Instabilities of galactic discs in the presence of star formation

Adi Nusser1,2
1Physics Department and the Asher Space Research Institute, Technion, Haifa 32000, Israel
2Astrophysics, Oxford University, Keble Rd, Oxford OX1 3RK

ABSTRACT

We discuss the stability of galactic discs in which the energy of interstellar clouds is gained in encounters with expanding supernova (SN) remnants and lost in inelastic collisions. Energy gain and loss processes introduce a phase difference between the pressure and density perturbations, making discs unstable on small scales for several recipes of star formation. This is in contrast to the standard stability analysis in which small-scale perturbations are stabilized by pressure. In the limit of small scales, the dispersion relation for the growth rate reduces to that of thermal instabilities in a fluid without gravity. If instabilities lead to star formation, then our results imply a secondary mode of star formation that operates on small scales and feeds on the existence of a primary mode on intermediate scales. This may be interpreted as positive feedback. Further, the standard stability criterion on intermediate scales is significantly modified.

Key words: instabilities – stars: formation – galaxies: structure.

1 INTRODUCTION

Instability and star formation in galactic discs are intimately linked. The observational support for this is the significant decline of star formation activity in regions with relatively low gas mass surface density, as expected from various stability analyses (e.g. Martin & Kennicutt 2001). Theoretically, the collapse of cold gas provides ripe conditions for the appearance of molecular clouds and, subsequently, star formation. (e.g. Spitzer 1968; Quirk 1972).

On large scales (of the order of the size of the disc), perturbations are stabilized by rotation. Discs are also stable on small scales if the pressure, \( P \), and density, \( \rho \), are related by \( P \propto \rho^\gamma \), where \( \gamma \) is a constant. On intermediate scales, neither rotation nor pressure can stop gravity from amplifying the perturbations if the surface density is above a critical value. The first detailed stability calculations for single fluid discs have been performed in two classic papers by Toomre (1964) and Goldreich & Lynden-Bell (1965). A similar type of analysis has been generalized to the more realistic case of discs containing stellar and gaseous components (Jog & Solomon 1984; Rafikov 2001; Griv, Gedalin & Yuan 2002). Viscous discs have also been considered and found to be always unstable, with small-scale perturbations growing at a low rate that is set by the viscosity (Gammie 1996). These studies have assumed that the perturbation in the pressure is proportional to that in the density.

Here, we consider the instability of the interstellar cloud component of the disc, when it is subject to local energy gain and loss processes. Clouds gain energy in encounters with expanding supernova (SN) remnants and lose energy in inelastic collisions among each other (McKee & Ostriker 1977, hereafter MO77). Both of these processes depend, in general, on the local surface density and velocity dispersion of the clouds. The star formation rate is enhanced above the global average in a region with a positive density fluctuation, resulting in a local increase of the velocity dispersion. A perturbation in the velocity dispersion changes the energy loss rate and may also affect the star formation rate. We perform a detailed stability analysis under the assumption that the cloud component be treated as a hydrodynamical fluid over sufficiently large length and temporal scales. The treatment here leads to a significant modification of the standard stability criteria. We get instabilities on much smaller scales than what is inferred from the standard analysis. In several important cases, instabilities extend down to the scale where the fluid treatment becomes invalid.

The paper is organized as follows. The equations are derived and expanded in Laplace transforms in Section 2. A stability analysis is presented in Section 3. This section includes a classification of stable and unstable modes by means of Nyquist diagrams, a discussion of a few limiting cases. In Section 4 we present a treatment of instabilities with specific form for the energy gain and loss rates. The validity of our approach is assessed in Section 5. A summary and discussion of the results are presented in Section 6.

