International Celestial Reference Frame). These ephemerides provide solar-system barycentric coordinates of position and velocity for the Earth-Moon barycenter, Sun and eight major planets, and geocentric coordinates of the Moon. In this way the solar system barycenter can be determined with great accuracy for any specified period epoch.

The transformation between barycentric time (t) and observed time (τ) takes the form:

$$t - \tau = \frac{\bar{r} \cdot \bar{s}}{c} + + \frac{(\bar{r} \cdot \bar{s})^2 - |\bar{r}|^2}{2cd} + \Delta_{S_\odot} + \Delta_{E_\odot} - \Delta\, D/f^2 \tag{1}$$

Here $\bar{r}$ is the position of the observatory with respect to the barycentre, $\bar{s}$ is a unit vector in the direction of the pulsar at a distance d and c is the speed of light. The first term on the right hand side of the equation is the light travel time from the observatory to the solar system barycentre. This term is usually called the "Roemer delay". The second term is the "annual parallax". Incoming pulses from all but the nearest pulsars can be approximated by plane wavefronts. The second term, which represents the delay due to spherical wavefronts, yields the parallax and hence the distance d. This has so far only been measured for five nearby millisecond pulsars.

The term $\Delta_{s_\odot}$ represents the Shapiro corrections due to general relativistic time delays in the solar system. This correction arises because of the extra delays due to the curvature of space-time caused by thre presence of masse sin the solar system. The Shapiro delay can be calculated as:

$$\Delta\, t = -\frac{2GM}{c^3} ln(1 - \bar{s} \cdot \bar{x}) \tag{2}$$

where $\bar{s}$ is again the vector pointing from the observed to the source and $\bar{x}$ is the vector pointing from the observer to the gravitating mass M. The delays are largest for a signal passing the limb of the Sun ($\sim 120 \mu\, s$) while Jupiter can contribute as much as 200 ns. In principle one has to sum over all the solar system bodies, with the effect being larger for the most massive bodies and for those bodies which are grazed by the radio beam. In practice, giving the timing precision usually achievable with typical pulsars, only the Sun and in a few occasions Jupiter, need to be considered for this correction. The third term, the "Einstein delay", describes the combined effect of time dilation due to the motion of the Earth and gravitational redshift caused by the other bodies in the solar system. This time varying effect takes into account the variation of an atomic clock on Earth in the changing gravitational potential as the Earth follows its elliptical orbit around the Sun. Since measurements can be carried out at different observing frequencies f with different dispersive delays D, TOAs are generally referred to the equivalent time that would be observed at infinite frequency. This transformation is the term $\Delta\, D/f^2$. The quantity $\Delta\, D$ is related to the dispersion measure DM via a constant $k = \frac{e^2}{2\pi\, m_e c}$ (cf. Eq. 10.4.13 on Shapiro & Teukolsky) and f is the frequency of the electromagnetic wave observed (*not* the pulsar frequency !).

Following the accumulation of a number of TOAs, a surprisingly simple model is usually sufficient to account for the TOAs during the time span of the observations and to predict the arrival times of subsequent pulses. The model is a Taylor expansion of the rotational frequency $\Omega = 2\pi P$ (with P equal to the pulsar spin period) about a model value Ω_0 at some reference epoch t_0. The model pulse phase is:

$$\phi(t) = \phi_0 + \Omega_0(t - t_0) + \frac{1}{2}\dot{\Omega}_0(t - t_0)^2 + \frac{1}{3!}\ddot{\Omega}_0(t - t_0)^3 + ... \tag{3}$$

where t is the barycentric time and ϕ_0 is the pulse phase at t_0. Based on this simple model, and using initial estimates of the position, dispersion measure and pulse period,

a timing residual is calculated for each TOA as the difference between the observed and predicted pulse phases.

Ideally, the residuals should have a zero mean and be free from any systematic trends. To reach this point, however, the model needs to be refined in a bootstrap fashion. Early sets of residuals will exhibit a number of trends indicating a systematic error in one or more of the model parameters, or a parameter not incorporated into the model (see Figure 2).

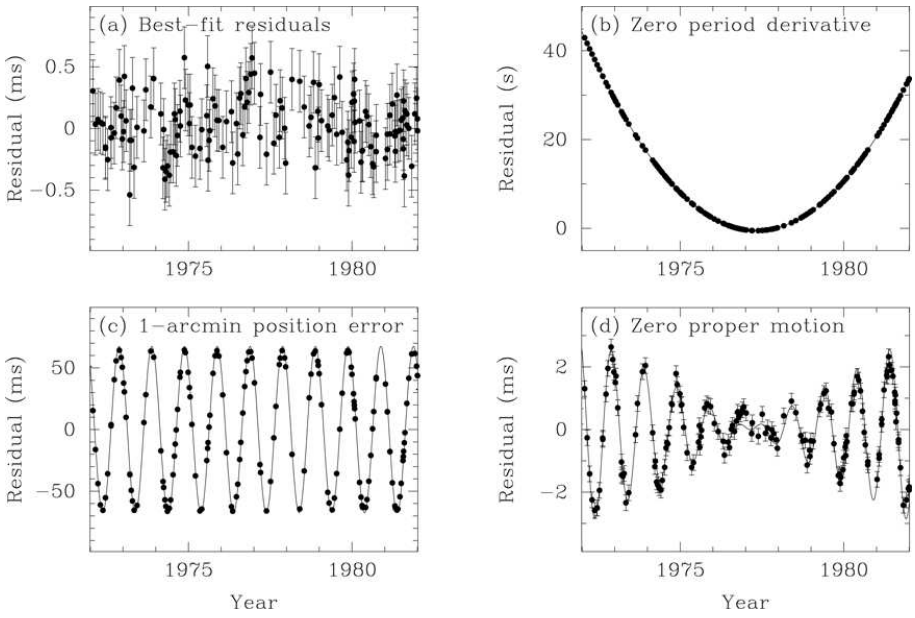

Figure 2: Timing model residuals for PSR B1133+16. Panel a: Residuals obtained from the best-fitting model which includes period, period derivative, position and proper motion. Panel b: Residuals obtained when the period derivative term is set to zero. Panel c: Residuals showing the effect of a 1-arcmin position error. Panel d: Residuals obtained neglecting the proper motion. The lines in Panels bd show the expected behaviour in the timing residuals for each effect.

From Equation (2), an error in the assumed Ω_0 results in a linear slope with time. A parabolic trend results from an error in $\dot{\Omega}_0$ (see Panel b of Figure 2). Additional effects will arise if the assumed position of the pulsar (the unit vector $\bar{s}$ in Equation 1) used in the barycentric time calculation is incorrect. A position error results in an annual sinusoid (see Panel c of Figure 2). A proper motion produces an annual sinusoid of linearly increasing magnitude (see Panel d of Figure 2).

After a number of iterations, and with the benefit of a modicum of experience, it is possible to identify and account for each of these various effects to produce a timing solution which is phase coherent over the whole data span. Phase coherent or "phase-connected" means that the model is able to describe each rotation of the neutron star at any time within your data span. Therefore given a certain arbitrary time $\hat{t}$ that lies in the data window where the timing model is accurate, it is possible to predict the exact position of the neutron star in its rotational motion with a precision better than 0.5 cycles (usually this precision is much better than 0.5 cycles, typical values being less than 0.01-0.1 cycles).

The resulting model parameters provide spin and astrometric information with a precision which improves as the length of the data span increases. For example, timing observations of the original millisecond pulsar B1937+21, spanning almost 9 years (exactly 165,711,423,279

rotations!), measure a period of $1.5578064688197945 \pm 0.0000000000000004$ ms defined at midnight Universal Time on December 5, 1988! Measurements of other parameters are no less impressive, with astrometric errors of 3 μarcsec being presently possible for the bright millisecond pulsar J04374715.

1.2 Braking index

Pulsars have strong dipole magnetic fields and according to classical electrodynamics the radiation power equals the spin-down luminosity:

$$\dot{E}_{dipole} = \frac{2}{3c^3}|m|^2\Omega^4 \, sin^2 \, \alpha = \dot{E}_{sd} = -I\Omega\dot{\Omega} \tag{4}$$

(see §10.5 on Shapiro & Teukolsky). By using the pulsar spin frequency $\nu = 1/P$ rather than Ω, it follows that:

$$\dot{\nu} = -K\nu^n \tag{5}$$

where n is the *braking index* and K is usually assumed to be constant. We expect that $n = 3$ for pure magnetic dipole spin down, since:

$$\dot{\Omega} = -\left(\frac{2|m|^2 sin^2\alpha}{3I\,c^3}\right)\Omega^3 \tag{6}$$

which follows from Eq. 1.2. However, other dissipation mechanisms may exist (e.g., a wind or outflow of particles) that carry away angular momentum from the neutron star other than the pure magneto dipole spin down. The braking index can be determined if, in addition to $\dot{\nu}$, a second spin frequency derivative $\ddot{\nu}$ can be measured. Differentiating the above expression for $\dot{\nu}$ we have:

$$n = \frac{\nu\ddot{\nu}}{\dot{\nu}^2} \tag{7}$$

Usually the observed $\ddot{\nu}$ is contaminated by the so-called "timing noise" (see next section). However, for a few pulsars, a determination of the braking index that reflects the true spin=down behaviour has been determined. Values range from $n = 1.4$ to $n = 2.9$. These deviations indicate that the assumption $n = 3$ is usually **not correct** ! Nevertheless, this supposition is usually made to *define* a number of useful quantities that are used to characterise the basic properties of a radio pulsar. It should always be kept in mind that the resulting quantities are rough model-dependent estimates.

1.3 Timing noise

Timing noise manifests itself in a quasi-random walk in one or more of the rotational parameters on timescales of months to years (see Figure 3).

It appears mostly for *young* pulsars and scales with some power of the period derivative $\dot{P}$. The older millisecond pulsars show much less timing noise, althougj it is observed in some sources, albeit with a musch smaller amplitude. It nevertheless affects the timing precision of some millisecond pulsars, with the fast rotating pulsar B1937+21 ($P = 1.56$ ms) as the most prominent example (see for example, Kaspi et al. 1994, The Astrophysical Journal, vol. 428, no. 2, pt. 1, p. 713-728). Usually, timing noise is recognized by the measurement of the second derivative of the pulsar spin frequency $\ddot{\nu}$, that is *hugely inconsistent* with that

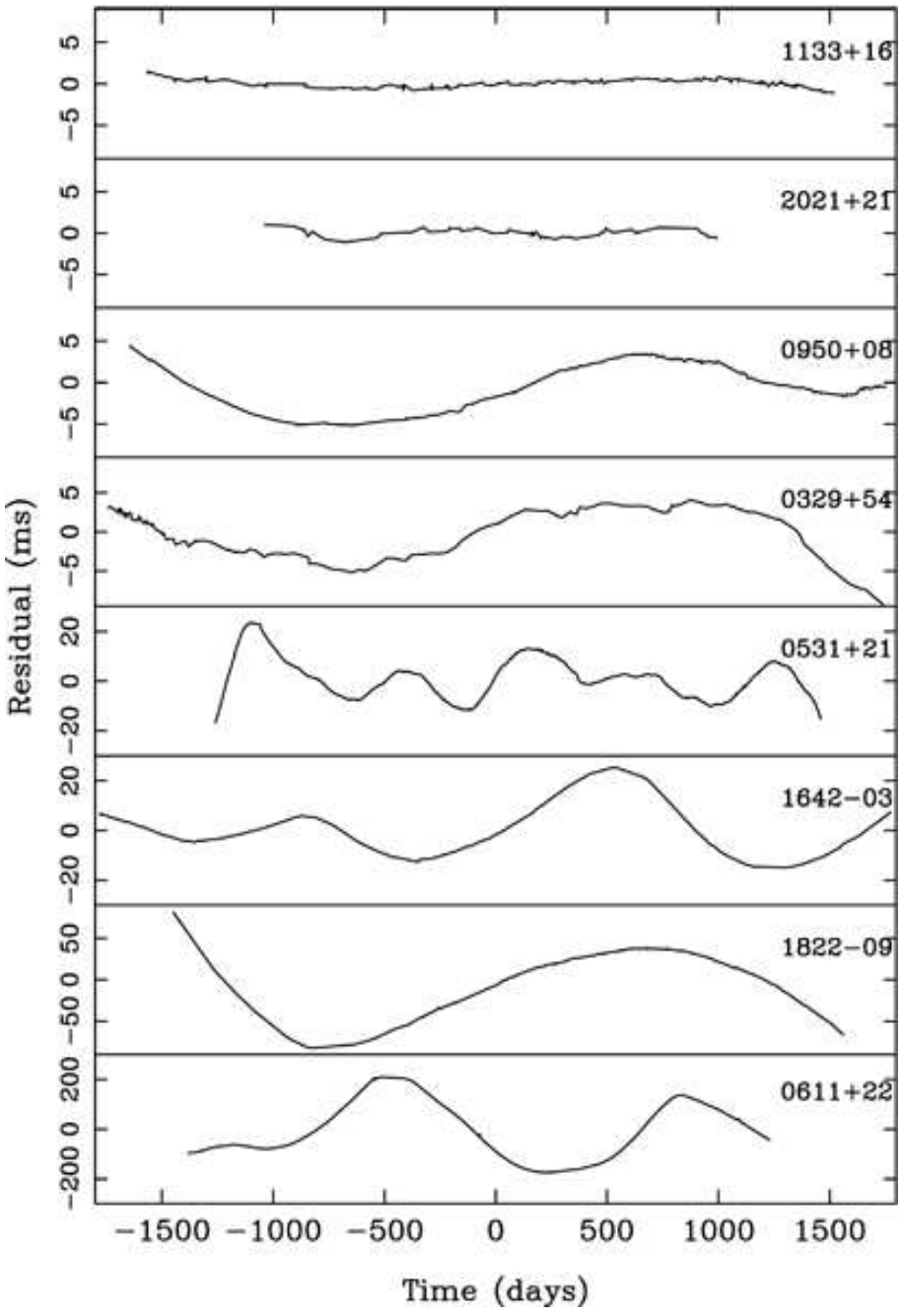

Figure 3: Examples of timing residuals for a number of young radio pulsars. Note the varying scale on the ordinate axis, the pulsars being ranked in increasing order of timing activity.

expected for a standard pulsar spin down evolution with a braking index of $n = 3$. By inverting the braking index equation, we can obtain an expected $\ddot{\nu}$ value as:

$$\ddot{\nu} = \frac{n\dot{\nu}^2}{\nu} \tag{8}$$

For a typical young pulsar with $\nu = 1$ Hz and $\dot{\nu} = 10^{-15}\,\mathrm{Hz\,s^{-1}}$, the expected value for $n = 3$ is $\ddot{\nu} \sim 10^{-30}\,\mathrm{Hz\,s^{-2}}$. Such values are measurable only for the youngest pulsars. Often, howeverm the timing analysis produces significant values for $\ddot{\nu}$, positive and negative, that are much larger than this. Unless other explanations are put forward, a measurament like this is a clear indication for timing noise.

