

MHD Simulation of the Magnetosphere

Antonius Otto, Geophysical Inst., Univ. Alaska, Fairbanks

October 20, 2004

Goals:

- Simulation of basic magnetospheric processes.
- Plasma and energy entry from the solar wind into the magnetosphere.
- Energy release and energy dissipation in the magnetosphere/ionosphere system.



Woodcut of Aurora.

Topics:

- Physics of the magnetosphere
 - Basic plasma physics and magnetic reconnection
 - Structure of the magnetosphere
 - Processes at the magnetospheric boundary: Magnetic reconnection and KH modes
 - Magnetotail properties: Equilibrium and magnetospheric substorms
 - Magnetosphere-Ionosphere Coupling
 - Inner magnetosphere and magnetic storms
- Numerical simulation
 - Different types of problems
 - Discrete representation
 - Stability and accuracy
 - Boundary and initial conditions
 - Efficiency
 - Computer performance and architecture
 - Parallelization
- Simulation package

1 Physics of the Magnetosphere

1.1 Structure of the Magnetosphere

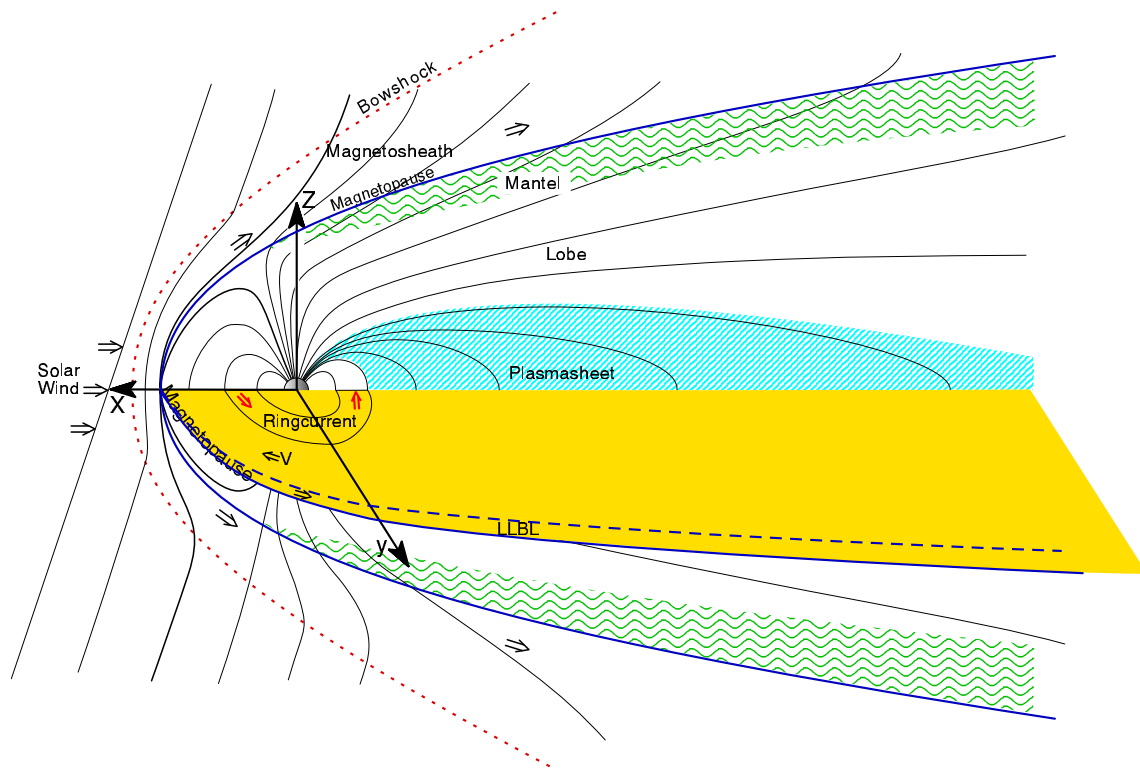


Figure 1: Sketch of the structure of the magnetosphere.

The magnetosphere: Large plasma cavity generated by the Earth's magnetic field and the solar wind plasma. The streaming solar wind compresses the dayside portion of the Earth's field and generates a tail which is many hundreds of Earth radii long. The basic mechanism for the formation of the magnetosphere: Magnetic dipole exposed to a stream of charged particles. The entire magnetosphere is subject to only two boundary conditions, explicitly the boundary between the magnetosphere on the streaming solar wind and the boundary of the plasma in the ionosphere. The basic elements of the magnetosphere are

- **The Bow Shock and the Magnetosheath**
- **The Magnetopause**
- **The Magnetotail - Magnetic Substorms**
- **The Inner Magnetosphere - Magnetic Storms**
- **Magnetosphere - Ionosphere Coupling**

Detailed properties

- **The Bow Shock and the Magnetosheath**

- While not part of the magnetosphere proper the magnetosheath is an outer layer embedding the magnetosphere. The solar wind plasma travels usually at super-fast speeds relative to the magnetosphere. Therefore a standing shock wave forms around the magnetosphere just as in front of an aircraft traveling at supersonic speeds. The bow shock is the shock in front of the magnetosphere and the magnetosheath is the shocked solar wind plasma. Therefore it is not directly the solar wind plasma which constitutes the boundary of the magnetosphere but the strongly heated and compressed plasma behind the bowshock. The region is rich in various wave phenomena, boundaries and shocks are often treated as discontinuities.

- **The Magnetopause**

- The magnetopause is the actual boundary between the shocked solar wind and the magnetospheric plasma. However, the magnetosphere is not closed in terms of the magnetic field but there is considerable magnetic flux crossing the magnetopause. Thus it is not easy to define this boundary precisely. Also the boundary does permit a certain amount of solar wind plasma entry. This entry is easier along magnetic field lines. The magnetopause is an highly important region because the physical processes at this boundary control the entry of plasma, momentum, energy and the redistribution of geomagnetic flux. Essential instabilities are reconnection (tearing modes) and Kelvin Helmholtz modes in addition to various micro-instabilities.

- **The Magnetotail**

- The magnetotail is the long tail-like extension of the magnetosphere on anti-sunward side of the magnetosphere. since the magnetic field points toward the Earth in the northern lobe and away in the southern lobe there is a current in the westward direction. Because of its structure there is considerable energy stored in the magnetic field in the magnetotail. During magnetically quiet times convection is typically low and energy in the plasma flow is only a tiny fraction of the overall energy density.
- **Magnetospheric Activity:** Two basic types of geomagnetic activity:
 - * Magnetic storms
 - * Magnetospheric substorms
- This terminology is misleading in that substorms are not small storms. Rather a storm can consist of several substorms but also of quiet periods.
- A substorm is characterized by a specific large scale auroral intensifications and corresponding magnetic perturbations at high latitudes typically close to magnetic midnight. Characteristic are also fast flows in the magnetotail, plasma ejection in the tailward direction, release of energy stored in the lobe magnetic field, energetic particle injections at geosynchronous distances, and strong intensifications of field-aligned current systems. A substorm

consists of the growth phase, the (auroral) expansion phase, and the recovery phase. Substorms are clearly related to periods of southward interplanetary magnetic field (IMF) which leads to reconnection on the dayside, transport of magnetic flux from the dayside to the tail, and storage of magnetic energy in the tail during the growth phase. There is a rapid release of magnetic energy which is partly deposited in the the ionosphere through Joule heating and precipitation and in part ejected tailward through so-called plasmoids.

- **The inner Magnetosphere**

- The inner magnetosphere is different from most of the magnetosphere in that the magnetic field is mostly dipolar and perturbations of the field are small compared to the average dipole field. However, there can still be large amounts of energy stored in this region in particular during so-called storm times. During such times the ring current (current due to gradient curvature drifts of charged particles) intensifies strongly and is responsible for strong magnetic perturbations at low geomagnetic latitudes on the Earth.
- A storm is a large and long duration perturbation of the magnetosphere which leads to a strong compression and a contraction of the magnetosphere. Characteristic is a strong amplification of the ring current and the associated magnetic field measured at equatorial latitudes. Aurora is typically visible at much lower latitudes. Storms are associated with larger solar flares and/or coronal mass ejections and storm durations are many hours.

- **Magnetosphere - Ionosphere Coupling**

- The ionosphere is the region where the atmosphere is partially ionized and plasma and neutrals strongly interact. This interaction exerts a drag on the plasma. The plasma density can be very high but also strongly variable such that the ionospheric conductance can vary by orders of magnitude. Magnetospheric plasma motion is transmitted into the ionosphere and forces ionospheric convection. This also implies the existence of strong currents along magnetic field lines which close through the ionosphere. In particular at high latitudes these currents lead to magnetic perturbations during times of strong magnetospheric activity (fast convection and changes of the magnetospheric configuration).

1.2 Planetary Magnetospheres

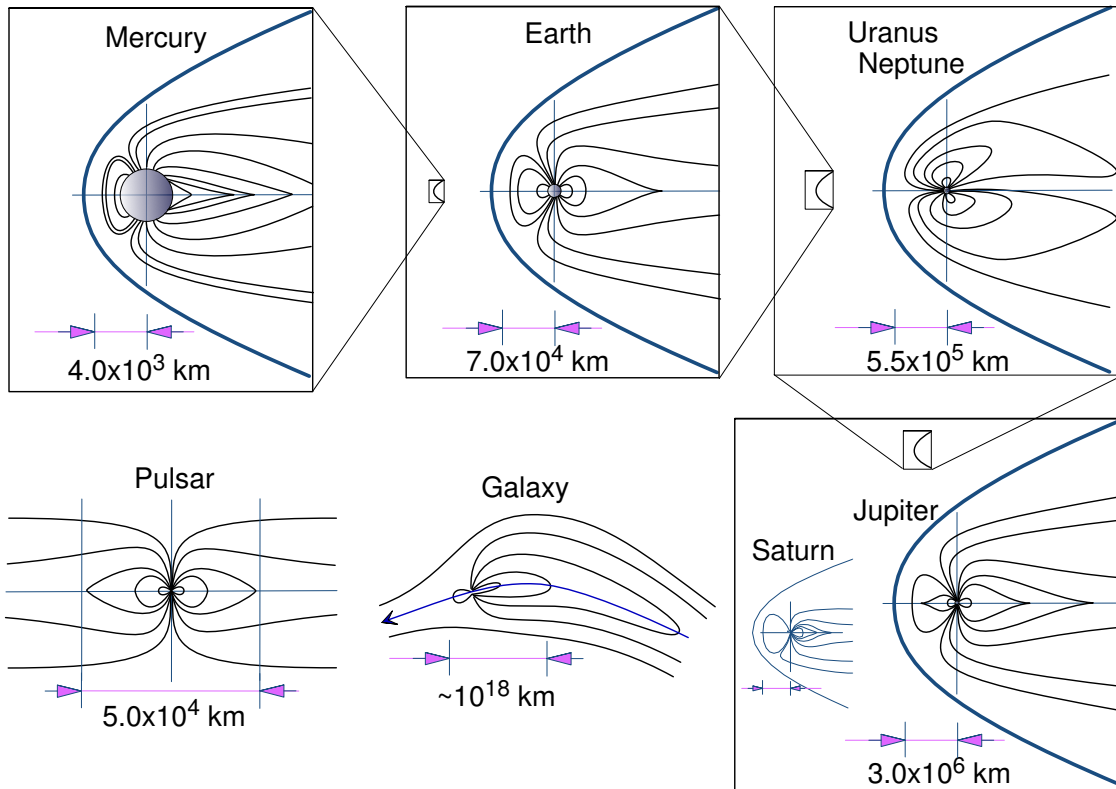


Figure 2: Comparison of different planetary magnetospheres.

