



The physical basis of ionospheric electrodynamics

V. M. Vasyliūnas

Max-Planck-Institut für Sonnensystemforschung, 37191 Katlenburg-Lindau, Germany

Correspondence to: V. M. Vasyliūnas (vasyliunas@mps.mpg.de)

Received: 23 August 2011 – Revised: 21 January 2012 – Accepted: 9 February 2012 – Published: 13 February 2012

Abstract. The conventional equations of ionospheric electrodynamics, highly successful in modeling observed phenomena on sufficiently long time scales, can be derived rigorously from the complete plasma and Maxwell's equations, provided that appropriate limits and approximations are assumed. Under the assumption that a quasi-steady-state equilibrium (neglecting local dynamical terms and considering only slow time variations of external or aeronomical-process origin) exists, the conventional equations specify how the various quantities must be related numerically. Questions about how the quantities are related causally or how the stress equilibrium is established and on what time scales are not answered by the conventional equations but require the complete plasma and Maxwell's equations, and these lead to a picture of the underlying physical processes that can be rather different from the commonly presented intuitive or ad hoc explanations. Particular instances include the nature of the ionospheric electric current, the relation between electric field and plasma bulk flow, and the interrelationships among various quantities of neutral-wind dynamo.

Keywords. Ionosphere (Electric fields and currents; Ionosphere-atmosphere interactions; Plasma convection)

1 Introduction

The conventional treatment of ionospheric electrodynamics expounded in standard textbooks and tutorial publications (e.g. Matsushita, 1967; Rishbeth and Garriott, 1969; Boström, 1973; Kelley, 1989; Volland, 1996; Rishbeth, 1997; Richmond and Thayer, 2000; Heelis, 2004; Fuller-Rowell and Schrijver, 2009, and many others) has a dual aspect: the equations actually used to carry out calculations, and the verbal descriptions intended to explain physical processes represented by the equations. The conventional equations of large-scale ionospheric electrodynamics constitute a well-defined

set and have been successfully applied to model correctly a multitude of observed phenomena (down to time scales as short as minutes in some cases), despite the fact (not always explicitly acknowledged) that the equations are based on assumptions valid only in the case of quasi-steady-state equilibrium and therefore cannot describe dynamic developments or provide information about causal relations. The verbal descriptions are intended to provide an intuitive understanding of results from calculations, but sometimes they go beyond their nominal purpose and are expanded into qualitative discussions of causal sequences and sometimes even of temporal developments, even though these are not really described by the conventional equations.

The physical basis on which the verbal descriptions have been formulated and the conventional equations derived, to a large extent, is that provided by electromagnetic theory at the ordinary textbook level (which I refer to as E&M, to distinguish it from the electromagnetism of the complete plasma and Maxwell's equations). The purpose of this paper is to show that, when the complete plasma and Maxwell's equations are applied, the conventional equations are obtained under appropriate well-defined approximations, but the more rigorous physical understanding of causal sequences and time variations may be different from, and sometimes inconsistent with, the conventional verbal descriptions. The conventional ionospheric equations may thus be accepted as valid under restricted conditions (it is therefore no surprise that, despite their deficient treatment of plasma physics, they have proved adequate to explain the observations), but the detailed description of some underlying physical processes needs to be revised. The revised description can draw attention to important unanswered questions, the significance (or in some cases even the existence) of which may not be apparent in the conventional approach.

The reaction of some ionospheric physicists to the above argument has been that, as long as the equations are correct within their range of applicability and predict results in

agreement with observations, the detailed physical picture can be ignored. This may be called (Vasyliūnas, 2010) the Ptolemaic approach, by analogy to astronomical theory in the 16th century, when the heliocentric theory could be ignored or rejected on the grounds that the Ptolemaic scheme explained the observations just as well if not better than the Copernican (which was true at the time of Copernicus and for some decades afterwards).

In this paper I consider ionospheric electrodynamics both from the conventional and from the more rigorous point of view, with emphasis on the differences between the two approaches and their consequences. Section 2 summarizes the basic equations as conventionally presented and lists some advantages as well as problems associated with them. In Sect. 3 the conventional equations are derived from rigorous complete equations of plasma physics, with particular attention to approximations required and to discrepancies from the conventional verbal descriptions. Section 4 discusses two processes where the underlying physics is viewed in a particularly different way depending on whether the conventional or the more rigorous picture is invoked: the relation between electric field and plasma bulk flow, and the origin of the electric current produced by a neutral wind. Finally, as an illustrative example, Sect. 5 sketches how the prototypical neutral-wind dynamo is treated in the two approaches.

2 Conventional ionospheric electrodynamics

2.1 Basic concepts and equations

Restricted to questions of electrodynamics (electric and magnetic fields, currents, plasma bulk flow) as distinct from aeronomy (ionization, recombination, diffusion, chemical processes) and in the quasi-static limit (leaving out short-period wave processes), the theory of the ionosphere can be reduced to five basic equations (Gaussian units are used throughout this paper, for reasons discussed by Parker, 2007, his Sect. 6.4).

2.1.1 Ionospheric Ohm's law

The electric current density \mathbf{J} is assumed to be related by an Ohm's law

$$\mathbf{J} = \boldsymbol{\sigma} \cdot \mathbf{E}' = \sigma_P \mathbf{E}'_{\perp} + \sigma_H \hat{\mathbf{B}} \times \mathbf{E}' + \sigma_{\parallel} \mathbf{E}'_{\parallel} \quad (1)$$

to the electric field \mathbf{E}' in the frame of reference of the neutral atmosphere

$$\mathbf{E}' \equiv \mathbf{E} + \frac{1}{c} \mathbf{V}_n \times \mathbf{B} \quad (2)$$

where \mathbf{V}_n is the neutral wind velocity and \mathbf{E} the electric field (in the "fixed" frame of reference in which the calculation is done, most commonly the frame of the rotating Earth). $\hat{\mathbf{B}}$ is the unit vector along the ambient magnetic field \mathbf{B} , with \perp or \parallel marking quantities perpendicular or parallel to $\hat{\mathbf{B}}$; $\boldsymbol{\sigma}$ is the

conductivity tensor, with σ_P , σ_H , σ_{\parallel} the Pedersen, Hall, and parallel conductivities, respectively.

2.1.2 Potential electric field

The electric field (in the "fixed" frame) is assumed to have negligible curl and hence can be expressed as the gradient of a scalar potential,

$$\mathbf{E} = -\nabla \Phi. \quad (3)$$

2.1.3 Continuity of current

The electric current density is assumed to have zero divergence,

$$\nabla \cdot \mathbf{J} = 0. \quad (4)$$

Combining Eq. (4) with Eqs. (1), (2), and (3) yields a second-order elliptic differential equation

$$\nabla \cdot \boldsymbol{\sigma} \cdot \nabla \Phi = \nabla \cdot \boldsymbol{\sigma} \cdot (\mathbf{V}_n \times \mathbf{B})/c \quad (5)$$

from which the potential Φ can be calculated, given its boundary conditions and the spatial distribution of neutral winds \mathbf{V}_n .

2.1.4 Near-equipotential field lines

The component of the electric field along the magnetic field $E_{\parallel} \equiv \mathbf{E} \cdot \hat{\mathbf{B}}$ (independent of frame of reference) is assumed to be negligibly small, hence the potential can be taken as constant along a field line,

$$\hat{\mathbf{B}} \cdot \nabla \Phi \simeq 0. \quad (6)$$

Although in principle this property should appear as the result of solving Eq. (5) for Φ when the conditions $\sigma_{\parallel} \gg \sigma_P$, $\sigma_{\parallel} \gg \sigma_H$ hold, usually it is introduced as a separate assumption, thereby greatly simplifying the calculation.

2.1.5 Plasma bulk flow equation

Ion and electron bulk flows do not appear explicitly in the electrodynamic equations presented thus far. Conventionally they are assumed to be determined by $\mathbf{E} \times \mathbf{B}$ drifts plus collision effects. The plasma bulk flow velocity \mathbf{V} can then be calculated (over most of the ionosphere) from the equation

$$0 = c \frac{\mathbf{E}}{B} + \mathbf{V} \times \hat{\mathbf{B}} - \frac{\nu_{in}}{\Omega_i} (\mathbf{V} - \mathbf{V}_n) \quad (7)$$

where ν_{in} is the ion-neutral collision frequency and Ω_i the ion gyrofrequency. (Corrections of order electron-to-ion mass ratio, $m_e/m_i \ll 1$, are neglected throughout this paper, except for mentioning a few special instances where the case $m_e = m_i$ serves to illuminate some aspect of physics.)

Not included among the basic equations is Ampère's law

$$c \nabla \times \mathbf{B} = 4\pi \mathbf{J} \quad (8)$$

which relates the disturbance magnetic field to the ionospheric currents. In the conventional approach, it is used solely to compute the disturbance fields for comparison with observations and plays no other role in the electrodynamics.

2.2 Advantages and problems

An undeniable advantage of the conventional approach is mathematical convenience. Under a wide range of conditions, Eq. (3) implies that the horizontal electric field is to a good approximation independent of altitude within the ionosphere. At high latitudes, where the magnetic field is nearly vertical, this follows from Eq. (6); elsewhere and more generally, this follows directly from Eq. (3) whenever the horizontal length scale is large in comparison to the vertical length scale of the ionosphere (which is the case in many problems of interest). The spatial dependence of the ionospheric electric potential is then effectively on latitude and longitude only, and Eq. (5) can be reduced to a much simpler two-dimensional form, with height-integrated conductivities.