2 THE EQUATIONS

We write the equations governing the evolution of perturbations in a thin gaseous disc. The gas represents interstellar clouds that, for simplicity, are assumed to have isotropic velocity dispersion. We assume that the gas has an ideal equation of state that after
integration over the height of the disc yields
\[ p = (\gamma - 1) \mu \varepsilon \] (1)
to relate the ‘two-dimensional’ pressure, \( p \), to the mass surface density, \( \mu \), and the ‘random’ kinetic energy per unit mass, \( \varepsilon \) = \( V_{\text{rms}}^2/2 \), where \( V_{\text{rms}} \) is the three-dimensional velocity dispersion of the clouds, and \( \gamma > 1 \). We will also refer to \( \varepsilon \) as internal energy. The adiabatic index \( \gamma \) relates the projected quantities and, therefore, is in general different from the physical three-dimensional index, \( \Gamma \). The relation between \( \gamma \) and \( \Gamma \) depends on the structure of the disc. For a self-gravitating disc \( \gamma = 3 - (2/\Gamma) = 9/5 \) for \( \Gamma = 5/3 \) (e.g. Gammie 2001). The difference between the two indices has little effect in our applications and so we present numerical results only for \( \Gamma = 5/3 \). We write energy conservation in the following form,
\[ \frac{\partial \varepsilon}{\partial t} = \frac{p}{\mu^2} \frac{\partial \mu}{\partial r} + G - L, \] (2)
where \( G(\mu, \varepsilon) \) and \( L(\mu, \varepsilon) \) are, respectively, the energy gain and loss rates per unit mass, assumed to be explicit functions of \( \mu \) and \( \varepsilon \) alone. In the meantime, we will study some general features of the stability of the system without referring to specific forms for \( G \) and \( L \).

Let the disc be in a steady state in which the disc is axisymmetric, gravity is balanced by rotation and energy gain is balanced by energy loss, i.e. \( G = L \). This steady state is characterized by the rotational speed, \( \Omega(r) \), the gas mass surface density, \( \mu_0(r) \), and the internal energy per unit mass, \( \varepsilon_0(r) \), as functions of the distance from the centre of the disc, \( r \). We are interested in the linear response of the system to small perturbations from this steady state. For simplicity, we consider only radial perturbations as the complication introduced by non-axisymmetric perturbations does not hide any additional relevant physical effects. Let \( u, v, \mu^1, \varepsilon^1, \mu^3, \varepsilon^3 \) and \( \phi^1 \) be, respectively, the perturbations in the radial velocity, the transverse velocity, the surface density, the internal energy, the pressure and the gravitational potential field. The first-order equations governing the evolution of these perturbations are:
\[ p^1 = (\gamma - 1) \mu_0 \varepsilon^1 + (\gamma - 1) \varepsilon_0 \mu^1, \] (3)
\[ \partial_t \varepsilon^1 = \frac{\rho_0}{\mu_0^2} \partial_r \mu^1 + C_\mu \mu^1 + C_\varepsilon \varepsilon^1, \] (4)
\[ \partial_t u - 2\Omega(r) v + \partial_r p^1/\mu_0(r) + \partial_r \phi^1|_{r=0} = 0, \] (5)
\[ \partial_t v - 2B(r) u = 0, \] (6)
\[ \partial_r \mu^1 + \partial_r (r \mu_0 u) = 0, \] (7)
\[ r \partial_r \phi^1|/r + \partial_r^2 \phi^1 = 4\pi G \mu_1 \delta(z), \] (8)
where \( \rho_0 = (\gamma - 1) \varepsilon_0 \mu_0 \), and
\[ C_\mu = \left( \frac{\partial (G - L)}{\partial \mu} \right)_{\mu_0, \varepsilon_0} \quad \text{and} \quad C_\varepsilon = \left( \frac{\partial (G - L)}{\partial \varepsilon} \right)_{\mu_0, \varepsilon_0}. \] (9)
The first two of these equations are obtained by linearization of equations (1) and (2), respectively. The equations (5) and (6) are the radial and transverse linear versions of the Euler equations. Mass conservation is represented by equation (7) and the Poisson equation by equation (8). We expand the perturbations in Fourier modes, \( \exp(i \mathbf{k} \cdot \mathbf{r}) \). Considering only modes satisfying \( kr \gg 1 \), the linear equations lead to,
\[ \varepsilon^1 = \frac{\rho_0}{\mu_0^2} \mu_1 + C_\mu \mu_1 + C_\varepsilon \varepsilon_1 \] (10)
\[ \mu^1 + [k^2 - 2\pi G \mu_0 k + (\gamma - 1) \varepsilon_0 k^2] \mu_1 + (\gamma - 1) \varepsilon_0 k^2 \varepsilon_1 = 0, \] (11)
where \( \mu_1 \) is defined by \( \mu_1(r, t) = \mu_1(t) \exp(i \mathbf{k} \cdot \mathbf{r}) \) and similarly for \( \varepsilon_1 \), and \( \kappa^2 = 2\Omega(1 + (d \ln \Omega/d \ln r)/2)^{1/2} \) is the epicyclic frequency. In deriving these equations, the relation \( \partial_r \phi^1|_{r=0} = i 2\pi G \mu_1 \exp(i \mathbf{k} \cdot \mathbf{r}) \) has been used (Toomre 1964).