- Terrestrial Magnetosphere is not unique!
- Any planet with a sufficient magnetic field exposed to plasma stream generates a magnetosphere

	Distance (AU)	Radius R_P (10^3 km)	Magn. Moment (ME)	Tilt	Polarity	Spin	MP Dist. (R_P)
Sun	0	700	2.2×10^{-9}	var	N	27d	-
Mercury	0.4	2.49	5.6×10^{-4}	10°	N	58.6d	1.6
Earth	1.0	6.37	1	11.5°	N	24h	11
Mars	1.5	3.38	2×10^{-4}	$<20^\circ$	R	24.5h	1.4
Jupiter	5.2	71.4	20,000	10°	R	10h	50
Saturn	9.5	60.4	580	1°	N	10.65h	21
Uranus	19.2	23.8	49	59°	N	17.3h	27
Neptune	30.0	22.2	27	47°	R	16h	26

AU = Astronomical Unit = 1.5×10^8 km, ME = Earth's dipole moment = 8.05×10^{22} Amp m², N = normal polarity, R = reverse polarity, d = day, h = hour.

1.3 Basic plasma physics and magnetic reconnection

1.3.1 Two-fluid and MHD equations

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \rho \mathbf{u} \quad (1)$$

$$\frac{\partial \rho \mathbf{u}}{\partial t} = -\nabla \cdot (\rho \mathbf{u} \mathbf{u}) - \nabla p + \mathbf{j} \times \mathbf{B} \quad (2)$$

$$\mathbf{E} + \mathbf{u} \times \mathbf{B} = \frac{m_e m_i}{e^2 \rho} \left[\frac{\partial \mathbf{j}}{\partial t} + \nabla \cdot (\mathbf{u} \mathbf{j} + \mathbf{j} \mathbf{u}) \right] - \frac{M}{e \rho} \nabla p_e + \frac{m_i}{e \rho} \mathbf{j} \times \mathbf{B} + \eta \mathbf{j} \quad (3)$$

$$\frac{\partial p}{\partial t} = \nabla \cdot p \mathbf{u} - (\gamma - 1) (p \nabla \cdot \mathbf{u} + \eta \mathbf{j}^2) \quad (4)$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} \quad (5)$$

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E} \quad (6)$$

1.3.2 Properties of the Two-Fluid and MHD equations:

The MHD equations:

- Conservation of mass, momentum, and energy.
- Valid on scales larger than the ion inertia scale c/ω_{pi} .
- Ideal MHD equation no intrinsic physical length scale => Self-similarity $\tau_{typ} = L_{typ}/u_A$ (with $u_A = B_0/\sqrt{\mu_0\rho_0}$). For a system which is identical except that it is 10 times larger the length scale is $10L_0$ and the time scale is $10\tau_0$. Thus a simple re-normalization yields exactly the same dynamics.

Frozen-in condition: Ideal MHD assumes Ohm's law of the form $\mathbf{E} + \mathbf{u} \times \mathbf{B} = 0$.

=> Plasma frozen to the magnetic field

Magnetic flux through the surface C is the surface integral

$$\Phi_C = \int_C \mathbf{B} \cdot d\mathbf{s}$$

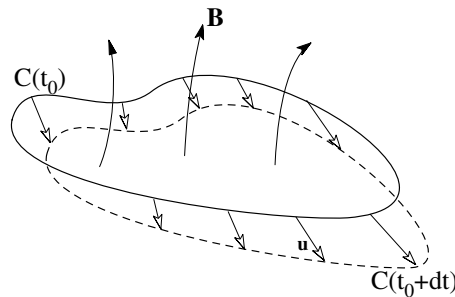


Figure 3: Illustration of the frozen-in condition.

Ideal Ohm's law =>

$$\frac{d\Phi_C}{dt} = 0$$

- The magnetic flux through any closed contour which moves with the flow \mathbf{u} is constant over time!
- Equivalent: Two fluid elements are always connected by a magnetic field line if they were connected at one time by a field line (defined by the direction of the magnetic field at any moment in time).
- Magnetic field line can be identified by fluid plasma elements (line conservation).

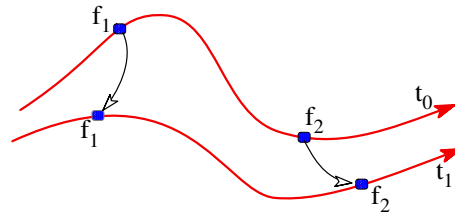


Figure 4: Illustration of line conservation

A more complete form of Ohm's law

$$\mathbf{E} + \mathbf{u}_e \times \mathbf{B} = -\frac{M}{e\rho} \nabla p_e \quad (7)$$

=> Frozen-in condition applies to the electron fluid

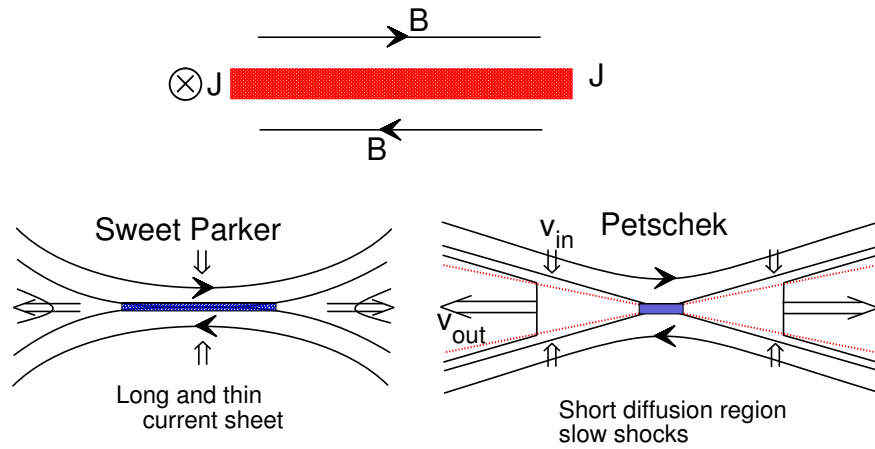
Definitions:

Steady state assumes time stationary solutions with nonzero velocity, $\partial/\partial t = 0$ and $\mathbf{u} \neq 0$.

Equilibrium solutions assume $\partial/\partial t = 0$ and $\mathbf{u} = 0$. Note that for kinetic systems the velocity in phase space is always nonzero for physical systems. Also the electron velocity is nonzero in current regions.

Electrostatic solutions assume $\partial\mathbf{B}/\partial t = 0$. This implies $\nabla \times \mathbf{E} = 0$ or $\mathbf{E} = -\nabla\phi$. In this case Ohm's law must be replaced by the Coulomb equation.

1.3.3 Reconnection models



- Stationary magnetic reconnection $E_z = const = E_r \Rightarrow \partial B / \partial t = 0$
 - Diffusion region: $E_r = \eta j_z$
 - Inflow region: $E = v_{in} B_0 = E_r$
 - Outflow region: $E = v_{out} B_{out} = E_r$ with $v_{out} = v_{A0}$
 - Aspect ratio of the diffusion region: $\epsilon = d/L$
 - Continuity equation: $v_{in} = \epsilon v_{out}$ and $E_r = \epsilon E_0$
 - with $E_0 = v_{A0} B_0$
- Sweet-Parker (1957):
 - Diffusion region through resistive diffusion of the magnetic field
 - $\Rightarrow E_r = \eta^{1/2} E_0 \Leftrightarrow$ long and thin diffusion region \Rightarrow slow reconnection
- Petschek (1963)
 - slow shocks \Rightarrow short diffusion region = fast reconnection, $\epsilon \approx 0.1$
- No selfconsistent analytic treatment of the diffusion region \Rightarrow numerical models

1.4 The Magnetospheric Boundary

1.4.1 The solar wind plasma

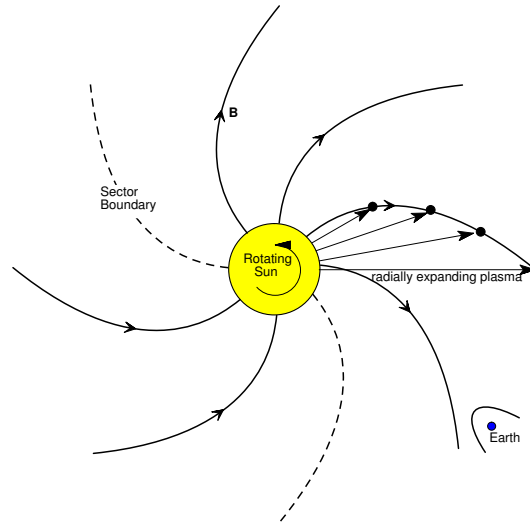


Figure 5: Illustration of spiral and sector structure of the solar wind.

Fast mode speed: $c_f = \sqrt{(B^2 + \gamma p) / \mu_0 m_i n} \approx 60 \text{ km/s} \ll v_{sw}$. The solar wind is much faster than the MHD waves in the frame of the Earth.

=> Shock in front of the magnetosphere - the bow shock

- deceleration and compression of the solar wind plasma to 'sub-fast' velocities
- Shocked plasma between the bow shock and the magnetosphere called magnetosheath

1.4.2 The Magnetopause:

The Magnetopause = Boundary between the magnetosphere and the shocked solar wind.

Problem: Solar wind plasma entry - Is the region where solar wind plasma enters not part of the magnetosphere?

More practical definition: The magnetopause is the region of highest current density and a mixture of particle of magnetospheric and magnetosheath origin.

Magnetic Topology: 3 different types of magnetic connection for a field line

- Closed geomagnetic field lines (or magnetic flux) have both 'end' points in the Earth
- Open magnetic field lines have one footpoint on the Earth and connect with the other side to the solar wind
- IMF field lines are not connected to the Earth

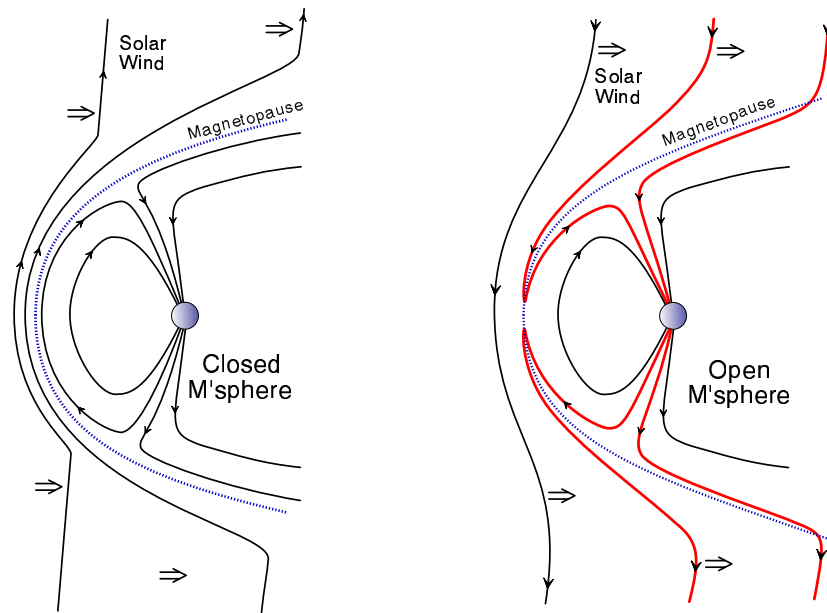


Figure 6: Sketch of closed and open magnetospheric configurations

Why is the magnetopause important?

- The MP controls the transport of mass, momentum, energy, and magnetic flux into the magnetosphere.
- The magnetic topology and the change of the magnetic topology are of major importance for the large scale transport (particles, momentum, energy) into the magnetosphere.

Can mass, momentum, energy be transmitted through a closed boundary into the magnetosphere?

- Mass: No, only very energetic testparticles can enter the magnetosphere through nonadiabatic motion.
- Momentum and energy: Yes, via waves.