But what is the physical meaning of timing noise in radio pulsars ? And why is it almost completely absent in old millisecond pulsars ? While the physical processes of this phenomenon are not well understood, it seems likely that they may be connected to superfluid processes and temperature changes in the interior of the neutron star, or to processes in the magnetosphere. In both cases the relative weak strength of the old millisecond pulsar magnetic fields might be connected with the relatively small activity of timing noise in these pulsars. However, after about 50 years of pulsar observations, the physical meaning of timing noise remains a big mistery of neutron star science.

1.4 Timing binary pulsars

For binary pulsars, the timing model introduced above needs to be extended to incorporate the additional motion of the pulsar as it orbits the common centre-of-mass of the binary system. Describing the binary orbit using Kepler's laws to refer the TOAs to the binary barycentre requires five model parameters: the orbital period P_b, projected semi-major axis $x = a_p \cdot sin(i)$ (a_p is the semimajor axis of the pulsar orbit), orbital eccentricity e, longitude of periastron ω, and the epoch of periastron passage T_0. This description, using five "Keplerian parameters", is identical to that used for spectroscopic binary stars.

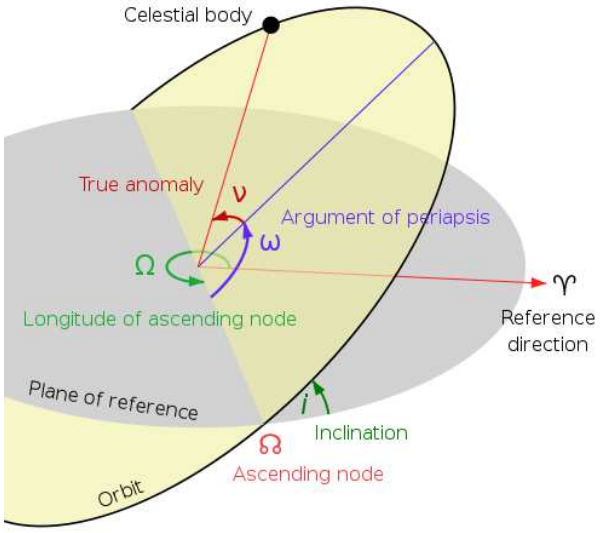

Figure 4

The orbital elements can be defined as follows:

- inclination (i): angle between the orbital plane and a reference plane.
- argument of pericenter (ω): the angle from the ascending node to the body in the orbital plane (denoted when measured from any point in the orbital plane).
- semimajor axis (a_p): half the major axis of the neutron star orbit's ellipse.
- eccentricity (e)
- epoch of periastron (T_0): time at which the pulsar passed through the periastron.

Analogous to the radial velocity curve in a spectroscopic binary, for binary pulsars the orbit is described by the apparent pulse period against time. An example of this is shown in Panel a of Figure 5. Alternatively, when radial accelerations can be measured, the orbit can also be visualised in a plot of acceleration versus period as shown in Panel b of Figure 5. This method is particularly useful for determining binary pulsar orbits from sparsely sampled data.

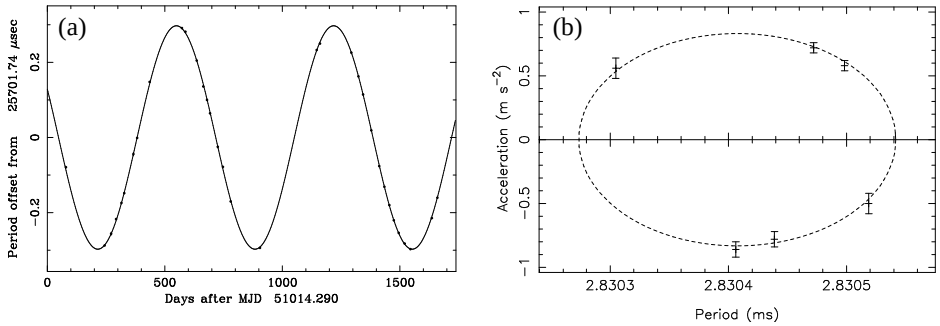

Figure 5: Panel a: Keplerian orbital fit to the 669-day binary pulsar J0407+1607. Panel b: Orbital fit in the period-acceleration plane for the globular cluster pulsar 47 Tuc S.

Constraints on the masses of the pulsar $m_{\rm p}$ and the orbiting companion $m_{\rm c}$ can be placed by combining x and $P_{\rm b}$ to obtain the mass function

$$f_{\rm mass} = \frac{4\pi^2}{G}\frac{x^3}{P_{\rm b}^2} = \frac{(m_{\rm c}\sin i)^3}{(m_{\rm p}+m_{\rm c})^2}, \tag{9}$$

where G is Newton's gravitational constant and i is the (initially unknown) angle between the orbital plane and the plane of the sky (i.e. an orbit viewed edge-on corresponds to $i = 90°$). The mass function can be easily derived from the Kepler laws.

In the absence of further information, the standard practice is to consider a random distribution of inclination angles. Since the probability that i is *less* than some value i_0 is $p(< i_0) = 1 - \cos(i_0)$, the 90% confidence interval for i is $26° < i < 90°$. For an assumed pulsar mass, the 90% confidence interval for $m_{\rm c}$ can be obtained by solving Equation (9) for $i = 26°$ and $90°$.

Although most of the presently known binary pulsar systems can be adequately timed using Kepler's laws, there are a number which require an additional set of "post-Keplerian" (PK) parameters which have a distinct functional form for a given relativistic theory of gravity. In general relativity (GR) the PK formalism gives the relativistic advance of periastron

$$\dot{\omega} = 3\left(\frac{P_{\rm b}}{2\pi}\right)^{-5/3}(T_\odot M)^{2/3}(1-e^2)^{-1}, \tag{10}$$

the time dilation and gravitational redshift parameter

$$\gamma = e \left(\frac{P_\mathrm{b}}{2\pi}\right)^{1/3} T_\odot^{2/3} M^{-4/3} m_\mathrm{c}(m_\mathrm{p} + 2m_\mathrm{c}), \tag{11}$$

the rate of orbital decay due to gravitational radiation

$$\dot{P}_\mathrm{b} = -\frac{192\pi}{5} \left(\frac{P_\mathrm{b}}{2\pi}\right)^{-5/3} \left(1 + \frac{73}{24}e^2 + \frac{37}{96}e^4\right) \left(1 - e^2\right)^{-7/2} T_\odot^{5/3} m_\mathrm{p} m_\mathrm{c} M^{-1/3} \tag{12}$$

and the two Shapiro delay parameters

$$r = T_\odot m_\mathrm{c} \tag{13}$$

and

$$s = x \left(\frac{P_\mathrm{b}}{2\pi}\right)^{-2/3} T_\odot^{-1/3} M^{2/3} m_\mathrm{c}^{-1} \tag{14}$$

which describe the delay in the pulses around superior conjunction where the pulsar radiation traverses the gravitational well of its companion. In the above expressions, all masses are in solar units, $M \equiv m_\mathrm{p} + m_\mathrm{c}$, $x \equiv a_\mathrm{p} \sin i/c$, $s \equiv \sin i$ and $T_\odot \equiv GM_\odot/c^3 = 4.925490947\ \mu$s. Some combinations, or all, of the PK parameters have now been measured for a number of binary pulsar systems. Further PK parameters due to aberration and relativistic deformation are not listed here but may soon be important for the double pulsar.

1.5 Testing general relativity

The key point in the PK definitions introduced in the previous section is that, given the precisely measured Keplerian parameters, the only two unknowns are the masses of the pulsar and its companion, m_p and m_c. Hence, from a measurement of just two PK parameters (e.g., $\dot{\omega}$ and γ) one can solve for the two masses and, using Equation (9), find the orbital inclination angle i. If three (or more) PK parameters are measured, the system is "overdetermined" and can be used to test GR (or, more generally, any other theory of gravity) by comparing the third PK parameter with the predicted value based on the masses determined from the other two.

The first binary pulsar used to test GR in this way was PSR B1913+16 discovered by Hulse & Taylor in 1974. Measurements of three PK parameters ($\dot{\omega}$, γ and $\dot{P}_\mathrm{b}$) were obtained from long-term timing observations at Arecibo. The measurement of orbital decay, which corresponds to a shrinkage of about 3.2 mm per orbit, is seen most dramatically as the gradually increasing shift in orbital phase for periastron passages with respect to a non-decaying orbit shown in Figure 8. This figure includes recent Arecibo data taken since the upgrade of the telescope in the mid 1990s.

The measurement of orbital decay in the B1913+16 system, obtained from observations spanning a 30-yr baseline, is within 0.2% of the GR prediction and provided the first indirect evidence for the existence of gravitational waves. Hulse and Taylor were awarded the 1993 Nobel Physics prize in recognition of their discovery of this remarkable laboratory for testing GR. A similar, though less precise, test of GR from this combination of PK parameters has recently been performed in the double neutron star binary PSR B2127+11C in the globular cluster M15. For this system, the measurement uncertainties permit a test of GR to 3% precision which is unlikely to improve due to various kinematic contaminations including the acceleration of the binary system in the cluster's gravitational potential.

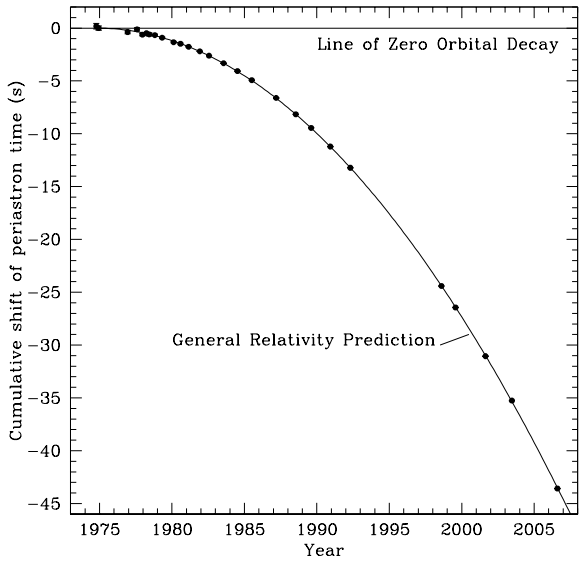

Figure 6: Orbital decay in the binary pulsar B1913+16 system demonstrated as an increasing orbital phase shift for periastron passages with time. The GR prediction due entirely to the emission of gravitational radiation is shown by the parabola. Figure provided by Joel Weisberg.

Five PK parameters have been measured for the double pulsar discussed in detail below and for the double neutron star system PSR B1534+12 where the test of GR comes from measurements of $\dot{\omega}$, γ and s. In this system, the agreement between theory and observation is within 0.7%. This test will improve in the future as the timing baseline extends and a more significant measurement of r can be made.

Currently the best binary pulsar system for testing GR in the strong-field regime is the double pulsar J0737−3039. In this system, where two independent pulsar clocks can be timed, five PK parameters of the 22.7-ms pulsar "A" have been measured as well as two additional constraints from the measured mass function and projected semi-major axis of the 2.7-s pulsar "B". A useful means of summarising the limits so far is Figure 7 which shows the allowed regions of parameter space in terms of the masses of the two pulsars. The shaded regions are excluded by the requirement that $\sin i < 1$. Further constraints are shown as pairs of lines enclosing permitted regions as predicted by GR. The measurement of $\dot{\omega} = 16.899 \pm 0.001$ deg yr^{-1} gives the total system mass $M = 2.5871 \pm 0.0002\, M_\odot$.

The measurement of the projected semi-major axes of both orbits gives the mass ratio $R = 1.071 \pm 0.001$. The mass ratio measurement is unique to the double pulsar system and rests on the basic assumption that momentum is conserved. This constraint should apply to any reasonable theory of gravity. The intersection between the lines for $\dot{\omega}$ and R yield the masses of A and B as $m_\mathrm{A} = 1.3381 \pm 0.007\, M_\odot$ and $m_\mathrm{B} = 1.2489 \pm 0.0007\, M_\odot$. From these values, using Equations (11–14) the expected values of γ, $\dot{P}_\mathrm{b}$, r and s may be calculated and compared with the observed values. These four tests of GR all agree with the theory to within the uncertainties. Currently the tightest constraint is the Shapiro delay parameter s where the observed value is in agreement with GR at the 0.05% level.

