Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **540,** 2975–2990 (2025) https://doi.org/10.1093/mnras/staf771
Advance Access publication 2025 May 13

How stellar winds reach supersonic speed

Paul Song,[1]* Jiannan Tu,[1] Stanley W. H. Cowley,[2] Chi Wang[3] and Hui Li[3]

[1]*Space Science Laboratory and Department of Physics, University of Massachusetts Lowell, Lowell, MA 01854, USA*
[2]*Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK*
[3]*National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China*

Accepted 2025 May 1. Received 2025 March 25; in original form 2025 January 4

ABSTRACT

In classical theory of stellar wind formation, supersonic stellar wind starts with an initial speed of the eigenspeed at the inner boundary of the corona, goes along a continuous eigenfunction, reaches the sonic point while requiring a critical condition to be satisfied, and becomes supersonic across the critical point. If the initial speed is smaller than the eigenspeed, the wind is subsonic, and if the initial speed is greater than the eigenspeed, stellar wind cannot form. Since the initial flow speed is determined by ionization processes at the top boundary of the chromosphere and chromospheric and coronal heating processes, the initial speed of the wind can often be greater than the eigenspeed, posing a dilemma to the classical stellar wind theory which predicts no stellar wind under such conditions. We examine the classical stellar wind evolution equation and find that when the initial speed is greater than the eigenspeed, it cannot hold at the sonic point. In 1D steady state gasdynamics, the evolution equation can be rewritten with two expressions, one at the sonic point and one for everywhere else. The Rankine–Hugoniot relations are used to connect the solutions across the sonic point. A discontinuity standing in the flow that travels at the local sonic speed can facilitate the sonic transition. Supersonic winds can form when the inner boundary speed is greater than the eigenspeed. The critical solution separates the parameter regimes of supersonic from subsonic winds, and most supersonic stellar winds do not go through the critical point or critical points.

Key words: Sun: corona – Sun: fundamental parameters – solar wind – Sun: transition region – star: coronae – star: winds, outflows.

1. INTRODUCTION

1.1. Critical solution

One of the fundamental problems in solar physics is the formation of the supersonic solar wind. The conventional theory is based on analysis of the types of solutions of the solar/stellar wind evolution equation which can be derived from the continuity and momentum equations (Parker, 1958, 1964a, b, 1965a, b) as shown in Fig. 1. In the figure, thin solid lines show family curves of the flow velocity solved from the evolution equation. The dashed curve shows the sonic line, and the sonic point is where a thin solid line crosses the dashed curve. Supersonic solar/stellar wind is indicated by the segments of thin solid lines above the dashed curve. The most important feature of this figure is the critical point where the sonic line, thick solid line AB, and thick line CD cross each other. Line AB is an eigenfunction or the critical solution. The speed at the inner boundary of the corona, point A in Fig. 1, is called the eigenspeed V_e, derived backward from the critical point along the eigenfunction. This specific initial flow speed at the inner boundary determines whether supersonic solar/stellar wind can form or not. If the solar/stellar outflow starting in the corona is subsonic, in the lower-left corner, only line AB represents a supersonic solar/stellar wind that can flow to the earth's orbit, far on the right of the chart. If the solar/stellar wind speed generated in the transition region and/or the temperature produced by the heating in the corona does not exactly satisfy the requirement at the critical point, supersonic solar/stellar wind cannot form. Therefore, the conventional theory concludes that all supersonic solar/stellar winds from the corona are formed by flow going through the critical point. Weber & Davis (1967) added the effects of the magnetic field into the theory and predicted additional critical points and more critical solutions, but the supersonic winds are still formed by flows going along the eigenfunctions through the critical points. The nature of requirements of critical points remains the same.

Because the solar winds observed at the earth's orbit, although fluctuating greatly, are supersonic, the conventional theory has been considered as a milestone in space/astro physics. All current solar/stellar wind models have been built based on this theory and reasoning.

An important feature in the conventional theory is that the eigenfunction is continuous. Shergelashvili et al. (2020) and Westrich et al. (2024) recently considered allowing a discontinuity at the critical point.

1.2. Inner boundary conditions

In the early years of conventional theory development, people knew little about the inner boundary conditions at the corona. Models

* Corresponding author. Email: Paul_Song@uml.edu

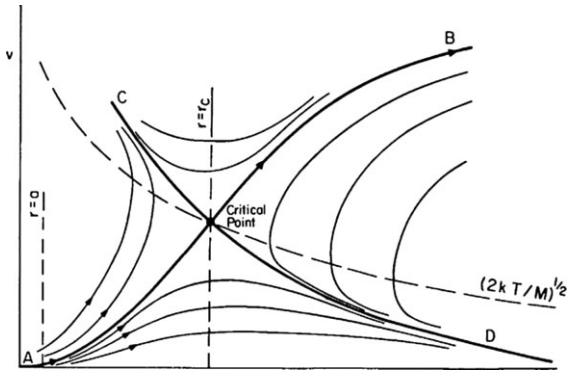

Figure 1. Conventional stellar wind theory (adapted from Parker, 1964b). The horizontal axis is radial distance from the center of the Sun/star. The sonic speed, dashed line, and illustrative flow velocities, solid lines. At the coronal base, the temperature is high, and the flow speed is very small, as shown in the lower-left corner. The flow speed becomes transonic at the critical point, r_c. According to the conventional solar/stellar wind theory, supersonic solar/stellar wind is exclusively formed along the critical flow line, the thick solid line from point A to point B, called the eigenfunction. The lines with arrowheads below the critical flow line are subsonic, and those above the critical flow line are deemed unphysical and cannot exist because the solutions are multivalued.

assumed the corona to be isothermal or polytropic. For an isothermal corona, the coronal base temperature needs to be million degrees. This leaves the question of how the million-degree temperature can be raised from a few thousand-degree chromosphere (e.g. Klimchuk 2006).

The potential mismatch between the inner boundary conditions and the requirement of critical solution had been noted and investigated (e.g. Brandt et al. 1969; Wolff et al. 1971) as 'incompatibility' between the critical point phenomena and possible independent physical choices for inner boundary conditions. These authors investigated such effects as by magnetic field, heat conductions, a small viscosity, and electron-ion two-fluid. They experimented and developed numerical methods to successfully find smooth solutions passing through the critical regions. Nevertheless, as summarized by Hundhausen (1972), the predicted solar wind speed and temperature are usually on the lower side to the observed ones, causing a solar wind 'energy problem' while each additional mechanism adds substantial complicity to the formalisms. The 'energy problem' remains outstanding and is often attributed to insufficient heating at the coronal base, but more heating would result in more severe incompatibility between the inner boundary conditions with the critical condition.

The eventual solar/stellar wind properties, such as the mass flux and energy flux, depend critically on two processes, not only one: the ionization in the transition region, which determines the available ionized mass flux for the solar/stellar wind, and the coronal heating around the coronal base which determines the total energy of the solar/stellar wind. These two processes cannot be solely determined by processes (or requirements) at the critical point which is several solar/stellar radii farther away.

Recently, a series of investigations (Song & Vasyliunas 2011; Tu & Song 2013; Song 2017) have shown semi-quantitatively that upward propagating Alfvénic perturbations from the photosphere can be efficiently damped by neutral-plasma collisions in the chromosphere to provide most of the required heat energy for the radiative losses which keep the chromosphere weakly ionized at lower temperatures. This mechanism in principle is able to address the ballpark part of the so-called coronal heating problem. Song et al. (2023) further investigated the ionization of the chromospheric gases in the transition region which produces the nearly fully ionized outflow into the corona forming the solar wind. They showed that to produce the fully ionized solar wind flux observed at the Earth's orbit, coronal heat sources other than dissipation of the heat flux are needed. Here, we recall that although the heat sources in the corona have been widely recognized and agreed upon (e.g. Parker, 1960, 1964a, b, 1965a; Hundhausen 1972; Withbroe & Noyes 1977; Aschwanden 2005; Zank et al. 2021), some models suggested that dissipation of downward conduction flux from the corona be sufficient to heat the medium and to launch the solar wind (e.g. Hundhausen, 1972) while it is two order of magnitude less than what is required to sustain the chromospheric radiative losses (Song & Vasyliunas, 2011; Song, 2017).

The analysis of the formation of the transition region (Song et al. 2023) has shown semiquantitative understanding of the processes at and below the transition region, removing much uncertainty about the inner boundary conditions of the corona. This also removes the possibility that inner boundary conditions are determined by the requirement at the critical point. Since the solar wind flux and, with a given density, the initial speed at the bottom of the corona is determined by the ionization processes in the transition region, the initial speed at the inner boundary of the corona may not always exactly equal the eigenspeed but has a large range of variations. This raises the question about the validity of conventional theory because most often the inner boundary flow speed is unlikely to match the critical condition as required by the conventional theory, and, however, the supersonic solar wind has been observed almost all the time. The inconsistency indicates that the conventional model of the formation of the supersonic solar/stellar wind needs to be fundamentally modified or revised.

1.3. Our approach

In this study, we investigate the whole notion about the critical point and significance of the critical solution in forming a supersonic wind. In this regard, if most supersonic winds are formed without going through the critical point, the critical point may not be important and the number of critical points even less important to understand the formation of the supersonic solar/stellar winds. Therefore, in this study, we concern only a classical problem that has been extensively studied for more than six decades; detailed mechanisms and effects are less important. In this initial study, we have to limit ourselves to a tractable theoretical framework because any additional effect may need one or more than one detailed study and validation.

The conventional theory we refer to in this paper is the fundamental theoretical framework originally proposed by Parker (e.g. Parker, 1958, 1965b; Hundhausen, 1972; Holzer, 1977; Holzer and Leer, 1980; Leer and Holzer, 1980). We focus on two major issues: sonic transonic processes (note that it is not only about the critical point transition) and the formation of supersonic flow. In the conventional theory, these two issues are resolved with a single mechanism/treatment – critical solution which offers one possible solution to the two issues. However, as we know from the above discussion that since the inner boundary flow conditions are determined by the processes below the inner boundary and seldom satisfy the critical conditions to form the supersonic solar wind according to the conventional theory, supersonic solar wind should not have been frequently observed at 1 au. Therefore, we set our goal for this study to provide a theoretical framework that can produce

supersonic wind with a relatively large range of inner boundary conditions.

There are many 'realistic models' developed over years in the community, such as, kinetic and various global simulation models, each claiming some advantages. It is obvious that if the foundation of the supersonic wind formation is modified, those 'realistic models', including the concepts and fundamental ideas discussed in these studies, will need to be modified accordingly. In this first step to fundamentally modify the conventional theory, we treat only the processes that can qualitatively change the nature of the wind solutions to provide a more solid foundation for further 'realistic' and application model developments. To avoid distractions from basic physical processes, we start with the simplest situation of a single-fluid, one-dimensional, steady-state hydrodynamic fluid with generic ad hoc heating sources and heat conduction and its dissipation. For example, although the magnetic field may make a single critical point to three critical points, it does not change the nature of the supersonic wind coming out along the eigenfunction assumed in the conventional theory. Therefore, with or without the magnetic field will not fundamentally change the conclusion of our analysis. The effects of the magnetic field are included qualitatively in the basic consideration but not formally treated or solved. We leave the inclusion of other effects in the formation of the solar/stellar coronal base and solar/stellar wind, such as geometry, asymmetry, multidimensionality, non-steadiness, electromagnetic field, viscosity, and specific heating mechanisms for future studies. However, because time-dependent effects have been widely considered as the solution to the problem in Parker's theory, we formally treat the issue theoretically in Appendices B and C and discuss extensively in Sections 3.3 and 5.3. Issues with numerical simulations in modeling the solar/stellar wind formation should be studied separately.

The theory we propose is more generally applicable to the conditions at a variety of stars, and hence, it is referred to as a stellar wind theory. However, to demonstrate validity and for quantitative control, we use observed average parameters from the solar chromosphere, coronal base, and solar wind to guide our discussion and semi-quantitative evaluations. Recent observations from the Parker Solar Probe and Solar Orbiter may provide more constraints on the details for further model development (e.g. Shi et al. 2022; Raouafi et al. 2023; Telloni et al. 2023).

In Section 2, we present the one-dimensional steady-state governing equations used in studying the space above the transition region with a focus on the solutions around the sonic point in a similar nature to Parker's original formalism. The details of the determination of the wind flux or outflowing velocity from the transition region due to the increased ionization are provided in Appendix A. In Section 3, we propose a model of the supersonic wind formation with a discontinuity at the transonic point. In Section 4, we show numerical solutions from the model for a range of initial speeds. In Section 5, we discuss some concepts in our theory and possible alternatives. Details for the wave analysis are provided in Appendix B and for formation of a standing structure in the flow are given in Appendix C.