Basic Processes:

- Viscous interaction: Momentum coupling through microinstabilities or Kelvin Helmholtz modes
- Magnetic reconnection: Generation of open magnetic field
- Pressure Pulses and impulsive penetration: Energy, momentum, mass transfer through SW perturbations

1.4.3 Basic Magnetopause Structure

The magnetopause as a MHD discontinuity $u_n = const$:

- closed magnetopause: $B_n = 0 \Rightarrow$ tangential discontinuity
- open magnetopause: $B_n = const \neq 0 \Rightarrow$ rotational discontinuity

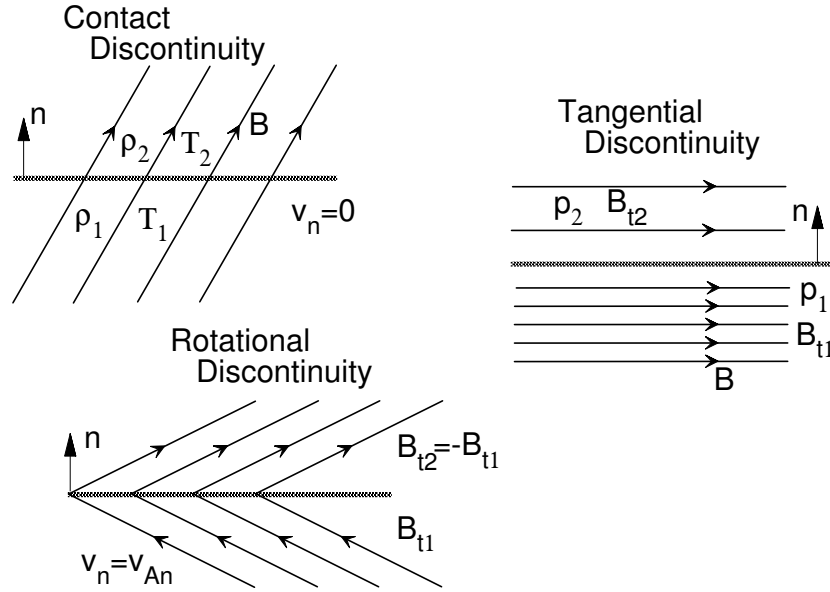


Figure 7: MHD discontinuities.

(a) **Tangential Discontinuity:** $u_n = 0, B_n = 0 \Rightarrow \left[p + \frac{B_t^2}{2\mu_0} \right] = 0$

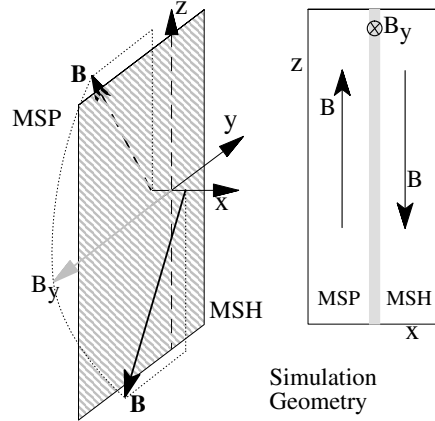
(b) **Rotational Discontinuity:** $u_n = const \Rightarrow [\rho] = 0, \left[p + \frac{B_t^2}{2\mu_0} \right] = 0$

$$\Rightarrow u_n^2 = u_{An}^2 \text{ and } [u_t] = \frac{u_{An}}{B_n} [B_t]$$

$$\text{and } B_{td} = -B_{tu}, p_d = p_u \text{ (or trivial solution } B_{td} = B_{tu}), [u_t] = \pm [v_{At}]$$

1.4.4 Simulation of dayside reconnection:

Dayside Magnetic Reconnection: Initial configuration



Simulation parameters

	MSH	MSP
B	50 nT	50 nT
n	1.9 cm^{-3}	13 cm^{-3}
β	1	1
v_A	800 km/s	250 km/s

Rot angle for $B = 130^\circ$

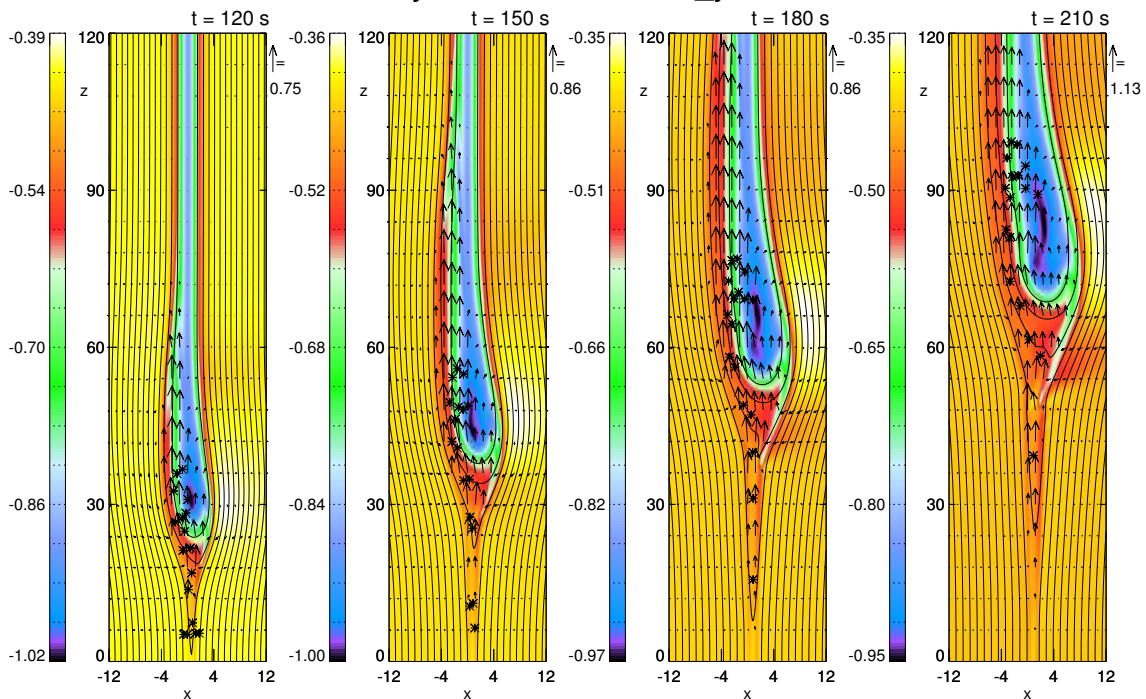
$L_0 = 400 \text{ km}$

$\langle v_A \rangle = 400 \text{ km}$

$t_0 = 1 \text{ s}$

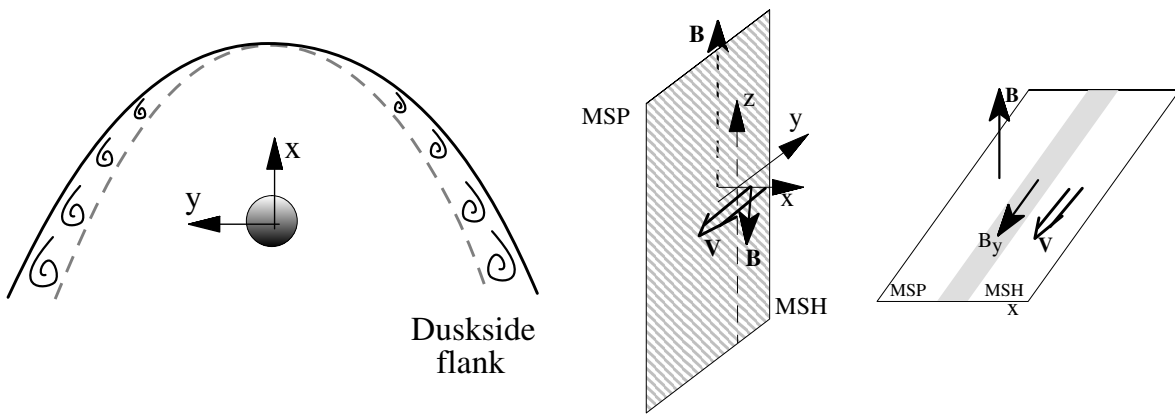
Single (bursty) X line reconnection:

Velocity, Field Lines and B_y



1.4.5 Viscous Interaction

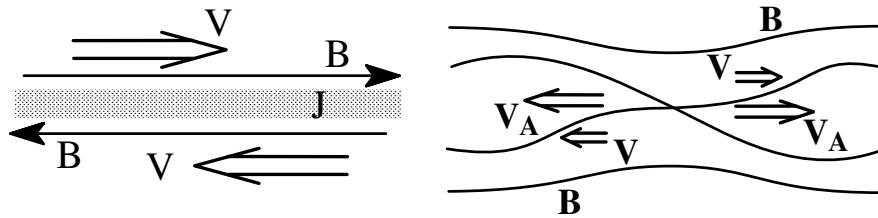
- Momentum diffusion at the Low Latitude Boundary Layer (LLBL)
- Mechanisms:
 - Microinstabilities: Lower hybrid drift, ion acoustic, ion cyclotron, etc
 - Kelvin Helmholtz modes
- Boundary layer width $\approx 0.5R_E$
- Diffusion coefficient: $D = 10^9 m^2 s^{-1}$



1.4.6 Kelvin Helmholtz vs magnetic reconnection

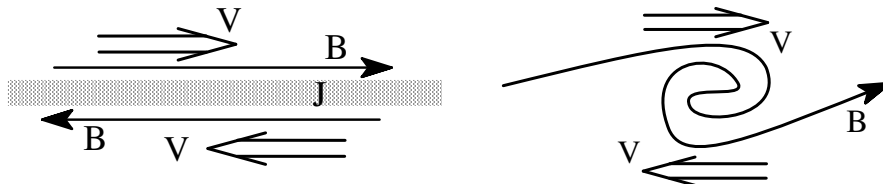
Magnetic Reconnection:

- Transport of plasma, momentum, energy, and magnetic flux
- Requires large anti-parallel magnetic field components $\Delta V_A > \Delta v$



Kelvin-Helmholtz mode:

- Ideal instability => Transport of momentum and energy (viscous coupling)
- Requires $\Delta v > V_{A,typ}$ along the k vector of the instability

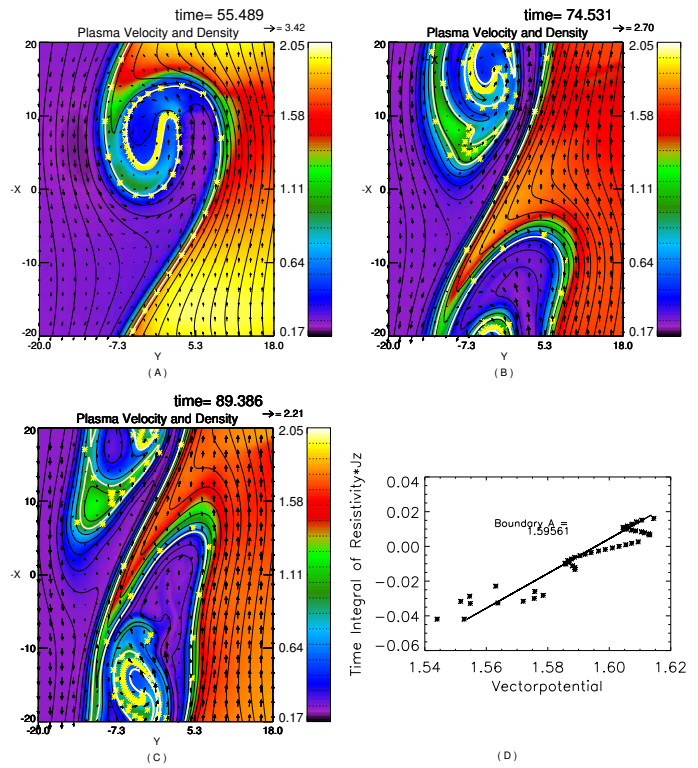
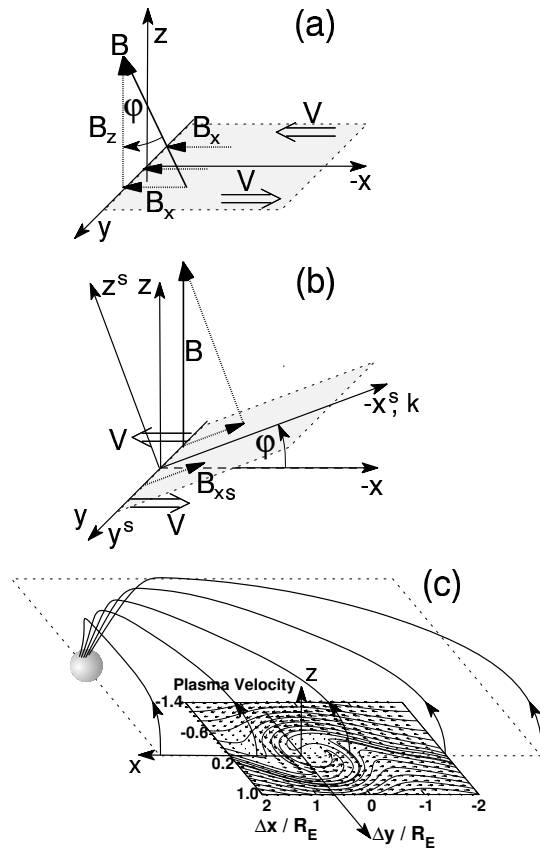


Combination of tearing/reconnection with KH (Vortex induced reconnection) in 2D??