Another unique feature of the double pulsar system is the interactions between the two pulsars' radiation beams. Specifically, the signal from A is eclipsed for 30 s each orbit by the

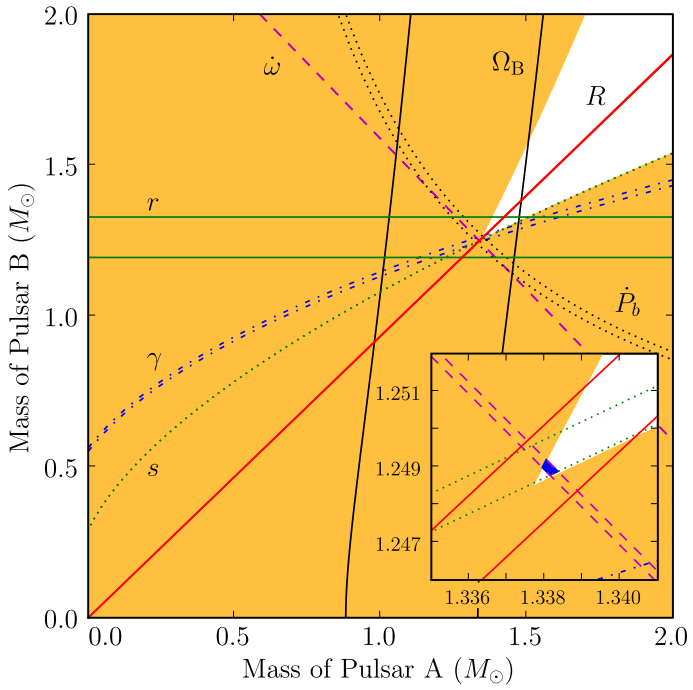

Figure 7: 'Mass–mass' diagram showing the observational constraints on the masses of the neutron stars in the double pulsar system J0737−3039. Inset is an enlarged view of the small square encompassing the intersection of the tightest constraints. The shaded region is excluded by classical Keplerian motion, i.e., it is excluded by the mass function. The white area is further constrained by using PK parameters as predicted by general relativity. The measurement of $\dot{\omega}$ gives the total system mass (equal to $\approx 2.59\,M_\odot$. The PK parameter R gives the mass ratio $m_A/m_B = 1.07$. The parameter γ gives the gravitational redshift/time dilation. The parameters r and s are the two Shapiro delay parameters. $\dot{P}_b$ is the orbital period decay due to gravitational wave emission. This particular pulsar represents so far (May 2011) the most precise test of general relativity in strong gravity field regime with a precision of 0.05%.

magnetosphere of B and the radio pulses from B are modulated by the relativistic wind from A during two phases of the orbit. These provide unique insights into plasma physics and, as shown shown in Figure 7, a measurement of relativistic spin–orbit coupling in the binary system and, hence, another constraint on GR. By careful modeling of the change in eclipse profiles of A over a four-year baseline, Breton et al. have been able to fit a remarkably simple model and determine the precession of B's spin axis about the orbital angular momentum vector. This remarkable measurement agrees, within the 13% measurement uncertainty, to the GR prediction.

It took only two years for the double pulsar system to surpass the tests of GR possible from three decades of monitoring PSR B1913+16 and over a decade of timing PSR B1534+12. On-going precision timing measurements of the double pulsar system should soon provide even more stringent and new tests of GR. Crucial to these measurements will be the timing of the 2.7-s pulsar B, where the observed profile is significantly affected by A's relativistic wind. A careful decoupling of these profile variations is required to accurately measure TOAs for this pulsar and determine the extent to which the theory-independent mass ratio R can be measured. This task is compounded by the fact that pulsar B's signal is now significantly weaker compared to the discovery observations five years ago. This appears to be a direct result of its beam precessing out of our line of sight.

The relativistic effects observed in the double pulsar system are so large that corrections to higher post-Newtonian order may soon need to be considered. For example, $\dot{\omega}$ may be measured precisely enough to require terms of second post-Newtonian order to be included in the computations. In addition, in contrast to Newtonian physics, GR predicts that the spins of the neutron stars affect their orbital motion via spin-orbit coupling. This effect would most clearly be visible as a contribution to the observed $\dot{\omega}$ in a secular and periodic fashion. For the J0737−3039 system, the expected contribution is about an order of magnitude larger than for PSR B1913+16. As the exact value depends on the pulsars' moment of inertia, a potential measurement of this effect would allow the moment of inertia of a neutron star to be determined for the first time. Such a measurement would be invaluable for studies of the neutron star equation of state and our understanding of matter at extreme pressure and densities.

Finally, in the neutron star–white dwarf binary J1141−6545, the measurement of $\dot{P}_b$, $\dot{\omega}$, γ and s permit a 6% test of GR. Despite this being nominally only a relatively weak constraint for GR, the significantly different self gravities of the neutron star and white dwarf permits constraints on two coupling parameters between matter and the scalar gravitational field. These parameters are non-existent in GR, which describes gravity in terms of a tensor field, but are predicted by some alternative theories of gravity which consider tensor and scalar field components. The constraints provided by PSR J1141−6545 for the strongly non-linear coupling parameter are currently the most stringent to date and are set to improve dramatically over the next five years of timing observations.

2 Accreting Magnetized Compact Objects

As we have seen in Lecture 10, magnetized compact objects like neutron stars or white dwarfs can accrete by channeling plasma along the magnetic field lines in the innermost regions of the disc. When the accretor is a white dwarf the binary system is called a *cataclysmic variable*, which shows outburst-quiescence cycles as accreting neutron stars (see Lecture 10). There is a very large sub-classification of cataclysmic variables often named after a bright prototype star characteristic of the class, so their names are not particularly enlightening. What is

important here is that the magnetic field of the white dwarf can be in some cases sufficiently intense to disrupt the accretion disc (intermediate polar) or even to prevent the formation of the accretion disc altoghether (polar). The term "polar" refers to the polsarization of the emitted optical light, which is highly polarized for strongly magnetized objects. Polars and intermediate polars do show pulsations in several wavelength, which unveil the rotational period of the white dwarf. The highest magnetic field observed so far in polsar can reach values up to 10^8 G.

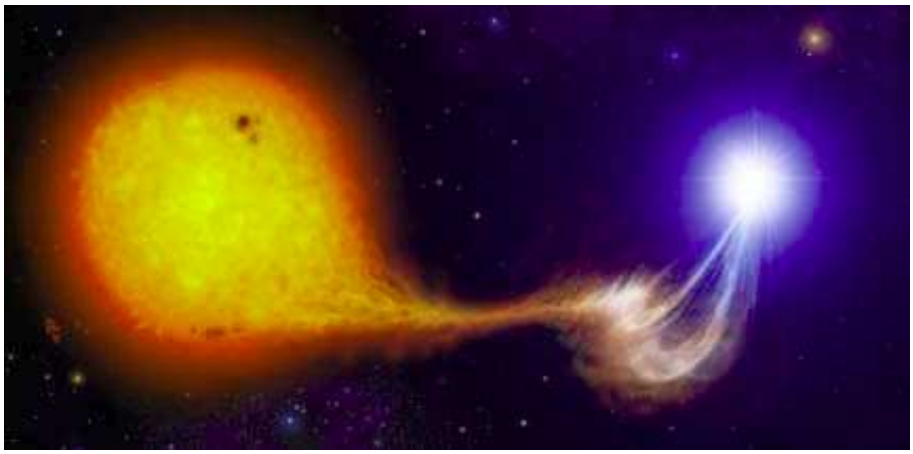

Figure 8: Artist illustration of a "polar". The magnetic field of the white dwarf is so intense to prevent the formation of an accretion disc.

When the accretor is a neutron star we can split the classification in two broad classes: low mass X-ray binaries (NS-LMXBs) and high mass X-ray binaries (NS-HMXBs). The "low" and "high" mass refers to the companion star. In this section we will focus on accreting neutron stars, mostly on NS-LMXBs.

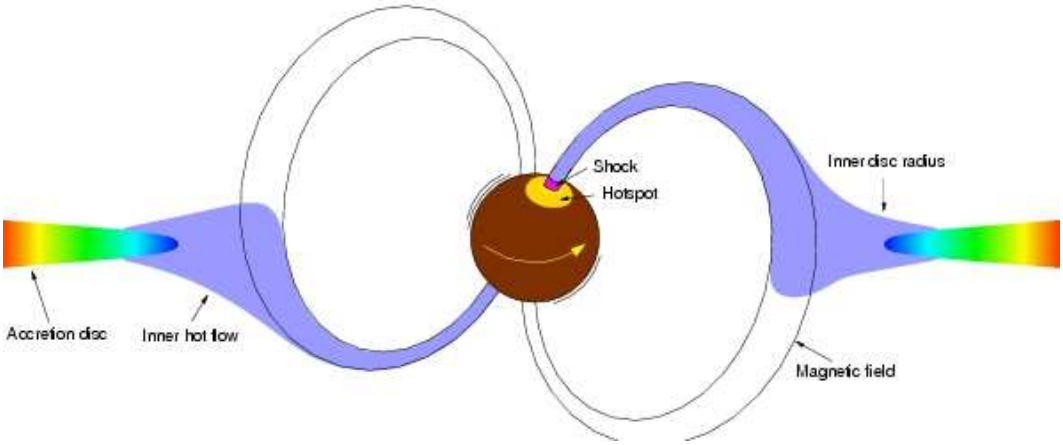

Figure 9: Artist illustration of the innermost region of an accreting neutron star in a NS-LMXB. The magnetic field of the neutron star truncates the accretion disc and channels the gas along the field lines. The gas flows along the "accretion columns" and lands on the "hot spot". Since the gas has a sudden deceleration from highly supersonic to zero on the surface, an "accretion shock" forms right above the hot spot.

Orbital periods have been measured for some 30 of these systems. They range from 11 min. to 17 days, similar to the orbital periods of cataclysmic variables. Only in the few widest of these systems one can observe the spectrum of the optical companion. In all other systems, the optical spectrum is that of the hot accretion disk. The NS-LMXBs are very seldom X-ray pulsars. The reason is not completely understood, and might be related with the evolution of their magnetic field (usually $10^8 10^{11}$ G), which is expected to result from accretion-induced field decay. On the other hand these sources show X-ray bursts (sudden thermonuclear fusion of accreted matter at the surface of the neutron star, which are suppressed if the magnetic field strength is $> 10^{11}$ G. For this reason such bursts are not observed in HMXBs. Among ~ 150 NS-LMXBs, only ~ 20 show X-ray pulsations. Of these, 14 show millisecond periods (accreting millisecond X-ray pulsars, AMXPs), whereas the others have much higher spin periods in the range of seconds.

There are dramatic differences in pulse shape and amplitude between one pulsar and another. In some cases, the pulse profile shows a dependence on energy and luminosity. AMXPs however always show an highly sinusoidal pulse profile which can be explained simply in terms of single hot spot with varying projected area (due to rotation).

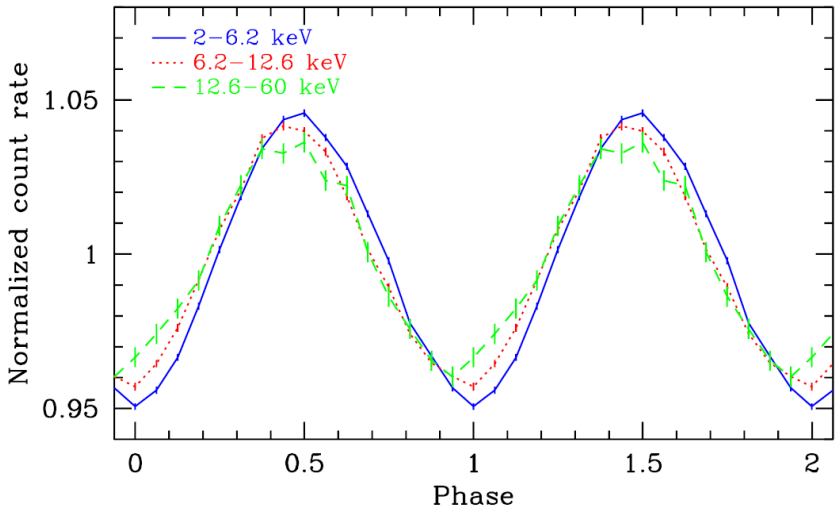

Figure 10: Observed pulse profile for the AMXP SAX J1808.4-3658 in three different X-ray energy bands (two cycles are plotted for clarity). Note that the pulse profiles are almost identical reagardless of the energy band, and show an almost purely sinusoidal shape which can be easily explained in terms of one single rotating hot spot.

NS-LMXBs show an X-ray spectrum that is remarkably similar for pulsating and for non pulsating sources. They are characterized by a continuum that peaks in the X-ray energy band, and extends up to several hundreds keV. At very low X-ray energies the continuum is dominated by the accretion disc, which emits as a multicolor disc blackbody. This is nothing but a superposition of blackbodies at different temperatures, which can be identified with different disc annuli. At higher temperatures there is a blackbody which is probably generated by the thermal emission of the plasma hitting the neutron star surface. The high energy part of the X-ray spectrum is instead probably generated by comptonization of soft (i.e., low energy) X-ray photons coming from the hot spot and upscattered via inverse compton up to several hundreds of keV (see Figure 11 for an example for the NS-LMXB SAX J1808.4-3658).

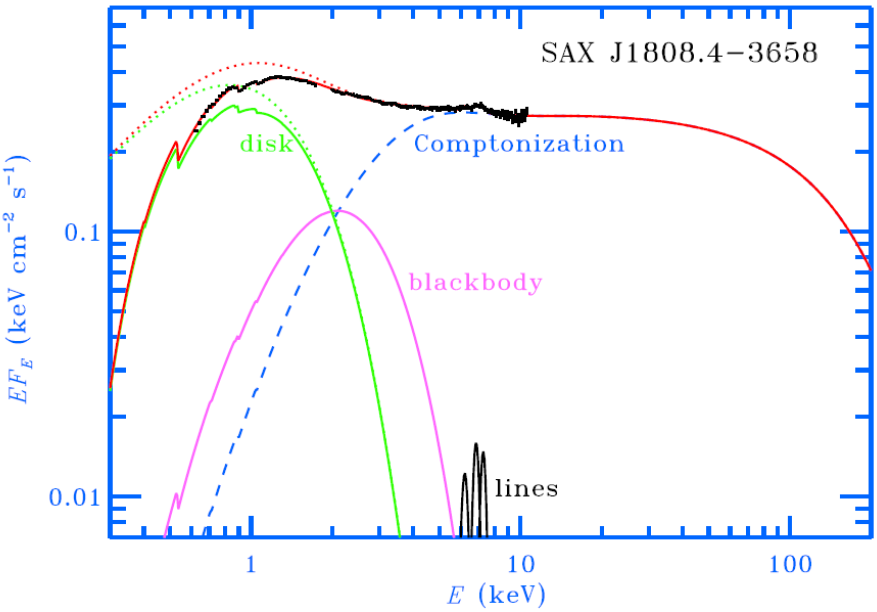

Figure 11: Observed broadband X-ray spectrum of the AMXP SAX J1808.4-3658 during its 2008 outburst. Three different component are visible: a disc component at soft energies (green line) a blackbody probably corresponding to the hot spot (pink line) and a hard X-ray component originating from comptonization of soft seed photons (dashed blue line). The red line is the sum of the three components. The red and green dotted line refer to the "unabsorbed" spectrum, i.e., to the spectrum in which the effect of galactic neutral hydrogen absorption has been corrected.