The differential equations (10) and (11) are linear with constant coefficients and they can be solved by means of Laplace transformation (see Appendix A). By taking the Laplace transform of equations (10) and (11), we get
\[ sE - \varepsilon_1(0) = (\gamma - 1) \frac{\varepsilon_0}{\mu_0^2} [sM - \mu_1(0)] + C_\mu M + C_\varepsilon E, \] (12)
\[ s^2M - s\mu_1(0) - \mu_1(0) + [k^2 - 2\pi G \mu_0 k + (\gamma - 1) \varepsilon_0 k^2]M + (\gamma - 1) \varepsilon_0 k^2 E = 0, \] (13)
where \( M(s) \) and \( E(s) \) are the Laplace transforms of \( \mu_1(t) \) and \( \varepsilon_1(t) \), respectively, and \( \mu_1(0), \mu_1(0) \) and \( \varepsilon_1(0) \) represent the initial conditions given at \( t = 0 \). Solving for \( M \), yields
\[ B(s)M(s) = (s - C_s)[s\mu_1(0) + \mu_1(0)] \] (14)\[ + (\gamma - 1) \mu_0 k^2 \varepsilon_1(0) - (\gamma - 1) \frac{\varepsilon_0}{\mu_0} \mu_1(0), \] (15)
where
\[ B(s) = s^3 - s^2 C_s + s\omega_0^2 - C_s \omega_0^2, \] (16)
with
\[ \omega_0^2(k) = k^2 - 2\pi G \mu_0 k + (\gamma - 1) \varepsilon_0 k^2, \] (17)
\[ \omega_0^2(k) = \omega_0^2 - (\gamma - 1) \varepsilon_0 k^2 \left( \frac{\mu_0 C_\mu}{\varepsilon_0 C_\varepsilon} + \gamma - 1 \right). \] (18)
The third-order polynomial \( B(s) \) has at most three distinct roots, \( s_j \). Because \( M(s) \propto 1/B(s) \) then according to the theory of Laplace transforms (see Appendix A) \( \mu_1(t) \) is a linear combination of \( \exp(s_j t) \). Therefore, an unstable (growing) mode corresponds to a root with a positive real part.

In the limit of small scales, i.e. large \( k \), the terms proportional to \( k^2 \) are dominant in equations (17) and (18). This means that the effects of rotation and self-gravity are negligible. In this limit, the characteristic equation \( B(s) = 0 \) reduces to the dispersion relation derived by Field (1965) for the growth rate of thermal instabilities (see equation 15 in Field 1965, without the thermal conduction term).

### 3 Stability Analysis

Our task is to establish the relevant time-scales for the evolution of the perturbations, and classify the stable and unstable modes. The three time-scales corresponding to the roots \( s_j \) of \( B(s) \) must be reflected in \( C_s, \omega_0 \), and \( \omega_1 \). We will find that \( C_s \) is negative for the specific forms of \( G \) and \( L \) we use below. The dependence on \( C_s \) can be scaled out by working with the variables \( \tilde{s} = s/(-C_s), \tilde{\omega}_0 = \omega_0/(-C_s) \) and \( \tilde{\omega}_1 = \omega_1/(-C_s) \). If \( \tilde{s} = j = 1, 2, 3 \) solves
\[ \tilde{s}^3 + \tilde{s}^2 + \tilde{s}\tilde{\omega}_0^2 + \tilde{\omega}_0^2 = 0, \] (19)
then \( s_j = -C_s \tilde{s}_j \). Therefore, the stability of a mode with a wave-number \( k \) depends only on \( \omega_0^2(k) \) and \( \omega_1^2(k) \). The time-scales, however, may depend non-trivially on \( C_s \).