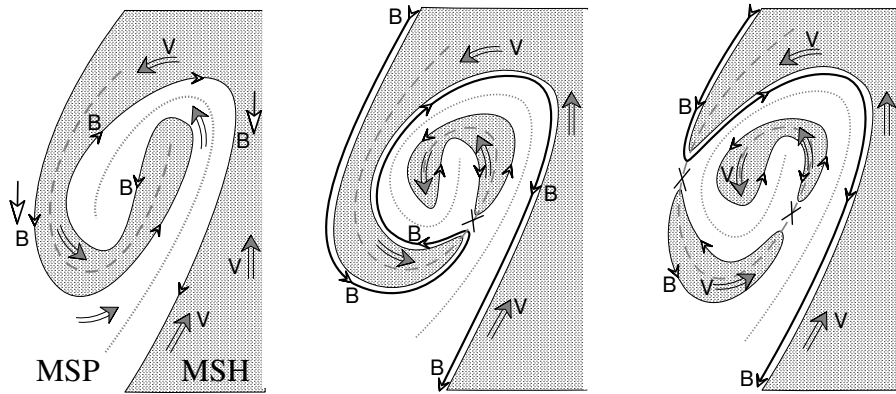
- Requires $\Delta v > V_{A,typ}$ for KH and $\Delta V_A > \Delta v$ for reconnection
- For strongly northward IMF: ΔB is small along the flanks of the magnetosphere

1.4.7 Simulation of Kelvin-Helmholtz modes and reconnection

Initial Geometry:

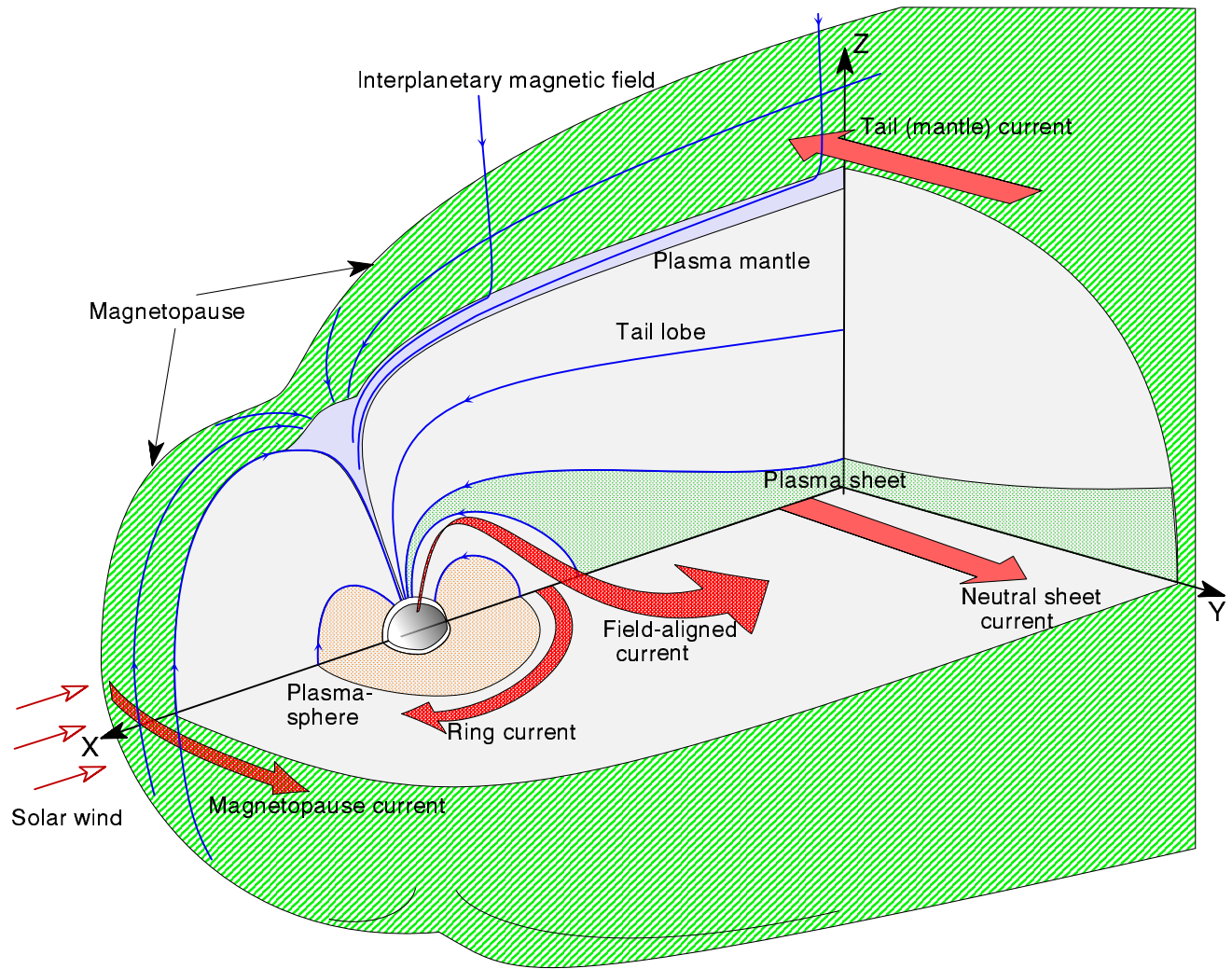


Reconnection in KH vortices:



=> mass diffusion coefficient $D = 10^9 m^2 s^{-1}$

1.5 The Magnetotail

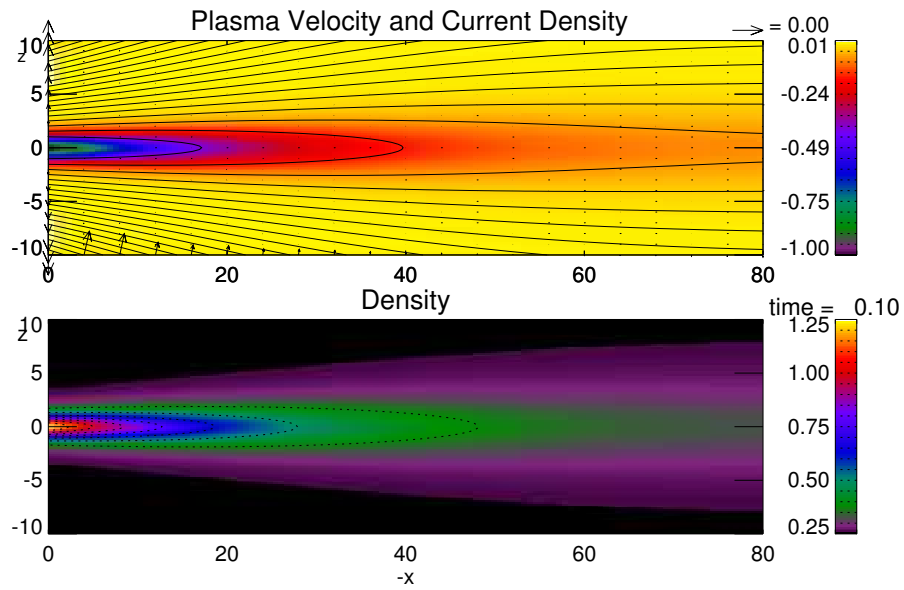


1.5.1 Magnetotail models

1.5.2 Equilibrium Configuration

Tail-like equilibrium: Modified Harris sheet (Harris (1960?))

$$\begin{aligned} A_y &= A_c \ln \cosh(z/l(x)) + f(x) \\ B_x &= -\partial A_y / \partial z = B_0(x) \tanh(z/l(x)) \\ B_z &= \partial A_y / \partial x \end{aligned}$$



1.5.3 Convection in the Magnetotail

Stationary convection $\partial/\partial t = 0$

$$\frac{\partial \mathbf{B}}{\partial t} = 0 \Rightarrow \nabla \times \mathbf{E} = 0 \text{ or } \mathbf{E} = -\nabla \phi$$

Southward IMF: $\mathbf{v}_{sw} = -v_0 \mathbf{e}_x$, $\mathbf{B}_{sw} = -B_0 \mathbf{e}_z \Rightarrow \mathbf{E} = -\mathbf{v} \times \mathbf{B} = v_0 B_0 \mathbf{e}_y = \text{const}$

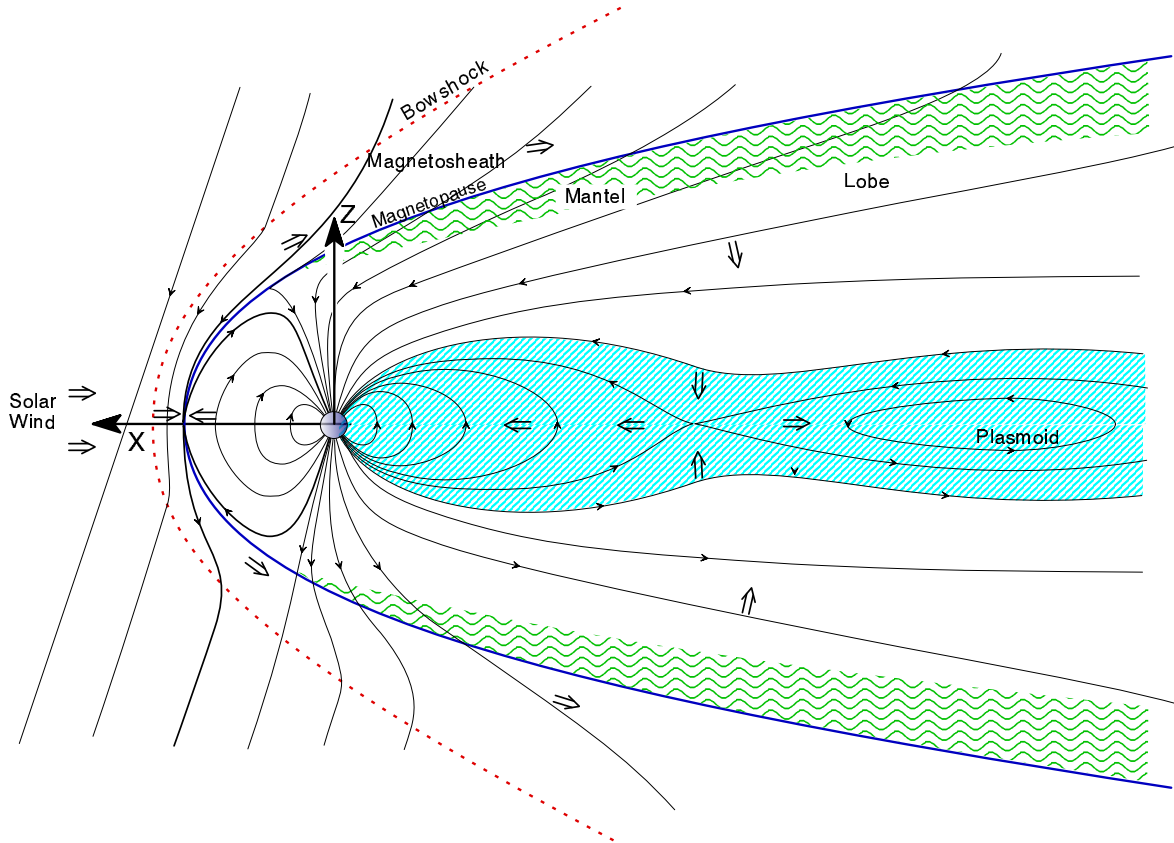


Figure 8: Convection in the magnetosphere for southward IMF.

- Plasma flow toward the plasma sheet in the tail.
- Earthward jetting in the plasma sheet.
- Plasmoid ejection (but tailward end of the plasmoid should move earthward if the model were accurate).
- Flow toward the dayside magnetopause.

ii) Convection in the inner magnetosphere due to co-rotation.