Outweighing this mathematical advantage are several problems of physical understanding, most of them related to the fact that the conventional theory applies only in situations of quasi-steady-state equilibrium.

1. The equations are approximate ones, valid (roughly speaking) for situations of slow change. The limits of validity – how slow is slow? – vary from equation to equation, but a clear unambiguous statement is generally lacking, partly perhaps because the equations have often been derived by ad hoc or semi-intuitive arguments, instead of by systematic approximation starting from complete equations of plasma physics (presented here in Sect. 3).
2. In theory, it is acknowledged that the equations are merely relations between quantities at a given time and do not say anything about cause and effect; occasionally one encounters, e.g. an explicit remark that the ionospheric Ohm's law just relates electric field and current without stating which is cause and which is effect (Boström, 1973; Richmond and Thayer, 2000). In practice, though, particularly in tutorial presentations, the equations may be invoked to draw quite specific and detailed conclusions about cause and effect – conclusions that in some cases are not supported, or are even contradicted, by a more rigorous treatment on the basis of the plasma equations.

Some specific instances of the problem noted above:

- (a) Concerning the interpretation of Eq. (7) for the plasma bulk flow, it seems to be assumed almost universally in the ionospheric/thermospheric community that the electric field is a cause producing a plasma bulk flow as an effect ($\mathbf{E} \times \mathbf{B}$ drift). For

plasma regimes that exist within most of the ionosphere and magnetosphere, however, the reverse can be shown to hold: an imposed plasma flow creates the corresponding electric field, but an imposed electric field by itself does not create a flow (discussed in Sect. 4.1).

- (b) Currents associated with the neutral-wind dynamo are often described as produced directly by neutral winds, through their larger effect on ion flow than on electron flow. This direct process, however, produces only a transient current that dies away very quickly; the main dynamo current results (on a much longer time scale) from velocity gradients and does not depend on differences between ion-neutral and electron-neutral collision effects (discussed in Sect. 4.2).
 - (c) Constancy of electric potential along a field line, expressed by Eq. (6), represents a well-understood *property* of a quasi-steady configuration, but it is frequently described as if it were a *process* of “electric field mapping along magnetic field lines,” supposedly acting to establish the property (discussed in Sect. 3.2.4).
3. Since the equations describe the quasi-steady-state equilibrium but not the processes by which it is established or modified, the conventional approach leaves out of consideration important physical questions: e.g. for 2a above, how the bulk flow is actually produced, in terms of dynamics (what forces act to impart the required linear momentum to the plasma?); for 2b, the role of magnetic stresses in the neutral-wind dynamo; for 2c, the role of MHD waves in communicating plasma flow and electric field changes along field lines.

3 Deriving the conventional equations

3.1 Fundamental equations of plasma electrodynamics

The complete and exact equations describing the relevant physics constitute the starting point for a proper derivation of any simplified approximate forms. For ionospheric electrodynamics at the level considered in this paper, the following set is necessary and sufficient.

3.1.1 Maxwell's equations

In their complete form, Maxwell's equations

$$\frac{\partial \mathbf{E}}{\partial t} = -4\pi \mathbf{J} + c \nabla \times \mathbf{B} \quad (9)$$

$$\frac{\partial \mathbf{B}}{\partial t} = -c \nabla \times \mathbf{E} \quad (10)$$

$$\nabla \cdot \mathbf{B} = 0 \quad \nabla \cdot \mathbf{E} = 4\pi \rho_c \quad (11)$$

provide the only exact and universal first-principles description of the electromagnetic field; other expressions, such as Coulomb's law, Biot-Savart law, Ampère's law, etc., are derived or approximated.

3.1.2 Generalized Ohm's law

The evolutionary equation for current density (obtained from the charge-weighted sum of velocity-moment equations of all the species)

$$\frac{\partial \mathbf{J}}{\partial t} = \frac{ne^2}{m_e} \left(\mathbf{E} + \frac{1}{c} \mathbf{V} \times \mathbf{B} - \frac{\mathbf{J} \times \mathbf{B}}{nec} \right) + e(n_i - n_e) \mathbf{g} + e \left(\frac{\nabla \cdot \boldsymbol{\kappa}_e}{m_e} - \frac{\nabla \cdot \boldsymbol{\kappa}_i}{m_i} \right) + \left(\frac{\delta \mathbf{J}}{\delta t} \right)_{\text{coll}} \quad (12)$$

describes the time development of \mathbf{J} , the primary source term in Maxwell's equations. Equation (12) is here given in a slightly simplified form that assumes a quasi-neutral plasma of electrons and one species of singly charged ions with $|n_i - n_e| \ll n$ and $m_i \gg m_e$ (for the exact multi-species form, see, e.g. Rossi and Olbert, 1970; Greene, 1973; Vasyliūnas, 2005a, 2011, and references therein). The effect of the gravitational acceleration \mathbf{g} , included for completeness, is in practice completely negligible here; $\boldsymbol{\kappa}_a$ is the kinetic tensor of species a , more commonly written as $\boldsymbol{\kappa} = \rho \mathbf{V} \mathbf{V} + \mathbf{P}$; and the collision term applicable in the ionosphere is

$$\left(\frac{\delta \mathbf{J}}{\delta t} \right)_{\text{coll}} = - \left(\nu_{ei} + \nu_{en} + \frac{m_e}{m_i} \nu_{in} \right) \mathbf{J} + (\nu_{en} - \nu_{in}) ne (\mathbf{V} - \mathbf{V}_n) \quad (13)$$

where ν_{in} , ν_{en} , and ν_{ei} are the ion-neutral, electron-neutral, and electron-ion collision frequencies, respectively.

3.1.3 Plasma momentum equation

The time development of plasma bulk flow \mathbf{V} , which plays an important role in the generalized Ohm's law Eq. (12) is described by

$$\frac{\partial}{\partial t} (\rho \mathbf{V}) + \nabla \cdot \boldsymbol{\kappa} - \rho \mathbf{g} = \frac{1}{c} \mathbf{J} \times \mathbf{B} + \left(\frac{\delta \rho \mathbf{V}}{\delta t} \right)_{\text{coll}} \quad (14)$$

(obtained from the mass-weighted sum of velocity-moment equations of all the species) where ρ is the mass density and $\boldsymbol{\kappa}$ the kinetic tensor of the entire plasma (not including the neutrals); the collision term in the ionosphere,

$$\begin{aligned} \left(\frac{\delta \rho \mathbf{V}}{\delta t} \right)_{\text{coll}} &= -n(m_i \nu_{in} + m_e \nu_{en}) (\mathbf{V} - \mathbf{V}_n) \\ &\quad - m_e (\nu_{in} - \nu_{en}) \frac{\mathbf{J}}{e} \\ &\approx -n m_i \nu_{in} (\mathbf{V} - \mathbf{V}_n), \end{aligned} \quad (15)$$

is universally approximated by the second expression (valid if $m_e \ll m_i$ and $|\nu_{in} - \nu_{en}| \ll \Omega_e$).

The time development of bulk flow of the neutrals is described by a neutral momentum equation, the counterpart of Eq. (14). In calculations where electrodynamic aspects are of primary interest, the neutral flow is often assumed to be known independently; also, the plasma density distribution is taken as given (from empirical or from independent theoretical modeling), and the kinetic-tensor terms are assumed negligible. (The vertical structure of the ionosphere, where pressure gradients and gravity cannot be neglected, is generally treated separately from electrodynamics, as a question of aeronomy.) Equations for the evolution of ρ and of $\boldsymbol{\kappa}_e$, $\boldsymbol{\kappa}_i$ as well as of the neutral flow thus need not be considered explicitly when (as in this paper) the emphasis is on ionospheric electrodynamics, although they must of course be included for self-consistent modeling of the ionosphere/thermosphere interaction (e.g. Richmond et al., 1992; Fuller-Rowell et al., 1996; Ridley et al., 2006; Song et al., 2009; Tu et al., 2011).

3.2 Approximations of the conventional approach

The conventional equations listed in Sect. 2.1 are in essence (although not necessarily derived in this way) the result of dropping the time derivatives from all the fundamental Eqs. (9), (10), (12), (14). These fundamental equations, however, determine uniquely the time derivatives of all the quantities involved, given the present values of the quantities and their spatial gradients. The (approximate) conventional equations thus cannot be derived simply by assuming a steady state (the more so when they are invoked to interpret observed ionospheric phenomena, every one of which is variable on some time scale); it is necessary to consider under what conditions the time derivative implied by each fundamental equation can be treated as negligibly small.