2. THEORETICAL FRAMEWORK

2.1. Governing equations

For simplicity, we treat the stellar wind as an electron-ion single fluid and consider 1D gas flowing radially out from the star; at radial distance r from the center of the star, the number density is N, and the flow speed is V. The steady-state governing equations are as follows.

Continuity

$$\frac{\partial}{\partial r}\left(r^2 N V\right) = 0 \tag{1}$$

Since we assume no source or sink (no ionization nor recombination) of the fluid, equation (1) is applicable to regions above the transition region where the fluid is nearly fully ionized.

Momentum equation

$$\rho \mathrm{V}\frac{\partial V}{\partial r} = -2\frac{\partial p}{\partial r} - \rho g + (\mathbf{J}\times\mathbf{B})\cdot\hat{\boldsymbol{r}} \tag{2}$$

Here, $2p = (p_\mathrm{i} + p_\mathrm{e})$ is the thermal pressure of the plasma, in which factor 2 is a conventional single fluid treatment of electron–ion fluids assuming that the ion pressure equals the electron pressure, $\rho = mN$ is the mass density, m is the average mass of an electron–ion fluid particle, and $g = \frac{g_s r_s^2}{r^2} = -\frac{\partial \Psi}{\partial r}$ is the gravitational acceleration; here, $\Psi = -\frac{g_s r_s^2}{r}$ is the gravitational potential of the star. Subscript s denotes the value at the stellar surface, i.e. $r = r_\mathrm{s}$. Vectors $\boldsymbol{J}$ and $\boldsymbol{B}$ are the electric current density and magnetic field, respectively, and $\hat{\mathbf{r}}$ is the unit vector in the radial direction. The $\boldsymbol{J}\times\boldsymbol{B}$ force in the radial direction is of the order of $\frac{\partial}{\partial r}(\rho c_{At}^2)$ where $c_{At}^2 = \frac{B_t^2}{\mu_0\rho}$ is the square of the Alfvén velocity based on the transverse component of the magnetic field, μ_0 is the permeability of the vacuum, and B_t is the magnetic field component transverse to the flow direction. Since in the region we are interested, the magnetic field is dominated by the radial component, in the following analysis we neglect the effects of the (large-scale) DC component of electromagnetic fields on the momentum equation and will discuss qualitatively the effects below in the Poynting vector and the effects on the formation of wave fronts in Appendix B.

The steady-state energy equation (e.g. Rossi & Olbert 1970) of the corona, when including cooling (e.g. Vasyliūnas & Song 2005; Song 2017), is

$$\nabla\cdot\left[\mathbf{V}\left(\frac{1}{2}\rho V^2 + 5p + \rho\Psi\right) + \mathbf{q} + \mathbf{S}\right] = -C \tag{3}$$

We have assumed the ratio of specific heats to be 5/3 in the thermal energy density. Vector $\boldsymbol{S} = \mathbf{E}\times\mathbf{B}$ is the Poynting vector, $\mathbf{E}$ is the electric field, $\mathbf{q}$ is the conduction heat flux, and C is the radiative cooling rate which is not commonly included in MHD theory. We assume that in the corona and the stellar wind, radiation is a pure loss of energy when it is not reabsorbed. The radial component of the Poynting vector is $S_\mathrm{r} = \frac{V B_t^2}{\mu_0} = V\rho c_{At}^2$. Although we do not treat electromagnetic effects quantitatively in this study, we include it here for potential heating sources to be discussed below.

The Poynting vector $\mathbf{S}$ is the electromagnetic energy flux carried by the medium, where in ideal MHD, the non-dissipative part of the Poynting vector satisfies $\nabla\cdot\mathbf{S}_{non} = 0$, and in an ideal gas with no dissipation, no heating and no cooling, $\nabla\cdot[\boldsymbol{V}(\frac{1}{2}\rho V^2 + 5p + \rho\Psi)]_{non} = 0$, where subscript *non* denotes the ideal or non-dissipative part of the fluid although the dissipative processes may feed energy flux into it. Heating rate Q can be derived from dissipative MHD and HD processes, which convert the fluid and EM energies to heat (e.g. Song & Vasyliunas 2011)

$$\nabla\cdot\mathbf{S}_\mathrm{diss} + \nabla\cdot\left[\boldsymbol{V}\left(\frac{1}{2}\rho V^2 + 5p + \rho\Psi\right)\right]_\mathrm{diss} = -Q \tag{4}$$

Here, we emphasize that one cannot simply add the heating rate from a perceived mechanism to the right-hand-side of equation (3). All heating mechanisms have to be derived according to first

principles from equation (4) and semi-quantitatively constrained by observations (e.g. Song & Vasyliunas 2011).

According to Fourier's law, the heat conduction flux is $\boldsymbol{q} = -\kappa \nabla T$ where T is the temperature of the plasma and κ is the heat conductivity. Here, we recall that heat conduction is maintained by electron collisions in our parameter regime in the presence of a temperature gradient. If there are no collisions, the divergence of the heat flux $\nabla \cdot \mathbf{q}_{\rm non}$ is zero. The dissipation of the heat conduction in our problem of a dissipative collisional plasma flow is dominated by electron collisions that produce the divergence of the heat flux $\nabla \cdot \mathbf{q}_{\rm d}$, where $\mathbf{q}_{\rm d} = -\kappa_0 T^{5/2} \nabla T$, where $\kappa_0 = 9.2 \times 10^{-7}$ in CGS units (Spitzer 1956; Aschwanden 2005) and subscript d denotes dissipative. Constant κ_0 is smaller than that used in other studies (Parker 1964b; Hundhausen 1972; Holzer & Leer 1980). *Spitzer*'s heat conductivity is based on Fokker–Planck electron collisions for distribution functions with no large gradients in velocity space. Although wave-particle collisions may modify the heat conductivities and hence dissipation rates, the features and conclusions of this study will unlikely be changed fundamentally and we will use *Spitzer*'s conductivity to semi-quantitatively evaluate the model results. The divergence of the heat flux is

$$\nabla \cdot \boldsymbol{q} = \nabla \cdot \boldsymbol{q}_d \tag{5}$$

It is worth mentioning that the dissipation of the heat flux, equation (5), is not directly related to the heating mechanisms (Q in equation 4). Many conductive heating models assumed $Q = 0$.

In this study, we neglect the convective DC electromagnetic energy, i.e. the DC Poynting vector, and combine dissipative MHD and HD effects as a heating source with an ad hoc heating rate Q around the coronal base where the temperature reaches a maximum. Dissipation of the heat flux is present in all regions where the temperature gradient is non-zero and is enhanced where the gradient is very large. In 1-D, energy equation (3), combined with equation (1), i.e. $r^2 F = {r_{\rm s}}^2 F_{\rm s}$ where $F = NV$ is the number flux per unit cross-section area and in a single fluid treatment with $p = NkT$, where k is the Boltzmann constant, becomes

$$F_{\rm s} r_{\rm s}^2 \frac{\mathrm{d}}{r^2 \mathrm{d}r}\left[\frac{1}{2} m V^2 + 5kT + m\Psi\right]_{\rm non} + \frac{\mathrm{d}\left(r^2 q_{\rm d}\right)}{r^2 \mathrm{d}r} = Q - C. \tag{6}$$

Subscript non will be dropped from now on for simplicity. The dissipative effects other than dissipation of the heat flux are combined to become Q based on equation (4) which includes the dissipation of the Poynting vector while assuming the Poynting vector carried by the flow to be non-dissipative. The outflowing plasma flux $F_{\rm s}$ from the star is primarily determined by the ionization/recombination process in the transition region (see Appendix A).

The physical meaning of equation (6) is clear. The quantity in the brackets is the sum of the fluxes of 'non-dissipative' kinetic energy, enthalpy, and gravitational energy that an average particle carries. Q and C are the local heat source and sink, respectively, of the energy flux of the medium. The energy from sources or to sinks is redistributed into the flow by heat conduction with (in the presence of dissipation) or without (in the absence of dissipation) conductive heating. Note that there is a possibility of $\nabla T \neq 0$ while $\nabla \cdot \boldsymbol{q} = 0$, namely, there is a temperature gradient and hence non-zero heat flux, but the dissipation is negligibly small. This is a situation where the heat flux supports the long-range energy redistribution as discussed later in Section 3.2. We note that equation (6) is the same as the energy equation of Parker (1964b) when $Q - C = 0$. In other words, in Parker's model, there is no heating other than the dissipation of the conduction flux as we mentioned in the introduction.

By contrast, in our treatment, there is an ad hoc heating source term, Q in equation (4), similar to that of Leer and Holzer (1980), due to dissipative processes in addition to the dissipation of the heat flux. The heat energy produced by the dissipation can feed back into the flow energy as indicated by equation (6). Heat conduction transfers the heat energy to areas both upward and downward from the heating source. The dissipation of the heat flux converts the heat into other forms of flow energy as a result of collisions. Although we do not specify the mechanism of the additional heating source, many mechanisms from previous theoretical investigations, such as shock heating, cyclotron resonance, kinetic Alfvén waves, resistive heating, and reconnection (see Tu et al. 1984; Aschwanden 2005; Zank et al. 2021; Tu & Marsch 1995; Hu et al. 1999; Vainio et al. 2003; Shergelashvili & Fichtner 2012), have provided a large range of heating rates. For example, the generation and growth of perturbations may draw energy from the flow energy and/or electromagnetic energy (Zank et al. 2021). Nevertheless, we should point out that although most of these mechanisms may not be able to produce enough heating to support the chromospheric radiative losses, they are most likely able to provide sufficient energy for the formation of the coronal base and supersonic wind, which, other than neutral-ion collisions, may be only a few per cent of the total required 'coronal heating' as discussed before (e.g. Song & Vasyliūnas 2011). A greater coronal heating rate Q may unlikely raise the temperature of the coronal base. Instead, it is more likely to produce greater downward heat fluxes resulting in more outward flow fluxes as shown in Appendix A. In Section 4, we assume the ad hoc $Q - C$ term to have the form of a Maxwellian function.

We integrate equation (6) from the inner boundary of the present model, r_0, to r and obtain

$$F_{\rm s}\left[\frac{m}{2}V^2 + 5kT + m\Psi\right]_{r0}^{r} - \left[\kappa_0 T^{\frac{5}{2}} \frac{r^2}{r_{\rm s}^2}\frac{\mathrm{d}T}{\mathrm{d}r}\right]_{r_0}^{r} = \int_{r_0}^{r} \frac{r^2}{r_{\rm s}^2}(Q - C)\,\mathrm{d}r \tag{7}$$

Let

$$E_{\rm r} = \left[\frac{m}{2}V^2 + 5kT + m\Psi\right] \tag{8}$$

$$q_{\rm r} = -\kappa_0 T^{\frac{5}{2}} \frac{\mathrm{d}T}{\mathrm{d}r} \tag{9}$$

$$H_{\rm r} = \int_{r_0}^{r} \frac{r^2}{r_{\rm s}^2}(Q - C)\,\mathrm{d}r \tag{10}$$

Equation (7) can then be rewritten as

$$F_{\rm s} E_{\rm r} + \frac{r^2}{r_{\rm s}^2} q_{\rm r} - H_{\rm r} = F_{\rm s} E_0 + \frac{r_0^2}{r_{\rm s}^2} q_0 - H_0 = U_0 \tag{11}$$

where U_0 is the first integral of the energy equation and can be determined at the inner boundary, and subscript r denotes values at radius r. $H_0 = 0$ when $r = r_0$ from equation (10).