$\mathbf{v} = \omega_E \mathbf{e}_\omega \times \mathbf{r}$ such that $\mathbf{E} = -\omega_E (\mathbf{e}_\omega \times \mathbf{r}) \times \mathbf{B}_{dip}$.

iii) Remarks on the assumption of a constant electric field or electric field mapping:

- Distinguish between a selfconsistent steady state solution and the superposition of an electric field to a model magnetic field.
- Often assumed or argued that the SW electric field penetrates into the magnetosphere to set up convection.
- Convection is truly a result of the forces in a system. For instance the tailward end of the plasmoid should move Earthward if the electric field was constant => counter intuitive but not consistent with any observations or quantitative models.

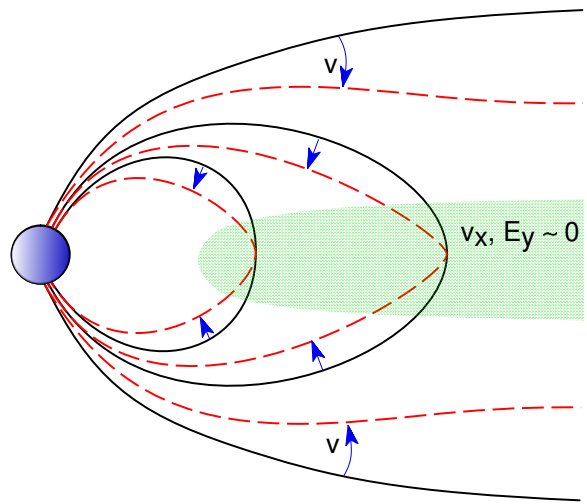
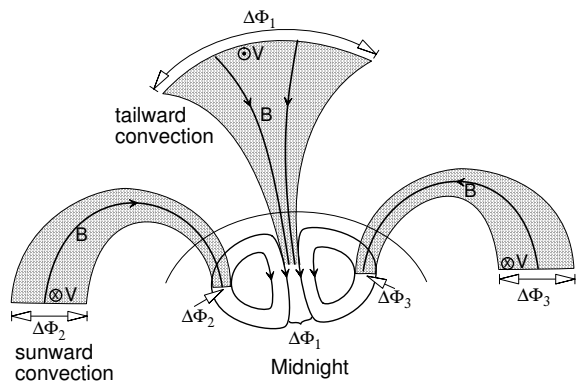
1.5.4 Magnetospheric Substorms

Three distinct phases:

- Growth phase
- Expansion phase
- Recovery phase

(a) Growth phase:

- Essential Nature
 - Increased transfer of energy and magnetic flux from solar wind to magnetosphere. The accumulated energy is stored in the magnetotail.
- DURATION: 30-60 minutes
- Properties
 - Southward turning of the interplanetary magnetic field (IMF)
 - Gradual increase in the size of the polar cap and equatorward expansion of the trapping boundary.
 - Change of dayside (and nightside) ionospheric convection
 - Stretching of the near-Earth magnetic field configuration (into a more tail-like configuration)
 - Formation and expansion of a thin near Earth current sheet.
 - (Slow intensification of the Birkeland currents)
 - Convergence of most equatorward discrete arc and diffuse arc.
 - Slow increase in the (magnitude of the) magnetotail lobe magnetic field



(b) Expansion phase

- Essential Nature
 - Release of energy stored in the magnetosphere during the growth phase.
- DURATION: 30-60 minutes
- Properties
 - Sudden brightening of the most equatorward (discrete) arc in premidnight sector
 - Development of the westward-travelling surge (poleward and westward expansion of the aurora)
 - Enhancement of:
 - * the (auroral) electrojet (across the midnight sector)
 - * ionospheric convection?
 - * Region 1 and 2 currents
 - Dipolarization of the magnetic field in the near-Earth magnetotail (current disruption)
 - Particle injection
 - Signatures of the passage of plasmoids and/or Travelling Compression Regions (TCR's) by satellites
 - "Disappearance" of the plasma sheet (in some locations)
 - Sudden decrease of the magnetic field in the magnetotail lobes?
- Timing of the specific events not yet clear.

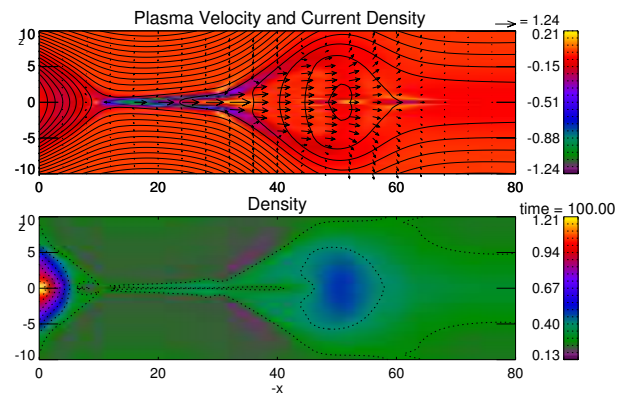
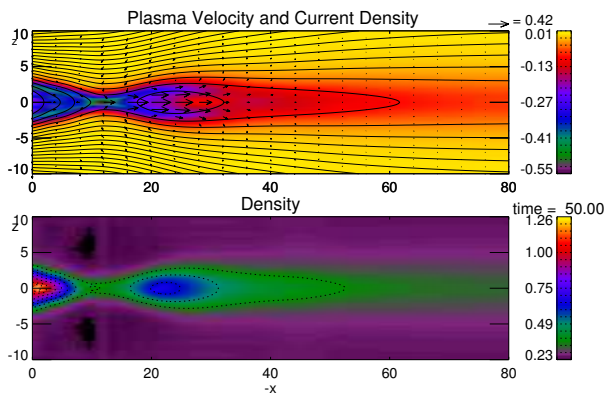
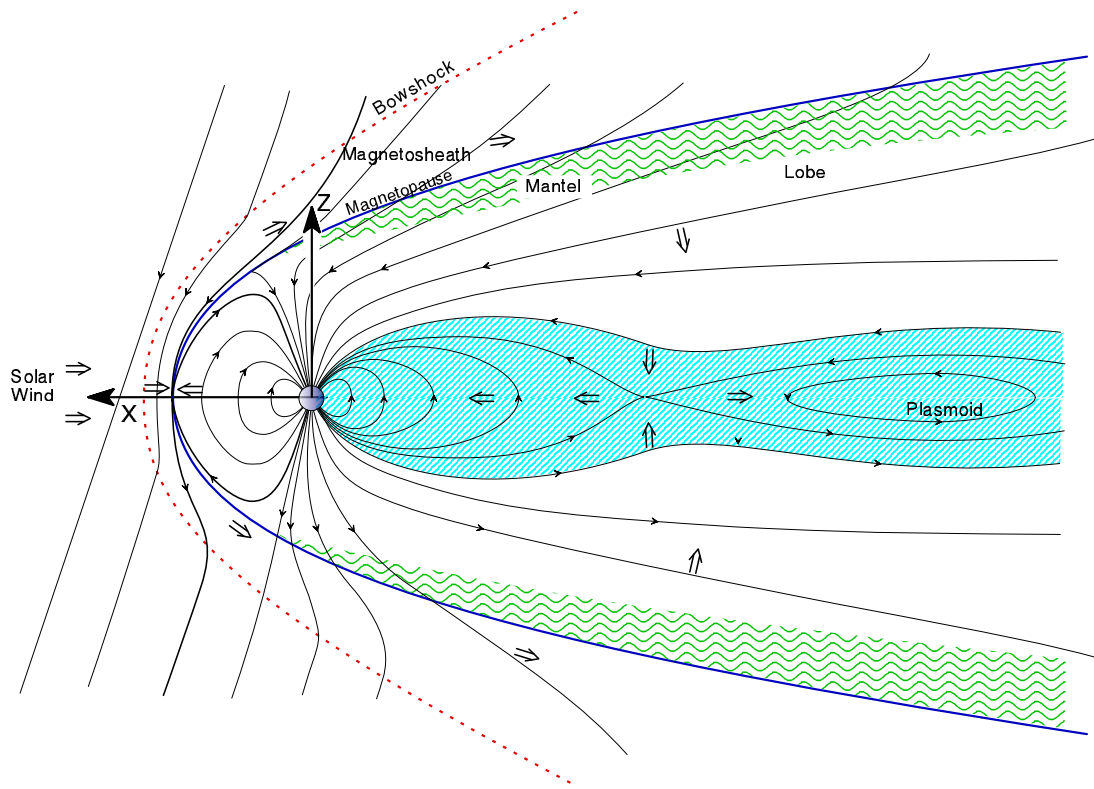
Within a minute there are various onset signature (auroral, plasma acceleration in the near Earth tail, particle injection).

Most tail expansion phase signatures

- Plasmoid formation
- Particle injection
- Dipolarization

are well explained by reconnection in the magnetotail.

Reconnection Model:



1.6 Magnetosphere - Ionosphere coupling

Steady magnetospheric convection $\Rightarrow \mathbf{E} = -\nabla\phi$ and for $\mathbf{E} \cdot \mathbf{B} = 0$ potential is constant on magnetic field lines and maps into the ionosphere

- Large magnetopause reconnection \Rightarrow large convection in the polar cap \Rightarrow large Joule heating
 $\mathbf{j}_p \cdot \mathbf{E}_\perp = \sigma_p \mathbf{E}_\perp^2 \Rightarrow$ expanded neutral atmosphere
- Ionospheric convection = image of convection in the magnetosphere

Problems:

- Magnetospheric convection is non-steady
- $\mathbf{E} \cdot \mathbf{B} \neq 0$
- Magnetospheric convection is compressible vs ionospheric convection is incompressible
- Finite propagation time (traveling with Alfvén speed)

\Rightarrow mismatch between magnetospheric dynamics and ionospheric dynamics

- Ionosphere provides important feedback to the magnetosphere:
- Ionization due to precipitation increases number of ionospheric particles (conductivity)
 - slower response of the ionosphere
 - ionosphere acts as drag for the magnetosphere
- Ion outflow (expansion phase!) changes inertia of the magnetosphere.
- Small scale physics may be very important for instance for Joule heating.

1.7 Inner magnetosphere

Mostly dipolar magnetic field => single particle drifts (adiabatic invariants).

Bounce Period: Time for a particle to move from one mirror point to the other and back.

$$\tau_b = 4 \int_0^{\lambda_m} \frac{ds}{v_{\parallel}} \approx \frac{LR_E}{(W/m)^{1/2}} (3.7 - 1.6 \sin \alpha_{eq}) \quad (8)$$

- 1 keV particle and equatorial pitch angle of 30° yields a few to 10 seconds for electrons and 2 to 6 minutes for the bounce period.

Loss cone: Particles with a mirror point at or below ionospheric heights will be absorbed in the atmosphere through collisions with the neutral atmosphere.

$$\sin^2 \alpha_l = \frac{B_{eq}}{B_E} = \frac{\cos^6 \lambda_E}{(1 + 3 \sin^2 \lambda_E)^{1/2}} \quad (9)$$

- Particles with $\alpha < \alpha_l$ and all particles with $\alpha > 180^\circ - \alpha_l$ will be absorbed in the Earth's atmosphere.

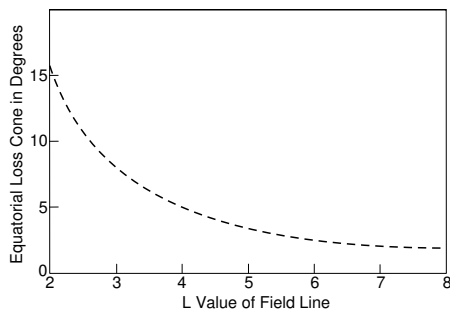


Figure 9: Equatorial loss cone as a function of L value.

1.7.1 Average drift velocity in the dipole field

During gyro and bounce motion a particle will also undergo gradient and curvature drift motion.