Neglect of time derivatives $\partial \mathbf{J} / \partial t$ in the generalized Ohm's law Eq. (12) and $\partial \mathbf{E} / \partial t$ (displacement-current term) in Maxwell's Eq. (9), reducing the latter to Ampère's law Eq. (8), can be shown (Vasyliūnas, 2005a,b, 2011) to be valid for all plasma phenomena characterized by spatial scales much larger than the electron inertial length (collisionless skin depth) $\lambda_e \equiv c / \omega_p$ and time scales much longer than the inverse plasma frequency ω_p^{-1} , i.e. length and time scales much larger than those of (electron) plasma oscillations. (Numerical values: $\lambda_e = 5.3 \text{ km} / \sqrt{n_e}$, $\omega_p = 2\pi \sqrt{n_e} \times 9.0 \text{ kHz}$, n_e in cm^{-3} .) In the ionosphere, this means length scales longer than some tens of meters and frequencies below the radio range, which includes most of the phenomena studied by the conventional approach. On these large scales, charge separation effects can be neglected; conversely, charge separation and accumulation effects, when significant, occur on times scales of order $\sim \omega_p^{-1}$ (cf. Sect. 3.2.3).

Below I discuss the derivation of each of the conventional equations of Sect. 2.1 in turn.

3.2.1 Ionospheric Ohm's law

With neglect of the time derivative as discussed above, and with further neglect of kinetic-tensor and gravitational terms, the generalized Ohm's law Eq. (12) can be rewritten as

$$0 \simeq \mathbf{E} + \frac{1}{c} \mathbf{V} \times \mathbf{B} - \frac{\mathbf{J} \times \mathbf{B}}{nec} + \frac{m_e}{ne^2} \left(\frac{\delta \mathbf{J}}{\delta t} \right)_{\text{coll}} \quad (16)$$

where the collision term is, to lowest order in m_e/m_i ,

$$\frac{m_e}{ne^2} \left(\frac{\delta \mathbf{J}}{\delta t} \right)_{\text{coll}} \simeq - \frac{m_e v_e}{ne^2} \mathbf{J} \quad v_e \equiv v_{ei} + v_{en} . \quad (17)$$

The ionospheric Ohm's law Eq. (1) is conventionally presented as analogous to the ordinary Ohm's law in a conducting medium: \mathbf{E} in the frame of reference of the medium drives \mathbf{J} that is limited by collisions. Inserting the collision term (17) into Eq. (16) does not, however, lead to Eq. (1) but yields a different equation

$$0 \simeq \mathbf{E} + \frac{1}{c} \mathbf{V} \times \mathbf{B} - \frac{\mathbf{J} \times \mathbf{B}}{nec} - \frac{m_e v_e}{ne^2} \mathbf{J} \quad (18)$$

which when solved for \mathbf{J} relates it linearly (by an equation of the same form as Eq. (1) but with different conductivity coefficients) to the electric field in the frame of reference of the plasma, with the conductivity depending primarily on electron collisions; in contrast, Eq. (1) relates \mathbf{J} to the electric field \mathbf{E}' in the frame of reference of the neutral atmosphere, with the conductivity depending primarily on ion-neutral collisions (the two alternatives were pointed out by Song et al., 2001, who give explicit expressions for both forms of the conductivity).

The root of the difference is most clearly seen by rewriting Eq. (18) in terms of \mathbf{E}' from Eq. (2) instead of \mathbf{E} :

$$0 \simeq \mathbf{E}' + \frac{1}{c} (\mathbf{V} - \mathbf{V}_n) \times \mathbf{B} - \frac{\mathbf{J} \times \mathbf{B}}{nec} - \frac{m_e v_e}{ne^2} \mathbf{J} . \quad (19)$$

Obviously, Eq. (19) will predict a linear relation between \mathbf{J} and \mathbf{E}' if and only if there exists a linear relation between the velocity difference $\mathbf{V} - \mathbf{V}_n$ and \mathbf{J} . Precisely such a relation, however, is provided by the plasma momentum Eq. (14) if its LH side (acceleration, gravitational, and pressure-gradient terms) is neglected, reducing Eq. (14) to

$$0 \simeq \frac{1}{c} \mathbf{J} \times \mathbf{B} - nm_i v_{in} (\mathbf{V} - \mathbf{V}_n) . \quad (20)$$

Inserting $\mathbf{V} - \mathbf{V}_n$ from Eq. (20) into Eq. (19) does yield Eq. (1). Alternatively (and more fundamentally), one may derive Eq. (1) by considering \mathbf{J} as given by Eq. (20) and invoking Eq. (16) solely to eliminate \mathbf{V} by expressing it in terms of \mathbf{E} .

The ionospheric current is thus primarily a stress-balance current, obtained by balancing the Lorentz force against the collisional drag force from relative bulk motion between plasma and neutrals (Vasyliūnas and Song, 2005; Vasyliūnas, 2005b, 2011); in this respect it is analogous rather to the

“ring current” in the magnetosphere, which is obtained by balancing the Lorentz force against plasma pressure gradient. That it is not an Ohmic current in the usual sense is also shown by lack of a unique “frame of reference of the medium”: \mathbf{J} can be related to the electric field equally well (Song et al., 2001) in the frame moving with bulk flow of the neutral atmosphere \mathbf{V}_n (the conventional approach), or with bulk flow of the plasma \mathbf{V} , or indeed (Vasyliūnas and Song, 2005) with any velocity $\mathbf{U} = \mathbf{V} + \zeta (\mathbf{V}_n - \mathbf{V})$ where ζ is an arbitrary constant (\mathbf{U} is any point in velocity space on the line defined by \mathbf{V} and \mathbf{V}_n); the conductivity coefficients relating \mathbf{J} to the electric field are different for each choice of frame. As further indication of the stress-balance nature of the current, energy dissipation calculated as $\mathbf{J} \cdot (\mathbf{E} + \mathbf{V}_n \times \mathbf{B}/c)$, conventionally called ionospheric Joule heating, arises predominantly from mechanical dissipation by plasma-neutral collisions with only a minor contribution from electromagnetic dissipation $\mathbf{J} \cdot (\mathbf{E} + \mathbf{V} \times \mathbf{B}/c)$ (Joule heating in the proper sense; see Vasyliūnas and Song, 2005, for a detailed discussion).

3.2.2 Potential electric field

Equation (3), representing the electric field as the gradient of a potential, is valid to the extent that $\nabla \times \mathbf{E}$ and hence $\partial \mathbf{B}/\partial t$ can be neglected. Faraday's law Eq. (10) is the only one of the fundamental equations in which $\nabla \times \mathbf{E}$ as well as $\partial \mathbf{B}/\partial t$ appear explicitly, and since it contains no other quantities to which they could be compared, there is no obvious general criterion for neglecting them. The approximation of potential \mathbf{E} can thus be justified only by indirect, order-of-magnitude arguments, involving plausible or observed time scales.

The very small ratio $\delta B/B$ of the observed magnetic fluctuations to the main geomagnetic (dipole) field is sometimes invoked to justify assuming a potential \mathbf{E} in the ionosphere. An order-of-magnitude relation between non-potential field $\delta \mathbf{E}$ and time-varying δB is readily obtained from Eq. (10),

$$\delta \mathbf{E} \sim \frac{\lambda}{c\tau} \delta B \quad (21)$$

where τ is the time scale of the variation and λ is the gradient length scale of $\nabla \times \mathbf{E}$, but there is no corresponding relation between potential \mathbf{E} and the main field B . (Splitting the electric field into potential and non-potential parts is most simply done with the Coulomb gauge.) The total \mathbf{E} (potential plus non-potential) appears, however, in Eq. (1) and can thus be related to \mathbf{J} , which in turn is related to δB by Ampère's law, Eq. (8). This yields the order-of-magnitude relation between δB and the total \mathbf{E} :

$$\delta B \sim \frac{4\pi}{c} \Sigma_P E \quad (22)$$

where Σ_P is the Pedersen conductance (height-integrated conductivity). Combining Eqs. (21) and (22) gives an estimate for the ratio of potential to total \mathbf{E} :

$$\frac{\delta E}{E} \sim \frac{4\pi \Sigma_P \lambda}{c^2 \tau} = \frac{4\pi \Sigma_P V_A}{c^2} \frac{\lambda}{V_A \tau} \quad (23)$$

(in SI units, replace $4\pi \Sigma_P/c^2$ by $\mu_0 \Sigma_P$).

Equation (23) shows that, contrary to what is sometimes supposed, the ratio $\delta E/E$ is *not* proportional to $\delta B/B$. For given λ and τ , the ratio is in fact approximately proportional to $1/B$ (because of the $\sim B^{-1}$ dependence of Σ_P) but does not depend on δB (a somewhat counter-intuitive result from Eqs. (21) and (22), which show that δE and E are both proportional to δB). The dimensionless quantity in the second expression of Eq. (23)

$$\frac{4\pi \Sigma_P V_A}{c^2} = \frac{\Sigma_P}{1 \text{ mho}} \frac{V_A}{796 \text{ km s}^{-1}} \quad (24)$$

need not be small in the ionosphere, hence the condition for negligible $\delta E/E$ is $\lambda \ll V_A \tau$. The approximation of potential E is thus valid primarily because the time scale τ of magnetic variations is assumed long compared to the spatial scale λ divided by the Alfvén speed; equivalently, only slow changes of a quasi-steady equilibrium are considered, neglecting any wave propagation effects.