2.2. Stellar wind evolution equation

Beyond the coronal heating source region, the net heating rate $H_{\rm r}$ from equation (10) reaches a constant $H_{\rm m}$ so that $H = H_{\rm m} + U_0$ becomes constant, and the energy equation is

$$F_{\rm s} E_{\rm r} + \frac{r^2}{r_{\rm s}^2} q_{\rm r} = H \tag{12}$$

From the definition of q_r, equation (9), we then have

$$\frac{dT}{dr} = \frac{F_s E_r - H}{\kappa_0 T^{\frac{5}{2}}} \frac{r_s^2}{r^2} = \frac{1}{\kappa_0 T^{\frac{5}{2}}} \left\{ F_s \left[\frac{m}{2} V^2 + 5kT + m\Psi \right] - H \right\} \left(\frac{r_s}{r} \right)^2, \quad (13)$$

which specifies the temperature gradient and hence the temperature profile. Here we should point out that although in regions far away from the coronal heating sources, such as near the sonic transition or critical point, local heating may be negligible, the total heating H in general cannot be set to 0 as did in the conventional theory discussed above. In theory, H includes all heating effects, such as turbulence, reconnection, and most wave heating mechanisms, as shown in equations (4) and (10). In steady state, by combining equations (1) and (2) to remove the density, we derive the stellar wind evolution, or momentum, equation (Parker, 1958, 1964b; Chamberlain 1961)

$$\left(1 - \frac{c^2}{V^2}\right) V \frac{dV}{dr} = \left(\frac{2c^2}{r}\right) - \frac{dc^2}{dr} + \frac{\Psi}{r} \quad (14)$$

Here $c^2 = \frac{2kT}{m}$ is the characteristic thermal speed, and, in the following discussion, is referred to as the sonic speed. Since the sonic speed and the temperature can be directly converted, we use them interchangeably in our equations and discussions, as used in the conventional theory. Combining equations (4), (10), (13), and (14), reconnection, turbulence, and almost all heating mechanisms around and below the coronal base are included in equation (14). Equation (14) may be written in a simple form, for later discussions, as

$$L \frac{dV}{dr} = R \quad (15)$$

$$\text{where } L = \left(1 - \frac{c^2}{V^2}\right) V \quad (16)$$

$$\text{and } R = \left(\frac{2c^2}{r}\right) - \frac{dc^2}{dr} + \frac{\Psi}{r} \quad (17)$$

Substituting energy equation (13) into equation (17) yields

$$R = \left(\frac{4kT}{mr}\right) - g_s r_s \frac{r_s}{r^2} + \frac{2k}{m\kappa_0 T^{\frac{5}{2}}} \frac{r_s^2}{r^2} [H - F_s E_r] \quad (18)$$

In principle, given the energy E_0 from the inner boundary, flux $F_0 = F_s \frac{r_s^2}{r_0^2}$, and the heating source above the coronal base H, the stellar wind velocity and temperature as functions of radial distance can be derived uniquely. As will be shown in subsection 5.1, the physical meaning of R is the effective driving force. Equation (18) shows that R is not a strong function of flow speed when the flow is subsonic. In the lower corona, R is negative and increases monotonically with distance. The second term on the right may be dominant when r is not very large. This feature is important for supersonic wind formation. In particular, it is not necessarily equal to zero at the distance where the sonic point is approached, because in general R depends on the flux from the transition region F_s and the heating process H. However, at the sonic point, $V = c$ and $L = 0$, so equation (15) becomes

$$0 = R \quad (19)$$

One can easily derive equation (19) from the original form of the continuity and momentum equations when $V = c$. Therefore, equations (15) and (19) can be combined as

$$\begin{cases} L \frac{dV}{dr} - R = 0 \, (V \neq c) \\ R = 0 \, (V = c) \end{cases}. \quad (20)$$

Equation (20) is the momentum equation combined with mass conservation to eliminate the density. Although it appears to be a simple mathematical maneuver by combining equations (15) and (19), it reveals a new physical feature. Since R is negative when closer to the star and $L < 0$ when the flow is subsonic, if R increases to 0 before $L = 0$, the first expression of (20) can hold by letting $dV/dr = 0$. Further out, from $R > 0$ and $L < 0$, $dV/dr < 0$. This gives a subsonic solution when the flow speed never reaches the sonic speed.

If the sonic point $L = 0$ is reached when $R = 0$, both expressions of (20) hold. This gives a critical point, which is a removable singularity described in the conventional stellar wind theory at $V = c$, that is

$$\frac{dV}{dr} = R/L \quad (21)$$

At the critical distance r_c, $R = 0$ and $L = 0$. The condition $R = 0$ will be referred to as the critical condition while the condition $L = 0$ is the sonic condition. The situation for the critical condition and sonic condition occurring at the same point is possible when the singularity is removed, and V and T remain continuous functions.

A real problem occurs if the inner speed is greater than V_e, the sonic condition $L = 0$ is reached when R is still negative. The first term in the first expression of equation (20) is zero, requiring $R = 0$ for the equation to hold, but R, defined by equation (18), is not equal to zero because R depends on the flux from the transition region F_s and the heating process H. This is no longer a simple problem of singularity but is a problem of validity of the conventional stellar wind momentum equation (14). Specifically, if $R < 0$ at the sonic point, no stellar wind solution can be found from either equation (14) or (21). This creates a theoretical difficulty in the conventional stellar wind theory and is the central issue to be resolved in this study.

2.3. Approaching the sonic point

From equation (20), close to the star in the parameter range of interest, both L and R are negative and increase monotonically with distance. However, since L and R are determined differently by the energy and heat fluxes at the inner boundary of the corona and the heating processes around the coronal base, in general, the sonic condition and critical condition do not necessarily occur at the same point. As shown in Table 1, there are three possibilities when the flow reaches the characteristic distance r_1, this being the radial distance at which either L or R first becomes zero in the outflow, with the other parameter remaining negative. The differing cases relate to whether the sonic condition or the critical condition occurs first at the characteristic distance r_1.

If the critical condition is reached at r_1, $R_1 = 0$, before L reaches zero, then $L_1 < 0$ downstream of point r_1, where subscript 1 denotes the value at r_1. R then becomes positive while $L < 0$, and hence, from equation (20), $dV/dr < 0$. The flow thus decelerates and the sonic point $L = 0$ will never be reached. This, shown as the first scenario in Table 1, gives the subsonic wind solution, corresponding to curves in the area below A–r_c–D in Fig. 1, as agreed upon by both Parker and Chamberlain (e.g. Parker 1965b). The flow speed at the inner boundary V_0 is less than the eigenspeed V_e.

The scenario shown in the second row in Table 1 occurs when the sonic condition and critical condition occur at the same point, e.g. equation (21) that *requires* $R_1 = 0$ when $L_1 = 0$, and is consistent with

Table 1. Three possible scenarios at the characteristic distance r_1.

	Inner Boundary r_0	Upstream of r_1 $r = r_1^-$		Downstream of r_1 $r = r_1^+$		Stellar wind @ $r >> r_1$
Only the critical condition No sonic point Parkers and Chamberlains subsonic solution	$v_0 < V_e$	$L_1 < 0$ $V' < c'$	$R_1 = 0$	$L_1' < 0$ V' < c'	$R_1' > 0$ T' and V' Continuity	Subsonic: $V' < c'$ $dV'/dr < 0$ $dT'/dr < 0$
Critical condition at sonic point Parker's model	$V_0 = V_e$	$L_1 = 0^- \langle sy \rangle$ V' < c'	$R_1 = 0^-$	$L_1' = 0^+$ V' > c'	$R_1' = 0^+$ T' and V' Continuity	Supersonic: $V' > c'$ $dV'/dr > 0$ $dT'/dr < 0$
Sonic point before critical condition Present study	$V_0 > V_e$	$L_1 = 0^-$ V' < c'	$R_1 < 0$	$L_1' = 0^+$ V' > c' $\Delta T > 0$, $\Delta V > 0$	$R_1' = 0^+$ T' and V' Discontinuity	Supersonic: $V' > c'$ $dV'/dr > 0$ $dT'/dr < 0$

Parker's supersonic stellar wind solution. We have referred to this scenario as conventional stellar wind theory or the 'eigenfunction solution', where the common point is the 'critical point'. The corresponding initial velocity at the inner boundary is referred to as the eigenspeed, V_e. At the critical point, this is then L'Hospital's problem in which the singular point in equation (14) is removable, and the solution for the velocity passes through the critical point smoothly and satisfies both expressions of momentum equation (20). The wind becomes supersonic beyond the critical point. Since the supersonic solar wind has been continuously observed, this scenario has been widely accepted to explain *all* solar winds observed in interplanetary space and at the Earth's orbit. As a result, it is believed that *all* supersonic solar winds come *only* along the eigenfunction, line A–B in Fig. 1.

Chamberlain disputed this scenario and hence the possibility of supersonic solar wind because the required initial velocity at the inner boundary, i.e. the eigenspeed, V_e, allows too small a range of initial velocity to be significant, or as Cranmer & Winebarger (2019) have written 'It seemed unlikely that the system would naturally choose this one critical solution out of an essentially infinite number of others that do not become supersonic (see e.g. Chamberlain 1961)'. As shown by equation (18), for a given initial speed, a slight difference in plasma flux F_s or in net heating rate H would make the solution of the velocity miss the critical point, so that the solution is either subsonic or there is no numerical result possible beyond r_1. To address this concern, the stability of the critical point has later been investigated as a solution to this serious problem (e.g. Parker 1966b). Therefore, one may question whether the observation of a supersonic solar wind can be used as evidence to *exclude* pathways from the inner boundary to interplanetary space *other* than that along the eigenfunction, line A–B in Fig. 1.

Parker dismissed the possibility of solutions when $R_1 < 0$ and $L_1 = 0$, shown in the lower row of Table 1. This corresponds to the situation when the initial speed V_0 is greater than V_e, or curves in the area between A–r_c–C in Fig. 1. When the initial speed is greater than the eigenspeed, the flow reaches the sonic point before the critical condition when the effective driving force is still negative and dominated by the negative gravitation potential term in equation (17). Equation (15) is no longer valid because the left-hand-side is zero, but the right-hand-side is not zero. This situation and hence these solutions are unphysical, as Parker argued, because there are two possible values for the velocity at a given distance $r < r_1$. This has indeed been shown by Parker and many others for the case of isothermal situations (e.g. Aschwanden 2005; Shi et al. 2022). In the isothermal case, i.e. when equation (13) or the last term in equation (18) is zero, the conventional stellar wind theory relies on the presence of two, not one, intercepting separatrices, A–B and C–D in Fig. 1, to argue against supersonic solutions in the area between A–r_c–C. However, in the isothermal case, the sonic speed line is horizontal, and not tilted as in Fig. 1 and the total energy is not conserved without continuous heating sources to support the continuous increase in speed while overcoming the gravitational force. More generally, with a decreasing temperature profile in equation (13) or an increasing last term in equation (18), the sonic line, the dashed curve in Fig. 1, may not intercept with the two separatrices at the same point as shown in Fig. 1. Therefore, the existence of the separatrix segment C–r_c is questionable. Without the segment C–r_c, it is possible to directly connect the two segments above the eigenfunction line A–B.

We should point out that when $V_0 > c_0$, corresponding to the region between the dashed line and line C–r_c in Fig. 1, the solution starts from the coronal base above the dashed line and stops at the sonic point. Therefore, the supersonic portion of the lines in the region between A–r_c–C does not imply that the flow turns back to the star leading to a multivalue problem, as argued by the conventional theory. In fact, that portion is derived from a different inner boundary condition at the coronal base flowing outward, decelerating, and eventually ending at the sonic point but without further solution beyond the sonic point. Therefore, the 'multivalue' argument discussed above is questionable.

Nevertheless, because the profiles of L and R are determined by the properties of the outflowing plasma flux from the chromosphere, the heating process in the corona, and heat conduction, the situation when $L_1 = 0$ and $R_1 < 0$ cannot be dismissed simply because no continuous solution can be found in the conventional analyses. This describes our motivation to express the momentum equation in equation (20). It is possible to have a discontinuous solution between the subsonic and supersonic regimes. In the next, we investigate a scenario in which an outgoing supersonic wind can form over a broader range of initial speeds.

3. FORMATION OF A DISCONTINUITY AT THE SONIC POINT

3.1. Discontinuity jump relations

At the sonic transition $L = 0$ when $R < 0$, equation (20) jumps from the first expression to the second expression, indicating a possible discontinuous solution. This is the scenario Parker dismissed. In our fluid treatment, the internal structure of discontinuity cannot be resolved or described in detail. Across the discontinuity, mass flux, momentum, and energy flux have to be conserved. Since equation (20) is derived from mass conservation (1) and momentum

conservation (2), we examine its jump condition. For inner boundary conditions that result in a velocity profile above the eigenfunction, the effective force R is negative when the sonic point is approached. However, at the sonic point, from the second expression of equation (20), $R = 0$, there is a jump in the effective force from $R_1 < 0$ to $R'_1 = 0$. Here, we follow the tradition by adding a prime to denote the downstream value. The change in the momentum equals the change in the effective driving force, $R'_1 - R_1$. From equation (17), we have

$$R'_1 - R_1 = \left(\frac{2c'^2_1}{r_1}\right) - \left(\frac{2c_1^2}{r_1}\right) - \frac{\mathrm{d}c'^2_1}{\mathrm{d}r} + \frac{\mathrm{d}c_1^2}{\mathrm{d}r} \tag{22}$$

Note that the gravity force is continuous across the discontinuity. Not only does the temperature change, but also its gradient. The two sides of the discontinuity each cannot be treated as the same in terms of the temperature gradient. The change in the temperature gradient is associated with a change in the heating flux from equation (9). The heat flux change can be determined from energy conservation (12),

$$\frac{r_s^2}{r_1^2} F_s \left[\frac{m}{2} V'^2_1 + 5kT'_1\right] + q'_1 = \frac{r_s^2}{r_1^2} F_s \left[\frac{m}{2} V_1^2 + 5kT_1\right] + q_1 \tag{23}$$

We should point out that the discontinuity proposed in the present model is not a shock. By definition, a shock, such as the bow shock and interplanetary shocks often discussed in space physics and astrophysics, refers to a type of discontinuity where the flow changes from supersonic to subsonic state.