The roots of \( B(s) \) can either be all real (with zero imaginary parts) or one real and two conjugate complex roots (i.e. having the same
Disc instabilities with star formation

3.1 Limiting cases

Nyquist diagrams do not contain any information on the instability time-scales. Those could be obtained by numerically finding the roots of \( B(s) \). Here, we discuss a few important limiting cases that can be studied analytically. Let \( \Delta \omega^2 = \omega^2 - \omega_0^2 \). First consider the limit \( |\Delta \omega^2| \ll |\omega_0^2| \) in which the roots differ by a small amount from the values \( \delta = -1 \) and \( \delta = \pm \sqrt{-\omega^2} \), which solve equation (19) with \( \Delta \omega^2 = 0 \). In this case,

\[
x_1 = -|C_i| \frac{1 + (\omega_0/C_i)^2}{1 + (\omega_0/C_i)^2},
\]

\[
x_{2,3} = |C_i| \frac{\mp \sqrt{\omega_0^2 - \omega_0^2 - \omega_0^2}}{\omega_0^2 + \omega_0^2} \pm \sqrt{-\omega_0^2}.
\]

Note that these expressions involve \( C_i \), in a complicated way. For \( \omega_0^2 > \omega_0^2 > 0 \), the system is unstable, in accordance with the Nyquist analysis, with a growth time-scale of \( |\Re(s_2)|^{-1} = |(1/2)C_i(\omega_0^2 - \omega_0^2)/(C_i^2 + \omega_0^2)|^{-1} \). For \( \omega_0^2 > \omega_0^2 > 0 \), there are no unstable modes.

The second and more important limit is when \( \omega_0^2 \gg 1 \) and \( \Delta \omega^2 = O(\omega_0^2) \). This holds for small-scale perturbations because for sufficiently large \( k \) the leading term in \( \omega_0^2 \) is the one involving \( k^2 \) (see equation 17). The roots of \( B(s) \) are approximated by

\[
x_1 = -|C_i| \frac{\omega_0^2}{\omega_0^2} \text{ and } x_{2,3} = |C_i| \frac{\omega_0^2 - \omega_0^2}{\omega_0^2} \pm i\omega_0.
\]

For \( \omega_0^2 > \omega_0^2 \), the root \( s_1 \) represents a decaying mode, while the conjugate roots \( s_2 \) and \( s_3 \) correspond to sound waves that grow at a rate proportional to \( |C_i| \) and oscillate with frequency of \( \omega_0/2\pi \).

The origin of the growth is heating that occurs during the compression of the waves (cf. Field 1965). In terms of \( C_\mu \) and \( C_\sigma \), the condition, \( \omega_0^2 > \omega_0^2 \), for the appearance of growing sound waves reads

\[
\frac{\mu_0 C_\mu}{\varepsilon_0 C_\sigma} + \sigma - 1 < 0,
\]

which is similar to the condition in equation (24) of Field (1965). For \( \omega_0^2 > \omega_0^2 > 0 \), all modes are decaying with \( s_1 \) and \( s_3 \) representing overstables (oscillating) modes. For \( \omega_0^2 > 0 \) but \( \omega_0^2 < 0 \), the mode \( s_1 \) is growing without oscillations, and \( s_2 \) and \( s_3 \) are overstable modes.

The mode described by \( s_1 \) corresponds to growth by condensation under nearly constant pressure conditions as described by Field (1965). It can result from an enhanced cooling efficiency as the density is increased. The condition \( \omega_1^2 < 0 \) holds for

\[
1 - \frac{\mu_0 C_\mu}{\varepsilon_0 C_\sigma} < 0.
\]

This is equivalent to the condition for a condensation mode as given by equation (23) of Field (1965).

3.2 Numerical solutions

To better visualize the time dependence of perturbations, we plot in Fig. 1 the functions \( \mu_1(t) \) and \( \varepsilon_1(t) \) for \( \omega_0^2 < 0 \) (top panel), \( \omega_0^2 > \omega_0^2 > 0 \) (middle) and \( \omega_0^2 > \omega_0^2 \) (bottom). The figure shows only cases with \( \omega_0^2 > 0 \), which is always true for sufficiently large \( k \).