3.2.3 Continuity of current

Equation (4) is an immediate mathematical consequence of Ampère's law Eq. (8) and thus holds whenever the displacement-current term in Maxwell's Eq. (9) can be neglected, which (as discussed in Sect. 3.2) is the case quite generally for plasma regimes found in the ionosphere and the magnetosphere, except for phenomena at frequencies $> \omega_p$ and/or length scales $< \lambda_e$. The current continuity assumption of the conventional approach is thus valid to a very high degree of approximation.

A problem arises when sometimes (particularly in tutorial presentations) the physical process that establishes current continuity is described entirely in terms of the conventional equations. A common argument is that if the initial distribution of electric potential and neutral-wind dynamo ($\mathbf{V}_n \times \mathbf{B}$) drives a current with $\nabla \cdot \mathbf{J} \neq 0$, the implied charge accumulation modifies the potential and hence \mathbf{E} , which in turn, through Eq. (1), modifies \mathbf{J} until the condition $\nabla \cdot \mathbf{J} = 0$ is satisfied. How fast does this occur? From

$$\nabla \cdot \mathbf{J} = -4\pi \frac{\partial \rho_c}{\partial t} = \nabla^2 \left(\frac{\partial \Phi}{\partial t} \right) \quad (25)$$

and with \mathbf{J} and Φ assumed related by Eqs. (1) and (3), the implied time scale τ can be estimated as

$$\frac{1}{\tau} \sim 4\pi \sigma_P = v_{\text{in}} \left(\frac{c^2}{V_A^2} \right) = v_{\text{in}} \left(\frac{\omega_p^2}{\Omega_e \Omega_i} \right) \sim O(\omega_p) \quad (26)$$

(in agreement with the general conclusion of Sect. 3.2 that anything involving charge separation happens on time scales of this order). On such a short time scale, however, neither Eq. (1), which presupposes neglect of plasma acceleration terms (Sect. 3.2.1), nor Eq. (3), which presupposes time

scales longer than Alfvén wave travel times (Sect. 3.2.2), can be assumed to apply.

The relation between \mathbf{J} and \mathbf{B} is discussed at the level of fundamental equations by Vasyliūnas (2005b, 2011), with the result that any difference between \mathbf{J} and $(c/4\pi)\nabla \times \mathbf{B}$ (which is a necessary condition for $\nabla \cdot \mathbf{J} \neq 0$) disappears on a time scale of order either $\tau_1 \simeq \mathcal{L}/c$ (light travel time, where \mathcal{L} is the gradient length scale) or $\tau_2 \simeq \omega_p^{-1}$ (inverse plasma frequency), whichever is the shorter; $\tau_1 = \tau_2$ corresponds to $\mathcal{L} = \lambda_e$. If $\tau_1 \ll \tau_2$ (typical of the ordinary E&M laboratory), $\nabla \times \mathbf{B}$ changes to match $(4\pi/c)\mathbf{J}$; if $\tau_1 \gg \tau_2$ (typical of the large-scale ionospheric and magnetospheric plasmas), \mathbf{J} changes to match $(c/4\pi)\nabla \times \mathbf{B}$.

3.2.4 Near-equipotential field lines

The component of the generalized Ohm's law (e.g. in the form of Eq. (18)) parallel to \mathbf{B} implies, if the electron-collision term is negligible, $E_{\parallel} \equiv \mathbf{E} \cdot \hat{\mathbf{B}} \simeq 0$; from this, Eq. (6) follows if and only if \mathbf{E} can be represented as the gradient of a scalar potential, conditions for which have been discussed in Sect. 3.2.2.

An argument frequently invoked to derive Eq. (6) is that charges can move freely along \mathbf{B} and therefore will short out any potential difference. The immediate result, however, of this process (which involves charge separation and therefore occurs on a very short time scale of order ω_p^{-1} , cf. Sect. 3.2.3) is to remove the parallel electric field only, leaving unchanged the perpendicular electric field which varies on much longer time scales. Even if the initial \mathbf{E} was a potential field, the altitude-dependent \mathbf{E}_{\perp} that remains after \mathbf{E}_{\parallel} has been "shorted out" has a non-zero curl (except in trivially simple geometries) and therefore acts to produce a changing \mathbf{B} , which in general also has a non-zero curl and therefore acts in turn (via the $\mathbf{J} \times \mathbf{B}$ force) to change the cross-field flow of the plasma. What is here described can be readily seen to be the velocity shear (implied by non-equipotential field lines) propagating as an Alfvén wave. The physical process is thus a mutual readjustment of differences in flow and \mathbf{E}_{\perp} between different locations on a field line by MHD (Alfvén) waves propagating back and forth along the field line; potential mapping is not the physical process but only its final result when and if a quasi-steady state is reached. The process itself has been widely studied particularly in the context of magnetosphere-ionosphere coupling (e.g. Holzer and Reid, 1975; Kan et al., 1982; Wright, 1996; Lysak, 2004, and many others).

3.2.5 Plasma bulk flow equation

Equation (7) can be derived most simply by using the reduced form of the plasma momentum Eq. (20) to eliminate the $\mathbf{J} \times \mathbf{B}$ term from the (reduced) generalized Ohm's law Eq. (18) (the electron-collision \mathbf{J} -term could also be eliminated the same way but is usually neglected as being of order

v_e/Ω_e compared to other terms). The essential step, here as well as in the analogous derivation (Sect. 3.2.1) of Eq. (1) by using Eq. (20) to eliminate the $\mathbf{V} - \mathbf{V}_n$ term from Eq. (19), is neglect of the acceleration and pressure-gradient terms in the full plasma momentum Eq. (14). Equation (7) thus represents the plasma bulk flow that results from stress balance between electromagnetic and collisional forces, just as the ionospheric Ohm's law Eq. (1) represents the current that results from this stress balance.

4 Underlying physics of some processes

4.1 Electric field and plasma flow: which drives which?

When collision effects are negligible, Eq. (7) predicts plasma bulk flow equal to $\mathbf{E} \times \mathbf{B}$ drift plus an unspecified (and sometimes overlooked) flow component parallel to \mathbf{B} . This flow can also be obtained by calculating the drift motion of individual charged particles in given \mathbf{E} and \mathbf{B} fields; such derivations, found in many textbooks, seem to constitute the primary argument for the view that the electric field produces the plasma flow. Taken by itself, however, Eq. (7) (or its simplification to the MHD condition $c\mathbf{E} + \mathbf{V} \times \mathbf{B} = 0$) is merely a relation between the quantities at a given time: it states that if the electric field exists, the plasma flow must also exist, and vice versa, a result that can be derived from the assumed condition of stress balance (Sect. 3.2.5).

If, instead, one views Eq. (7) as a statement that the electric field gives rise to or produces the plasma flow, one is presupposing (explicitly or implicitly) that the relation between the two is the result of a time evolution, with the electric field specified first and the plasma flow then developing as a consequence. Whether the electric field produces the plasma flow in this way, or the plasma flow produces the electric field, or both are produced by something else, can be determined unambiguously on purely physical grounds by calculating the time evolution of all the relevant quantities. The simplest way of answering the question "which quantity produces which?" is to consider an initial-value problem: assume that at $t = 0$ only one of the quantities is present and solve the equations to determine whether and how the other quantity evolves at subsequent times. (The alternative is a philosophical discussion, which involves subtle distinctions, e.g. between efficient and formal cause, and is not likely to yield a concrete physical result.) The conventional textbook derivation of $\mathbf{E} \times \mathbf{B}$ drift is in fact the initial-value calculation of a single particle in fixed given fields; the difference in a plasma is that time evolution is now governed by the full set of fundamental equations, and single-particle calculations that ignore the effect of plasma on fields (particularly on \mathbf{E}) are no longer adequate.

Self-consistent plasma calculations of how \mathbf{E} and \mathbf{V} evolve from some arbitrarily specified (unconstrained) initial values (Buneman, 1992; Vasyliūnas, 2001) show that,

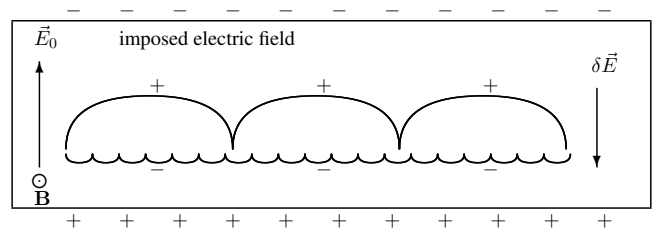


Fig. 1. Sketch of trajectories and charge locations for initially imposed electric field with no plasma flow. + – outside the box are charges associated with the imposed electric field; + – inside the box indicate charge separation resulting from the drift motion, with $\delta\mathbf{E}$ the associated change of electric field. The trajectories are drawn for the limit $|\delta\mathbf{E}| \ll |\mathbf{E}_0|$; see text for discussion of the more general case.

as long as $V_A^2 \ll c^2$ (i.e. the inertia of the plasma is dominated by the rest mass of the plasma particles and not by the relativistic energy-equivalent mass of the magnetic field), flows produce electric fields but electric fields do not produce flows, in a precise sense: an initially imposed bulk flow creates the corresponding \mathbf{E} , while an initially imposed electric field does not create the corresponding \mathbf{V} but dissolves into fluctuations. In both cases, the change occurs on a time scale of order ω_p^{-1} and is therefore (since $v_{in}^2 \ll \omega_p^2$ when $V_A^2 \ll c^2$) not affected by collisions. Why this happens can be understood in several intuitive ways, some independent of the initial-value approach.