3 Isolated Neutron Stars

Beside radio pulsars and accreting neutron stars, a third important class of objects is the isolated neutron star class. Observations of isolated neutron stars (and sometimes neutron stars in quiescent X-ray binaries) are extremely important in fundamental physics, as thermal emission from the surface of a neutron star carries signatures of itsgravitational field, which may be used to infer its mass and radius. Detection of absorption lines corresponding to elements on the neutron star atmosphere and measurement of their gravitational redshift would provide rather accurate data. From the gravitational redshift at the surface of the neutron star, the ratio between its mass and radius may be measured, providing a very strong constraint on neutron star models. Such models give physics an experimentally testable handle on properties of matter at (supra-)nuclear densities.

In particular, direct emission from the neutron star surface layers (the only source of information on the physical conditions of the star) is only accessible in a handful of sources. This is because the surface of old neutron stars (age $> 10^6$ yr) is too cold to emit X-rays, while the radiation from young active radio pulsars (age $< 10^4$ yr) is dominated by nonthermal emission from the magnetosphere surrounding the star. In the last decade ROSAT satellite observations led to the discovery of seven dim sources, which were later associated with radio-silent, isolated, middle-age (age 10^5–10^6 yr) neutron stars. They are characterized by clean thermal emission at energies of about 0.1 keV without any trace of contamination from a surrounding supernova remnant or magnetospheric activity. They are located within a few hundred pc, close enough to be studied with the latest X-ray orbiting observatories and therefore represent important targets for the study of neutron star surface emission.

Observations of neutron stars with detection of thermal emission are summarized in Table 1 and 2.

Star	$\log t_{sd}$ yr	$\log T_\infty$ K	Distance kpc	$\log L_\infty$ erg/s
RX J0822–4247	3.90	6.24 ± 0.04	1.9–2.5	33.85–34.0
1E 1207.4–5209	5.53	6.21 ± 0.07	1.3–1.9	33.27–33.74
RX J0002+6246	–	6.03 ± 0.03	2.5–3.5	33.08–33.33
PSR 0833–45 (Vela)	4.05	5.83 ± 0.02	0.22–0.28	32.41–32.70
PSR 1706–44	4.24	5.8 ± 0.13	1.4–2.3	31.81–32.93
PSR 0538+2817	4.47	6.05 ± 0.10	1.2	32.6–33.6

Table 1: Neutron stars with hydrogen atmospheres.

These tables display four inferred quantities: the spin-down age t_{sd},the surface temperature T_∞, the total thermal luminosity L_∞ and the estimated distance d. The subscript ∞ refers to quantities observed at Earth which are redshifted relative to their values at the stellar surface. Table 1 refers to models incorporating atmospheres dominated by hydrogen, while Table 2 presents properties inferred from black-body atmospheres. There are also sources available with only upper limits. The estimation of T_∞ and L_∞ from the observed spectral fluxes requires atmospheric modelling with the inclusion of three additional factors: the composition of the atmosphere, the column density of X-ray absorbing material between the star and Earth, and the surface gravitational redshift. The column density is important, since the bulk of the emitted flux from neutron stars is absorbed by interstellar hydrogen,

Star	$\log t_{sd}$ yr	$\log T_\infty$ K	R_∞ km	Distance kpc	$\log L_\infty$ erg/s
RX J0822–4247	3.90	6.65 ± 0.05	1–1.6	1.9–2.5	33.60–33.90
1E 1207.4–5209	5.53	6.48 ± 0.01	1.0–3.7	1.3–3.9	32.70–33.88
RX J0002+6246	–	6.15 ± 0.11	2.1–5.3	2.5–3.5	32.18–32.81
PSR 0833–45 (Vela)	4.05	6.18 ± 0.02	1.7–2.5	0.22–0.28	32.04–32.32
PSR 1706–44	4.24	6.22 ± 0.04	1.9–5.8	1.8–3.2	32.48–33.08
PSR 0656+14	5.04	5.71 ± 0.04	7.0–8.5	0.26–0.32	32.18–32.79
PSR 0633+1748 (Gem)	5.53	5.75 ± 0.04	2.7–8.7	0.123–0.216	30.85–31.51
PSR 1055–52	5.43	5.92 ± 0.02	6.5–19.5	0.5–1.5	32.07–33.19
RX J1856.5–3754	–	5.6–5.9	> 16	0.105–0.129	31.44–31.68
RX J0720.4–3125	6.0 ± 0.2	5.55–5.95	5.0–15.0	0.1–0.3	31.3–32.5

Table 2: Neutron Stars with black-body atmospheres. The uncertainty in the radius are still too big for confrontation with theory.

before it reaches the Earth (Fig. 12, see also Fig. 12). Since no narrow spectral lines are observed in the neutron stars in Table 1 and 2, the atmospheric composition of these stars is unknown. Neutron star atmospheres can be either light-element (i.e. H and He) or heavy-element dominated. The latter ones have spectral distributions more closely resembling a black-body distribution, due to higher opacities of heavy elements. If the distance is known, the neutron star emission radius can be inferred. It seems that the inferred neutron star radii for stars younger than 100,000 years are consistent with canonical values only if they are modelled with light-element atmospheres. On the other hand, stars older than 100,000 years have inferred radii close to the canonical range only if modelled with heavy-element atmospheres. It seems therefore that the atmospheric composition of a neutron star evolves from light elements to heavy elements on a time-scale of about 100,000 years.

interstellar absorption to nearby stars. These dim isolated neutron stars (DINs) are key objects in compact object astrophysics. They offer a unique laboratory for investigating the properties of matter under extreme conditions, such as the equation of state at supranuclear densities, or the interaction of highly relativistic plasmas with radiation in the presence of gigaor teragauss magnetic fields. Detailed X-ray spectra of DINSs have been recently obtained with Chandra and XMMNewton, and show quite unexpected characteristics. The prototype of the class, RX J1856.53754, exhibitsa featureless spectrum extremely close to a pure black-body (Fig. 12). Broad absorption features have been detected in four pulsating sources with evidence of a spectral variation with phase. Very recently, spectral evolution on timescale of years have been reported for the second most luminous source, RX J0720.43125. In addition, when detected the optical counterpart lies a factor about 510 above the extrapolation of the X-ray black-body at optical wavelengths. All these new findings represent a challenge for conventional atmospheric models, typically based on surface temperature distributions induced by a dipolar magnetic field. interstellar absorption to nearby stars. These dim isolated neutron stars (DINs) are key objects in compact object astrophysics. They offer a unique laboratory for investigating the properties of matter under extreme conditions, such as the equation of state at supranuclear densities, or the interaction of highly relativistic plasmas with radiation in the presence of gigaor teragauss magnetic fields. Detailed X-ray spectra of DINSs have been recently obtained with Chandra and XMMNewton, and show quite unexpected characteristics. The prototype of the class, RX J1856.53754, exhibits a featureless spectrum extremely close to a pure black-body (Fig. 12). Broad absorption features have been detected in four pulsating sources with evidence of a spectral variation with phase. Very

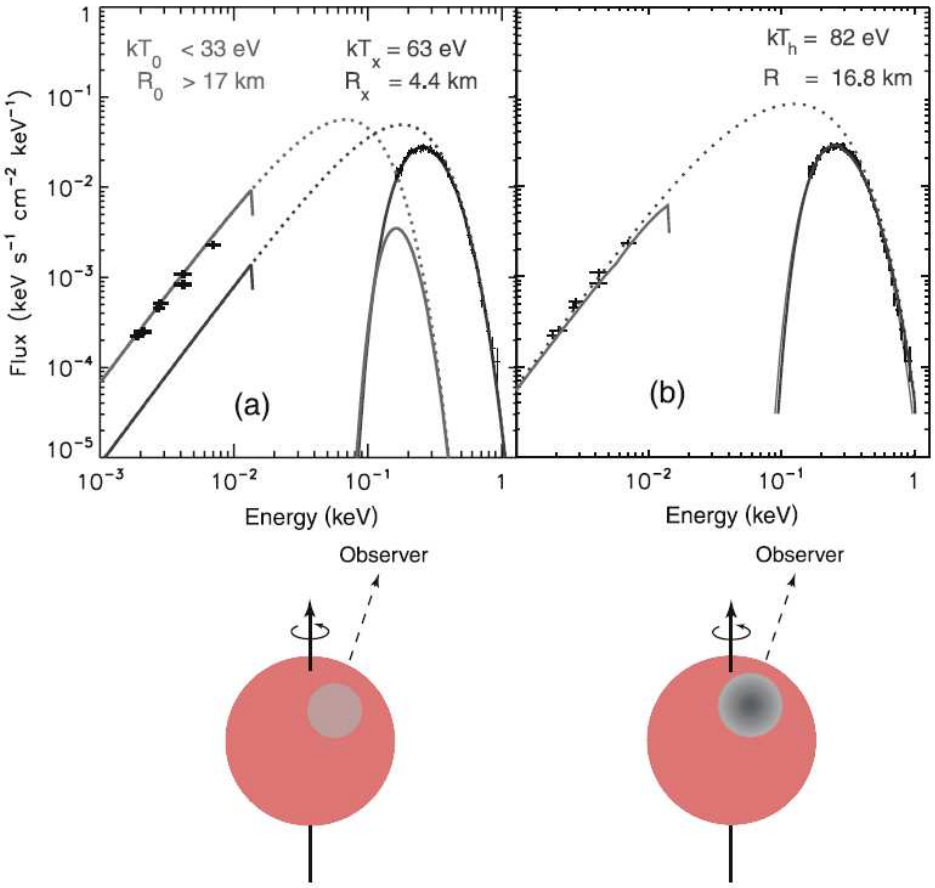

Figure 12: The X-ray spectrum of the isolated neutron star RJ 18563754 for (a) a twocomponent model and (b) a model with continuous temperature distribution. The X-ray and optical data show no spectral features and are best fitted with a two-component black-body model with $kT^X_{\infty,\,bb} \simeq 63.5$ eV and $R^X_{\infty,bb} \simeq 4.4(d/120\,\mathrm{pc})$ km for the hot X-ray emitting region, and $kT^{opt}_{\infty,bb} < 33$ eV and $R^{opt}_{\infty,bb} > 17(d/120\,\mathrm{pc})$ km for the rest of the neutron star surface responsible for the optical flux. The optical emission is then the RayleighJeans spectrum emitted by the neutron star surface and the hot spot. The neutron star radius would be 14 km. This would rule out quark stars.

recently, spectral evolution on timescale of years have been reported for the second most luminous source, RX J0720.43125. In addition, when detected the optical counterpart lies a factor about 510 above the extrapolation of the X-ray black-body at optical wavelengths. All these new findings represent a challenge for conventional atmospheric models, typically based on surface temperature distributions induced by a dipolar magnetic field.

3.1 Compact Stellar Radii from Thermal Emission

The observed temperature $T_{eff,\infty}$ determines the observed luminosity:

$$L_\infty = 4\pi\, {}_{SB} T_{eff,\infty}^4\, R_\infty^2 = L\, \alpha^2(R) \tag{15}$$

The last equality follows from the fact that the energy is redshifted and time is also redshifted with respect to the timescale at the surface of the star, $dt_\infty = d\tau/\alpha(R)$. The BB spectrum is redshifted when it leaves the gravitational field of the star, i.e.

$$T_{eff,\infty} = \alpha(R) T_{eff} = \sqrt{1 - 2GM/(c^2 R)} \cdot T_{eff} \tag{16}$$

Since $L = 4\pi\, R^2 \sigma_{SB} T_{eff}^4$, this has the consequence that the observed radius at infinity is apparently bigger than the true radius of the neutron star $R_\infty = R/\alpha(R) \simeq R/\sqrt{1 - 2GM/(c^2 R)}$. An intrinsic radius for a neutron star of 1.4 solar masses corresponds to an observed radius between 13 and 14.5 km. If massive neutron stars would have quark cores, we would expect observed radii between 12 and 13 km.

Lecture Notes 7

1 Quasi Periodic Oscillations

The principal motivation for studying accreting neutron stars and black holes is that these objects provide a unique window on the physics of strong gravity and dense matter. One of the most basic expressions of the compactness of these compact objects is the short (0.1–1 msec) dynamical time scale characterizing the motion of matter under the influence of gravity near them. Millisecond variability will naturally occur in the process of accretion of matter onto a stellar-mass compact object, an insight that dates back to at least Shvartsman (1971). For example, hot clumps orbiting in an accretion disk around black holes and neutron stars will cause quasi-periodic variability on time scales of about a millisecond (Sunyaev 1973). Accreting low-magnetic field neutron stars will reach millisecond spin periods, which can be detected when asymmetric emission patterns form on the star's surface during X-ray bursts (Radhakrishnan & Srinivasan 1984, Alpar et al. 1982; Shara 1982, Livio & Bath 1982, see also Joss 1978). These early expectations have finally been verified in a series of discoveries with NASA's Rossi X-Ray Timing Explorer (RXTE; Bradt et al. 1993) within 2.5 years after launch on 30 December 1995.

Here we will concentrate on **quasi-periodic** millisecond oscillations in X-ray flux with frequencies exceeding $10^{2.5}$ Hz, but we also discuss their relations to slower variability and X-ray spectral properties. The term **quasi-periodic** distinguishes these phenomena from the periodic (or coherent) millisecond oscillations seen for example in accreting X-ray pulsars, which reflect the spin rate of the neutron star. Quasi periodic oscillations have so far been seen nearly exclusively from low-magnetic-field neutron stars and black-hole-candidates.