Average drift with $W = mv^2/2$

$$\langle v_d \rangle \approx \frac{3WL^2}{qB_E R_E} (0.35 + 0.15 \sin \alpha_{eq}) \quad (10)$$

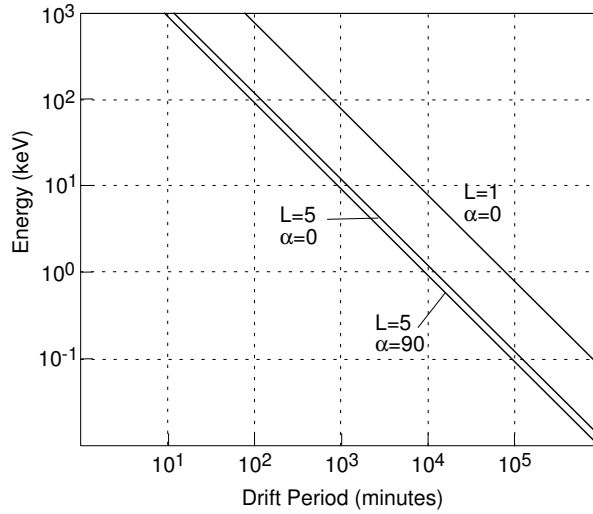


Figure 10: Drift period in the Earth's dipole field.

Weak dependence on the pitch angle α_{eq} . The drift period is

$$\langle \tau_d \rangle \approx \begin{cases} 43.8/LW & \alpha_{eq} = 90^\circ \\ 62.7/LW & \alpha_{eq} = 0^\circ \end{cases} \quad (11)$$

Notes:

- 1 keV particle has drift periods of 100 to several hundred hours. On these time scale the third adiabatic invariant is likely violated.
- The drift velocity scales with $L^2 \Rightarrow$ drift periods are shorter for larger L values.
- Energy dependence leads to a separation of particles with different energies (if injected at the same location).
- Pitch angle dependence also generates some separation of particles with different pitch angles.
- Typical ring current particles (10 keV) do not complete a full revolution.
- Only MeV particles orbit fast enough to complete closed orbits
- Electrons and ions drift in opposite directions thereby generating the so-called ring current.

1.7.2 Ring Current and Magnetic Storms

The drift current for particles with an equatorial pitch angle of $\alpha_{eq} = 90^\circ$, energy W , and density n on a given L shell is

$$j_d = \frac{3L^2 n W}{2B_E R_E}$$

Magnetic disturbance:

The magnetic perturbation at the Earth's center can be computed from Biot-Savart's law. With the total energy of all drifting particles U_R

$$\Delta B = -\frac{\mu_0}{2\pi} \frac{U_R}{B_E R_E^3}.$$

Magnetic Storms

Large solar flares or coronal mass ejections => strong perturbation of the terrestrial magnetosphere for long periods of time (days).

- Caused by a strong increase in ring current energy
- Most easily measured at equatorial latitudes
- The average magnetic perturbation near the equator establishes the *Dst index*.
- A geomagnetic storm: strong depression of the equatorial magnetic field by several 100 nT.
- This corresponds often to disturbances of more than 1 % of the dipole magnetic field.

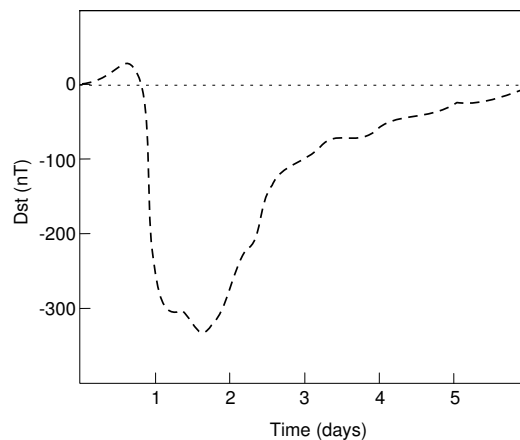


Figure 11: Equatorial magnetic field depression (Dst) during a magnetic storm.

Storms have two phases:

- 1st strong increase increase of the ring current particles.
 - long duration (many hours) of southward IMF conditions.
 - much larger average electric fields in the magnetosphere
 - increase of ring current particles and particle energies
- Ring current decays on time scales of a few days.
- Assume that at least half of the depression during storms is caused by the intensification of the ring current => estimate of the associated energy: $\Delta U_R \geq (\pi/\mu_0) B_E R_E^3 Dst$
- Increase in ring current energy of $2 \cdot 10^{13}$ J per 1 nT depression of the magnetic field. Assuming current at $L = 5$ a current of $4 \cdot 10^4$ A produces a 1 nT depression.
- Large magnetic storm => the ring current energy increase by more than $5 \cdot 10^{15}$ J and total current of more than 10^7 A.
- With an average energy of 10 keV per particle more than $3 \cdot 10^{30}$ particles are injected into the ring current at such times.

2 Numerical Simulation

2.1 Motivation for Numerical Simulation

Traditional approaches:

- Experiment and observation
- Analytical theory

All interesting problems (physics, engineering, biology, and other sciences) are nonlinear and analytical solutions exist only in extremely rare cases.

- Classical gravitationally interacting three-body problem (has no analytical solution except for rather special cases).
- Weather prediction.
- Aerodynamics of airplanes, cars etc.
- Solar and magnetospheric physics.

In many systems observations or experiments are not available, difficult, or and/or incomplete to describe the state:

- Combustion problems
- Oil reservoirs
- Ground water flow
- Meteorology, weather and climate systems
- Solar and magnetospheric physics.

Problems must be well posed in terms of governing equations and boundary conditions. -> Numerical Models!

2.2 Important Background for the Numerical Experiment

2.2.1 Well Posed Problems:

All well posed problem are described by a corresponding set of basic equations.

Basic types of equations

$$\begin{aligned} \frac{d\mathbf{r}}{dt} = v, \quad \frac{d\mathbf{v}}{dt} = \frac{\mathbf{F}}{m} & \quad \text{ordinary diff. equation} \\ \frac{\partial f}{\partial t} + \frac{\partial v f}{\partial x} = 0 & \quad \text{hyperbolic partial diff. eq.} \\ \frac{\partial f}{\partial t} - \nu \frac{\partial^2 f}{\partial x^2} = 0 & \quad \text{parabolic partial diff. eq.} \\ \nabla^2 f = s(\mathbf{r}) & \quad \text{elliptic partial diff. eq.} \end{aligned}$$

The different types of differential equations correspond to different physical problems:

- Hyperbolic PDE's \leftrightarrow convection or wave propagation.
- Parabolic PDE's \leftrightarrow diffusion.
- Elliptic PDE's \leftrightarrow force fields generated by sources s .

Hyperbolic PDE's have only real characteristics which implies that information (waves) is always propagating with finite speed. In contrast parabolic and elliptic equation are always global, meaning a change of source or boundary conditions has an influence on the entire system.

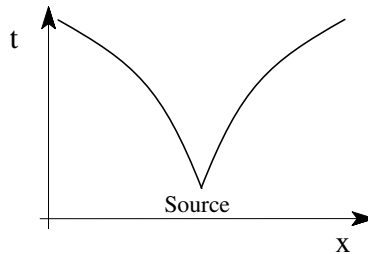


Figure 12: Illustration of characteristics branching from a point source for hyperbolic PDE's.

Local vs. nonlocal dynamics: Because of the finite propagation speed any changes at one point of a system can be computed from the immediate vicinity for hyperbolic equations (local). Vice versa changes at one point of a system can be caused from remote locations (boundary) for parabolic and elliptic PDE's.

=> *Different types of differential equations have different mathematical **and** physical properties!*

A mathematical problem is called well posed if

- the solution exists
- the solution is unique
- the solution depends continuously on the auxiliary data (e.g., boundary and initial conditions)

In the same terms one can define a numerical problem as well posed through existence, uniqueness, and the continuous dependence on initial and boundary conditions.

2.2.2 Discretization of differential equations

Systematic approach: Weighted Residual Method

Minimization of the error (residual) caused by the discretization measured

Methods of discretization:

Finite volume method: Introduce subdomains and store values of f on central point of each subdomain. Weight function is $W_m = 1$ inside subdomain and 0 outside.

Finite element methods: Special sets of base functions which are $\neq 0$ only in a small subdomain and 0 otherwise.

Spectral methods: Base functions are Fourier series, Legendre, or Chebyshev polynomials. Note that base functions are global, i.e., nonzero almost everywhere!

Finite difference methods: Special cases of finite volume or finite element methods. Values of f are stored on grid points and derivatives are approximated through appropriate differences.

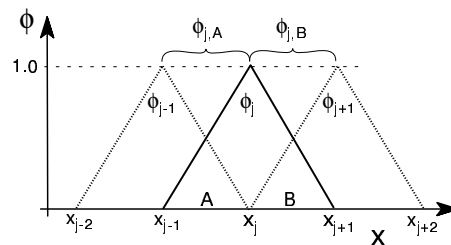


Figure 13: Illustration of one-dimensional linear finite elements.

2.2.3 Example for the Finite Difference Method

Consider the equation

$$\frac{\partial f}{\partial t} = -v \frac{\partial f}{\partial x} \quad (12)$$

which represents for instance one-dimensional heat conduction or diffusion.

- f is the density or concentration of a substance and
- v is the convection velocity.

Store the value of f on a discrete spatial grid and advance f in discrete time increments.

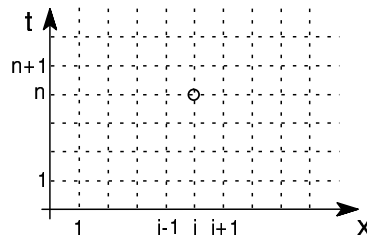


Figure 14: Spatial and temporal grid for the finite difference method.

Assume: spatial grid separation Δx , a total number of N_x grid point for the range $0 \leq x \leq 1$ and a time increment of Δt

Discretize the convection equation 12 for instance in the following manner

$$\frac{f_i^{n+1} - f_i^{n-1}}{2\Delta t} = -v \frac{f_{i+1}^n - f_{i-1}^n}{2\Delta x} \quad (13)$$

with f_i^n representing the value of the density at the (i,n)th node. This is the so-called **Leapfrog** method. The algebraic equation can be used to advance the the solution f_i^n explicit in time

$$f_i^{n+1} = f_i^{n-1} - \frac{v\Delta t}{\Delta x} (f_{i+1}^n - f_{i-1}^n) \quad (14)$$

to determine the solution at time t_{n+1} from the solutions at time levels t_n and t_{n-1} .

Note: the solution to f_i^n is different from the exact solution f_e due to the numerical approximation.

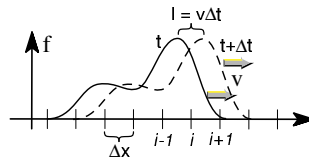


Figure 15: Propagation of f by the convection equation ($v = \text{const}$).

The spatial derivative in Leapfrog scheme is centered at grid point i . Other examples of the first order derivatives $\partial f / \partial x$ are

$$\left. \frac{\partial f}{\partial x} \right|_{forward} = \frac{f_{i+1} - f_i}{\Delta x}$$

$$\left. \frac{\partial f}{\partial x} \right|_{backward} = \frac{f_i - f_{i-1}}{\Delta x}$$

for a forward and a backward (non-centered) difference approximations.

2.3 Limitations, Accuracy and Stability

Considerations for all Computer models:

- Physical limitations
- Numerical limitations
 - Errors and accuracy
 - Stability
- Efficiency

2.3.1 Physical limitations:

The model equations can never represent an exact model of real systems because of the usually large degree of freedom for instance

Weather simulation: Assuming the size of a simulation box of $(100\text{km})^3$ (this may be a little high in altitude but also a bit low on area) with the goal to resolve air turbulence (which has a length scale of about 10 cm) requires a grid of about 10^{18} . Thus modern computers are short by more than a factor of 10^8 to resolve the fine scale air turbulence in a weather simulation.

↪ mathematical description is always an approximation and one has to understand the quality of this approximation in other words:

Physical (chemical ...) model must appropriately describe the system under study.

2.3.2 Accuracy

The discretization always introduces errors and there are various types of such error sources. Typical errors and limitations for a discretization are due to

- Truncation and rounding errors.
- The discrete representation and the limited computer resources (memory and speed).
- Accumulation of errors in a systematic manner for certain numerical methods.