4.1.1 Effect of \pm particle trajectories

Charge separation from the cycloidal trajectories of drifting particles modifies the electric field (this argument, related to the initial-value calculation, is perhaps the one most consonant with the conventional approach). Figure 1 illustrates, in a simple slab geometry, the trajectories of ions and electrons (equal initial concentrations $n_e \simeq n_i \equiv n$) injected with zero velocity (no initial flow) into an initial \mathbf{E}_0 . As the result of $\mathbf{E} \times \mathbf{B}$ drift, charge layers (of density $\pm ne$ and thickness \simeq gyroradius of the corresponding (\pm) drifting particles) appear at the boundaries of the (locally homogeneous) region, as shown. The implied change of the electric field is

$$\delta\mathbf{E} = -4\pi ne \left[\frac{c\mathbf{E}}{B} \right] \left[\frac{(m_i + m_e)c}{eB} \right] = -\mathbf{E} \frac{c^2}{V_A^2} \quad (27)$$

where

$$\mathbf{E} = \mathbf{E}_0 + \delta\mathbf{E} \quad (28)$$

is the actual electric field inside the region, which determines the $\mathbf{E} \times \mathbf{B}$ drift. If the plasma flow is to be essentially the $\mathbf{E} \times \mathbf{B}$ drift of the imposed electric field (as assumed in conventional ionospheric theory and in drawing Fig. 1), $\delta\mathbf{E}$ must be negligible in comparison to \mathbf{E}_0 , which requires that the charged-particle concentration n be sufficiently low. The

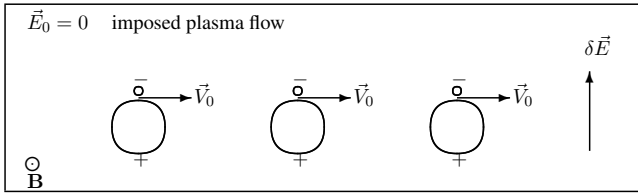


Fig. 2. Sketch of trajectories and charge locations for initially imposed plasma flow with no electric field. + – indicate charge separation resulting from the gyromotion, with δE the associated electric field. The trajectories are drawn for the limit $|\delta E| \ll |V_0 \times B/c|$; see text for discussion of the more general case.

quantitative relation of the actual to the initial E is given by Eqs. (27) and (28) as

$$E = \frac{V_A^2}{c^2 + V_A^2} E_0 \tag{29}$$

(where V_A^2 is merely shorthand for $B^2/4\pi n(m_i + m_e)$ and hence not subject to the constraint $\leq c^2$), showing that only if the density is so low that (nominally) $V_A^2/c^2 \gg 1$ can δE be neglected. When $V_A^2/c^2 \ll 1$, the electric field is reduced to a small fraction of its initial value, with corresponding reduction of the $E \times B$ drift. (For more detailed discussion and numerical simulation, see Tu et al., 2008).

Figure 2 illustrates the trajectories if ions and electrons are injected with common initial velocity V_0 but no initial electric field. As the result now of gyromotion, charge layers appear at the boundaries as before, with implied electric field

$$\delta E = -4\pi n_e e \left[V \times \hat{B} \right] \left[\frac{(m_i + m_e)c}{eB} \right], \tag{30}$$

the $E \times B$ drift of which modifies the initial velocity to

$$V = V_0 + \frac{c \delta E \times B}{B^2} \tag{31}$$

From Eqs. (30) and (31), the relations of the actual E ($= \delta E$ in this case) and V to the initial velocity are

$$E = \frac{c^2}{c^2 + V_A^2} \left[\frac{-V_0 \times B}{c} \right] \quad V = \frac{c^2}{c^2 + V_A^2} V_0. \tag{32}$$

The conventional assumption (implicit in drawing Fig. 2) that, without an imposed electric field, the flow velocity is reduced to a negligible value (the initial V_0 turning into gyromotion) proves to be valid, again, only in the low-density limit $V_A^2/c^2 \gg 1$. For $V_A^2/c^2 \ll 1$, the flow velocity remains at almost its initial value, and the requisite corresponding electric field is created.

Equations (29) and (32) agree with the qualitative inferences from momentum conservation (Sect. 4.1.3). Since charge separation occurs because the positive and the negative particles have gyromotions of opposite sense, regardless of whether they do or do not have different masses, the results do not depend on the ratio m_e/m_i , provided the Alfvén speed is defined on the basis of mass density $\rho = n(m_i + m_e)$.

4.1.2 Maintaining charge quasi-neutrality

On spatial and temporal scales large compared to those of electron plasma oscillations, the electric field is determined by the requirement that differential acceleration of ions and electrons must not separate charges too much (this is the physical meaning of dropping the $\partial J/\partial t$ term in the generalized Ohm’s law Eq. (12), discussed in Sect. 3.2). The bulk flow of the plasma is essentially that of the ions, which are much heavier than electrons, hence the electric field primarily changes the flow of electrons to match that of ions. This is the simplest argument, quite adequate for the real ionosphere but not a completely general one; contrary to the impression it might give, the relation between electric field and plasma flow does not depend on smallness or otherwise of m_e/m_i (redoing the calculations of Buneman (1992) and Vasyliūnas (2001) with $m_e = m_i$ does not change their results, consistent with the qualitative description in Sect. 4.1.1).

4.1.3 Momentum conservation

This is perhaps the most fundamental argument. Bulk flow carries linear momentum and thus can be produced only by adding linear momentum to the plasma. The linear momentum density of the plasma medium is

$$\rho V + \frac{1}{4\pi c} E \times B \tag{33}$$

where the first term represents the momentum in bulk flow of the plasma and the second represents the momentum in the electromagnetic field. When bulk flow equals $E \times B$ drift, the ratio (in magnitude) of the second term to the first is V_A^2/c^2 . If a given electric field E is to produce a bulk flow, its linear momentum suffices only for a flow velocity of order V_A^2/c^2 times the $E \times B$ drift. Contrariwise, a given flow V needs to be reduced by subtracting only a small fraction V_A^2/c^2 , in order to supply the linear momentum for $E = -V \times B/c$.

If an electric field is not merely imposed initially but is maintained at a steady value by some external agency, a current must in general be supplied, and the $J \times B$ force of that current (not E) is what adds the linear momentum to the plasma and thereby produces the flow. Note that the current in question is *not* that described by the ionospheric Ohm’s law Eq. (1), the $J \times B$ force of which is completely balanced by plasma-neutral collisions (as discussed in Sect. 3.2.1) and thus does not add linear momentum to the plasma.

The arguments in Sects. 4.1.2 and 4.1.3, it should be noted, are completely independent of any initial-value considerations.

4.2 How do wind-driven currents arise?

A primary task of ionospheric electrodynamics is to describe electromagnetic fields and electric currents that result from motions of the neutral atmosphere (commonly called the

“neutral-wind dynamo” problem). This is a topic of historical significance (much of the conventional theory was first developed to explain observed quiet-time geomagnetic variations as the result of a neutral-wind dynamo) but also one in which the gap between the usual qualitative explanation and the physics of the equations may be especially wide.

How does bulk flow of neutral particles give rise to an electric current? Since the neutral particles themselves do not interact with electromagnetic fields, any electrodynamic effect can result only from collisions between neutrals and plasma particles. In the conventional approach, it is generally assumed that a current is produced because, when plasma collides with a flowing neutral medium, different bulk velocities are imparted to ions and to electrons. Qualitative explanations of how the process works have been suggested in several different forms:

1. The current is produced directly by the difference between ion-neutral and electron-neutral collision frequencies (not proposed explicitly, but hinted at in some tutorial presentations). This process is easily identified in the full generalized Ohm’s law Eq. (12) as the contribution to $\partial \mathbf{J} / \partial t$ of the collision term in the second line of Eq. (13), proportional to $\nu_{en} - \nu_{in}$ as expected. What is not at all apparent, however, is how this leads to the expression from which the neutral-wind dynamo is calculated in practice, the $\mathbf{V}_n \times \mathbf{B}$ term in the ionospheric Ohm’s law Eq. (1) (note that the current given by this term does *not* vanish when $\nu_{en} = \nu_{in}$).
2. The current results from modification of $\mathbf{E} \times \mathbf{B}$ drift by collisions, and the modification depends on the ratio of collision frequency to gyrofrequency – obviously very different for ions and for electrons. This explanation is generally derived from calculations of single-particle trajectories in given \mathbf{E} and \mathbf{B} fields, similar to those invoked to discuss the relation between \mathbf{E} and \mathbf{V} (Sect. 4.1) and suffers from the same limitation: it may be a correct description of the relation between quantities under stress equilibrium but does not specify how the relation arises physically.
3. The current is described in conventional textbook fashion as arising from motion of a conductor through a magnetic field. This does lead in a straightforward way to the $\mathbf{V}_n \times \mathbf{B}$ term in Eq. (1), provided one assumes that the conductor is moving with the neutral-wind velocity \mathbf{V}_n . The problem is that in the ionosphere the conductor is the plasma, which in general does *not* move with \mathbf{V}_n .