Accreting neutron stars and black holes occur in X-ray binaries (e.g., Lewin et al. 1995a). In these systems matter is transferred from a normal ('donor') star to a compact object. Thermal X-rays powered by the gravitational potential energy released are emitted by the inner

regions of the accretion flow and, if present, the neutron star surface. For a compact object with a size of order 10^1 km, 90% of the energy is released in the inner $\sim 10^2$ km. It is with this inner emitting region that we shall be mostly concerned here. Because accreting low-magnetic-field neutron stars are mostly found in low-mass X-ray binaries (in which the donor star has a mass of $<1 M_\odot$) these systems will be the ones we focus on.

The mass transfer usually occurs by way of an accretion disk around the compact object. In the disk the matter moves in near-Keplerian orbits, i.e., with an azimuthal velocity that is approximately Keplerian and a radial velocity much smaller than this. The disk has a radius of 10^{5-7} km, depending on the binary separation. The geometry of the flow in the inner emitting regions is uncertain. In most models for accretion onto low-magnetic-field neutron stars (e.g., Miller et al. 1998a) at least part of the flow extends down into the emitting region in the form of a Keplerian disk. It is terminated either at the radius R of the star itself, or at a radius r_{in} somewhat larger than R, by for example the interaction with a weak neutron-star magnetic field, radiation drag, or relativistic effects. Within r_{in} the flow is no longer Keplerian and may or may not be disk-like. Both inside and outside r_{in} matter may leave the disk and either flow in more radially or be expelled. Particularly for black holes advective flow solutions are discussed where the disk terminates and the flow becomes more spherical at a much larger radius (e.g., Narayan 1997).

Whatever the geometry, it is clear that as the characteristic velocities near the compact object are of order $(GM/R)^{1/2} \sim 0.5c$, the dynamical time scale, the time scale for the motion of matter through the emitting region, is short; $\tau_{dyn} \equiv (r^3/GM)^{1/2} \sim 0.1$ ms for r=10 km, and $\sim$2 ms for r=100 km near a $1.4 M_\odot$ neutron star, and $\sim$1 ms at 100 km from a $10 M_\odot$ black hole. So, the significance of millisecond X-ray variability (both periodic and quasi-periodic) from X-ray binaries is clear: milliseconds is the natural time scale of the accretion process in the X-ray emitting regions, and hence strong X-ray variability on such time scales is nearly certainly caused by the motion of matter in these regions. Orbital motion, neutron-star spin, disk- and neutron-star oscillations are all expected to happen on these time scales.

The inner flow is located in regions of spacetime where strong-field general-relativity is required to describe the motion of matter. For that reason one expects to detect strong-field general-relativistic effects in these flows, such as for example the existence of a region where no stable orbits are possible (see for example the ISCO on Shapiro & Teukolsky). The precise interactions between the elementary particles in the interior of a neutron star which determine the equation of state (EOS) of supra-nuclear-density matter are not known. Therefore we can not confidently predict the radius of a neutron star of given mass, or the maximum spin rate or mass of neutron stars (e.g., Cook et al. 1994). So, by measuring these macroscopic quantities one constrains the EOS and tests basic ideas about the properties of elementary particles. In summary, the main motivation for studying millisecond variations in X-ray binaries is that their properties depend on untested, or even unknown, properties of spacetime and matter.

Three different millisecond phenomena have now been observed in X-ray binaries. Historically, the first to be discovered were the twin kilohertz quasi-periodic oscillations (kHz QPOs), widely interpreted now as due to orbital motion in the inner accretion flow. Then came the burst oscillations, probably due to the spin of a layer in the neutron star's atmosphere in near-corotation with the neutron star itself. Finally RXTE detected the first true spin frequency of an accreting low-magnetic field neutron star, the long-anticipated accreting millisecond pulsar.

Here we will examine the kHz QPO phenomenon, in the mostly uncharted territory of the innermost accretion flows around neutron stars and black holes, which obviously is "where the monsters are", but also where the greatest rewards wait.

2 Techniques

Most of the variability measurements discussed here rely on Fourier analysis of X-ray count-rate time series with sub-millisecond time resolution (van der Klis 1989). The reason why Fourier analysis is used is due to the low count rate and the short timescales at which these X-ray binaries are observed. The brightest X-ray sources show count rates not exceeding 10000 ct/s. Therefore on the timescales of interest ($\sim$ 1 ms)

the X-ray lightcurve will contain about 10 photons in the best case (see for example Fig 1).

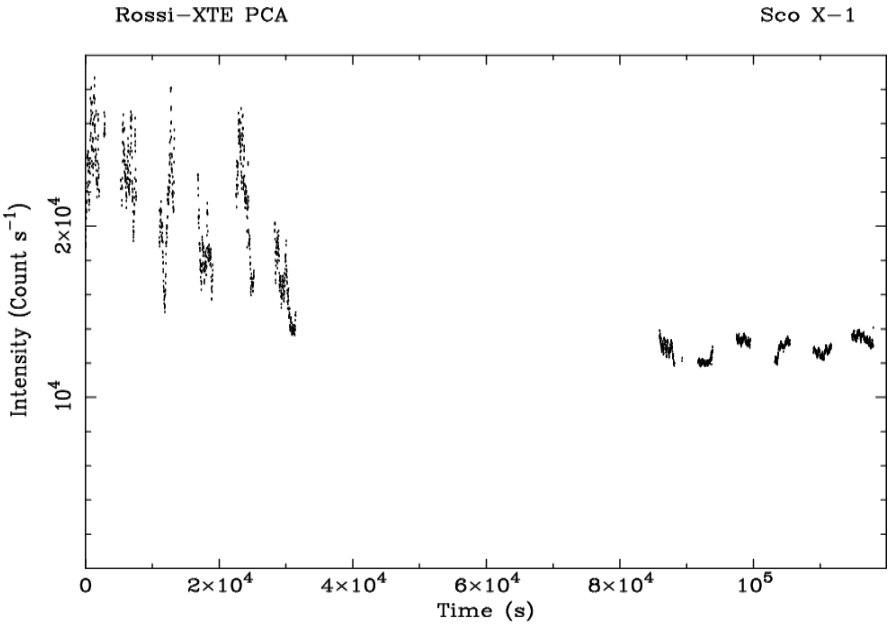

Figure 1: X-ray lightcurve observed with the Proportional Counter Array onboard RXTE. The X-ray source is "Sco X-1", the brightest persistent X-ray source known in the sky. It is believed to be an accreting neutron star system in a Low Mass X-ray Binary. The X-ray lightcurve shows a count rate of a few tens of thousands photons per seconds. Therefore X-ray variability at sub-ms timescales cannot be directly observed at such short timescales. Sco X-1 is one of the sources that exibit the kHz QPO phenomenon. As can be verified, there is no signature of such fast variability in the lightcurve since the kHz QPOs can only be observed in a power spectrum by using Fourier techniques.

Typical X-ray sources show count rates of $\sim$ 100 ct/s, meaning a total of less than a photon in 1 ms. Therefore any study of X-ray flux variability at ms timescales is impossible given the actual count rates (which depend on the source brightness but also on the detector capability). Therefore the best way to proceed is to study these timescales with Fourier techniques.

3 Timing

Fourier analysis. A number of *variability components* or *power-spectral components* together make up the power spectrum. An aperiodic component by definition covers several, usually many, frequency resolution elements. Broad structures are called *noise* and narrow features *quasi-*

periodic oscillations (QPOs); 'broad-band noise' and 'QPO peaks' are common terms (Fig. 2). Least-squares fitting techniques are used to measure these components. When a series of power spectra is calculated from consecutive chunks of data, usually the components change: they move through the power spectrum (change frequency from one power spectrum to the next), vary in width and strength, etc., but they remain identifiable as the changes are gradual. This is the empirical basis for the concept of a power-spectral component. The shortest time scale on which changes can typically be followed is seconds to minutes.

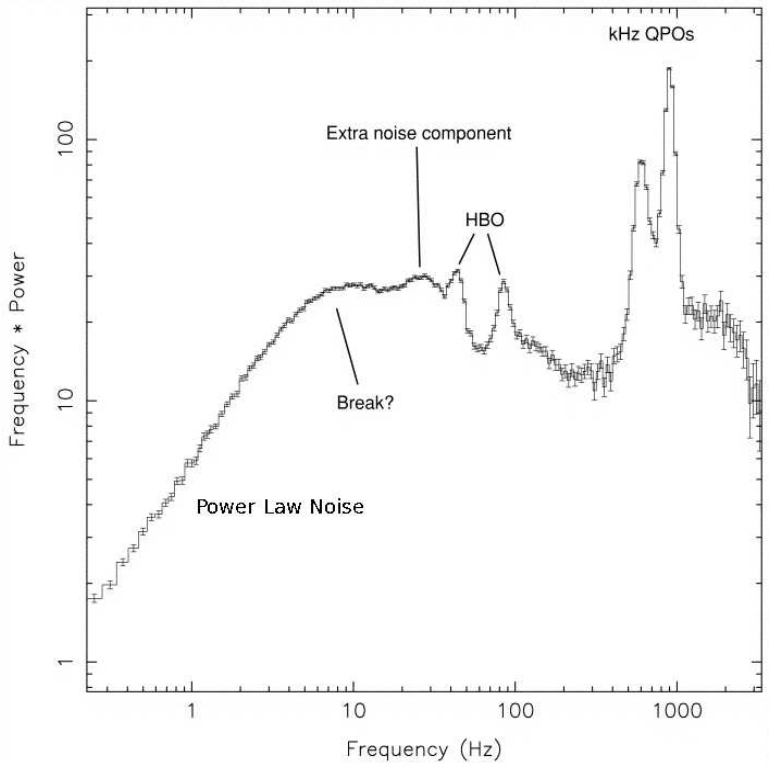

Figure 2: Power spectrum of an X-ray binary. It can be clearly seen the power law noise at low frequencies, the twin kHz QPOs and some extra noise at intermediate frequencies. The two QPO like features between 10-100 Hz are the so called Horizontal Branch QPOs, which often appear in power spectra and whose origin is still unknown.

The signal in the time series is not completely specified by the power spectrum (and the signals are usually too weak to recover the Fourier phases). The same QPO peak could be due to, e.g., a damped harmonic oscillator, randomly occurring short wave trains, a frequency-modulated oscillation, an autoregressive signal, white noise observed through a narrow passband filter or even a closely spaced set of periodic signals.

Power-law noise is noise that (in the frequency range considered) follows a power law $P_\nu \propto \nu^{-\alpha}$. The power-law index (also 'slope') α is typically between 0 and 2; for $|\alpha| > 2$ Fourier analysis suffers from power leakage, so measurements of noise steeper than that are suspect (e.g., Bracewell 1986, Deeter 1984). '1/f noise' has $\alpha = 1$, and *white noise* is constant ($\alpha = 0$). *Red noise* is a term variously used for either $\alpha = 2$ power-law noise or any kind of noise whose P_ν decreases with ν.

A *quasi-periodic oscillation* (QPO) is a finite-width peak in the power spectrum. It can usually be described with a Lorentzian $P_\nu \propto \lambda/[(\nu - \nu_0)^2 + (\lambda/2)^2]$ with *centroid frequency* ν_0 and full width at half maximum (FWHM) λ. This is the power spectrum of an exponentially damped sinusoid $x(t) \propto e^{-t/\tau}\cos(2\pi\nu_0 t)$, but the underlying signal may well be different from this. λ is related to the *coherence time* $\tau = 1/\pi\lambda$ of the signal, and is often reported in terms of the *quality factor* $Q \equiv \nu_0/\lambda$, a measure for the *coherence* of the QPO. Conventionally, signals with $Q > 2$ are called QPOs and those with $Q < 2$ peaked noise. A *sharp* QPO peak is one with high Q.

The strength (variance) of a signal is proportional to the integrated power $P = \int P_\nu d\nu$ of its contribution to the power spectrum, and is usually reported in terms of its *fractional root-mean-squared (rms) amplitude* $r \propto P^{1/2}$, which is a measure for signal amplitude as a fraction of the total source flux. It is often expressed in percent, as in "2% (rms)".

The signal-to-noise of a weak QPO or noise component is $n_\sigma = \frac{1}{2} I_x r^2 (T/\lambda)^{1/2}$ (van der Klis 1989b, see van der Klis 1998 for more details), where I_x is the count rate and T the observing time (assumed $\gg 1/\lambda$). As n_σ is proportional to signal amplitude *squared*, if a clear power-spectral feature "suddenly disappears" it may have only decreased in amplitude by a factor of two (and gone from, say, 6 to 1.5σ).

Power spectra are in practice presented in a variety of ways, each more suitable to emphasize particular aspects (displaying either P_ν or νP_ν, linearly or logarithmically, subtracting the white Poisson background noise or not; e.g., Figs. 3).

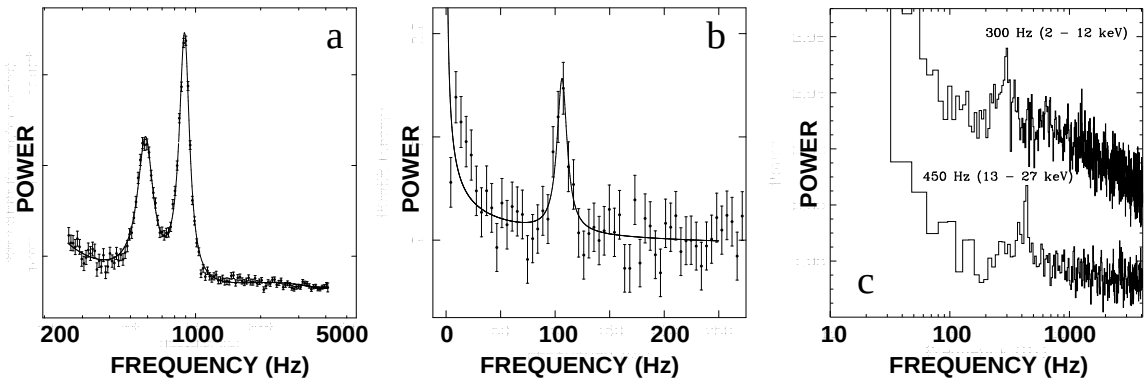

Figure 3: *(a)* Twin kHz QPOs in Sco X-1 (van der Klis et al. 1997), *(b)* hectohertz QPO in 4U 0614+09, *(c)* HF QPOs in GRO J1655–40 (Strohmayer 2001a).