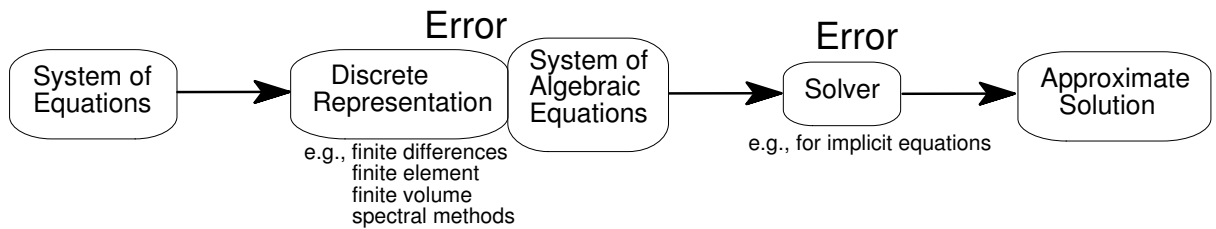


Figure 16: Steps to arrive at the numerical solution .

2.3.3 Finite difference approximation of derivatives

Taylor series expansion for the value of f :

$$f_{i+1}^n = f^n(x_i + \Delta x) = \sum_{m=0}^{\infty} \frac{\Delta x^m}{m!} \left[\frac{\partial^m f}{\partial x^m} \right]_i^n \quad (15)$$

and

$$f_i^{n+1} = f_i(t + \Delta t) = \sum_{m=0}^{\infty} \frac{\Delta t^m}{m!} \left[\frac{\partial^m f}{\partial t^m} \right]_i^n \quad (16)$$

The series can be truncated for any number of terms with the error determined by the next higher order term in the expansion. Note that this usually requires $\Delta x \ll 1$ and $\Delta t \ll 1$. For second order accuracy we obtain

$$f_{i+1}^n = f_i^n + \Delta x \left[\frac{\partial f}{\partial x} \right]_i^n + \frac{\Delta x^2}{2} \left[\frac{\partial^2 f}{\partial x^2} \right]_i^n + O(\Delta x^3)$$

This truncation implies that the error of the approximation is smaller than $k\Delta x^3$ for a suitable finite value of k . Note that this is strictly true if the radius of convergence is larger than the choice for Δx .

Forward difference approximation: Using the Taylor expansion one can determine the error from finite differences

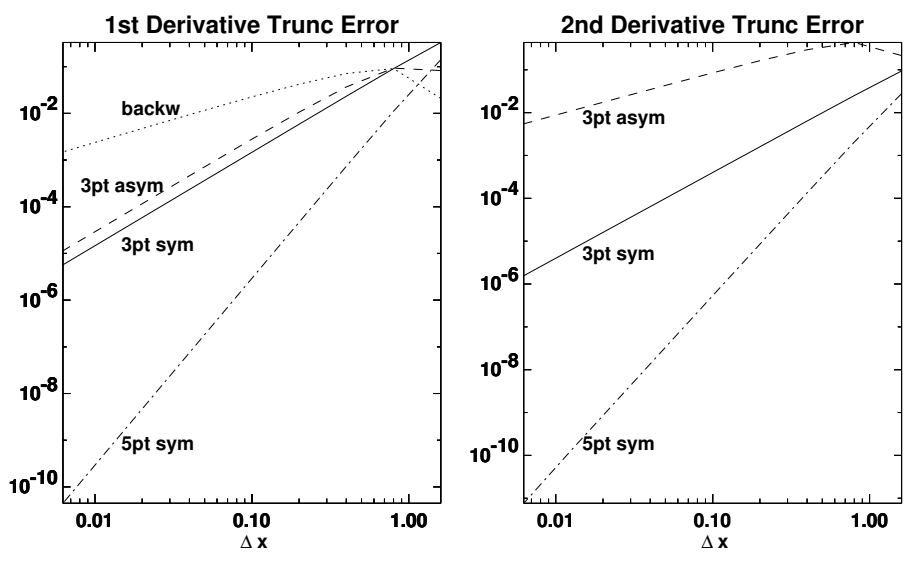
$$\left[\frac{\partial f}{\partial x} \right]_i^n = \frac{f_{i+1}^n - f_i^n}{\Delta x} - 0.5\Delta x \left[\frac{\partial^2 f}{\partial x^2} \right]_i^n + \dots \quad (17)$$

which demonstrates that the forward differencing is accurate to $O(\Delta x)$. The additional terms in the finite difference approximation are the truncation error.

- The centered Leapfrog scheme is second order accurate: Truncation error = $O(\Delta x^2)$
- Similar to the forward differencing other difference formulas can be examined for the truncation error.

Case	Difference formula	Leading truncation error term
3pt sym	$(f'_{i+1} - f'_{i-1}) / 2\Delta x$	$\Delta_x^2 f_{xxx} / 6$
Forw diff	$(f'_{i+1} - f'_i) / \Delta x$	$\Delta_x f_{xx} / 2$
Back diff	$(f'_i - f'_{i-1}) / \Delta x$	$-\Delta_x f_{xx} / 2$
3pt asym	$(-1.5f'_i + 2f'_{i+1} - 0.5f'_{i+2}) / \Delta x$	$-\Delta_x^2 f_{xxx} / 3$
5pt sym	$(f'_{i-2} - 8f'_{i-1} + 8f'_{i+1} - f'_{i+2}) / 12\Delta x$	$-\Delta_x^4 f_{xxxxx} / 30$

Case	Difference formula	Leading error term
3pt sym	$(f'_{i-1} - 2f'_i + f'_{i+1}) / \Delta x^2$	$\Delta_x^2 f_{xxxx} / 12$
3pt asym	$(f'_i - 2f'_{i+1} + f'_{i+2}) / \Delta x^2$	$\Delta_x f_{xxx}$
5pt sym	$(-f'_{i-2} + 16f'_{i-1} - 30f'_i + 16f'_{i+1} - f'_{i+2}) / 12\Delta x^2$	$\Delta_x^4 f_{xxxxx} / 90$



Note: The Taylor series expansion can also be used in a systematic manner to achieve higher order approximations by using a general expression for instance

$$\left[\frac{\partial f}{\partial x} \right]_i^n = a f_{i-2}^n + b f_{i-1}^n + c f_i^n + d f_{i+1}^n + e f_{i+2}^n + O(\Delta x^m) \quad (18)$$

and expanding each term as a Taylor series to yield the best approximation to the first derivative.

The way to improve on the numerical model results is to reduce the respective error source. This can be done for instance by:

- Better spatial and/or temporal resolution (smaller grid space, more finite elements, larger number of base functions, smaller time step).
- An increase of the number of particles.
- A change to a better and more appropriate method.

2.3.4 Stability

A particularly poor error evolution is caused by numerical instabilities. The discreteness of the numerical algorithm always introduces modifications to the equations which are to be solved. Explicit finite difference methods for hyperbolic equations are always subject to a stability limit for the time step

$$\Delta t \leq O(\Delta x/v_{typ})$$

known as the Courant condition (or the CFL=Courant-Friederichs-Levi condition).

Reason for numerical instability:

Hyperbolic PDE's have always real characteristics \leftrightarrow finite maximum speed for information transport (waves). **Explicit differencing** implies that the update is obtained with quantities known in the vicinity of the particular node i , i.e., usually requires nodes $i - 1$, i , and $i + 1$.

Assume that information is carried in the simulation with a maximum velocity v_{typ} implying that this information is carried a distance $l = v_{typ} \cdot \Delta t$ in a single time step Δt . If l is larger than the grid separation Δx , then information (density front) travels more than a single grid spacing in one time step. Since the update of density or temperature (or similar) uses only local information, the arrival of a wave front cannot be predicted correctly at any given grid point or node. In other words, the maximum velocity that is resolved by an explicit scheme is given by $v_{max} \cong \Delta x/\Delta t \geq v_{typ}$ yielding the condition $\Delta t \leq \Delta x/v_{typ}$ as in the Courant condition above. While this consideration does not necessarily imply an instability it makes clear that bad thing must happen for a too large time step because the method is not able to resolve the proper propagation of information in the system.

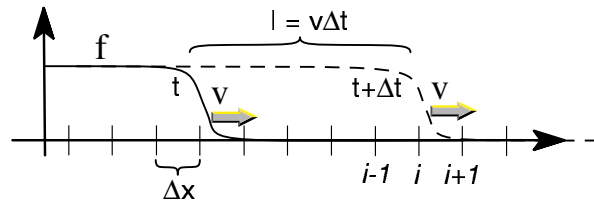


Figure 17: Illustration of a wave front traveling with a speed faster than v_{max} across a uniform grid.

2.3.5 Systematic approach: Von Neumann method

Assume that the numerical error is

$$\xi = f_e - f$$

with the exact solution f_e and the numerical solution f . The numerical solution is unstable if

$$g = \frac{\xi_j^{n+1}}{\xi_j^n} > 1$$

for any grid point j . If the governing equation is linear ξ obeys the same equation as f . Thus one can use a normal mode analysis where ξ is expanded in a Fourier series

$$\xi_j^n = \xi_0 \exp [i (kx_j + qt_n)] \quad (19)$$

and

$$\xi_j^{n+1} = \xi_0 \exp [i (kx_j + qt_n) + iq\Delta t]$$

Thus

$$\frac{\xi_j^{n+1}}{\xi_j^n} = \exp [iq\Delta t] = g \quad (20)$$

Similarly we can express

$$\xi_{j+1}^n = \xi_0 \exp [i (kx_j + qt_n) + ik\Delta x]$$

Leapfrog method for convection:

$$\Delta t \leq \Delta x/v$$

Example: Diffusion equation with forward time centered space differencing

$$\frac{f_i^{n+1} - f_i^n}{\Delta t} = \alpha \frac{f_{i-1}^n - 2f_i^n + f_{i+1}^n}{\Delta x^2} \quad (21)$$

=>

$$f_i^{n+1} = f_i^n + s (f_{i-1}^n - 2f_i^n + f_{i+1}^n) \quad (22)$$

$$s = \frac{\alpha \Delta t}{\Delta x^2} \quad (23)$$

Stability condition:

$$s \leq \frac{1}{2} \quad (24)$$

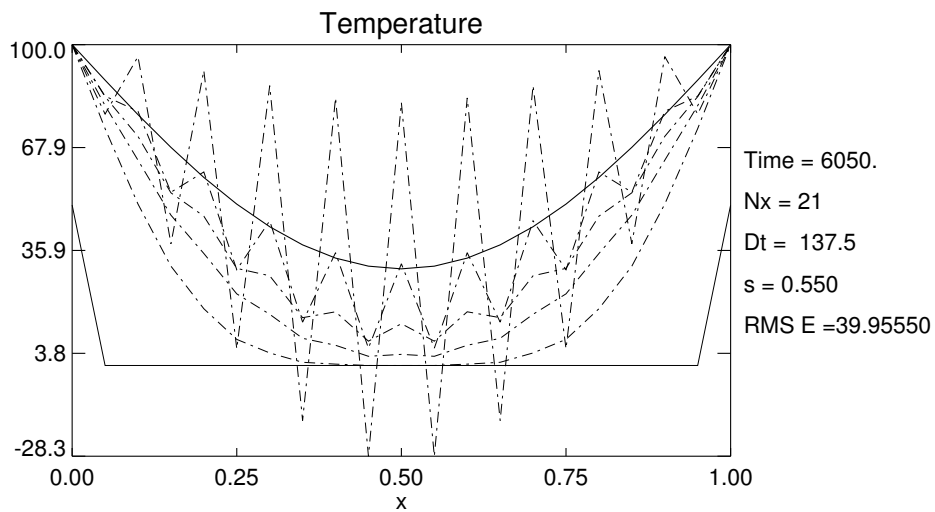


Figure 18: Grid oscillations increasing with time are typical for numerical instabilities.