Since $R'_1 = 0$ and $R_1 < 0$, and since the right-hand-side of equation (22) is dominated by the temperature terms, across the transonic discontinuity, the temperature needs to increase, $\Delta T > 0$. For a subsonic upstream flow to become a supersonic wind, the velocity jump also needs to be positive, $\Delta V > 0$, which leads to a negative density jump, $\Delta N = N'_1 - N_1 < 0$. Here, symbol Δ denotes the difference of the downstream value from the upstream one. Similarly, the velocity jump is $\Delta V = V'_1 - V_1$, temperature jump is $\Delta T = T'_1 - T_1$, and the heat flux difference is $\Delta q_r = q'_1 - q_1$.

The heat flux terms in energy flux conservation equation (23), which affects the temperature gradient term in the effective force, make our jump conditions very different from the conventional Rankine–Hugoniot relations which often assume adiabatic or polytropic processes across a discontinuity (e.g. Habbal and Tsinganos, 1983; Velli, 1994). Without dissipation, the heat flux terms cancel out. As shown in equation (22), with positive jumps in the temperature, the discontinuity requires a reduction of the heat flux, $\Delta q_\mathrm{r} < 0$, to heat up the medium.

The intensity of the discontinuity, the ratio of the downstream to upstream values, depends on how much heat flux is dissipated within the discontinuity which is ultimately determined by the negative value of R upstream of the discontinuity. The jump conditions and heat flux dissipation within the discontinuity are self-consistently determined by the conservation laws and the requirement of the downstream supersonic flow condition. Since in our discontinuity treatment, there is no spatial scale for the thickness of the sonic transition layer, the rates of heat dissipation and flow energization are inversely proportional to the actual thickness of the discontinuity layer. The detailed structure and processes therein can be resolved by solving the original form, not the integrated form, of the energy equation (6). To gain a better physical understanding of the processes, which may determine the thickness of the discontinuity, within the transonic region, more theoretical and numerical investigations are needed.

3.2. Dissipation of conduction flux

Although we have assumed that there are no external heating Q and cooling C processes, a fraction of a stellar radius beyond the coronal base in the region we are considering, from equation (6), the conduction flux carries heat energy from the coronal heating source region to regions of lower temperature over a long range. However, heat conduction flux has been a confusing concept in the context of coronal heating and solar wind formation (Hollweg 1976; Holzer and Leer 1980). We adapt *Spitzer*'s formula given in Section 2.2. From equation (5), one can show that if T drops with distance at $r^{-2/7}$, the dissipation is zero while heat flux is $q_\mathrm{r} \sim r^{-2}$ and the divergence of the heat flux vanishes. This temperature drop-off rate may be the asymptotic temperature profile at large distances in the heliosphere before encountering the termination shock. We notice that the heliospheric temperature drop-off rate (Maruca et al. 2023), with a slope of power index -0.273, from 4.4 au to the termination shock appears to be related to this drop-off rate of a power index -0.286. In regions where the temperature drops faster than this rate, the heat flux dissipates and constantly feeds energy into the flow. Before encountering the sonic transition, as the temperature decreases at a fast rate, the acceleration is accomplished with energy from the dissipation of the heat flux and the reduction of enthalpy. The insight we gain from this discussion is that the asymptotic temperature profile cannot be a constant temperature, as we discussed earlier, because an isothermal profile requires energy input to maintain.

Upstream of the sonic transition, under the conditions for supersonic winds to form, the energy of the flow, which is the term within the brackets in equation (23), is still negative because of the large negative gravitational energy. The outward energy flux from the coronal base is mostly carried by heat flux. We recall that the function of the discontinuity in the sonic transition is to increase R from a negative value to 0 as shown in the last scenario of Table 1. From equation (17), there are two apparently opposing possibilities to increase R, either to raise the temperature T or to decrease the temperature more rapidly which increases the negative temperature gradient $|\mathrm{d}T/\mathrm{d}r|$. Although a larger negative temperature gradient may indeed lead to supersonic conditions, the resulting temperature may further reduce the effective force R making it remain negative. As a result, the flow speed will decrease, the supersonic state cannot be sustained, and the possibility is unphysical.

The remaining possibility would then be a jump in the temperature at the discontinuity followed by a flatter decrease. The combination of the two may decrease the heat flux if the temperature jump is not too large. Across the sonic transition, from equation (14), $L = 0^-$, the velocity increases rapidly. From energy conservation equation (12) or (23), the change in the heat flux is

$$-\Delta q_\mathrm{r} = \left(\frac{m}{2}\Delta V^2 + 5k\Delta T\right) F_\mathrm{s} \left(\frac{r_\mathrm{s}}{r}\right)^2 \tag{24}$$

With the increases in temperature and velocity, the heat flux decreases across the discontinuity. Physically, the higher downstream temperature is energized self-consistently by heat flux dissipation. Therefore, dissipation of the heat flux is very strong within the discontinuity and the heat flux decreases across it. This explains the cause and the location of the discontinuity.

3.3. Formation of a stable standing discontinuity in the flow

There are three requirements to produce the proposed discontinuity. First, it needs to be able to form; secondly, it needs to stay in the flow

and not be carried away by the flow; and thirdly, it needs to be stable and not to be dispersed by fluctuations (Kennel, 1994).

The formation of a discontinuity first needs a wave steepening process. A well-known wave steepening process is the nonlinear steepening formation of wave fronts, such as in the shallow water wave, but the conventional non-linear wave steepening process is not applicable in our case. In the corona-wind system, neglecting the electromagnetic fields, the perturbations propagate as sonic waves, see details of wave analysis in Appendix B, propagating both forward (along the flow direction) and backward (against the flow direction) and each propagates at sonic speed in frame of reference of the flow. Note that the two oppositely propagating waves are generated at each point in space in order to satisfy the required perturbation relations. The resulting net perturbation δV, which is the summation of the forward and backward propagations, modifies the flow toward the profile of the large-scale solution. However, if a perturbation propagates with a non-zero speed in the frame of reference rest to the star, it will propagate and/or be carried away. These propagations will not be able to form a quasi-steady state structure. An observable structure, such as the bow shock, has to be formed by a standing wave, a perturbation that has zero propagation velocity in the reference frame of the star. As shown in Appendix C, all forward propagations are travelling and cannot form a standing structure. The backward propagation in the subsonic (supersonic) outflow will propagate inward (outward) back to (away from) the star. A standing structure in the flow can be formed at the sonic point by the backward propagation. Note that, to form a standing structure, the perturbations of the waves have to have the correct phases. In a quasi-steady state, the backward propagating waves in principle are not the reflected waves from the incident waves which may have random phases. Therefore, we follow only the backward propagation next. When the effects of the electromagnetic field are included, there are three wave modes and six possible propagations in the flow: the slow magnetoacoustic, Alfvenic and fast magnetosonic modes, each of which can propagate either forward or backward (e.g. Weber and Davis, 1967). The sonic transition treated in this study corresponds to the slow mode transition.

The backward propagation follows equation (C6), similar to equation (14) by replacing $\mathrm{d}V/\mathrm{d}r$ with $\mathrm{d}\delta V/\mathrm{d}r$ with an additional factor of 1/2. When $\mathrm{d}\delta V/\mathrm{d}r$ is positive (negative), the flow accelerates (decelerates) with distance. For $R < 0$, in the subsonic regime, the net perturbation speed gradually increases while the sonic speed decreases as the temperature decreases. At the sonic point, the inward propagation ceases to propagate because the backward propagation is carried back by the outward flow and the perturbation can steepen to form a standing wave.

As discussed in Section 3.1, both the velocity and temperature have to increase across the standing structure. The outward perturbation is able to continuously propagate outward beyond the sonic point but cannot form a standing structure. However, as discussed in Section 3.2, enhanced dissipation of the heat flux takes place in the sonic transition leading to a higher downstream flow speed and higher temperature which is proportional to the square of the propagation speed. Further out from the discontinuity, the flow speed continues increasing while the temperature and hence the sonic speed decrease. A standing discontinuity is then formed to separate the parameter regimes of subsonic and supersonic winds.

A stable discontinuity structure requires the process that forms it to be irreversible; otherwise, it will dissolve or disperse easily even after it is formed. Since a process with increasing entropy is irreversible, often an increasing entropy, instead of irreversibility, is used as a proxy for a stable discontinuity/shock jump condition. We then analyse the change in the entropy. In an ideal gas, with added heat from the dissipation of the heat flux $-\Delta q_\mathrm{r}$, the entropy change is

$$\Delta S = -\Delta q_\mathrm{r}/T \tag{25}$$

Combining equations (24) and (25), at the sonic transition, yields

$$\Delta S = \left[\left(\frac{\Delta V}{c_1}\right)^2 + 5\frac{\Delta T}{T_1}\right] k F_\mathrm{s} \left(\frac{r_\mathrm{s}}{r_1}\right)^2 \tag{26}$$

With the accelerated flow $\Delta V > 0$ and the increased temperature $\Delta T > 0$, the entropy increases $\Delta S > 0$ across the discontinuity. Therefore, the processes that form the discontinuity are irreversible, and the structure of the discontinuity is stable. The stable discontinuity appears to be a rarefaction, which invalidates the overgeneralized conception that rarefaction jumps are inadmissible in the Rankine–Hugoniot conditions. This is not because the Rankine–Hugoniot condition failed but because the generalization of it from a specific situation to a more complicated situation is not valid. In our more complicated situation, there is conduction heat flux that is discontinuous across the discontinuity. There is substantial heat released within the discontinuity due to the strong dissipation of the heat flux.

4. SUPERSONIC WIND SOLUTIONS

The supersonic stellar wind temperature and velocity profiles start from the inner boundary conditions at the top of the transition region, assumed to be at $r_0 = 1\ r_\mathrm{s}$, see Appendix A and Song et al. (2023). We assume that the ad hoc heating function is a normal distribution centered at 2 r_s, with a half-width of 0.15 r_s. Above 2.5 r_s, the heating source, $Q \approx C \approx 0$, and the (integrated) total heating reaches its maximum value $H_\mathrm{r} = H_\mathrm{m}$. The solutions are derived directly from equations (13) and (20). Fig. 2 shows solutions for five initial velocities, V_0, at the inner boundary with other parameters, such as the temperature T_0 at r_0 (Song et al. 2023), heating rate H_m, density N_s, and eigenspeed V_e, held fixed.

We have tested several standard algorithms to solve these two relatively simple coupled ordinary differential equations, and the results are consistent, repeatable, and unambiguous. The location and the energy of the flow are affected by the inner boundary conditions, V_0, V_e, and F_s, and by the location, width, and magnitude of the heating source Q. We have also tested different spatial step lengths. The step length shown in Fig. 2 is 60 km. For the cases with initial speeds greater than the eigenspeed, $V_0 > V_\mathrm{e}$, the results all unambiguously show that the calculation stops at the sonic point. No solution can be found further in contrast to those solid line segments above the dashed line in Fig. 1. As discussed in subsection 2.3, the supersonic segment of the solutions in region between A–r_e–C in Fig. 1 can be obtained only when the inner boundary speed is supersonic, $V_0 > c_0$, not shown because they are not derived from the subsonic inner boundary speed and, hence, is unphysical.

When the step length is large, for example, a fraction of a stellar radius, corresponding to the commonly used step length in global corona-wind models, and when R is not negatively very large, the calculation is able to jump over the singularity at the sonic speed, the dashed line in Fig. 1, producing a small bump around the critical point. However, we do not show these results because we think the downstream solutions result from large numerical dissipations.