4 Spectroscopy

Variations in broad-band ($\Delta E/E > 0.1$, usually continuum-dominated) X-ray spectral shape usually just become detectable on the same seconds to minutes time scales on which power-spectral changes are detectable. Two different techniques of broad-band spectroscopy are used to diagnose these changes: multi-band photometry and spectral fitting. The spectral band used varies somewhat but is usually in the 1–60 keV range and nearly always covers at least the 3–8 keV band.

Photometric method One approach to quantifying broad-band X-ray spectral shape uses *X-ray colors*. An X-ray color is a 'hardness' ratio between the photon counts in two broad bands; it is a rough measure for spectral slope. By calculating two X-ray colors (a *hard color* in a higher energy band and a *soft color* in a lower band) as a function of time, a record is obtained of the broad-band X-ray spectral variations that is well-matched to the power-spectral variations. Plotted vs. one another in a *color-color diagram* (CD), one can observe the source to move through the diagram and, nearly always, create a pattern. It is then possible to study the relation between timing and location in the CD. A *hardness-intensity diagram* (HID) or color-intensity diagram is a similar diagram with a color vs. 'intensity'; *X-ray intensity* in this context is nothing but a count rate in some broad X-ray spectral band. Figs. 4 and 5 provide examples of CD and HID patterns. Whether CD

or HID presents the 'cleanest' pattern depends on source, and on the quality of the data. There are distinct advantages to working with the logarithm of colors and intensities, like magnitudes in the optical, but this is not general practice.

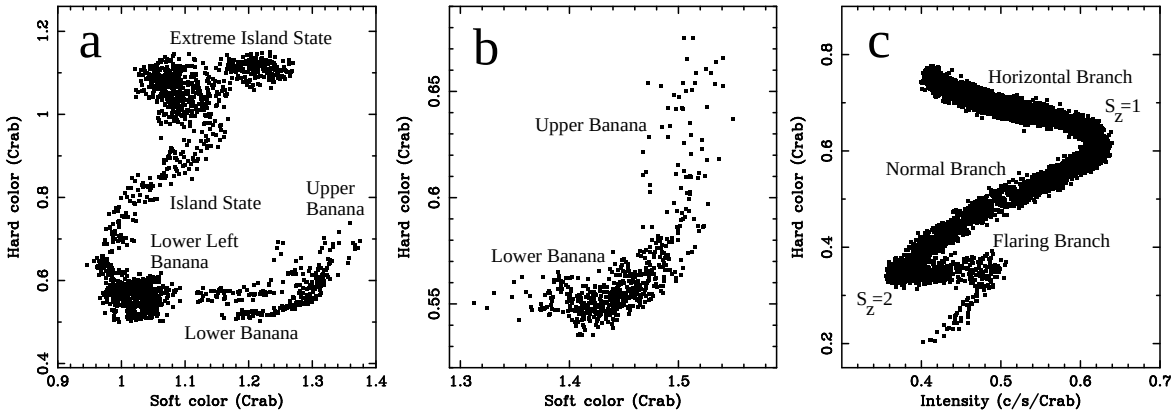

Figure 4: Spectral branches of neutron stars. *(a)* CD of the atoll source 4U 1608−52, *(b)* CD of the 'GX' atoll source GX 9+1, and *(c)* HID of the Z source GX 340+0. RXTE/PCA data; soft colour 3.5–6/2–3.5 keV, hard colour 9.7–16/6–9.7 keV, intensity: 2–16 keV, all normalized to Crab. Conventional branch names are indicated; Compare van Straaten et al. (2003), Reerink et al. (2004), Jonker et al. (2000a).

For distinguishing between source states the photometric method combined with timing performs well, and the method provides excellent sensitivity to subtle spectral variations. However, this is at the expense of detector dependence. All X-ray detectors are different and change over time, and contrary to optical photometry, in X-rays there is only one bright standard star (Crab). For the low spectral resolution detectors typically used for timing (proportional counters) it is not possible to completely correct X-ray colors for the detector response: In the absence of *apriori* knowledge of the intrinsic spectral shape all correction methods intended to derive 'intrinsic' colors (scaling by Crab colors, unfolding through the detector response matrix either directly or by fitting an arbitrary model) are mathematically imperfect (Kahn & Blisset 1980, Kuulkers et al. 1994, Kuulkers 1995, Done & Gierliński 2003).

Spectral fitting method There are considerable interpretative advantages in describing the X-ray spectral variations instead in terms of physical models fitted to the observed spectra. The drawback is that this involves a description of spectral shape in terms of more numbers (the

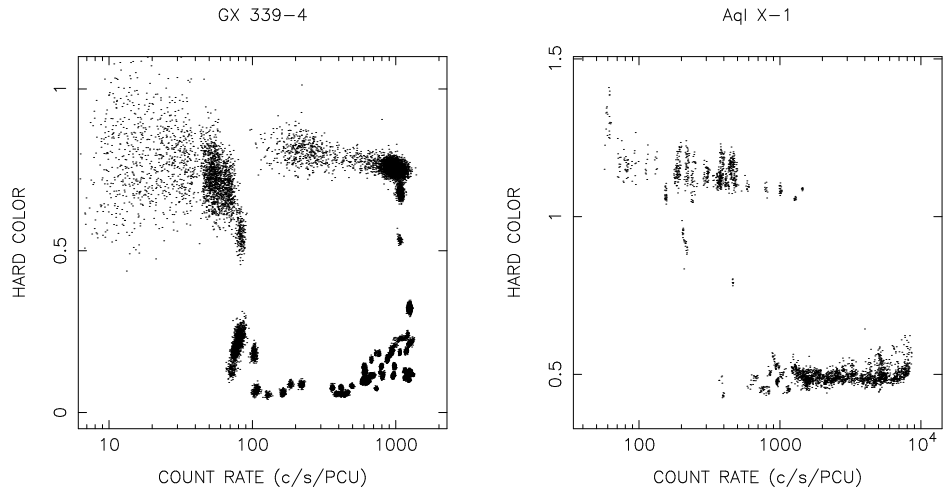

Figure 5: Hardness-intensity diagrams of the black hole GX 339–4 and the neutron-star LMXB transient Aql X-1. Note the orientation of these diagrams, which is similar to that of Fig. 4. Compare Belloni (2004), Reig et al. (2004).

spectral parameters) than just two X-ray colors, and that the correct models are unknown. This, plus the tendency towards spurious results related to statistical fit-parameter correlations masking true source variations, usually means that to obtain sensible results, fitting requires longer integration times than the time scale on which we see the power spectra, and the X-ray colors, change, and that measured parameter-value changes obtained are hard to interpret. Nevertheless, for sufficiently long integration times, and particularly in the case of the black-hole candidates, where the X-ray spectral variations are clearer than in neutron stars, this approach has met with some success. Because both methods have limitations, in practice the interpretation of X-ray spectral variations tends to occur in a back-and-forth between CD/HIDs and spectral fitting. Techniques such as plotting the predicted colors of model spectra in CDs together with the data, fitting spectra obtained by averaging data collected at different epochs but in the same part of the CD, and approximate intrinsic colors (e.g., Belloni et al. 2000, Done & Gierliński 2003), are all being employed to make the link between the two methods.

5 Source types

Different sources exhibit similar patterns of timing and spectral properties, allowing to group them into a number of *source types*. Both spec-

troscopy and timing, preferably in several source states, are necessary to reliably identify source type. The primary distinction is not between neutron stars and black holes but between high and low magnetic-field strengths. Black holes and low magnetic-field ($\lesssim 10^{10}$ G) neutron stars potentially have gravity-dominated flows down to the strong-field region (although radiative stresses can also be important). It is these objects, mostly found in low-mass X-ray binaries (LMXBs), that we shall be mostly concerned with.

5.1 Neutron stars *vs.* black holes

A neutron star is only a few times larger than its Schwarzschild radius, so accretion onto black holes and neutron stars is expected to show similarities. Indeed, based on just the inner-flow diagnostics (timing and spectrum), the distinction between neutron stars and black holes is notoriously difficult to make, with some examples of black-hole candidates turned neutron stars (e.g., Cir X-1, Jones et al. 1974, Tennant et al. 1986; V0332+53, Tanaka et al. 1983, Stella et al. 1985; GS 1826–238, Tanaka 1989, Ubertini et al. 1999; 4U0142+61, White & Marshall 1984, Israel et al. 1994), and in practice it is not easy to prove a compact object is a black hole. There are two levels of proof: (i) showing that, assuming general relativity, the compact object must be a black hole, and (ii) showing that such an object indeed has the properties general relativity predicts for a black hole. For (i) it is sufficient to prove that a mass is concentrated within its Schwarzschild radius; currently, this mostly relies on dynamical mass estimates combined with theoretical arguments about the maximum mass of a neutron star (e.g., Srinivasan 2002) and not on direct, empirical measurements of radius. For (ii) it is necessary to observe the interaction of the compact object with its surroundings, empirically map out its exterior spacetime, and demonstrate properties such as extreme frame dragging, the existence of an innermost stable orbit, and an event horizon.

In recent work, for well-studied sources, clear differences are discerned between the patterns of correlated spectral and timing behaviour of neutron stars and black holes which agree with distinctions based on X-ray bursts, pulsations, and dynamical mass estimates. In particular the most rapid variability, presumably that produced closest to the compact ob-

ject and most affected by its properties, is clearly quite different between neutron stars and black-hole candidates. Evidently we are learning to distinguish between black holes and neutron stars based on the properties of the flow in the strong-field gravity region, so there is progress towards the goal of testing general relativity. On the other hand, remarkable spectral and timing similarities exist between certified neutron stars and black-hole candidates, particularly in low luminosity states. To demonstrate the existence of black holes in the sense of general relativity based on an understanding of the accretion phenomena near them is a goal that has not yet been reached. The study of rapid X-ray variability of low-magnetic field compact objects in X-ray binaries is one of the programs contributing towards this end. Both the actual measurement of compact-object radius, which is part of a level (i) proof, as well as level (ii) proofs are addressed by the research described in this chapter: timing to diagnose motion very near compact objects. I shall use the term *black hole* both for objects whose black-hole candidacy is based on a measured mass as well as for those whose patterns of timing and spectral behavior put them into the same phenomenological category as these objects.

5.2 Low magnetic-field object types

The low magnetic-field neutron-star systems are subdivided into Z sources, atoll sources (Hasinger & van der Klis 1989) and what I shall call the 'weak LMXBs'. These three main sub-types are closely related. *Z sources* are the most luminous (Fig. 6), and accrete at an appreciable fraction of the Eddington critical rate (perhaps 0.5–1 L_{Edd}). *Atoll sources*, many of which are X-ray burst sources, cover a much wider range in luminosities, from perhaps 0.001 L_{Edd} (much lower in transients, but below this level timing becomes difficult) all the way up to the range of the Z sources (e.g., Ford et al. 2000; L_x overlaps may well occur; distances are uncertain). Ordinary atoll sources are usually in the 0.01–0.2 L_{Edd} range, while the 'GX' atoll sources in the galactic bulge usually hover at the upper end (perhaps 0.2–0.5 L_{Edd}), and the *weak LMXBs* at the lower end ($<$0.01 L_{Edd}) of that range. Weak LMXBs comprise the overlapping groups of faint burst sources, millisecond pulsars and low-luminosity transients; many of them appear to be just atoll sources stuck

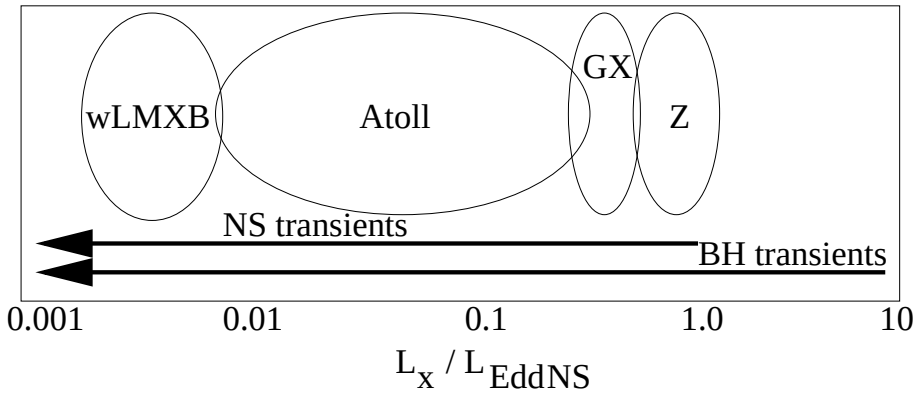

Figure 6: Luminosities attained by Z sources, GX atoll sources, ordinary atoll sources and weak LMXBs, respectively, as well as by neutron-star and black-hole transients. The extent of the L_x overlaps between these source types is undecided in detail, but those shown here are likely.

at low L_x. Some faint LMXBs have luminosities that are not well known due to uncertain distances or emission anisotropies (e.g., in dippers), but many of these are probably weak as well.

While nearly all LMXBs whose X-ray emission is persistent contain a neutron star, nearly all with a black hole (and many with a neutron star as well) are transients, showing intermittent activity intervals called outbursts usually lasting weeks to months and separated by long quiescent intervals. The black holes have no well-defined subdivisions; although one might expect differences in spin (and hence, frame dragging), accretion mode (wind or Roche-lobe overflow) or between transient and persistent systems to show up, none of these lead to obvious differences in the observable properties attributable to the inner flow. Some black-hole transients remain in the low hard state during their entire outburst (e.g., Nowak 1995, Brocksopp 2004), but this can vary from one outburst to another (e.g., Belloni et al. 2002b).