Note that all three explanations make no reference to any spatial gradients; in principle, the processes as envisaged should produce a current equally well in a strictly homogeneous medium.

4.2.1 Initial-value development

As in the case of the relation between \mathbf{E} and \mathbf{V} (Sect. 4.1), the simplest way of identifying what produces what (and in what sequence) is to work through an idealized initial-value problem as a thought experiment. Consider a small quasi-homogeneous region (the qualitative explanations, as noted above, do not require spatial gradients) and assume that locally the electric field \mathbf{E} and the plasma bulk flow velocity \mathbf{V} (in the “fixed” frame of reference) as well as the current \mathbf{J} and the magnetic disturbance field $\delta \mathbf{B}$ are all zero at the initial time $t = 0$, but there is a non-zero neutral wind \mathbf{V}_n . For $t > 0$, \mathbf{V}_n is assumed to remain constant at its initial value, but the time histories of all other quantities are determined by the fundamental Eqs. (12), (14), (9), (10), which (with the simplifications mentioned in Sects. 3.1.2 and 3.1.3) are conveniently rewritten to show the time scales:

$$\frac{\partial \mathbf{J}}{\partial t} = \frac{\omega_p^2}{4\pi} \mathbf{E} + ne\mathbf{V} \times \hat{\mathbf{B}}\Omega_e - \mathbf{J} \times \hat{\mathbf{B}}\Omega_e - \nu_e \mathbf{J} + (\nu_{in} - \nu_{en})ne(\mathbf{V}_n - \mathbf{V}) \quad (34)$$

$$\frac{\partial \mathbf{V}}{\partial t} = \frac{\mathbf{J}}{ne} \times \hat{\mathbf{B}}\Omega_i + \nu_{in}(\mathbf{V}_n - \mathbf{V}) \quad (35)$$

$$\frac{\partial \mathbf{E}}{\partial t} = -4\pi \mathbf{J} + c\nabla \times \delta \mathbf{B} \quad (36)$$

$$\frac{\partial \delta \mathbf{B}}{\partial t} = -c\nabla \times \mathbf{E} . \quad (37)$$

Here $\omega_p = \sqrt{4\pi ne^2/m_e}$ is the plasma frequency, Ω_e and Ω_i are electron and ion gyrofrequencies (both defined as positive), and it is assumed that $|\delta \mathbf{B}| \ll |\mathbf{B}|$.

Note that only Eqs. (36) and (37) contain spatial gradients (with light travel time as the associated time scale); also, they are the only equations in which $\delta \mathbf{B}$ appears. If spatial gradients of all quantities are zero (assumption of local homogeneity) at $t = 0$, they remain zero (or negligibly small) for all $t <$ effective propagation time over the gradient scale length (i.e. departure from homogeneity); over the same time interval, an initially zero $\delta \mathbf{B}$ remains zero. The effective propagation speed cannot exceed the speed of light and is usually much slower; in the ionosphere, it is related to the Alfvén speed significantly modified by the presence of plasma-neutral collisions (Song et al., 2005).

At $t = 0$, the only non-zero quantity on the RH sides of all the equations is \mathbf{V}_n , hence initially from Eqs. (34) and (35)

$$\mathbf{J} \simeq (\nu_{in} - \nu_{en})t \mathbf{J}_0 \quad \mathbf{J}_0 \equiv ne\mathbf{V}_n \quad (38)$$

$$\mathbf{V} \simeq \nu_{in}t \mathbf{V}_n . \quad (39)$$

As soon as \mathbf{J} is non-zero it produces, according to Eq. (36), an increasing \mathbf{E} of opposite sign (this is a general consequence of Maxwell’s equations, including \mathbf{E} from charge accumulation as a special instance when $\nabla \cdot \mathbf{J} \neq 0$). The oppositely directed \mathbf{E} acts, according to Eq. (34), to reduce \mathbf{J} , resulting in oscillatory behavior of both, as can be seen by

eliminating \mathbf{E} between Eqs. (34) and (36) to obtain an equation for \mathbf{J} alone:

$$\begin{aligned} \frac{\partial^2 \mathbf{J}}{\partial t^2} + \omega_p^2 \left(\mathbf{J} - \frac{c}{4\pi} \nabla \times \delta \mathbf{B} \right) + \Omega_i \Omega_e \mathbf{J}_\perp \\ + (\nu_{in} - \nu_{en}) \Omega_i \mathbf{J} \times \hat{\mathbf{B}} + \Omega_e \frac{\partial \mathbf{J}}{\partial t} \times \hat{\mathbf{B}} + \nu_e \frac{\partial \mathbf{J}}{\partial t} \\ = ne \nu_{in} \left[(\nu_{en} - \nu_{in}) (\mathbf{V}_n - \mathbf{V}) + \Omega_e (\mathbf{V}_n - \mathbf{V}) \times \hat{\mathbf{B}} \right] \end{aligned} \quad (40)$$

(Eq. (35) has been used to eliminate the $\partial \mathbf{V} / \partial t$ terms). Equation (40) is clearly a wave equation, with frequency equal to ω_p modified by gyrofrequency terms (see, e.g. Vasyliūnas, 2001).

Averaged over the oscillations, \mathbf{J} is given by Eq. (40) as

$$\begin{aligned} \mathbf{J} = \frac{c}{4\pi} \nabla \times \delta \mathbf{B} + \frac{V_A^2}{c^2} \frac{\nu_{in}}{\Omega_i} \left(\mathbf{J}_1 \times \hat{\mathbf{B}} + \frac{\nu_{en} - \nu_{in}}{\Omega_e} \mathbf{J}_1 \right) \\ + \dots \quad \mathbf{J}_1 \equiv ne (\mathbf{V}_n - \mathbf{V}) \end{aligned} \quad (41)$$

to lowest order in V_A^2/c^2 and with neglected terms $+\dots$ of order m_e/m_i or ν_e/ω_p (the last term could also be neglected, since usually $(\nu_{en} - \nu_{in})/\Omega_e \ll 1$). Without the $\nabla \times \delta \mathbf{B}$ term (which is negligible in the absence of initial spatial gradients, as noted above), the averaged \mathbf{J} is smaller than $ne(\mathbf{V}_n - \mathbf{V})$ by a factor of order V_A^2/c^2 . The $\mathbf{J} \times \hat{\mathbf{B}}$ term in Eq. (35) is smaller than the collision term by the same factor, so that \mathbf{V} evolves according to

$$\mathbf{V} \simeq \mathbf{V}_n [1 - \exp(-\nu_{in} t)] \quad (42)$$

as long as spatial-gradient effects remain negligible. The averaged \mathbf{E} , obtained from Eq. (34), is

$$\mathbf{E} = -\frac{1}{c} \mathbf{V} \times \mathbf{B} - \frac{m_e}{e} (\nu_{in} - \nu_{en}) (\mathbf{V}_n - \mathbf{V}) + \dots \quad (43)$$

where the neglected terms $+\dots$ are of order V_A^2/c^2 . The second term on the RH side, dominant initially when $\mathbf{V} \simeq 0$, imparts to electrons (by direct electric force) the same acceleration that collisions with neutrals impart to plasma as a whole. The first term represents the electric field driven by plasma flow, as discussed in Sect. 4.1; it becomes dominant when \mathbf{V} has increased sufficiently, which occurs for $t \gg \sim \Omega_e^{-1}$ according to Eqs. (42) and (43).

The process of collision between a plasma and a neutral wind with different frequencies for ion-neutral and for electron-neutral collisions thus does indeed produce a current proportional to $\nu_{in} - \nu_{en}$ and in the direction of \mathbf{V}_n . In the absence of spatial gradients, however, this current lasts for only a very short time, of order ω_p^{-1} , after which the electric field resulting from the displacement-current term in Maxwell's equations has become sufficiently large, on the average, to accelerate the electrons to the flow velocity of the ions. Collisions with neutrals then continue to accelerate the plasma as a whole, producing an increasing bulk flow described by Eq. (42), with associated $-\mathbf{V} \times \mathbf{B}/c$ electric field after a time of order Ω_e^{-1} , but only a small and decreasing

current (determined by $\mathbf{J} \simeq -(1/4\pi) \partial \mathbf{E} / \partial t$), predominantly in the direction of $\mathbf{V}_n \times \mathbf{B}$ and proportional to ν_{in} rather than $\nu_{in} - \nu_{en}$ (cf. Eq. 41). If spatial gradients remain negligible, for $t \gg \nu_{in}^{-1}$ the plasma flow approaches the neutral flow \mathbf{V}_n (with *no* additional $\mathbf{V}_n \times \hat{\mathbf{B}}$ term, regardless of the ratio ν_{in}/Ω_i), and the current approaches zero; these limiting values are easily seen to be the unique steady-state solution of Eqs. (34)–(37) in the homogeneous limit.