The solid (dashed) black line in Fig. 2 shows the flow speed (sonic speed) for the critical/eigenfunction solution, $V_0 = V_\mathrm{e}$. It is worth mentioning that to derive the critical solution, V_e has to be

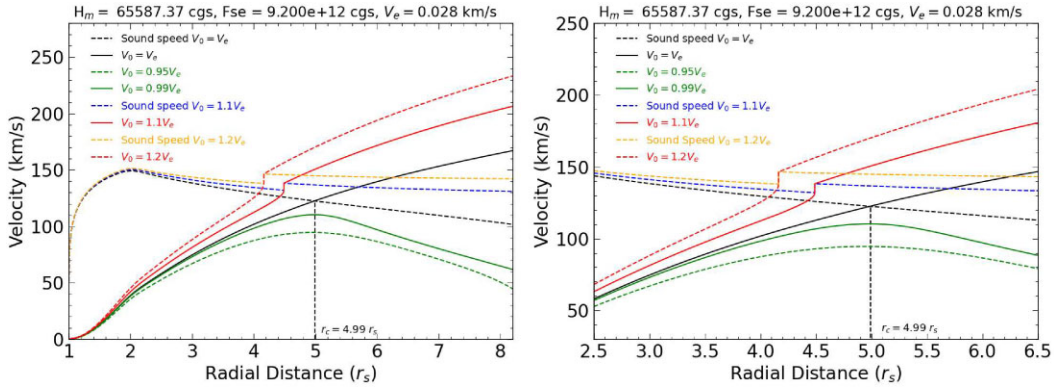

Figure 2. Illustration of stellar wind solutions from this study in the same format as Fig. 1 with the same density $N_0 = N_e$ and F_{se} at the top of the transition region $r_0 = 1r_s$ and total heating rate H_m, given above the frame. (a) from r_0 to 8.2 r_s, and (b) zoom-in around the sonic transition. The black, dark blue, and purple dashed lines show the sonic speeds. The black solid line shows the conventional critical stellar wind solution. Red lines (green lines) show two supersonic (subsonic) stellar wind solutions for coronal base speeds greater (smaller) than the eigenspeed V_e, given above the frame.

extremely accurate – down to the 11th digit, which may be correlated to the spatial resolution, 10^{-4} r_s, used in the calculation. If starting from the critical point and tracing back to the coronal base as many previous studies did, one would not encounter this issue. Otherwise, the solution is either subsonic or stops at the sonic speed. Therefore, the results are consistent with Parker's theory. Strictly speaking, the only difference is the missing supersonic segment between the sonic line and line C–r_c in Fig. 1. However, it also implies that Parker's critical solution may not be able to provide a meaningful amount of supersonic wind without substantial physical or artificial dissipation. If the spatial resolution is low, on the other hand, say, 0.1 r_s, the required accuracy of the critical point condition would not be as acute, and the solutions could step over the critical point with some fluctuations or a small bump around the critical point.

We are interested mostly in the supersonic winds. Winds with greater initial speeds, $V_0 > V_e$, shown by solid and dashed red lines, reach sonic speeds, blue and yellow dashed lines, respectively, starward of the critical distance r_c. The solution stops at r_1 where the sonic point, $L = 0$, is encountered while $R < 0$, as shown in the third scenario in Table 1 where the first expression of equation (20) encounters the singularity. However, the second expression of equation (20) is valid at the singularity. A discontinuity is needed to connect with the solution downstream of the sonic point. The quantities from the upstream, including R_1, are used to determine the downstream values using equations (22) and (23) assuming $R'_1 = 0^+$ and $L'_1 = 0^+$. Equations (13) and (15) are then used to calculate the temperature and velocity profiles further out.

Parker's critical solution and the subsonic solutions correspond only to situations for $R_1 \geq 0$ at the critical distance. The solutions are shown by the green and black lines in Fig. 2. For the subsonic solutions (green lines), the solutions diverge rapidly before encountering the critical point when the initial speed is merely slightly smaller (1 per cent) than the eigenspeed. In the subsonic solutions, the total heat flux, which is the heat flux over the whole 4π sphere, dominates the energy flux and eventually reaches an asymptote so that the kinetic energy and the enthalpy decrease to provide energy to counter the gravitational potential energy.

The jump condition relations in equations (22) and (23) only relate the upstream to downstream flow conditions of a discontinuity, but the discontinuity does not necessarily result in a supersonic downstream flow. The supersonic solution is guaranteed by the heating flux dissipation. From equation (13) $dT/dr < 0$, R' will increase to $R' > 0$. With the decrease of sonic speed, L increases to become positive, $L' > 0$, and then from equation (15), $dV/dr > 0$, forming the supersonic wind.

In Fig. 2, the critical condition $R = 0$ with different inner boundary speeds can be seen by connecting the second kink of the red lines with the critical point r_c. The most important feature shown in Fig. 2 is that most of the supersonic winds, e.g. solid and dashed red lines, are formed by solutions without going through the critical point where solid and dashed black lines cross. Therefore, the critical solution, the black line, marks the boundary between two parameter regimes but does not play an essential role in forming the supersonic sonic winds that were assumed by the conventional theory.

5. DISCUSSION

5.1. Physical description of the formation of a supersonic stellar wind

The physical processes learned from the analysis presented in this study are as follows.

In the simple example of a vertical flux tube, an isotropic fluid is driven by two opposite forces: the upward pressure gradient force and downward gravitational force. The pressure gradient force has two contributors, the temperature gradient, and the density gradient. If the flow is compressible, the density gradient results in an opposite flow velocity gradient and a geometric effect associated with the expansion in the cross-section area of the flux tube. The momentum equation, combined with the continuity equation, can be written as

$$mN\frac{V\,\mathrm{d}V}{\mathrm{d}r} = mN\frac{c^2}{V}\frac{\mathrm{d}V}{\mathrm{d}r} + mNR \tag{27}$$

where R, defined in equations (17) or (18), can now be understood as the effective driving force including the temperature gradient force, gravity, and the expansion effect. The first term on the right of equation (27) is due to the effect of the density gradient. Since the formation of the supersonic winds is essentially a process which increases the kinetic energy, it is clearer to express equation (27) in terms of the change in the kinetic energy flux on the left and the work done on the right, by multiplying equation (27) by V

$$F\frac{\mathrm{d}\left(\frac{mV^2}{2}\right)}{\mathrm{d}r} = F\frac{c^2}{V^2}\frac{\mathrm{d}\left(\frac{mV^2}{2}\right)}{\mathrm{d}r} + mFR \tag{28}$$

where $F = NV$ is the number flux. The first term on the right is due to the density decrease which accelerates the flow. Where $R < 0$, the effective driving force R does negative work to slow down the expansion of the flow. When $R > 0$, on the other hand, the effective driving force does work to accelerate the flow in addition to the expansional effect of the density decrease. Therefore, the real flow acceleration starts when $R = 0$, which is the critical condition.

Physically, if the outgoing stellar wind starts with a greater initial speed and reaches sonic speed before reaching the critical condition $R = 0$, the wind cannot flow faster than the sonic speed. A discontinuity can accomplish the task by making the downstream region satisfy the critical condition for the supersonic stellar wind to form. Therefore, the eigenfunction in Parker's stellar wind theory is actually the boundary separating the stellar winds that can reach the supersonic condition from those that cannot at the transition region. A higher outgoing speed results in a higher wind speed. For speeds less than that of the eigenfunction, the fluid flows more slowly and forms a subsonic wind.

By analogy, in the conventional supersonic stellar wind theory, the critical point functions more like a security guard who allows only the stellar wind with correct IDs, which is the initial speed V_0 at the inner boundary that matches the eigenspeed V_e, to pass. Numerical tests have shown that the correct ID match has to be exact up to 11 digits which can occur only in an extremely small range, $\sim 10^{-9}$ per cent, depending on the spatial resolution and numerical diffusive/viscous effects. If not matched, the output is either subsonic solutions (pass to slow lanes) or no solution (no pass). This is in support of Chamberlain's dispute (1961).

5.2. Critical point and eigenfunction solution

Conventional supersonic stellar wind theory has been based on the assumption that the singularity in the evolution equation (14) is removable by restricting the inner boundary condition of the system. However, this restriction is so strong that no significant flux of supersonic wind can be produced. Furthermore, a more serious issue is that the inner boundary condition is derived 'in reverse' by requiring knowledge of downstream conditions. That is, the required velocity at the inner boundary is mapped back from the critical point several stellar radii away along the eigenfunction, i.e. line A–B in Fig. 1. If the initial velocity cannot satisfy the critical condition at the sonic point, the wind either cannot form physically or has to be subsonic as discussed in Section 2.3 and shown in Fig. 1. The most problematic conclusion is in the former case – if in reality the initial speed is greater than the eigenspeed, there would be no stellar wind (solution), neither supersonic nor subsonic! One may feel that the argument seems problematic or 'abductive'. Actually, as discussed in Section 5.5 and Appendix D, it is known as abductive reasoning.

By contrast, in our model, the outflowing flux of corona/wind is determined in the transition region (assumed to be the stellar surface $r_0 = r_s$), $F_s = N_0 V_0$. The wind then flows out without much pre-knowledge about the existence of the critical point. T_0 is determined by the first impact ionization potential (Song et al. 2023), but V_0 cannot be determined without knowing the outflowing flux and density at r_0. Therefore, there is a potential mismatch between V_0 and the eigenspeed V_e, a situation when the conventional theory eliminates possible wind solutions when $V_0 > V_e$ using the 'unphysical' argument. Under this condition, because $L = 0$ and $R < 0$, the conventional momentum equation (15) is not able to describe the process. We have rewritten the momentum equation as equation (20) by defining the condition at the singularity with the sonic condition $R = 0$ while allowing a discontinuity to connect the two regimes of solutions. Parker's critical point condition, $R = 0$ when $L = 0$, allows only a continuous solution which is only one of the possible solutions to the singularity issue which is valid in a very small parameter range, i.e. it is a singular solution. In the present model, a large range of possible velocities at the inner boundary may be able to pass through the sonic condition and become a supersonic wind as shown in Fig. 2.

Stellar atmospheric processes and conditions should determine the properties and locations of the sonic point and critical condition and not be determined by the downstream requirement. What is not apparent in equation (14) in conventional theory is the absence of the mass flux or density. Although density N_0 can be given at the inner boundary according to mass conservation, the density does not affect the eigenfunction solution in conventional theory. Parker included the density later in his series of investigations. When including the heat flux, he concluded that there is a critical or eigen density N_B which may determine whether the wind is supersonic or subsonic (Parker 1965a). That is, when $N_0 < N_B$ ($N_0 > N_B$), the wind will be supersonic (subsonic). His conclusion is qualitatively consistent with our critical condition scenario in Table 1 since for a given initial flux F_s, a lower density results in a higher speed.

The present model is based on the thermal conduction formalism and is characteristically different from many stellar wind models that make either an isothermal or polytropic assumption to simplify the energy equation (e.g. Parker 1958; Velli et al. 1994, 2001; Shergelashvili et al. 2020) although the conductive energy equation has been used in the formation of the stellar wind theory since Parker (1964b). Here, we should point out that the energy conservation equation (11) holds at any reference point above the transition region when the fluid is fully ionized. For example, the reference point can be set above the heating sources. In this case, although there is no active heating, the total heating $H_0 = H_m$ in equation (11) can be non-zero. In fact, we believe that implicitly dropping H_m, which represents the heating around the coronal base, in the energy conservation may have led to the energy problem, according to which conventional solar wind models predict lower energy at 1au than observed (Holzer 1977; Hundhausen 1972).

There is an argument held among the global simulation community that if a magnetic field exists in stellar wind, the transonic problem discussed here does not exist. Unless the sonic transition occurs in the region of active reconnection, the magnetic effect will in principle modify the present theory by replacing sonic speed with the magnetic sonic speed with transverse magnetic field, ${c_{ms}}^2 = c^2 + {c_{At}}^2$, and hence the sonic point becomes the magnetosonic point. If the magnetosonic speed is much greater than the sonic speed, the sonic transition may be referred to as the Alfvén transition (Wexler et al. 2021). Therefore, an important conclusion drawn here is that the presence of the magnetic field in the stellar wind cannot remove the singularity at the sonic transition. Similarly, the numerical simulationalists often argue that 3-dimensionality may help remove the critical point. They should carefully investigate the possibility and prove it is indeed the case.

There is also an argument that trivializes the critical point issue studied in this work. It argues that Parker's theory was meant to be a 'toy' and should not be taken seriously. When there are many effects, such as turbulence, reconnection, time-dependence, 3D effects, electromagnetic fields, one argues, the critical point would never occur in reality. This argument is not based on sound physical reasoning and proof but is an unscientific speculation. Parker's theory was based on a careful order analysis to neglect the secondary effects. For example, when the background magnetic field is mostly radial, the magnetic force is negligibly small along the radial direction; near

the sonic transition, to the lowest order, flow is radially outward; and the time dependence, to be formally discussed in the next subsection, only decelerates the flow with time around the sonic transition. There has been no proposed physical process that can dominate gravity and pressure expansion at the sonic transition, as Parker originally identified, other than substantial numerical diffusion. A secondary effect in general cannot determine the overall characteristics of the solution other than secondary modifications. Therefore, this argument needs to first identify the effects, in mathematical terms, that are of the same order of magnitude as those in equation (14) and then falsify Parker's wind evolution equation, but neither of which has been done.