These curves have been obtained by numerical integration of the following dimensionless form of equations (10) and (11),

\[
\tilde{\mu}_1 + \left(1 - \frac{2}{Q_0} \tilde{k} \right) \tilde{\mu}_1 + \tilde{k}^2 \tilde{\varepsilon}_1 = 0,
\]

\[
\tilde{\varepsilon}_1 = (\gamma - 1) \tilde{\mu}_1 + \tilde{C}_\mu \tilde{\mu}_1 + \tilde{C}_\varepsilon \tilde{\varepsilon}_1,
\]

where the prime symbol denotes the differential with respect to the variable \( \tau = \kappa t_k = \pi G \mu_0/(\gamma - 1) \varepsilon_0 Q_0 \), \( \tilde{k} = k/k_0 \), \( Q_0 = k/(\gamma - 1) \varepsilon_0 \varepsilon_0 \). \( \tilde{\varepsilon}_1 = \varepsilon_1/\varepsilon_0 \), \( \tilde{\mu}_1 = \mu_1/\mu_0 \), \( \tilde{C}_\mu = \mu_0 C_\mu/\varepsilon_0 \varepsilon_0/k \) and \( \tilde{C}_\varepsilon = C_\varepsilon/k \). All solutions are for \( \tilde{C}_\mu = -1, \tilde{k} = 5, Q_0 = 1 \) and \( \gamma = 5/3 \). The solutions in the top, middle and bottom panels correspond to \( \tilde{C}_\mu = -1.3, 0 \) and 2, respectively. The initial conditions of all solutions are \( \tilde{\mu}_1(0) = \tilde{\varepsilon}_1(0) = 1 \) and \( \tilde{\mu}_1' = 0 \). The behaviour of the solutions is in agreement with Nyquist diagrams.

Figure 1. The time evolution of the perturbations in the surface density (solid line) and internal energy (dotted) in arbitrary units. The curves have been obtained by direct numerical integration of the equations of motion.
and the limiting cases. For \( \omega_0^2 > \omega_0^2 > 0 \) (bottom panel), the solution is oscillatory with a growing envelope. These growing sound waves arise because of the slight heating during compression. The case \( \omega_0^2 > \omega_0^2 > 0 \) is overstable. Note that the derivative of \( \tilde{e}_1 \) at \( \tau = 0 \) is positive although the dotted curve may appear as declining at \( \tau = 0 \).

The phase difference between the dotted and solid curves, however, becomes more evident as \( \tau \) increases.

For \( \omega_1 < 0 \) and \( \omega_2 > 0 \), the solution is unstable without oscillations. This mode is driven by thermal instability of a condensation mode (Field 1965). Note the phase difference between \( \mu_1 \) and \( \epsilon_1 \). This phase difference is the drive behind the instabilities shown in the top and bottom panels.

4 STABILITY WITH SPECIFIC FORMS OF \( G \) AND \( L \)

A key parameter is the ratio \( \mu_0 C_\mu / \varepsilon_0 C_\varepsilon \), which determines the difference between \( \omega_2^2 \) and \( \omega_0^2 \) (see equation 18). This ratio depends on the specific forms chosen to describe the energy gain and loss functions.

The function \( L(\mu, \varepsilon) \) represents energy dissipation per unit mass, via inelastic cloud collisions. Hereafter, we adopt the following form (MO77; Efistathiou 2000).

\[
L = \eta_1 \mu^2 e^{\kappa^2/2}.
\]

The energy gain function, \( G \), is directly proportional to the star formation rate (MO77). We work with two distinct choices, the first is a Schmidt–Kennicutt (e.g. Kennicutt 1989) form and the second assumes that the star formation rate depends explicitly on the Toomre parameter, which governs instabilities on intermediate scales.

4.1 Stability with a Schmidt–Kennicutt star formation law

Here, we write the energy gain function as

\[
G = \eta_0 \mu^n.
\]

where \( \eta_0 \) is a numerical factor having the appropriate units. Equating this \( G \) with \( L \) given by equation (25) gives

\[
C_\mu = (n - 2) \eta_0 \mu_0^{n-1} \quad \text{and} \quad C_\varepsilon = -\frac{1}{2} \frac{\eta_0^2}{\varepsilon_0} \mu_0^{4-n}.
\]

Therefore, using equation (18),

\[
\omega_1^2 = \omega_0^2 + (\gamma - 1)(2n - 3 - \gamma)\kappa^2.
\]

Of particular interest is the small-scale behaviour. For sufficiently large \( k \), we write \( \omega_0^2 \approx \gamma(\gamma - 1)\kappa^2 \). Therefore, for \( n < 3/2 \), we have \( \omega_2^2 < 0 \), while for \( n > 3 + \gamma/2 \), we have \( \omega_2^2 > \omega_0^2 \). The disc is unstable in these two cases. However, the nature of the instability is different: the former is described by a single non-oscillatory mode, while the latter has two growing oscillatory modes (see Fig. 1 and the discussion of Nyquist diagrams at the end of Section 3).