6 Black Hole QPOs

The fastest black-hole phenomenon are the high-frequency (HF) QPOs (Remillard et al. 1999). Frequencies range from 100 to 450Hz, and are reported to usually occur at fixed values different in each source, perhaps inversely proportional to black-hole mass. *In Black Hole X-ray binaries the observations of kHz QPO has never occurred and the fastest*

phenomena reach $\sim$ 450 *Hz.*

Although their properties are very different, suggesting different underlying physics explains them, black-hole HF and neutron-star kHz QPOs might possibly still be reconciled within a single explanation. The differences are clear: the neutron-star QPOs (i) have strongly tunable frequencies, (ii) come in pairs with a separation perhaps related to the stars spin and (iii) have low harmonic content, while the black-hole ones (i) have a much more stable frequency, and (ii) are single, but (iii) with high harmonic content. Note: the term "harmonic content" refers to the fact that some excess power is observed at values which are twice the peak frequency of the main QPO (and sometimes even higher multiples).

To reconcile these observations one might postulate that (i) the variable frequencies in neutron stars occur because the phenomenon occurs at a variable (e.g., inner disk) radius, set by interaction of the disk flow with either a magnetic field or radiation from the stellar surface, while in black holes, in the absence of these influences, the same phenomenon occurs at a more constant radius (perhaps, close to the ISCO), (ii) the second QPO occurs only in neutron stars because it is due to an interaction with the spin, and (iii) the high harmonic content in blackhole QPOs is due to relativistic effects on the flow and its emission that become important only near the ISCO (e.g, as a simple example, extreme Doppler boosting near the ISCO of a spinning black hole).

7 Kilohertz quasi-periodic oscillations

The kilohertz quasi-periodic oscillations (kHz QPOs) were discovered at NASA's Goddard Space Flight Center in February 1996, just two months after RXTE was launched (in Sco X-1: van der Klis et al. 1996a,b,c, and 4U 1728−34: Strohmayer et al. 1996a,b,c; see van der Klis 1998 for a historical account). Two simultaneous quasi-periodic oscillation peaks ("twin peaks") in the 300–1300 Hz range and roughly 300 Hz apart (Fig. 7) occur in the power spectra of low-mass X-ray binaries containing low-magnetic-field neutron stars of widely different X-ray luminosity L_x. The frequency of both peaks usually increases with X-ray flux. In 4U 1728−34 the separation frequency of the two kHz peaks is close to ν_{burst} (Strohmayer et al. 1996b,c). This commensurability of frequencies

provides a powerful argument for a beat-frequency interpretation.

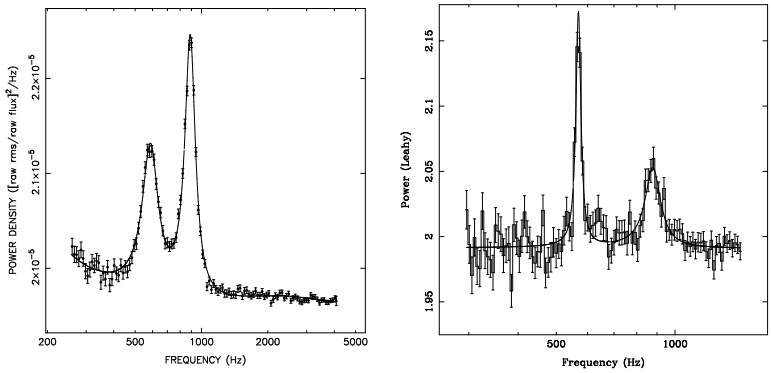

Figure 7: Twin kHz peaks in Sco X-1 (left; van der Klis et al. 1997b) and 4U 1608−52 (right; Méndez et al. 1998b).

7.1 Orbital and beat frequencies

Orbital motion around a neutron star occurs at a frequency of

$$\nu_{orb} = \left(\frac{GM}{4\pi^2 r_{orb}^3}\right)^{1/2} \approx 1200\,\mathrm{Hz}\,\left(\frac{r_{orb}}{15\,\mathrm{km}}\right)^{-3/2} m_{1.4}^{1/2},$$

and the corresponding orbital radius is

$$r_{orb} = \left(\frac{GM}{4\pi^2 \nu_{orb}^2}\right)^{1/3} \approx 15\,\mathrm{km}\,\left(\frac{\nu_{orb}}{1200\,\mathrm{Hz}}\right)^{-2/3} m_{1.4}^{1/3},$$

where $m_{1.4}$ the star's mass in units of $1.4\mathrm{M}_\odot$ (Fig. 8). In general relativity, no stable orbital motion is possible within the innermost stable circular orbit (ISCO), $R_{ISCO} = 6GM/c^2 \approx 12.5 m_{1.4}\,\mathrm{km}$. The frequency of orbital motion at the ISCO, the highest possible stable orbital frequency, is $\nu_{ISCO} \approx (1580/m_{1.4})\,\mathrm{Hz}$.

These expressions are valid for a Schwarzschild geometry, i.e., outside a non-rotating spherically symmetric neutron star (or black hole). Corrections to first order in $j = cJ/GM^2$, where J is the neutron-star angular momentum have been given by, e.g., Miller et al. (1998a) and can be several 10%. For more precise calculations see Morsink & Stella (1999).

In spin-orbit beat-frequency models some mechanism produces an *interaction* of ν_{orb} at some preferred radius in the accretion disk with the

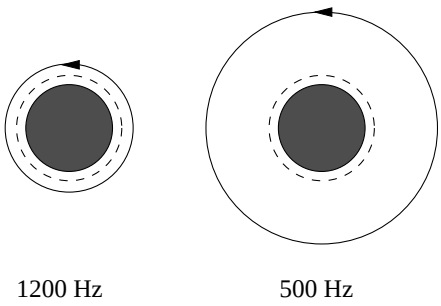

Figure 8: A 10-km radius, $1.4 M_\odot$ neutron star with the corresponding innermost circular stable orbit (ISCO; dashed circles) and orbits (drawn circles) corresponding to orbital frequencies of 1200 and 500 Hz, drawn to scale.

neutron star spin frequency ν_s, so that a beat signal is seen at the frequency $\nu_{beat} = \nu_{orb} - \nu_s$. As ν_{beat} is the frequency at which a given particle orbiting in the disk overtakes a given point on the spinning star, it is the natural disk/star interaction frequency. In such a rotational interaction (with spin and orbital motion in the same sense) no signal is produced at the sum frequency $\nu_{orb} + \nu_s$ (it is a "single-sideband" interaction).

7.2 Early interpretations

It was immediately realized that the observed high frequencies of kHz QPOs could arise in orbital motion of accreting matter very closely around the neutron star, or in a beat between such orbital motion and the neutron-star spin (van der Klis et al. 1996a, Strohmayer et al. 1996a; some of the proposals to look for such rapid QPOs with RXTE had in fact anticipated this). A magnetospheric spin-orbit beat-frequency model (Alpar & Shaham 1985, Lamb et al. 1985) was already in use in LMXBs for slower QPO phenomena so when three commensurable frequencies were found in 4U 1728−34, a beat-frequency interpretation was immediately proposed (Strohmayer et al. 1996c). Let us call ν_2 the frequency of the higher-frequency (the "upper") peak and ν_1 that of the lower-frequency kHz peak (the "lower peak"). Then the beat-frequency interpretation asserts that ν_2 is ν_{orb} at some preferred radius in the disk, and ν_1 is the beat frequency between ν_2 and ν_s, so $\nu_1 = \nu_{beat} = \nu_{orb} - \nu_s \approx \nu_2 - \nu_{burst}$, where the approximate equality follows from $\nu_{burst} \approx \nu_s$. That only a single sideband is observed is a strong argument for a rotational interaction. Later this was worked out in de-

tail in the form of the "sonic point beat-frequency model" by Miller et al. (1996, 1998a), where the preferred radius is the sonic radius (essentially, the inner edge of the Keplerian disk).

The beat-frequency interpretation implies that the observed kHz QPO peak separation $\Delta\nu = \nu_2 - \nu_1$ should be equal to the neutron star spin frequency ν_s, and should therefore be constant and nearly equal to ν_{burst}. It turned out that $\Delta\nu$ is not actually exactly constant nor precisely equal to ν_{burst} and this triggered the development of other models for the kHz QPOs. Stella & Vietri (1998) noted that the frequency ν_{LF} of low-frequency QPOs in the 15–60 Hz range that had been known in the Z sources since the 1980's (cf. van der Klis 1995), and that were being discovered with RXTE in the atoll sources as well is approximately proportional to ${\nu_2}^2$. This triggered a series of papers (Stella & Vietri 1998, 1999, Stella et al. 1999b) together describing what is now called the "relativistic precession model", where ν_2 is the orbital frequency at some radius in the disk and ν_1 and ν_{LF} are frequencies of general-relativistic precession modes of a free particle orbit at that radius. With the exception of the photon bubble model (Klein et al. 1996b), all models are based on the interpretation that one of the kHz QPO frequencies is an orbital frequency in the disk.

7.3 Dependence on source state and type

More than twenty sources have now shown kHz QPOs. Sometimes only one peak is detectable, but 18 of these sources have shown two simultaneous kHz peaks; the exceptions with only a single peak are the little-studied XTE J1723−376 and EXO 0748−676. Tables 1 and 2 summarize the results. There is a remarkable similarity in QPO frequencies and peak separations across a great variety of sources.

However, at a more detailed level there turn out to be differences between the different source types. The two main types are the Z sources and the atoll sources (Hasinger & van der Klis 1989, see van der Klis 1989a, 1995a,b for reviews). Z sources are named after the roughly Z-shaped tracks they trace out in X-ray color-color and hardness-intensity diagrams on a time scale of hours to days (Fig. 9). They are the most luminous LMXBs, with X-ray luminosity L_x near the Eddington luminosity L_{Edd}. Atoll sources produce tracks roughly like a wide U or C

Table 1: Observed frequencies of kilohertz QPOs in Z sources

Source	ν_1 (Hz)	ν_2 (Hz)	$\Delta\nu$ (Hz)	ν_{burst} (Hz)	References
Sco X-1	565	870	307±5		Van der Klis et al. 1996a,b,c,1997b
	845	1080	237±5		
		1130			
GX 5−1	215	505	298±11		Van der Klis et al. 1996e
	660	890			Wijnands et al. 1998c
	700				
GX 17+2		645			Van der Klis et al. 1997a
	480	785	294±8		Wijnands et al. 1997b
	780	1080			
Cyg X-2		730			Wijnands et al. 1998a
	530	855	346±29		
	660	1005			
GX 340+0	200	535	339±8		Jonker et al. 1998, 1999c
	565	840			
	625				
GX 349+2	710	980	266±13		Zhang et al. 1998a
	1020[a]				Kuulkers & van der Klis 1998

Values for ν_1 and ν_2 were rounded to the nearest 5, for ν_{burst} to the nearest 1 Hz. Entries in one column not separated by a horizontal line indicate ranges over which the frequency was observed or inferred to vary; ranges from different observations were combined assuming the ν_1, ν_2 relation in each source is reproducible (no evidence to the contrary exists). Entries in one uninterrupted row refer to simultaneous data (except for ν_{burst} values). Values of $\Delta\nu$ straddling two rows, or adjacent to a vertical line refer to measurements made over the range of frequencies indicated. Note: [a] Marginal detection.

(e.g., Fig. 9) which are somewhat reminiscent of a geographical map of an atoll because in the left limb of the U the motion through the diagram becomes slow, so that the track is usually broken up by observational windowing into "islands". The bottom and right-hand parts of the U are traced out in the form of a curved "banana" branch on a time scale of hours to a day. In a given source, the islands correspond to lower flux levels than the banana branch. Most atoll sources are in the 0.01–0.2 L_{Edd} range; a group of four bright ones (GX 3+1, GX 9+1, GX 13+1 and GX 9+9) that is nearly always in the banana branch is more luminous than this, perhaps 0.2–0.5 L_{Edd} (the distances are uncertain). Most timing and spectral characteristics of these sources depend in a simple way on position along the Z or atoll track. So, the phenomenology is essentially one-dimensional. A single quantity, usually referred to as "inferred accretion rate", varying on time scales of hours to days on the Z track and the banana branch, and more slowly in the island state must govern most of the phenomenology.