Nevertheless, let us analyse the consequences of the effects given in the above argument by assuming an effect that is indeed too large to be ignored. There are three types of possible consequences. First, this (additional) effect will affect the perturbation propagation. Therefore, this effect will end up on the left-hand-side of equation (14) to modify the dispersion relation and hence the sonic speed to a different or to multiple characteristic speeds. For example, the inclusion of the background magnetic field may result in three wave modes (e.g. Weber and David, 1967). At a transition point of these characteristic speeds, the left-hand-side of the equation (14) goes to zero as we have analysed in this study. Secondly, the effect produces additional acceleration or deceleration, such as the non-1D geometry of the background magnetic field as discussed by Holzer and Leer (1980) and Leer and Holzer (1980). This effect will then become a term on the right-hand-side of equation (14) and hence can be included in R. Thirdly, the effect can generate additional heating or cooling, such as turbulence, reconnection, and all wave heating mechanisms. In the regions far away from the heating processes near the sonic transition, the effect is already then included in the total heating rate H and then in $\mathrm{d}T/\mathrm{d}r$ by equation (12) to become a contributor to R. Combining these possible consequences, it is clear that at a trans-characteristic speed point, when the left-hand-side of equation (14) is zero, but R is not necessarily equal to zero. This is the critical point problem studied in this work at the characteristic speed transition point when $L = 0$ but $R \neq 0$, i.e. equation (15) cannot hold at the trans-characteristic speed point. If the effect is significant at the characteristic-transition point, it would broaden the discontinuity, of which a detailed treatment is beyond the scope of this study.

5.3. Time-dependent effects and spatial oscillation solutions

It has been argued that in a non-steady state process, there is no critical point (e.g. Parker, 1965b; Suess, 1982; Velli, 1994, 2001; Holzer and Leer, 1997; Keto, 2020) because the critical point may not be stable. Although the time-dependent theory of the supersonic wind formation deserves several full-length-papers to study, the time-dependent effect alone may not be able to remove the single point and produce supersonic winds when $V_0 > V_e$.

In this subsection, we focus on the situation for an initial speed greater than the eigenspeed when Parker's theory does not provide a solution, i.e. $R < 0$ at the sonic transition. This is a different problem from the instability of the critical point theory because in our problem the critical point condition $R = 0$ is not involved. As seen in Fig. 2, the red dashed line does not reach the critical point when the solution stops. A more accurate way to state our problem is about the stability of the transonic point, not about the critical point. This is where the argument about the stability of the critical point is confused in the first place.

The time dependent continuity equation and momentum equation are provided in equations (B1) and (B2) in Appendix B. Combining the two we have

$$\left(\frac{\partial V}{\partial t} - \frac{c^2}{V}\frac{\partial N}{N\partial t}\right) + L\frac{\partial V}{\partial r} = R \tag{29}$$

The temporal perturbations in either density and/or velocity in the ideal gas will be mitigated by sonic waves. For coherent sonic perturbations, the velocity and density are related by the perturbation relation equation (B4) in Appendix B. Substituting equation (B6) into (29) yields

$$\left(1 \mp \frac{c^2}{V^2}\right)\frac{\partial V}{\partial t} + L\frac{\partial V}{\partial r} = R \tag{30}$$

where the upper (lower) sign is for forward (backward) propagating perturbation. At the sonic point $L = 0$ and the value in the parentheses is either 0 or 2 which is non-negative. For $R < 0$, we have either $\partial V/\partial t = R/2 < 0$ or $\partial V/\partial t = -\infty$. Namely, at the transonic speed the flow speed cannot increase with time and the supersonic condition cannot be reached or maintained. Therefore, we conclude that the transonic point (not critical point) is stable to temporal fluctuations. We note that when $R = 0$, $\partial V/\partial t$ is undetermined and hence the critical point may or may not be stable.

Here, we recall that the conventional instability analysis studies whether an equilibrium is stable or not and an instability analysis is meaningless if no equilibrium can be found. Therefore, the analysis must start with an equilibrium. However, in the situation of interest, for $V_0 > V_e$, equation (15) does not provide such an equilibrium because $R < 0$. Therefore, the method of the instability analysis on the critical point reported before is not valid and cannot contribute to addressing the problem that little supersonic stellar wind can be produced in the conventional theory even if the critical point is indeed unstable. There is widespread confusion in the community about the instability analysis of the critical point.

In numerical solutions of the transonic models, spatial oscillations are often found starting at the sonic point. These spatial oscillations, in contrast to temporal oscillations, can be understood as follows. We start with a subsonic flow from the star with $L < 0$ and $R < 0$. From equation (13), in the region where the heating rate H is sufficiently large in order to produce supersonic winds from the corona (the inner brackets in the second expression), T decreases. Before L reaches zero, from equation (15), $\mathrm{d}V/\mathrm{d}r > 0$, so that the flow accelerates to reach $L = 0$. When the flow speed crosses the sonic point at r_1 and becomes supersonic, from equation (16), $L > 0$. If R remains negative, equation (15) then requires $\mathrm{d}V/\mathrm{d}r < 0$, so that the flow decelerates to become subsonic. Then, since $L < 0$, $\mathrm{d}V/\mathrm{d}r > 0$, and the subsonic flow will accelerate to become supersonic again. The same acceleration–deceleration-then-acceleration–deceleration process repeats. In general, this situation thus results in a flow speed that spatially oscillates around the sonic speed. This pattern is similar to that shown in fig. 3(d) of Holzer (1977) which presented one cycle of oscillations. The sonic point $L = 0$ in the present model is equivalent to his first saddle point and the critical condition $R = 0$ equivalent to the second saddle point. However, the wavelength of spatial oscillation would depend on the resolution of a numerical solution. We should point out that without invoking artificial dissipation, the oscillations cannot be damped. A finer resolution would result in more oscillations until the second saddle point, $R = 0$, is reached.

Rather than explicitly considering the situation in which the velocity oscillates about sonic speed, one may idealize and approximate it as a 'sonic layer' through which $V = c$. Within this layer, the

momentum equation (15) is then replaced by $V = c$ because equation (15) can no longer describe the situation at $L = 0$ if $R < 0$, a situation that can occur in reality. Numerical tests have shown that the sonic layer approximation can produce good results in which the momentum across the sonic transition is conserved up to $\sim 10^{-5}$ of the net momentum. However, this approach, although providing very good results, may be considered only as an approximation because one may not be able to prove that the solution within the sonic layer satisfies the momentum equation.

5.4. Inevitability of unremovable singularity

The proposed discontinuity is needed when the singularity at the sonic point in equation (14) is not removable. One may challenge our model by arguing that the singularity may be always removable. For example, they argue that if neglecting dissipation, and assuming the flow is subsonic at the base, there are only 2 outcoming characteristics into the system, a sound wave and an entropy wave. Therefore, only two boundary conditions may be imposed, either density N and temperature T, or velocity V and temperature T, or density N and velocity V. The system will then relax to a stationary state. Now the physically reasonable conditions to impose are N and T in the chromosphere. V will therefore adapt itself to go into the Parker-like critical solution automatically. If we impose N and V, on the other hand, then T will adjust to create the critical solution.

Let us first consider the simple isothermal system analysed in Parker's original proposal, without the energy equation, with two inner boundary conditions for N and V. One can indeed find solutions for V as a function of r. However, the solution may stop at the sonic point depending on whether the critical condition $R = 0$ can be reached or not exactly at this point. However, this depends on the temperature T, an unspecified yet. If T is smaller than a certain value so that $R < 0$, no supersonic solution can be found. Similarly in other possible inner boundary conditions as described above. Therefore, the above argument is not true. Theoretically, the argument overlooked the possibility that there are two, not one, sound waves, one propagating forward and one backward as shown in Appendix B. As discussed in Section 3.3, the key agent in forming the discontinuity is the backward sonic wave that the argument overlooked. The most important ingredient in our model is the 'dissipation' of the heat flux. With the dissipation, the energy conservation equation has to be included. We, therefore, have three dependent variables and three equations so that the temperature cannot be 'adjusted' at one's wish. We recall that if one employs an isothermal or polytropic assumption, T is no longer a real dependent variable but a function of N. With three dependent variables, it is more obvious that one cannot guarantee a continuous sonic transition as discussed in the text around equation (19). This is the central issue of our model.

The conventional stellar wind theory has been based on a critical point in the evolution equation at the transonic speed. For continuous solutions, the separatrices from the critical point separate different regimes of solution. One of the two separatrices, the eigenfunction, is able to connect the subsonic coronal base speed to the supersonic solution with smooth transit through the critical point and with the pressure going to zero at infinity. One may take issue with discussions on technical/mathematical grounds, but our question here is about the nature of the reasoning employed to reach conclusions. The conventional theory reasons that, *according to the mathematical analysis,* because only this eigenfunction can provide a solution spanning from subsonic to supersonic conditions, *and* the solar wind is observed to be supersonic while originating from the subsonic corona, (a) the solar wind must pass through the critical point along the eigenfunction from the coronal base, (b) all other solutions do not describe reality or are unphysical, and (c) the analysis must therefore be valid and correct. This reasoning seems very suspicious and potentially flawed because it draws three conclusions based on one observed fact and one analytical result, similar to being able to derive 3 unknowns from two equations.

This reasoning is called 'abductive' reasoning and is known to be potentially flawed and unscientific (Song, 2022). It is true that the conventional theory proved 'a' possible solution for the formation of a supersonic stellar wind which is observed. However, it has not disproved other possibilities while *abductively* eliminated/excluded them. The conclusion that the conventional theory has been *proven* observationally rejects other possibility of higher initial velocities at the inner boundary. In fact, the observed supersonic solar wind may not be significantly produced according to the critical condition along a smooth eigenfunction in most common situations as we discussed in the paper.

Furthermore, the conventional theory proved the existence of two separatrices only for the isothermal case and 'generalized' it to more general situations. It has not been shown for a more general temperature profile. As we have shown in our study, this generalization may not be valid because the segment of a separatrix from point C to the critical point in Fig. 1 may not exist under more general conditions, e.g. with heat conduction. Stellar winds are formed by processes in the photosphere, chromosphere, transition region, and lower corona, which are less likely to be controlled by the requirement at the critical point along a smooth eigenfunction.

Of course, our model is only a possible solution and cannot exclude other possibilities.

5.5. Observation

In this preliminary stage of model development, it may be too early for us to predict quantitatively the observational differences between the present model and the conventional one using currently available observations. Nevertheless, we may discuss conceptually and semi-quantitatively the observational differences between the two models.

Conceptually, let us assume that with the same solar/stellar wind density and temperature, the initial velocity, V_0, has a range of possible values. Both models would predict that when $V_0 < V_e$, the solar/stellar wind is subsonic. The conventional model predicts that all supersonic solar/stellar winds arise with $V_0 = V_e$, while speeds with $V_0 > V_e$ cannot go through the sonic point (because no single-valued mathematical solution can be found). Accordingly, a supersonic solar/stellar wind cannot form easily because most often the theory does not allow a solution. If there is a finite range of distribution in the speed at the inner boundary, only a small amount of the flux will be able to get through. Consequently, a supersonic solar/stellar wind may be observed either only occasionally or with extremely small average fluxes. By contrast, in the present model, all flow speeds with $V_0 \geq V_e$ can become supersonic solar/stellar wind. If the initial speed is greater, both the solar/stellar wind speed and the temperature are higher. Therefore, supersonic winds can be continuously observed most of the time with large fluxes and a range of variable speeds, temperatures, and densities, as shown in Fig. 2. Comparing red/blue/orange lines with black lines on the right side of Fig. 2, the wind temperature, velocity, and Mach number predicted by the present model may also tend to be higher than that of Parker's model for the same coronal temperature and Parker's eigenfunction should correspond to the lower bound of the wind speed. In principle the solar wind 'energy problem' can be resolved.

Recent observational reports from the Parker Solar Probe (Raouafi et al 2023) suggest that although the solar wind speed has a large range variability, there is a possible lower bound for the supersonic solar wind speed as a function of the radial distance. Furthermore, this lower bound has a shape and value similar to that of Parker's eigenfunction. This observation may be interpreted as to be consistent with the conventional theory; the lower bound may indicate the lower limit of the coronal base temperature, and the range of the velocities above the bound may be explained by the temperature fluctuations in the coronal base. Indeed, if both the temperature and velocity both fluctuate at the coronal base, the solar wind at Parker Solar Probe orbits can be observed with temperature and velocity fluctuations. However, according to the conventional theory, at a given time only, a small fraction of the flux and time has the velocity that matches the required critical condition. Therefore, the observed flux should be intermittently and on average much smaller than that available at the coronal base and there should be periods of time when the supersonic solar wind is absent.