We plot in Fig. 2 the real parts of the roots as a function of the wave number \( k \) for \( n = 1.5 \) (top panel) and \( n = 2 \) (bottom). The dashed line is the instability growth rate according to the standard analysis à la Toomre (1964). This is given by \( \sqrt{-\omega_0^2} \) for \( \omega_0^2 < 0 \), and zero otherwise. A value of \( \gamma = 5/3 \) is used, and \( \kappa \), \( \mu_0 \) and \( \varepsilon_0 \) have been tuned so that the maximum of the dashed curve is consistent with the growth rate inferred from the lower curve of fig. 1(a) of Jog & Solomon (1984). We use \( |C_\varepsilon| = 20 \text{km} \text{s}^{-1} \text{kpc}^{-1} \), corresponding to a time-scale of \( 5 \times 10^7 \text{yr} \) (MO77). The solid line shows the positive real parts of the roots. According to the solid curve, the system with \( n = 1.5 \) remains unstable on small scales, albeit with a longer time-scale than the instabilities at intermediate scales. For \( n = 2 \), the instability extends to smaller scales than what would be inferred from the standard analysis (dashed line).

4.2 \( Q \)-dependent star formation rate

Assume now that the star formation rate depends on the time-scale defined by a standard stability analysis (e.g. Toomre 1964). We work with a gain function of the form (e.g. Wang & Silk 1993; Elmegreen 1999)

\[
G = \eta G \mu F(Q),
\]

where

\[
Q = \frac{\kappa \varepsilon^{1/2}}{\pi G \mu} \quad \text{and} \quad F = \left(1 - \frac{Q^2}{2}ight)^{1/2}.
\]

Note that the meaning of \( \eta G \) is different from equation (26); the same symbol is used only for the sake of brevity. Energy balance \( (G = L) \) at \( \mu = \mu_0 \) and \( \varepsilon = \varepsilon_0 \) yields

\[
F(Q) = Q/K,
\]

where \( 1/K = (\eta G / \eta_0)(2\pi G / \kappa^2) \mu_0^2 \). This relation gives

\[
Q^2 = -\frac{K^2}{2} + \frac{K^2}{2} \left(1 + \frac{4}{K^2}\right)^{1/2}.
\]

For \( K \ll 1 \) and \( K \gg 1 \), we have \( Q \approx K^{1/2} \) and \( Q = 1 \), respectively. The first-order variation of \( (G - L) \) is

\[
\delta G - \delta L = \eta G \left(F \mu_1 + \mu_2 \frac{\partial F}{\partial Q} \frac{\partial Q}{\partial \mu} + \mu_1 \frac{\partial F}{\partial Q} \frac{\partial Q}{\partial \varepsilon}ight) - 2 \eta \mu \mu_1 \varepsilon_1.
\]
Because $d \ln F/d \mu = -Q/\mu$, $d \ln F = Q/(2 \epsilon)$ and $Q^2 = 2\pi G \mu Q/k$, we get

$$C_\mu = -\eta G k \left( 1 + \frac{d \ln F}{d \ln Q} \right) F, \quad (35)$$

$$C_\epsilon = -\frac{\eta G}{2 \epsilon} \left( 1 - \frac{d \ln F}{d \ln Q} \right) F. \quad (36)$$

Because $d \ln F/d \ln Q < 0$, the coefficient $C_\epsilon$ is negative. By evaluating the logarithmic derivative in terms of $Q$, we find $\mu C_\mu/\epsilon C_\epsilon = 2Q^2/(Q^2 - 2)$ and

$$\omega_i^2 = \omega_o^2 - (\gamma - 1) \epsilon k^2 \left( \gamma - 1 + \frac{2Q^2}{Q^2 - 2} \right). \quad (37)$$

For $Q^2 > 2(\gamma - 1)/\gamma + 1)$, we have $\omega_i^2 > \omega_o^2$, yielding unstable modes even on small scales. For lower $q$, the system is stable on small scales but, as in the case with $n = 2$ of the previous subsection, unstable modes extend to scales smaller than in the standard analysis. The ratio $\omega_i^2/\omega_o^2$ as a function of $Q$ is plotted in Fig. 3, in the limit of very large $k$. A value of $\gamma = 5/3$ is used in this figure.