In all Z sources and in 4U 1728−34 the kHz QPOs are seen down to the lowest inferred $\dot{M}$ levels these sources reach. The QPOs always become undetectable at the highest $\dot{M}$ levels. In the atoll sources, where

Table 2: Observed frequencies of kilohertz QPOs in atoll sources

Source	ν_1 (Hz)	ν_2 (Hz)	$\Delta\nu$ (Hz)	ν_{burst} (Hz)	References
4U 0614+09		450			Ford et al. 1996, 1997a,b; Van der Klis et al. 1996d; Méndez et al. 1997; Vaughan et al. 1997,1998; Kaaret et al. 1998; van Straaten et al. 1999
	418	765	312±2		
	825	1160			
		1215			
		1330			
EXO 0748−676		695			J. Homan 1999, in prep.
4U 1608−52	415				Van Paradijs et al. 1996; Berger et al. 1996; Yu et al. 1997; Kaaret et al. 1998; Vaughan et al. 1997,1998; Méndez et al. 1998a,b, 1999; Méndez 1999; Markwardt et al. 1999b
	440	765	325±7		
	475	800	326±3		
	865	1090[b]	225±12[b]		
	895				
4U 1636−53	830				Zhang et al. 1996a,b,1997a; Van der Klis et al. 1996d; Vaughan et al. 1997,1998; Zhang 1997; Wijnands et al. 1997a; Méndez et al. 1998c; Méndez 1999; Markwardt et al. 1999b; Kaaret et al. 1999a
	900	1150			
	950	1190	251±4[b]	291,582	
		1230			
	1070				
4U 1702−43	625				Markwardt et al. 1999a,b
	655	1000[b]	344±7[b]		
	700	1040[b]	337±7[b]	330	
	770	1085[b]	315±11[b]		
	902				
4U 1705−44	775	1075[a]	298±11		Ford et al. 1998a
	870				
XTE J1723−376		815			Marshall & Markwardt 1999
4U 1728−34		325			Strohmayer et al. 1996a,b,c; Ford & van der Klis 1998; Méndez & van der Klis 1999; Méndez 1999; Markwardt et al. 1999b; di Salvo et al. 1999
	510	845			
	875	1160	349±2	364	
	920		279±12		
KS 1731−260	900	1160	260±10	524	Wijnands & van der Klis 1997
		1205			
4U 1735−44	630	980	341±7		Wijnands et al. 1996a, 1998b; Ford et al. 1998b
	730	1025	296±12		
	900[a]	1150	249±15		
		1160			
4U 1820−30		655			Smale et al. 1996, 1997; Zhang et al. 1998b; Kaaret et al. 1999b; Bloser et al. 1999
	500	860	358±42		
	795	1075	278±11		
		1100			
Aql X-1	670	1040[a]	241±9[a]	549	Zhang et al. 1998c; Cui et al. 1998a; Yu et al. 1999; Reig et al. 1999; M. Méndez et al. 1999 in prep.
	930				
4U 1915-05		820			Barret et al. 1997,1998; Boirin et al. 1999
	515				
	560	925			
	655	1005	348±11		
	705[a]	1055			
	880				
		1265[a]			
XTE J2123−058	845	1100	255±14		Homan et al. 1998b,1999a; Tomsick et al. 1999
	855	1130	276±9		
	870[a]	1140	270±5[a]		

Caption: see Table 1. Notes: [a] Marginal detection. [b] Shift and add detection method, cf. Méndez et al. (1998a).

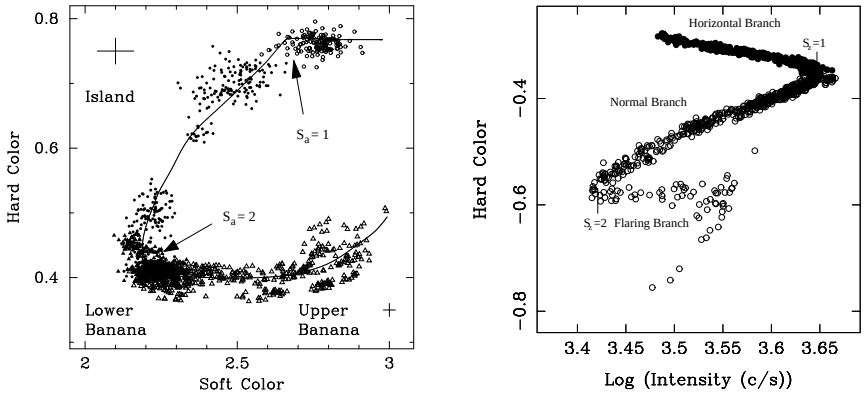

Figure 9: Left: X-ray color-color diagram of the atoll source 4U 1608−52. (Méndez et al. 1999) Right: X-ray hardness vs. intensity diagram of the Z source GX 340+0. (after Jonker et al. 1999c). Conventional branch names are indicated. Color is the log of a count rate ratio (3.5–6.4/2.0–3.5 and 9.7–16.0/6.4–9.7 keV for soft and hard color respectively); intensity is in the 2–16 keV band. kHz QPO detections are indicated with filled symbols.

the count rates are higher at higher inferred $\dot{M}$, this can not be a sensitivity effect. In most atoll sources, the QPOs are seen in the part of the banana branch closest to the islands, i.e., near the lower left corner of the U (Fig. 9); that they are often not detected in the island state may be related to low sensitivity at the low count rates there, but in one island in 4U 0614+09 the undetected lower kHz peak is really much weaker than at higher inferred $\dot{M}$ (Méndez et al. 1997). No kHz QPOs have been seen in the four bright atoll sources (Wijnands et al. 1998d, Strohmayer 1998, Homan et al. 1998a), perhaps because they do not usually reach this low part of the banana branch, and in several faint LMXBs, probably also atoll sources in the island state (SAX J1808.4−3658; Wijnands & van der Klis 1998c and XTE J1806−246; Wijnands & van der Klis 1999b, SLX 1735−269; Wijnands & van der Klis 1999c, 4U 1746−37; Jonker et al. 1999a, 4U 1323−62; Jonker et al. 1999b, 1E 1724−3045, SLX 1735−269 and GS 1826−238; Barret et al. 1999).

The QPO frequencies increase when the sources move along the tracks in the sense of increasing inferred $\dot{M}$ (this has been seen in more than a dozen sources, and no counterexamples are known). On time scales of hours to a day increasing inferred $\dot{M}$ usually corresponds to increasing X-ray flux, so on these time scales kHz QPO frequency usually increases with flux. When flux systematically decreases with inferred $\dot{M}$, as is the case in some parts of these tracks, the frequency is expected to maintain

its positive correlation with inferred $\dot{M}$ and hence become anticorrelated to flux, and indeed this has been observed in the Z source GX 17+2 (Wijnands et al. 1997b).

So, the kHz QPOs fit well within the pre-existing Z/atoll description of LMXB phenomenology in terms of source types and states, including the fact that position on the tracks in X-ray color-color or hardness-intensity diagrams ("inferred $\dot{M}$") and not X-ray flux drives the phenomenology.

8 Strong gravity and dense matter

Potentially kHz QPOs can be used to constrain neutron-star masses and radii, and to test general relativity. Detection of the predcited innermost stable circular orbit (ISCO) would constitute the first direct detection of a strong-field general-relativistic effect and prove that the neutron star is smaller than the ISCO. This possibility has fascinated since the beginning and was discussed well before the kHz QPOs were found (Kluźniak and Wagoner 1985, Paczynski 1987, Kluźniak et al. 1990, Kluźniak & Wilson 1991, Biehle & Blandford 1993). If a kHz QPO is due to orbital motion around the neutron star, its frequency can not be larger than the frequency at the ISCO. Kaaret et al. (1997) proposed that kHz QPO frequency variations that then seemed uncorrelated to X-ray flux in 4U 1608−52 and 4U 1636−53 were due to orbital motion near the ISCO and from this derived neutron star masses of $\sim 2 M_\odot$. However, from more detailed studies it is clear now that on short time scales frequency does in fact correlate to flux in these sources.

The maximum kHz QPO frequencies observed in each source are constrained to a narrow range. The 12 atoll sources with twin peaks have maximum ν_2 values in the range 1074–1329 Hz; among the Z sources there are two cases of much lower maximum ν_2 values ($<$900 Hz in GX 5−1 and GX 340+0), while the other four fit in. Zhang et al. (1997b) proposed that this narrow distribution is caused by the limit set by the ISCO frequency, which led them to neutron star masses near $2 M_\odot$ as well. It is in principle possible that the maximum is set by some other limit on orbital radius (e.g., the neutron star surface), or that it *is* caused by the ISCO, but the frequency we observe is not orbital, in which cases

no mass estimate can be made

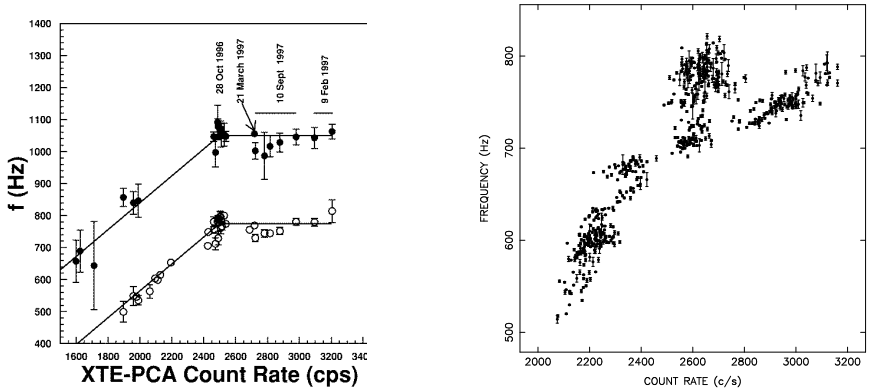

Figure 10: Left: evidence for a leveling off of the kHz QPO frequency with count rate in 4U 1820−30. (Zhang et al. 1998b) Right: a larger data set (see also Kaaret et al. 1999b) of 4U 1820−30 at higher resolution showing only the lower peak's frequency ν_1. (M. Méndez et al. in prep.).

Miller et al. (1996, 1998a) suggested that when the inner edge of the accretion disk reaches the ISCO, the QPO frequency might level off and remain constant while $\dot{M}$ continues rising. Later, an apparent leveling off at ν_2=1060±20 Hz with X-ray count rate was found in 4U 1820−30 (Zhang et al. 1998b; Fig. 10, left). If this is the orbital frequency at the ISCO, then the neutron star has a mass mass of $\sim$2.2M$_\odot$ and is smaller than its ISCO, and many equations of state are rejected. The leveling off is also observed as a function of X-ray flux and color (Kaaret et al. 1999b) and position along the atoll track (Bloser et al. 1999). However, the frequency vs. flux relations are known in other sources to be variable, and in 4U 1820−30 the leveling off seems not to be reproduced in the same way in all data sets (M. Méndez et al. in prep.; Fig. 10, right). It may also be more gradual in nature than originally suggested. No evidence for a similar saturation in frequency was seen in other sources, and most reach higher frequencies. Possibly, the unique aspects of the 4U 1820−30 system are related to this (it is an 11-min binary in a globular cluster with probably a pure He companion star; Stella et al. 1987a).

If a kHz QPO peak at frequency ν corresponds to stable Keplerian motion around a neutron star, one can immediately set limits on neutron star mass M and radius R (Miller et al. 1998a). For a Schwarzschild geometry: (1) the radius R of the star must be smaller than the radius r_K of

the Keplerian orbit: $R < r_K = (GM/4\pi^2\nu^2)^{1/3}$, and (2) the radius of the ISCO must *also* be smaller than r_K, as no stable orbit is possible within this radius: $r_{ISCO} = 6GM/c^2 < (GM/4\pi^2\nu^2)^{1/3}$ or $M < c^3/(2\pi 6^{3/2} G\nu)$. Condition (1) is a mass-dependent upper limit on R, and condition (2) an upper limit on M; neither limit requires detection of orbital motion at the ISCO. Fig. 11 (left) shows these limits in the neutron star mass-radius diagram for $\nu = 1220\,\mathrm{Hz}$, plus an indication of how the excluded area (*hatched*) shrinks for higher values of ν. The currently highest value of ν_2, identified in most models with the orbital frequency, is $1329\pm4\,\mathrm{Hz}$ (van Straaten et al. 1999), so the hardest equations of state are beginning to be imperiled by the method. Corrections for frame dragging (shown to first order in j in Fig. 11) expand the allowed region. They depend on the neutron star spin rate ν_s, and somewhat on the neutron star model, which sets the relation between ν_s and angular momentum. For 1329 Hz the above equations imply $M < 1.65\mathrm{M}_\odot$ and $R_{NS} < 12.4\,\mathrm{km}$; with corrections for a 300 Hz spin these numbers become $1.9\mathrm{M}_\odot$ and 15.2 km (van Straaten et al. 1999). Calculations exploring to what extent kHz QPOs constrain the EOS have further been performed by Miller et al. (1998b), Datta et al. (1998), Akmal et al. (1998), Kluźniak (1998), Bulik et al. (1999), Thampan et al. (1999), Li et al. (1999b), Schaab & Weigel (1999) and Heiselberg & Hjorth-Jensen (1999). If $\Delta\nu$ is near ν_s then the

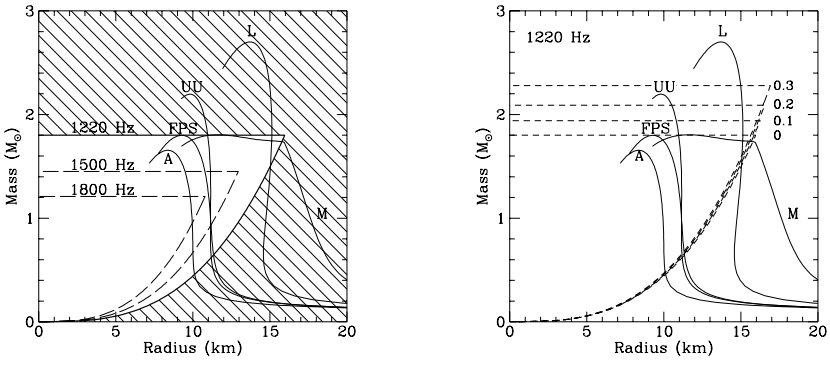

Figure 11: Constraints on the mass and radius of neutron stars from the detection of orbital motion with the frequencies indicated. Graphs are for negligible neutron star angular momentum (left) and for the values of $j = cJ/GM^2$ indicated (right). Mass-radius relations for some representative EOSs are shown. (Miller et al. 1998a)

18 neutron stars where this quantity has been measured all spin at frequencies between ∼240 and ∼360 Hz, a surprisingly narrow range. If the

stars spin at the magnetospheric equilibrium spin rates (e.g., Frank et al. 1992) corresponding to their current luminosities L_x, this would imply an unlikely, tight correlation between L_x and neutron-star magnetic-field strength (White & Zhang 1997; note that a similar possibility came up in the discussion about the uniformity of the QPO frequencies themselves. White & Zhang (1997) propose that when r_M is small, as is the case here, it depends only weakly on accretion rate (as it does in some inner disk models; cf. Ghosh & Lamb 1992, Psaltis & Chakrabarty 1999). Another possibility is that the spin frequency of accreting neutron stars is limited by gravitational radiation losses (Bildsten 1998a, Andersson et al. 1999, see also Levin 1999). If so, then gravitational radiation is transporting angular momentum out as fast as accretion is transporting it in, making these sources the brightest gravitational radiation sources in the sky. They would produce a periodic signature at the neutron star spin frequency, which would facilitate their detection.