By contrast, our model would interpret the observed range of the velocity and temperature result from the varying heating rate in the corona and chromosphere and the lower bound results from the energy bound required for supersonic solar wind described by Parker's eigenfunction.

Semi-quantitatively, as shown in Fig. 2, the increases in temperature and velocity at the discontinuity are not as dramatic as the bow shock which is about a factor of 2 to 4. For example, in Fig. 2, a small initial speed above the eigenspeed of, say, 10 per cent, may produce a jump of 5 per cent in the sonic speed and 10 per cent in temperature jump. From equation (8), this 50 per cent enthalpy gain plus the 10 per cent gain in kinetic energy is eventually mostly converted to the solar kinetic energy gain, translating to 30 per cent increase in velocity. The net speed gain in Fig. 2 at 6.5 r_s is already about 20 per cent. Therefore, the discontinuity effectively amplifies the increase of velocity. This is consistent with the commonly observed solar wind properties – the solar wind velocity fluctuation is often greater than the temperature fluctuation. Maruca et al. () have shown with a large data base over the entire supersonic solar wind domain that the solar wind temperature fluctuation is about 5 per cent and the velocity fluctuation is about 45 per cent. Given the observations were made by a long range of distances with many different phenomena from different satellites, it is fair to say that a significant part of the fluctuations in the velocity is produced by the fluctuations in the source of the solar wind and the heating rate in the coronal base.

Quantitatively, however, the small temperature and velocity changes at the sonic transition may be difficult to observe remotely by the coronagraphs. The in-situ observation around the sonic transition and critical point, say $\sim 5\, r_s$, is needed to observe these 10–20 per cent changes. Such observations are not available yet because the closest distance of Parker Solar Orbiter's has been at more than 10 r_s. It is of great interest to have *in situ* measurements.

For coronagraph observations, of which the brightness may be proportional to the density, the proposed discontinuity at the sonic transition may be observed as a drop in brightness because both the speed and temperature increase across the discontinuity.

5.6. Comments on Shergelashvili et al. and Westrich et al. models

Recently, Shergelashvili et al. (2020) and Westrich et al. (2024) also made attempts to remove the singularity by invoking a discontinuity at the critical point in order to derive supersonic solutions that are consistent with observations around the solar surface and at 1 au. Their momentum and energy equations are different from ours. Noting that as discussed in equation (20), the momentum equation needs two expressions and in general the single expression (14), which is the momentum equation of the Shergelashvili et al. (2020) and Westrich et al. (2024) models, becomes invalid at the upstream of the discontinuity. More importantly, their energy equation describes adiabatic or polytropic processes with a heating source at the discontinuity. Both models require adjustable heating sources in the middle of the region of concern. The amount of the heating is adjusted and constrained empirically from at the solar surface and 1 au.

Quantitatively, Shergelashvili et al. (2020) predicts that, in their Fig. 1 at the sonic transition, the velocity increases by about 50 times, the temperature increases by about 4000 times and the density decreases by about 50 times. Furthermore, they predict that the temperature drops to 1000 K at the upstream boundary of the sonic transition. With such low temperature, the gases would recombine to neutrals and need to be reheated and re-ionized in similar processes in the chromosphere and transition region but within the discontinuity.

6. CONCLUSIONS

To describe the stellar wind, the evolution of the momentum equation has a singularity at the sonic point where the flow speed equals the sonic speed. In the conventional stellar wind theory, the supersonic stellar wind solution is based on a continuous solution, which requires the critical condition to be satisfied at the sonic point in order to remove the singularity with a continuous function. However, the critical condition and sonic point are in general two independent conditions. We demonstrate that except in extremely rare situations, the two conditions do not occur at the same point and the singularity at the sonic point is not always removable with a continuous function. This is because the critical condition is determined by the upstream conditions for the formation of the stellar wind, i.e. the ionization process that determines the mass flux of the stellar wind and the heating processes around the coronal base that determine the total energy of the solar/stellar wind. Therefore, the conventional stellar wind theory needs fundamental modification or reformulation.

We then revisited the conventional stellar wind theory, especially its line of reasoning. In the conventional stellar wind theory, the requirement for the two conditions to be satisfied at the same point functions more like an ID checkpoint at the sonic point. The problem is that the check is not about the plasma parameters locally at the sonic point but about the initial velocity at the inner boundary of many stellar radii upstream. By analogy, the ID check is not about the characteristics of the person who is passing the gate but about that of their parents or grandparents. Specifically, if the initial speed is greater than the eigenspeed, according to the conventional theory, the flow is not allowed to pass through the checkpoint with no option of how to go further, the most problematic conclusion of all. We questioned whether the initial speed at the inner boundary should be quantitatively checked at the critical point as the condition for formation of the supersonic stellar wind. Although the properties of the flow close to the sonic point may be influenced and modified by the existence of the transonic condition, there should be no question that the flow will pass across the sonic distance either supersonically or subsonically. That no physical solution can be found only means the solution has not been found yet and should not be used as proof that there is no solution. Whether the initial speed at the inner boundary is greater than the eigenspeed or not should not change the answer to whether flow from the corona will go forward.

With the insight gained from the processes forming the transition region (Song et al. 2023), the stellar wind plasma flux is controlled primarily by ionization of the neutral particles from the chromosphere over the transition region with the energy primarily from the dissipation of the downward heat flux produced from heating in the coronal base. Therefore, the stellar wind flux is not determined by the critical point condition. When the flow properties are determined by the processes well upstream in the chromosphere, the singularity cannot be removed, and the singular solution provided by the conventional theory is not valid. The inability to find a solution under these conditions when the initial speed is greater than the eigenspeed cannot be used to prevent these possible physical conditions from occurring. It may be interpreted as a failure of the expression of the momentum equation (14) at the singular point. After examination of the stellar wind evolution equation (14), we found that in the range of inner boundary conditions when the conventional theory fails to provide a supersonic wind solution or any kind of solution, the evolution equation itself becomes invalid because it requires a variable of negative values to be equal to zero. To resolve the dilemma, we rewrite the equation with two expressions, equation (20), one for the sonic transition and one for everywhere else. The solution of such an equation in general has two segments which may be connected by a discontinuity in the fluid treatment. In this new treatment, the singularity issue is addressed completely. After investigating possible instabilities, possible sonic layer, and time-dependence effects at the critical point, we validated the discontinuous solution.

In our model, the stellar wind energy flux is determined by the heating processes in the corona, and the stellar wind flux is determined by the ionization process from the transition region. The coronal base temperature is a characteristic measure of energy. When encountering the sonic point, depending on the flux and energy, the stellar wind will become either supersonic or subsonic, as predicted by the conventional theory. However, only a negligibly small fraction may pass smoothly across the sonic point along the eigenfunction line A–B in Fig. 1 as described by the conventional theory, and most stellar wind will pass through a discontinuity at the sonic point to gain enough energy becoming supersonic without going through the critical point as shown in Fig. 2, a possibility that has not been considered in the conventional theory. The eigenfunction in the conventional theory is actually the separation line between the supersonic and subsonic wind regimes and it is not the sole source for all supersonic solar/stellar wind. The sonic transition problem has not drawn much attention from large-scale numerical modeling, most likely because of the coarse spatial resolution. When the spatial resolution is a fraction of a stellar radius, numerical effects may be large enough to damp the oscillations within the sonic transition layer. In comparison, the spatial resolution shown in this study is $10^{-4}\ r_s$. With the availability of faster computers, the transonic issue is becoming more and more obvious (e.g. Adhikari et al. 2022). The detailed physical processes within the discontinuity need further theoretical and numerical investigations.

ACKNOWLEDGEMENTS

We thank Drs. V. M. Vasyliunas, M. G. Kivelson, and G. P. Zank for their valuable comments and suggestions.

DATA AVAILABILITY

The code for solving the governing equations and data produced can be accessed at this URL under Supersonic Stellar Winds.

REFERENCES

Adhikari L., Zank G. P., Telloni D., Zhao L.-L., 2022, ApJ, 937, L29
Aschwanden M., 2005, Physics of the Solar Corona: an Introduction with Problems and Solutions. Springer, New York
Brandt J. C., Wolff C. L., Cassinelli J. P., 1969, ApJ, 156, 1117
Chamberlain J. W., 1961, ApJ, 133, 675
Cranmer S. R., Winebarger A. R., 2019, Ann. Rev. Astron. Astrophys., 57, 157
Habbal S. R., Tsinganos K., 1983, J. Geophys. Res., 88, 1965
Hollweg J. V., 1976, J. Geophys. Res., 81, 1649
Holzer T. E., Leer E., 1980, J. Geophys. Res., 85, 4665
Holzer T. E., Leer E., 1997, in Wilson A., ed., Fifth SOHO Workshop: The Corona and Solar Wind Near Minimum Activity, 404, ESA, Noordwijk, p. 65
Holzer T., 1977, J. Geophys. Res., 82, 23
Hu Y. Q., Habbal S. R., Li X., 1999, JGR, 104**,** 24565
Hundhausen A. J., 1972, Coronal Expansion and Solar Wind, Springer, New York, p. 238
Hundhausen A. J., 1972, Coronal Expansion and Solar Wind. Springer, p. New York
Kennel C. F., 1994, AIP Conference Proceedings, 314, 180
Keto E., 2020, MNRAS, 493, 2834
Klimchuk J. A., 2006, Sol. Phys., 234, 41
Leer E., Holzer T. E., 1980, J. Geophys. Res., 85, 4681
Maruca B. A. et al. 2023, A&A, 675, A196
Maruca B. A. et al. 2023, A&A, 675, A196
Parker E. N., 1958, ApJ, 128, 664
Parker E. N., 1960, ApJ, 132, 821
Parker E. N., 1964a, ApJ, 139, 72
Parker E. N., 1964b, ApJ, 139, 93
Parker E. N., 1965a, ApJ, 141, 1463
Parker E. N., 1965b, Space Sci. Rev., 4, 666
Parker E. N., 1966, ApJ, 143, 32
Raouafi N. E. et al., 2023, ApJ, 945, 28
Rossi B., Olbert S., 1970, Introduction to the Physics of Space. McGraw-Hill, New York
Shergelashvili B. M. et al. 2020, MNRAS, 496, 1023
Shergelashvili B. M., Fichtner H., 2012, ApJ, 752, 142
Shi C., Velli M., Bale S. D., Réville V., Maksimović M., Dakeyo J.-B., 2022, Phys. Plasmas, 29, 122901
Shi C., Velli M., Bale S. D., Réville V., Maksimovic M., Dakeyo J.-B., 2022, Phys. Plasmas, 29, 122901
Song P., 2017, ApJ, 846, 92
Song P., 2022, Philosophy o f Science: Perspectives from Scientists. World Scientific Publishing Co, Singapore
Song P., Tu J., Wexler D. B., 2023, ApJ, 948, L4
Song P., Vasyliūnas V. M., 2011, J. Geophys. Res., 116, A09104
Southwood D. J., Kivelson M. G., 1995, GRL, 22, 3275
Spitzer L., Jr., 1956, Physics of Fully Ionized Gases, Interscience. New York
Suess S. T., 1982, ApJ, 259, 880
Telloni D. et al., 2023, ApJ, 954, 108
Tu C.-Y., Marsch E. 1995, Space Sci. Rev., 73, 1
Tu C.-Y., Pu Z.-Y., Wei F.-S., 1984, J. Geophys. Res., 89, 9695
Tu J., Song P., 2013, ApJ, 777, 53
Vainio R., Laitinen T., Fichtner H., 2003, A&A, 407, 713
Vasyliūnas V. M., Song P., 2005, J. Geophys. Res., 110, A02301
Velli M., 1994, ApJ, 432, L55
Velli M., 2001, Astro. Space Sci., 277, 157
Weber E. J., Davis L., 1967, ApJ, 148, 217
Westrich L., Shergelashvili B. M., Fichtner H., 2024, A&A, 686, A113
Wexler D. B., Stevens M. L., Case A. W., Song P., 2021, ApJ, 919, L33
Withbroe G. L., Noyes R. W., 1977, Ann. Rev. Astron. Astrophys, 15, 363
Wolff E. L., Brandt J. E., Southwick R. G., 1971, ApJ, 165, 181
Zank G. P., Zhao L. L., Adhikari L., Kasper J. C., Bale S. D., 2021, Phys. Plasma, 28, 080501

APPENDIX A: FORMATION OF STELLAR WIND PLASMA FLUX

Because the chromospheric density is high and the temperature is low, most of the energy from chromospheric heating is lost to radiation when the temperature is below the first impact ionization potential of dominant species and the ionization fraction is low. If, on the other hand, the heat source is up in the corona, ionization can take place more efficiently on the top surface of the chromosphere, i.e. the transition region, by dissipation of the downward heat flux in a process similar to sunlight evaporating a moist land surface. Heating, ionization, and radiative losses in the chromosphere and transition region have been investigated (Song and Vasyliunas 2011; Song 2017; Song et al. 2023).