5 VALIDITY OF THE APPROACH

The treatment of the clouds as a hydrodynamical fluid is valid over scales larger than the mean free path, $l_f$, for cloud–cloud collisions. In the solar neighbourhood, $l_f \approx 100$ pc for the cold neutral cores and about 10 pc if the warm envelopes are included (MO77). Although the warm envelopes should contribute to the collisions, we take $l_f \approx 100$ pc as an upper limit.

Further, there are two time-scales to be considered. First, there is the mean time between collisions, $t_{\text{coll}} = l_f/V_{\text{rms}} \approx 3 \times 10^6$ yr for $V_{\text{rms}} = 10$ km s$^{-1}$. Second, there is the mean time, $t_{\text{reheat}}$, between two successive ‘heating’ events of a cloud. This comes about because we assume that multiple encounters with expanding SN remnants eventually amount to an increase in random kinetic energy rather than bulk motions. According to Cox & Smith (1974), $t_{\text{reheat}} \sim 10^6$–$10^7$ yr.

There is also the length-scale over which the velocity dispersion can equilibrate in a time $t = \max(t_{\text{reheat}}, t_{\text{coll}})$. This scale arises from the finite mean free path of the clouds (see also Gammie 1996) and is similar to the usual heat conduction (e.g. Zel’dovich & Raizer 2002). We estimate this scale as $(V_{\text{rms}} l_f)^{1/2}$. Taking $t = t_{\text{coll}}$ yields a scale of $(t_{\text{coll}} V_{\text{rms}} l_f)^{1/2} = l_l$.

Therefore, for an environment like the solar neighbourhood, our approach is valid over scales larger than 100 pc. The scale is sufficiently small that our treatment remains interesting. For example, this scale is smaller than the scale of unstable modes found by means of a standard stability analysis (e.g. Toomre 1964; Jog & Solomon 1984).

Finally, there is a continuous process of cloud destruction and production. We assume that this process is rapid and always produces a fixed spectrum for the cloud size distribution.

6 SUMMARY AND DISCUSSION

We have considered the instability of gaseous discs subject to local energy gain and loss processes in the presence of star formation. The gas represents interstellar clouds and the energy is gained in repeated encounters of the clouds with expanding SN remnants and is lost in inelastic cloud–cloud collisions. These energy exchange processes introduce a phase difference between the density and pressure perturbations. In several interesting situations this phase difference causes the pressure to amplify the density perturbation. For a star formation rate proportional to $\mu^2$, the instabilities extend to much smaller scales than what is inferred from standard stability analyses à la Toomre (1964) and Goldreich & Lynden-Bell (1965).

The small-scale instabilities may be responsible for triggering further star formation. Therefore, there may be two modes of star formation, a primary mode on intermediate scales and a secondary mode operating on small scales at a rate that is determined by the coefficient $C_\epsilon$ (see equation 9). This can be interpreted as positive feedback in which the intermediate scale mode is driving star formation on smaller scales through the development of instabilities. It is unclear what gain function one should use in equation (2). A fully self-consistent stability analysis should incorporate the energy gain resulting from the star formation mode that is induced by the instabilities. This complicates the problem substantially because the form of the gain function in this case must be derived from the stability analysis self-consistently. Nevertheless, we have found that small-scale instabilities develop at some level for most generic forms of the gain function. We do not expect a fully self-consistent treatment to change our conclusions significantly.

We have not included the coupling of gas to the stellar component. This could easily be done, but does not affect the main conclusions of the present work. Including the stellar component would tend to destabilize small-scale modes even further (Jog & Solomon 1984).

ACKNOWLEDGMENTS

The author thanks Joe Silk for stimulating discussions, and an anonymous referee for pointing out an error in a previous version of the paper and for several useful comments and clarifications.