4.2.2 Condition for permanent current

A permanent, non-transient electric current can thus be produced by a neutral wind only if spatial gradients are present from the outset. The starting point is Eq. (37): if \mathbf{E} given by Eq. (43) has non-zero curl because of spatial gradients in any quantities on the RH side, $\delta \mathbf{B}$ is produced, and if *it* has non-zero curl, \mathbf{J} given by Eq. (41) is modified, which in turn modifies \mathbf{V} through Eq. (35), which then further modifies \mathbf{E} , and so on. The process becomes one of wave propagation, similar to that discussed already in Sect. 3.2.4. The associated wave equation for $\delta \mathbf{B}$ can be derived (with the assumption that time scales now of interest are long compared to ω_p^{-1}) by differentiating Eqs. (37) and (43) with respect to time, inserting $\partial \mathbf{V} / \partial t$ from Eq. (35), and invoking Ampère's law, to obtain as the first step

$$\begin{aligned} \frac{\partial^2 \delta \mathbf{B}}{\partial t^2} + \nabla \times \left[V_A^2 (\nabla \times \delta \mathbf{B})_\perp \right] + \\ \nabla \times \nu_{in} [(\mathbf{V} - \mathbf{V}_n) \times \mathbf{B}] + \dots = 0 \end{aligned} \quad (44)$$

with neglected terms $+\dots$ of order $(\nu_{in} - \nu_{en})/\Omega_e$. As the second step, \mathbf{V} in Eq. (44) must be expressed in terms of $\delta \mathbf{B}$, which in general is complicated and frequency-dependent. If the time variations are slow compared to ν_{in}^{-1} , however, Eq. (35) (with $\partial \mathbf{V} / \partial t$ now neglected) can be invoked together with Ampère's law, yielding after some manipulation

$$\begin{aligned} \frac{\partial^2 \delta \mathbf{B}}{\partial t^2} + \nu_{in} \frac{\partial \delta \mathbf{B}}{\partial t} + \nu_{in} \nabla \times \left[(\nu_{in}^{-1}) V_A^2 (\nabla \times \delta \mathbf{B})_\perp \right] \\ + \dots = \nu_{in} \nabla \times (\mathbf{V}_n \times \mathbf{B}) + \dots \end{aligned} \quad (45)$$

The RH side of Eq. (45), the source term for $\delta \mathbf{B}$, indicates the nature of the required spatial gradient: if the neutral wind is to produce a permanent electric current (which presupposes $\nabla \times \delta \mathbf{B} \neq 0$), a necessary condition is

$$\nabla \times (\mathbf{V}_n \times \mathbf{B}) \neq 0 \quad (46)$$

somewhere within the dynamo region. (If there are no sources external to the ionosphere and $\nabla \times (\mathbf{V}_n \times \mathbf{B})$ is zero everywhere, any currents produced will be initial transients which die away.)

4.2.3 Condition for permanent current (from conventional approach)

Condition (46) for neutral-wind dynamo action can also be easily derived mathematically from the conventional

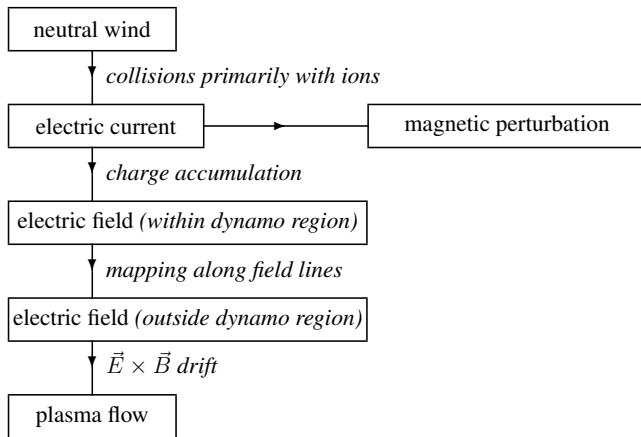


Fig. 3. Schematic diagram of how a neutral-wind dynamo develops, as understood in the conventional approach.

equations alone: note that if $V_n \times B/c = \nabla \Psi$ where Ψ is some scalar, Eq. (5) can be written as

$$\nabla \cdot \sigma \cdot \nabla (\Phi - \Psi) = 0 \tag{47}$$

which, for boundary conditions that represent no external input to the ionosphere ($J \cdot \hat{n} = 0$ at the boundaries, where \hat{n} is the normal to the boundary), has the solution $\Phi - \Psi = 0$, with $V_n \times B/c$ cancelled everywhere by an electrostatic field, hence producing no currents. Condition (46) is thus valid in the quasi-steady-state as well, independently of any initial-value arguments. The advantage of the derivation in Sect. 4.2.2 is to make clear its physical basis.

5 Neutral-wind dynamo

Neutral-wind dynamo models that are applied to the actual inhomogeneous ionosphere do, of course, necessarily contain spatial gradients from the beginning. The canonical problem specifies the neutral winds within some restricted range of altitudes (typically the E or else the F layer of the ionosphere) and then calculates electric field, current, and plasma flow patterns at all ionospheric altitudes (a series of simple specific examples is listed in Richmond and Thayer, 2000). The calculation proceeds from the conventional Eqs. (1), (3), (4), (6), (7) and can therefore be assumed to describe correctly the properties of a quasi-steady equilibrium configuration, but it does not describe the process by which the configuration has been established. It is nevertheless not uncommon (particularly in tutorial presentations, less so perhaps in research papers) to supplement the calculated model with a qualitative sketch of what is presumed to be the physical process.

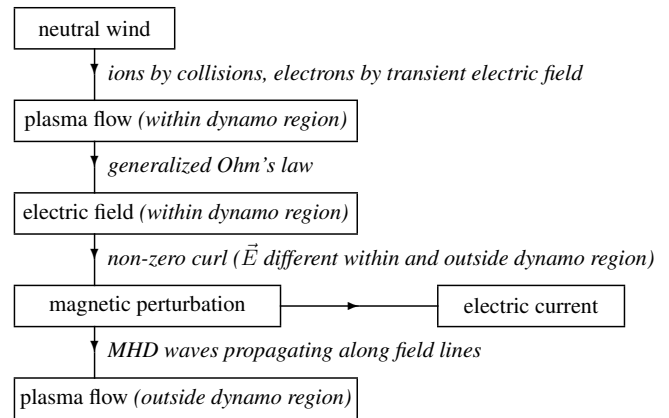


Fig. 4. Schematic diagram of how a neutral-wind dynamo develops, as understood on the basis of the complete physical equations.

5.1 Conventional description

A qualitative description of how a neutral-wind dynamo is established, according to the conventional approach, is sketched in Fig. 3. The assumed neutral wind within the dynamo region sets into flow, by collisions, ions but not electrons, thereby creating an electric current. If this current has a non-zero divergence, the resulting charge accumulation creates an electric field within the dynamo region, which is then extended to other regions by potential mapping along magnetic field lines. The $E \times B$ drift creates, together with collision effects, plasma bulk flow both within and outside the dynamo region. The magnetic perturbation can be calculated from the current (e.g. for comparison with observations) but otherwise plays no particular role in the model. In most discussions, time scales of the various steps in Fig. 3 are not mentioned (they are, of course, not specified by the conventional equations).

5.2 Physical description

As counterpoint to the conventional view of Sect. 5.1, the qualitative description deduced from the fundamental physical approach of Sect. 4.2 is sketched in Fig. 4. The assumed neutral wind within the dynamo region sets the plasma there as a whole into flow with $V = V_n$, on a time scale $\sim v_{in}^{-1}$ (after an initial transient on time scale $\sim \omega_p^{-1}$). The plasma flow creates the $-V \times B/c$ electric field, on a time scale between $\sim \omega_p^{-1}$ and $\sim \Omega_e^{-1}$. The spatial gradient of this electric field, specifically the difference between the fields within and outside the dynamo region, implies in general a non-zero $\nabla \times E$, which produces a magnetic disturbance field, the curl of which determines the electric current. The resulting Lorentz force changes the plasma flow, which further changes the electric field to continue the process. The final result is that the entire pattern of plasma flow and electromagnetic field changes propagates along the magnetic field;

in this way the plasma flow is established outside the dynamo region, on a time scale that is roughly the Alfvén wave travel time (albeit significantly modified by collision effects, cf. Song et al., 2005). The magnetic perturbation plays an essential role in the physical process.

This approach shows that what ultimately creates the (non-transient) dynamo current is an imbalance between the frictional force of plasma-neutral collisions, exerted by the neutral wind in the dynamo region, and the mechanical stresses exerted at other locations along the magnetic field line. If a quasi-steady state exists, the mechanical and the magnetic stresses must be in balance at each point, which requires an appropriate deformation of the magnetic field, and the curl of that deformation is what constitutes the current. Without opposing stresses elsewhere, the neutral wind would simply carry the plasma with it, giving

$$\mathbf{V} = \mathbf{V}_n \quad c\mathbf{E} = -\mathbf{V}_n \times \mathbf{B} \quad \mathbf{J} = 0 \quad (48)$$

which is a solution of Eqs. (1), (3), (4), (6), and (7). To have dynamo action ($\mathbf{J} \neq 0$, by definition) Eq. (48) must *not* be a valid solution, which requires that either (1) $\nabla \times (\mathbf{V}_n \times \mathbf{B}) \neq 0$ (noted already in Sects. 4.2.2 and 4.2.3), or (2) the boundary conditions on Φ are incompatible with $\mathbf{J} = 0$, or both. The first implies that the imposed plasma flow deforms the magnetic field and creates stresses within the ionosphere; the second implies that there are stresses exerted on the magnetic field outside the ionosphere.