The temperature at the top boundary of the transition region may be determined by the first impact ionization potential $\Phi_{\rm i}$ of dominant species there, which corresponds to $T_{\rm ts} \sim 158$kK for hydrogen. Based on the Song et al. (2023) model, the transition region is formed from weakly ionized gas to fully ionized plasma with the heating mostly arising from conversion of the coronal heat conduction flux and cooling from the energy losses due to radiation and ionization. This is where the ionization fraction increases from 0.5 to 1.0. The net ionized plasma flux is, from equation (20) of Song et al. (2023),

$$F_{\rm s} = \frac{-\Delta q_{\rm ts} + (Q_{\rm ts} - C_{\rm ts})\,\Delta r_{\rm ts}}{0.5\Phi_i + \Delta E_{\rm ts}} \tag{A1}$$

where subscript *ts* denotes the value at the top boundary of the transition region, $\Delta r_{\rm ts}$ is the thickness of the transition region, and $\Delta q_{\rm ts}$ is conduction flux from the coronal base flowing back into the transition region. It is a larger fraction of the net coronal base heating rate and is dissipated to provide the energy for ionization in the transition region. A higher heating rate in the coronal base produces a greater outgoing wind flux. $Q_{\rm ts}$ is the net heating rate, $C_{\rm ts}$ is the radiative cooling rate, and $\Delta E_{\rm ts} \sim 5kT_{\rm ts}$ is the energization in the transition region from the chromosphere. To the lowest order, the ionized wind energy flux is produced by a small fraction of the conduction flux from the corona, $q_{\rm ts}$, but a significant part of the heat flux is radiated away by $C_{\rm ts}$. Since the radiative cooling is approximately proportional to the square of the density, the wind flux is, though not explicitly expressed, a strong function of the density which is determined by the properties of the star. At the top of the transition region, i.e. the inner boundary of solar corona, the density is $N_{\rm ts}\sim 2.95 \times 10^9\,{\rm cm}^{-3}$ and the velocity is $V_{\rm ts}\sim 3.1 \times 10^3\,{\rm cm\,s}^{-1}$, giving the number flux $F_{\rm s}\sim 10^{13}\,{\rm cm}^{-2}\,{\rm s}^{-1}$, in the same range as the solar wind flux observed at the Earth's orbit (Hundhausen 1972).

APPENDIX B: PERTURBATION ANALYSIS

As we noted earlier in Section 1, when the fluid accelerates or decelerates, perturbations within the fluid are generated propagating in different directions. These perturbations will interact, either adding to each other or canceling out; coherent waves will develop and are described by the dispersion relation, which governs the propagation speed, and the perturbation relations, which governs the perturbations among different physical quantities. Only these coherent waves are of interest to our problem. A steady state flow solution describes these coherent waves and is able to accommodate all of the perturbations when the non-coherent perturbations and coherent oscillations are averaged out. For demonstration purposes, we neglect the effect of the magnetic field of which the analysis can be tedious (e.g. Weber and David, 1967). We start with the time-dependent governing equations.

Continuity

$$\frac{\partial N}{\partial t} + \frac{\partial}{r^2\partial r}\left(r^2 N V\right) = 0 \tag{B1}$$

Momentum equation

$$\rho\left(\frac{\partial V}{\partial t} + {\rm V}\frac{\partial V}{\partial r}\right) = -2\frac{\partial p}{\partial r} - \rho g \tag{B2}$$

In a radial 1D description, if the amplitudes of the perturbations δN and δV are produced in a small radial distance from r to $r + \delta r$, the density and velocity become $N + \delta N$ and $V + \delta V$. The velocity and density perturbations can propagate both radially forward, V_+, δV_+ and δN_+, and backward, V_-, δV_-, and δN_- relative to the flow speed, where the plus (minus) sign for the forward (backward). The net flow velocity is $V = V_+ + V_-$. The velocities and the perturbations in general have opposite signs. Noting that when the flow is supersonic, the terms forward/backward becomes ambiguous; we will use outward/inward for propagation relative to the star, in the supersonic wind, backward propagation is outward.

Coherent perturbations in each direction are governed by the perturbation relation of the propagation which can be derived from the conservation laws at each location with local parameters. The propagating perturbations of the velocity and density in the stellar frame of reference can be described by the form $e^{i(\omega t - k_\pm r)}$, where the subscript upper (lower) sign of $k_\pm$ relates to the forward (backward) propagations, respectively, relative to the flow.

Next, we derive the perturbation relationship between δN and δV. Parker's theory, which directly combines the steady state conservation laws for a simple single fluid, describes one of the possibilities. However, it encounters a singularity when the flow speed crosses the sonic speed as discussed above. Physically, the relationship for each propagation has to satisfy the relevant conservation laws. The propagation direction relative to the star depends on the value of the wavevector, k: positive (negative) is for outward (inward). The inclusion of the frequency is essential because it constrains the wavevector with the propagation speed as well as the perturbation relations. When neglecting the oscillation factor in a perturbed quantity, $e^{i(\omega t - k_\pm r)}$, the amplitudes are

$$\delta V = \delta V_+ + \delta V_- \tag{B3}$$

$$\delta N = \delta N_+ + \delta N_- \tag{B4}$$

The amplitudes are functions of distance; however, within δr, they are assumed constant since δr can be infinitely small when we do not consider nonlinear effects or discontinuities.

From the continuity and momentum equations (B1) and (B2), when neglecting the gravity waves, it is easy to derive the dispersion relation

$$v_{\rm ph\pm} = \frac{\omega}{k_\pm} = V \pm c \tag{B5}$$

where $v_{ph\pm}$ is the phase velocity in the frame of reference of the star, while the perturbation relation is

$$\frac{\delta N_\pm}{N} = \frac{\delta V_\pm}{v_{\rm ph\pm} - V} = \pm\frac{\delta V_\pm}{c} \tag{B6}$$

This is consistent with the conventional expectation that in the frame of reference at rest with respect to the flow, there are a forward and a backward propagating sonic waves.

We are interested in the transonic process, especially the backward propagating one. At the sonic point, $V = c$, the backward perturbation is not propagating, i.e. the phase velocity $v_{\rm ph-} = 0$ in the frame of reference of the star. In fact, the singularity in Parker's theory is

caused by the backward propagation at the transonic point. The forward perturbations propagate outward at twice the sound speed in the frame of reference of the star. They will propagate away quickly.

In the supersonic regime, both forward and backward propagations are outward with respect to the surface of the star.

When the effect of the magnetic field cannot be neglected, the JxB force will have to be added to equation (B2) which couples to Maxwell's equations. One can show that there exist three wave modes, the slow, Alfven, and fast modes and each mode has two (forward and backward) propagation directions (e.g. Weber and Davis, 1967). As we discussed earlier, three critical points or three standing discontinuities do not change the nature of the conclusion of this study that most supersonic winds are formed without going through the critical points.

APPENDIX C: FORMATION OF A STANDING DISCONTINUITY IN THE FLOW

From the discussion in Appendix B, there are many waves everywhere propagating either forward or backward and their density perturbations can be either in phase or antiphase. We now ask a different question about these perturbations. What perturbation will stay in a fixed location in the wind and corona with respect to the star, i.e. a standing structure against the background flow (e.g. Southwood and Kivelson, 1995)? In this problem, all time dependent processes in the stellar frame of reference are not important although the dispersion relation and perturbation relation remain in effect to constrain the allowed perturbations as well as the motion speed in the stellar frame of reference. When the magnetic field is neglected, subtracting (1) and (2) from (B1) and (B2) combined with (B5), and (B6) for steady state obtain

$$\frac{\mathrm{d}\,(\delta N_+)}{N\mathrm{d}r}+\frac{\mathrm{d}\,(\delta N_-)}{N\mathrm{d}r}+\frac{\mathrm{d}\,(\delta V_+)}{V\mathrm{d}r}+\frac{\mathrm{d}\,(\delta V_-)}{V\mathrm{d}r}+\frac{2}{r}=0 \tag{C1}$$

$$\frac{\mathrm{d}\,(\delta V_+)}{\mathrm{d}r}+\frac{\mathrm{d}\,(\delta V_-)}{\mathrm{d}r}=-\frac{c^2}{NV}\left(\frac{\mathrm{d}\,(\delta N_+)}{\mathrm{d}r}+\frac{\mathrm{d}\,(\delta N_-)}{\mathrm{d}r}\right)-\frac{c^2}{TV}\frac{\mathrm{d}T}{\mathrm{d}r}-\frac{g}{V} \tag{C2}$$

Substituting (B6) into (C1) and (C2) to remove the density perturbations we find

$$\left(\frac{1}{V}+\frac{1}{c}\right)\frac{\mathrm{d}\,(\delta V_+)}{\mathrm{d}r}+\left(\frac{1}{V}-\frac{1}{c}\right)\frac{\mathrm{d}\,(\delta V_-)}{\mathrm{d}r}=-\frac{2}{r} \tag{C3}$$

$$(V+c)\frac{\mathrm{d}\,(\delta V_+)}{\mathrm{d}r}+(V-c)\frac{\mathrm{d}\,(\delta V_-)}{\mathrm{d}r}=-\frac{c^2}{T}\frac{\mathrm{d}T}{\mathrm{d}r}-g \tag{C4}$$

Combining (C3) and (C4) to remove either the forward or backward propagation yields

$$v_{\mathrm{ph}+}\frac{\mathrm{d}\,(\delta V_+)}{\mathrm{d}r}=-\frac{1}{2}\left(\frac{2cV}{r}+\frac{c^2}{T}\frac{\mathrm{d}T}{\mathrm{d}r}+\frac{w^2r_s}{r^2}\right) \tag{C5}$$

$$v_{\mathrm{ph}-}\frac{\mathrm{d}\,(\delta V_-)}{\mathrm{d}r}=\frac{1}{2}\left(\frac{2cV}{r}-\frac{c^2}{T}\frac{\mathrm{d}T}{\mathrm{d}r}-\frac{w^2r_{\mathrm{s}}}{r^2}\right)=\frac{1}{2}R \tag{C6}$$

Note that v_{ph} is the propagation speed of a perturbation in the frame of reference of the star. When $v_{\mathrm{ph}}\neq 0$, the perturbation will propagate away and cannot form a stable structure in the flow. Therefore, a structure stands in the wind flow that is formed by the backward propagating perturbation at the sonic point when $v_{\mathrm{ph}-}(V=c)=0$. This forms a transonic structure between the subsonic and supersonic regimes. In the following, we discuss only the backward propagation (C6) near the sonic transition.

Subsonic flow. Before the sonic point, $v_{\mathrm{ph}-}=V-C<0$. There are two possible regimes of subsonic flow: $R>0$ and $R<0$. Since R starts with $R<0$, if R becomes positive before the sonic point, δV_- becomes negative, the flow speed may not be able to reach the sonic speed. This gives a subsonic wind solution.

If R remains negative as the sonic point is approached, δV_- will increase rapidly until the sonic point is reached. At the sonic point, since the LHS $=0$ but RHS <0, this solution is impossible except if $\delta V_-=\infty$. This forms a discontinuity.

Supersonic flow. Since the dT/dr term in equation (C6) is related to the heat flux, we neglect it for the moment in our discussion. From the subsonic flow discussion, the velocity has to increase rapidly as the sonic point is encountered. In order to maintain the transonic condition, the sonic speed and the temperature increase, which in turn increases R within the discontinuity. When the $R=0^+$ condition is reached, the supersonic condition can be maintained with continuous increase of the flow velocity.

Therefore, when $R<0$ as the sonic point is approached, supersonic flow can be formed downstream of the discontinuity at the sonic point. The process can be understood as follows. When neglecting the electromagnetic effects, the information in a HD flow is carried by the sound waves. Since the two propagations are at the sonic speed in the frame of reference of the flow, the forward propagation will propagate away and does not contribute significantly to the formation of a standing structure in the wind. By contrast, the backward propagation in the flow frame of reference propagates inward (outward) in the subsonic (supersonic) region, a standing structure may be able to form in the wind flow. As discussed above, within this structure, which is a discontinuity but not a shock, both the flow speed and temperature most likely increase. In Section 3.3, we show that these jump conditions form a stable discontinuity, i.e. across it the entropy increases.

When the magnetic field is included, in principle, three wave fronts can be formed for the slow, Alfvenic, and fast transitions, respectively (Weber and Davis 1967).

This paper has been typeset from a TEX/LATEX file prepared by the author.