On the relation between discretization and the actual exact solution of a mathematical problem:

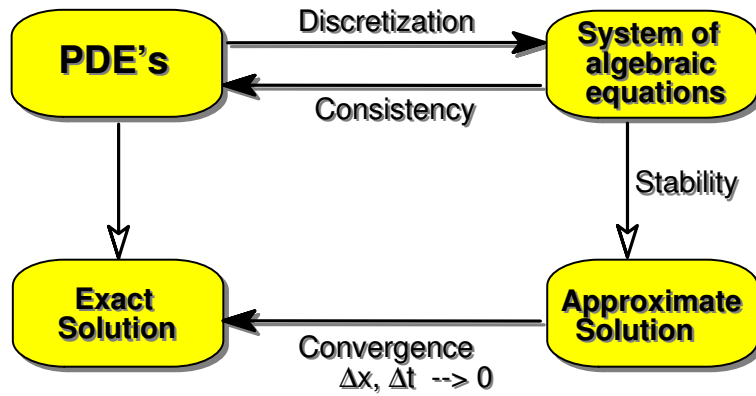


Figure 19: Relation between consistency, stability, and convergence.

2.4 Efficiency, Vector and Parallel Computation

2.4.1 Efficiency Considerations:

- Method must be appropriate
- High vs. low order approximation
 - method should be second order accurate for time dependent problems
- Explicite or implicate methods
 - explicite advantage: only local information is needed
 - implicate advantage: no time step restriction (no stability limit)
- Hardware: scalar, vector, or parallel computater environment.

2.4.2 Operation counts:

Basic computational operations have a typical execution time. The principle categories:

- floating point operations (fl), fixed point operations (fx), assignments (=), logical operations, e.g., if, and, or (l), mathematical library functions, e.g., sin, exp, etc (m)

Estimate for the actual computation time requires times for each of the typical operations in a program.

Table 1: Relative execution times for basic operations.

Operation	Microcomputer NEC-APC IV	Supermicro SUN Sparc St1	Mainframe IBM-3090
add (fl)	1.0	1.0	1.0
subtr (fl)	1.0	1.0	0.8
multipl (fl)	1.1	1.0	0.8
div (fl)	1.4	5.8	3.1
assign (=)	0.1	1.0	0.1
if (l)	0.1	0.9	0.2
add (fx)	0.05	0.6	0.2
subtr (fx)	0.05	0.6	0.2
multipl (fx)	0.4	0.6	0.6
div (fx)	0.5	5.4	3.2
power	2.7	20	16
sqrt	2.0	29	16
sin	10.0	29	17.6
exp	6.7	33	15

- Avoid divisions in particular on RISC (reduced instruction set) processors.
- Also library calls to mathematical functions are usually expensive and should be carefully chosen or avoided.
- Always estimates on the execution time for a simulation/modeling program.
 - determine possible inefficiencies due to inefficient programming
 - planning of larger size programs
 - understanding of the potential for a program, i.e., whether it can be scaled up to treat a compelling application.

2.4.3 Vector processing concept:

A vector processor can load an entire array of numbers or operate on such an array in a single cycle, different from a scalar processor which always loads or operates on single numbers (often the width of the processor, e.g., 32 bit (= 4 byte real numbers)).

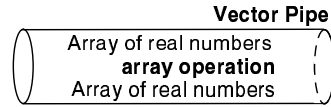


Figure 20: Illustration of the vector processing concept

Assuming

- a total number of operations $\equiv N$
- a vector speed (number of ops/sec in vector mode) $\equiv V(N)$
- a scalar speed $\equiv S$

\Rightarrow scalar computation time $\tau_S = N/S$

\Rightarrow vector computation time if a fraction P of the problem can be vectorized:

$$\tau_V = \frac{N}{S}(1 - P) + \frac{N}{V}P$$

Speed-up between scalar and vector operation for a given problem: **Amdahl's law**

$$\frac{\tau_S}{\tau_V} = \left(1 - P + \frac{S}{V}P\right)^{-1} \quad (25)$$

Fraction of a problem that is vectorizable must be very high to reach execution speed-ups close to the maximum of V/S .

Exercise: Plot the speed-up as a function of P for $V/S = 100$.

2.4.4 Parallel processing concept

A parallel processing with distributed memory requires inter-processor communication for almost all problems.

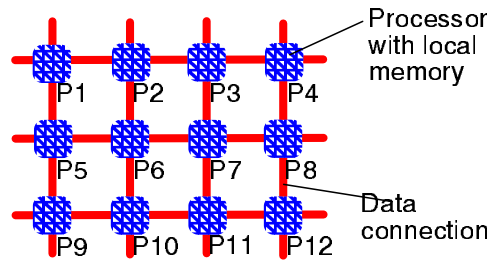


Figure 21: Illustration of the parallel processing concept

Distribute a given numerical problem to N_P processors in a manner that all processors have an approximately equal load (**Exercise:** What would happen if they don't?). Insure that all processors have the information they require to perform their individual portion of the problem. Defining

- $S \equiv$ speed of the individual processor
- $F \equiv$ efficiency of the parallelized code (dependent on parallelization method, number of processors, and problem size). For instance an unequal load for the processors may cause some processors to idle (F does not account for a situation where all processors except for one are idle. This is accounted for by $P < 1$)
- $N_p \equiv$ number of processors
- $P \equiv$ proportion of problem which can be parallelized
- $\tau_C \equiv$ time for inter-processor communication (problem dependent)

the total execution time for a problem of size N is

$$\tau_P = \frac{N}{S}(1 - P) + \frac{N}{FSN_p}P + \tau_C$$

resulting in a speed-up of

$$\frac{\tau_S}{\tau_P} = \left(1 - P + \frac{P}{FN_p} + \frac{\tau_C S}{N} \right)^{-1} \quad (26)$$

for the parallel execution compared to a scalar execution. Maximum speed-up: N_P .

Important aspects are the achieved efficiency, the fraction of the problem that cannot be parallelized, and a considerable inter-processor communication time.

Domain decomposition:

Distribution of a given numerical problem to many processors each of which dealing with a certain portion of the overall computation.

- Many ways to distribute any given problem to multiple processors and actual distribution depends on
 - problem to be solved,
 - the software (operating system and compilers)
 - the hardware (i.e. how many processors are available, how they are communicating, and the distribution of memory on the processors).

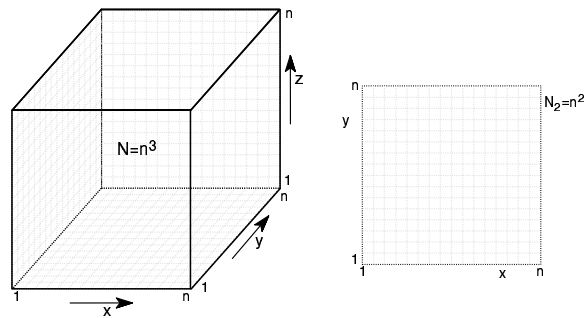


Figure 22: Illustration of a three-dimensional gridded system (left) and a cut in the x, y plane.

Different ways to slice a three-dimensional domain:

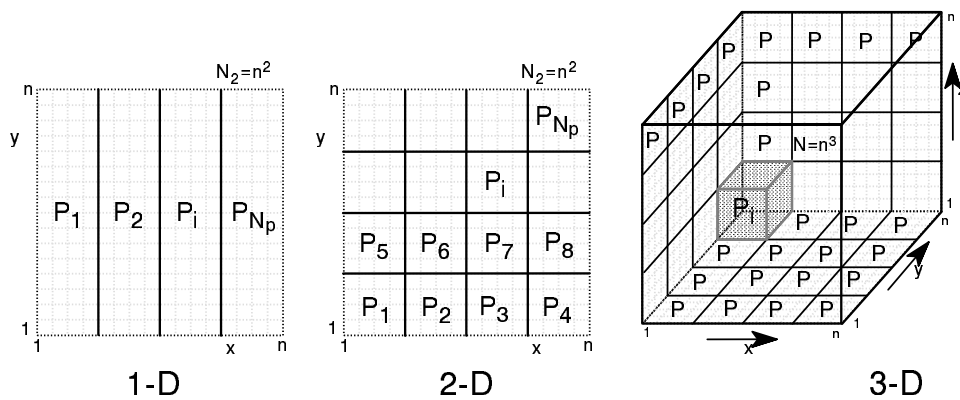


Figure 23: Illustration of three different (one-, two-, and three-dimensional) domain decompositions.

2.5 Computer performance through the years

Table 2: Computer performance for different systems from 1984 to 2004.

Year	PC-Type	Workstation	Mainframe	Vector SC	Parallel SC
1984	0.02 Intel 8086		3 Cyber 855	17 Cyber 205	
	0.037 Mot 68000				
	0.005 Atari ST				
1987	0.13 Intel 386	0.36 IBM RT		100 Cray XMP	
		0.51 Apollo			
1990-91	1 Intel 486	3 Sun Sparc 1		200 Cray YMP	
1993-94	7 Intel P5	20 Dec Alpha		450 Cray C90	
1997-98	40 Intel P6	100 Dec+others			$2 \cdot 10^4$ T3E
2000-01	250 Pent 1 GHz	500 SGI+others		1500 Cray SV1	$2 \cdot 10^5$ T3E
2002-03	1000 P4 3 GH				10^6 Cray X1

Except for the parallel machines all performance number reflect the speed of single processor systems (even though most vector computer have between 4 and 32 processors). Each entry gives the speed measured in so-called MFLOPS (million floating point operations per second) and the particular system line or processor name. The numbers quoted do not refer to the often used peak performance but to a performance which can be realized for reasonably well written programs. The real performance numbers are often between a factor of 2 to 10 below the so-called peak performance.

Computer Performance

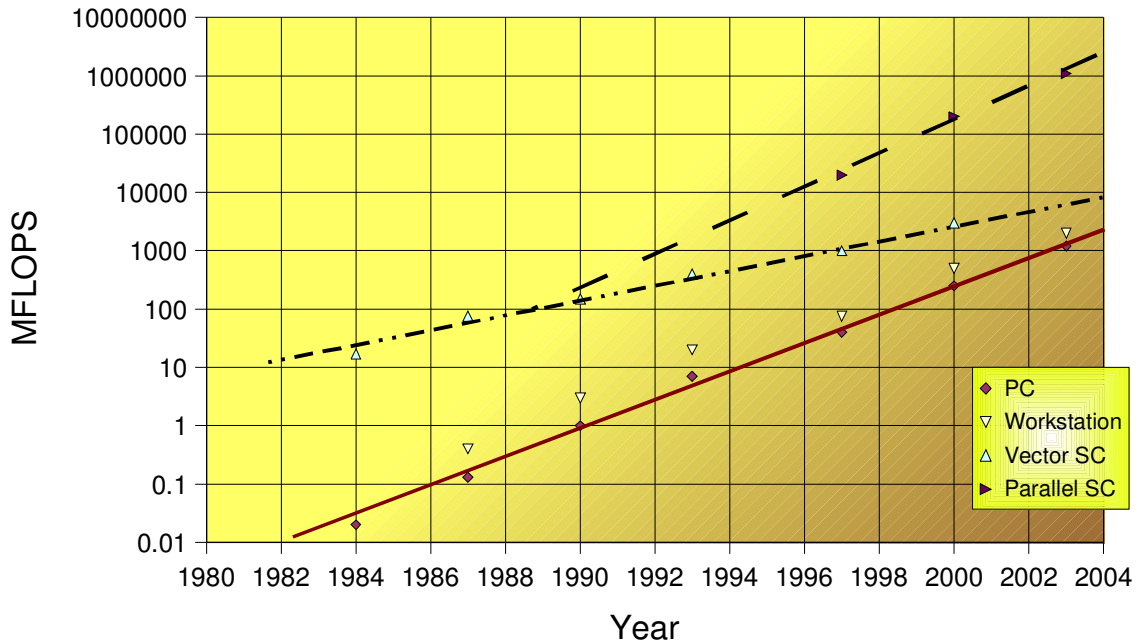


Figure 24: Computer performance throughout the years

- PC's demonstrate an increase in computer speed of a factor of 10 every 4 years
- Computer memory and typical disk space would yield a slightly lesser increase.
- Supercomputers offer a 1000fold advantage over PC's in the mid eighties.

But: How about efficiency

Table 3: Execution time for different computers and the problem $7 * 13$. *The time for the Cray T3E includes time to pass the security check for an user application on ARSC supercomputers.

Computer	Time
Human	3 seconds
Pocket Calculator	30 seconds
PC	5 minutes
Cray T3E	3 months*