REFERENCES

Cox D. P., Smith B. W., 1974, ApJ, 189, L105
Efstathiou G., 2000, MNRAS, 317, 697
Elmegreen B. G., 1999, in Nakamoto T., ed., Proc. Star Formation 1999.
Nobeyama Radio Observatory, Nobeyama, p. 3
Field G., 1965, ApJ, 142, 531
Gammie C. F., 1996, ApJ, 462, 725
Gammie C. F., 2001, ApJ, 553, 174
Goldreich P., Lynden-Bell D., 1965, MNRAS, 130, 97
Griv E., Gedalin M., Yuan C. 2002, A&A, 383, 338
Jog C. J., Solomon P. M., 1984, ApJ, 276, 114
the integration path. Therefore, by the residue theorem, we have
\begin{equation}
\int_C f(z) \, dz = 2\pi i \sum \text{Res}(f, a_i),
\end{equation}

where \( a_i \) are the poles of \( f(z) \).

As an example, consider \( f(z) = \frac{1}{z^2(z-1)} \). This function has a pole of order 2 at \( z = 1 \) and a simple pole at \( z = 0 \).

We will need the Laplace transforms of first and second derivatives of a function. Using equation (A1), these transforms can be related to the \( f(t) \) by
\begin{align}
\mathcal{L}\{f'(t)\} &= s f(s) - F(0) \\
\mathcal{L}\{f''(t)\} &= s^2 f(s) - s F(0) - F'(0),
\end{align}

where the prime and double prime denote first- and second-order derivatives, respectively. The Bromwich integral expresses \( F(t) \) in terms of \( f(s) \) as
\begin{equation}
F(t) = \frac{1}{2\pi i} \int_{t-i\infty}^{t+i\infty} \exp(st) f(s) \, ds,
\end{equation}

where \( i = \sqrt{-1} \) and \( y \) is a real number chosen so that all poles of \( f(s) \) lie, in the complex plane, to the left of the vertical line defining the integration path. Therefore, by the residue theorem, we have
\begin{equation}
F(t) = \sum \text{Res} \{\exp(st) f(s)\}.
\end{equation}

As an example, consider \( f(s) = \frac{1}{s(s-1)(s-2)} \), which has two simple poles at \( s = 1 \) and \( s = 2 \). The residues of \( \exp(st) f(s) \) at these poles are \( \exp(s_1 t) / (s_2 - s_1) \) and \( -\exp(s_2 t) / (s_1 - s_2) \) so that, by equation (A4),
\begin{equation}
F(t) = \exp(s_1 t) / (s_2 - s_1) - \exp(s_2 t) / (s_1 - s_2).
\end{equation}

\( F(t) \) has a pole of order 2 at \( s_1 \). The residue in this case is
d \( [s - s_1]^2 \exp(st) f(s)] / ds \) evaluated at \( s = s_1 \). Therefore,
\begin{equation}
F(t) = t \exp(s_1 t).
\end{equation}

**APPENDIX B: STABILITY BY MEANS OF NYQUIST DIAGRAMS**

Given a polynomial, \( P_N(s) \), of degree \( N \), Nyquist diagrams provide an easy and elegant method for determining the number of its roots that lie to the right of the imaginary axis, i.e. roots with a positive real part. The method can be summarized as follows:

(i) plot a contour of the value of the polynomial in the complex plane when its argument, \( s \), is varied from \( s = +i\infty \) to \( s = -i\infty \) along the imaginary axis; and

(ii) use the contour diagram to compute the increase, \( \Delta \psi \), in the phase of \( P_N(s) \) as \( s \) moves along the imaginary axis.

The number of roots with positive real parts is then \( \Delta \psi + N\pi \). (2\pi).

Fig. B1 shows Nyquist diagrams for the polynomial \( B(s) \) given in equation (16) for three cases as indicated in the figure. In the top panel, the phase changes from \( -\pi/2 \) to \( -3\pi/2 \) and therefore the number of roots with positive real parts is one. Further, in this case this root has zero imaginary part because otherwise its conjugate would also be a root and there would be two roots with positive (and equal) real parts. In the middle panel, the change is \( -3\pi \) and no roots lie to the right of the imaginary axis. In the bottom panel, there are two roots with positive real parts.

This paper has been typeset from a TeX/\LaTeX\ file prepared by the author.

\[ \omega_1^2 > 0 \]
\[ \omega_2^2 < 0 \]
\[ \Delta \psi = -\pi \]

\[ \omega_1^2 < 0 \]
\[ \Delta \psi = -3\pi \]

\[ \omega_0^2 > 0 \]
\[ \Delta \psi = +\pi \]

Figure B1. Nyquist diagrams of the polynomial \( B(s) \).