6 Conclusions

The conventional equations of ionospheric electrodynamics are obtained by neglecting all acceleration terms in the momentum equations and all time derivatives in Maxwell's equations; only slow time variations are considered, e.g. from varying external influences, or from density or temperature profile changes due to aeronomic processes. The equations state how the various quantities must be related numerically *if* a quasi-steady-state stable stress balance exists; they say nothing about how the quantities are related causally, or how the stress equilibrium is established and on what time scales.

Despite the above limitation, an extensive folklore appears to have grown up within the ionospheric community, supplying qualitative or even semiquantitative descriptions of causal relations and temporal sequences that are supposed to be the underlying physics of the conventional equations – descriptions derived mostly from ordinary E&M arguments. A more rigorous treatment, however, starting from the complete fundamental equations, shows that the regime of ionospheric, magnetospheric, space, and astrophysical plasmas can be quite different from the ordinary E&M environment (e.g. Parker, 2007; Vasylūnas, 2011, and references therein). The essential difference is the large number (because of the

large spatial scales) of free charged particles and the consequent overriding importance of self-consistency between their distributions and the electromagnetic fields, described by Maxwell's equations.

For the electrodynamics of the ionosphere, the main differences between the conventional approach and the more rigorous physical approach are the following:

1. Electric fields do not produce plasma flows; they are a consequence of plasma flow and other terms in the generalized Ohm's law.
2. The electric current is determined, in a quasi-steady state, by the requirement that the magnetic stress balance the mechanical stress (specifically for the ionosphere, the collisional friction between plasma and neutrals).
3. There is no "mapping" process along magnetic field lines; changes of plasma flow and electric field are propagated by appropriate (mostly MHD) waves.
4. Neutral winds do not create electric currents directly; they create plasma motions, which then deform the magnetic field, and the current arises to match the curl of the deformed field.
5. Conventionally, ionospheric electrodynamics is viewed as a problem of electrostatic/magnetostatic balance, arising out of mechanical stresses. In reality it is more a problem of mechanical stress balance, mediated by electromagnetic fields.

Although the two viewpoints usually give essentially the same results in many cases where quasi-steady equilibrium exists, the radically different understanding of the underlying physics may imply different predictions for shorter time scales and more detailed phenomena. (For modeling of some specific aspects, see, e.g. Song et al., 2009; Tu et al., 2011).

Acknowledgements. Much of the research underlying this paper was done in collaboration with Paul Song at the Center for Atmospheric Research, University of Massachusetts Lowell. I am grateful to Tim Fuller-Rowell, Rod Heelis, Bela Fejer, and George Siscoe for useful discussions and to Alex Dessler and the referees for penetrating comments that led to significant clarification.

The service charges for this open access publication have been covered by the Max Planck Society.

Topical Editor K. Kauristie thanks two anonymous referees for their help in evaluating this paper.

References

- Boström, R.: Electrodynamics of the ionosphere, in: *Cosmical Geophysics*, edited by: Egeland, A., Holter, Ø, and Ohmholt, A., 181–192, Universitetsforlaget, Oslo, 1973.

- Buneman, O.: Internal dynamics of a plasma propelled across a magnetic field, *IEEE Trans. Plasma Sci.*, 20, 672–677, 1992.
- Fuller-Rowell, T. J. and Schrijver, C. J.: On the ionosphere and chromosphere, in: *Heliophysics: Plasma Physics of the Local Cosmos*, edited by: Schriver, C. J. and Siscoe, G. L., 324–359, Cambridge University Press, N.Y., 2009.
- Fuller-Rowell, T. J., Rees, D., Quegan, S., Moffett, R. J., Codrescu, M. V., and Millward, G. H.: A coupled thermosphere-ionosphere model (CTIM), in: *STEP Handbook on Ionospheric Models*, edited by: Schunk, R. W., 217–238, Utah State Univ., Logan, 1996.
- Greene, J. M.: Moment equations and Ohm's law, *Plasma Phys.*, 15, 29–36, 1973.
- Heelis, R. A.: Electrodynamics in the low and middle latitude ionosphere: a tutorial, *J. Atmos. Solar-Terr. Phys.*, 66, 825–828, 2004.
- Holzer, T. E. and Reid, G. C.: The response of the day side magnetosphere-ionosphere system to time-varying field line reconnection at the magnetopause 1. Theoretical model, *J. Geophys. Res.*, 80, 2041–2049, 1975.
- Kan, J. R., Longenecker, D. U., and Olson, J. V.: A transient response model of Pi2 pulsations, *J. Geophys. Res.*, 87, 7483–7488, 1982.
- Kelley, M. C.: *The Earth's Ionosphere*, Academic Press, New York, 1989.
- Lysak, R. L.: Magnetosphere-ionosphere coupling by Alfvén waves at midlatitudes, *J. Geophys. Res.*, 109, A07201, doi:10.1029/2004JA010454, 2004.
- Matsushita, S.: Solar quiet and lunar daily variation fields, in: *Physics of Geomagnetic Phenomena*, Vol. 1, edited by: Matsushita, S. and Campbell, W. H., 301–424, Academic Press, New York, 1967.
- Parker, E. N.: *Conversations on Electric and Magnetic Fields in the Cosmos*, Princeton University Press, Princeton, New Jersey, 2007.
- Richmond, A. D. and Thayer, J. P.: Ionospheric dynamics: A tutorial, in: *Magnetospheric Current Systems*, edited by: Ohtani, S.-I., Fujii, R., Hesse, M., and Lysak, R. L., 131–146, *Geophysical Monograph 118*, American Geophysical Union, Washington, D.C., 2000.
- Richmond, A. D., Ridley, E. C., and Roble, R. G.: A thermosphere/ionosphere general circulation model with coupled electrodynamics, *Geophys. Res. Lett.*, 19, 601–604, 1992.
- Ridley, A. J., Deng, Y., and Toth, G.: The global ionosphere-thermosphere model, *J. Atmos. Solar-Terr. Phys.*, 68, 839–864, 2006.
- Rishbeth, H.: The ionospheric E-layer and F-layer dynamos – a tutorial review, *J. Atmos. Solar-Terr. Phys.*, 59, 1873–1880, 1997.
- Rishbeth, H. and Garriott, O. K.: *Introduction to Ionospheric Physics*, Academic Press, New York, 1969.
- Rossi, B. and Olbert, S.: *Introduction to the Physics of Space*, Chapters 10 and 12, McGraw-Hill, New York, 1970.
- Song, P., Gombosi, T. I., and Ridley, A. J.: Three-fluid Ohm's law, *J. Geophys. Res.*, 106, 8149–8156, 2001.
- Song, P., Vasyliūnas, V. M., and Ma, L.: Solar wind-magnetosphere-ionosphere coupling: Neutral atmosphere effects on signal propagation, *J. Geophys. Res.*, 110, A09309, doi:10.1029/2005JA011139, 2005.
- Song, P., Vasyliūnas, V. M., and Zhou, Z.-X.: Magnetosphere-ionosphere/thermosphere coupling: Self-consistent solutions for a one-dimensional stratified ionosphere in three-fluid theory, *J. Geophys. Res.*, 114, A08213, doi:10.1029/2008JA013629, 2009.
- Tu, J., Song, P., and Reinisch, B. W.: On the concept of penetration electric field, *AIP Conf. Proc.*, 974, 81–85, doi:10.1063/1.2885036, 2008.
- Tu, J., Song, P., and Vasyliūnas, V. M.: Ionosphere/thermosphere heating determined from dynamic magnetosphere-ionosphere/thermosphere coupling, *J. Geophys. Res.*, 116, A09311, doi:10.1029/2011JA016620, 2011.
- Vasyliūnas, V. M.: Electric field and plasma flow: What drives what?, *Geophys. Res. Lett.*, 28, 2177–2180, 2001.
- Vasyliūnas, V. M.: Time evolution of electric fields and currents and the generalized Ohm's law, *Ann. Geophys.*, 23, 1347–1354, doi:10.5194/angeo-23-1347-2005, 2005a.
- Vasyliūnas, V. M.: Relation between magnetic fields and electric currents in plasmas, *Ann. Geophys.*, 23, 2589–2597, doi:10.5194/angeo-23-2589-2005, 2005b.
- Vasyliūnas, V. M.: The Ptolemaic approach to ionospheric electrodynamics, Abstract SA11A-1558, presented at 2010 Fall Meeting, AGU, San Francisco, Calif., 13–17 December 2010.
- Vasyliūnas, V. M.: Physics of magnetospheric variability, *Space Sci. Rev.*, 158, 91–118, doi:10.1007/s11214-010-9696-1, 2011.
- Vasyliūnas, V. M. and Song, P.: Meaning of ionospheric Joule heating, *J. Geophys. Res.*, 110, A02301, doi:10.1029/2004JA010615, 2005.
- Volland, H.: Electrodynamic coupling between neutral atmosphere and ionosphere, in: *Modern Ionospheric Science*, edited by: Kohl, H., Ruster, R., and Schlegel, K., 102–135, European Geophysical Society, Katlenburg-Lindau, Germany, 1996.
- Wright, A. N.: Transfer of magnetosheath momentum and energy to the ionosphere along open field lines, *J. Geophys. Res.*, 101, 13169–13178, 1996.