Quo vadis, GGCM?

J. Raeder

Space Science Center, University of New Hampshire

GEM Summer Workshop, Zermatt, UT, June 19 2007

Overview

- Some GGCM history in GEM.
- Fundamental properties of MDH “spines”.
- The issues:
  - Reconnection and Convection
  - substorms and storms.
- Computing: A Space Weather Appliance.

GGCM history within GEM

- The original idea of a GGCM envisioned a regional modular approach: coupling of bow shock, magnetosheath, magnetopause, ring current, tail, …. Models (empirical or numerical). The GEM campaigns were to produce these modules.
- In the mid 90's this approach was found impractical. At the same time "global MHD" models showed promising results (beyound reproducing cartoons).
- Since then MHD models are used as GGCM "spines", coupled with other models where MHD is (clearly) not applicable: ionosphere, thermosphere, ring current, plasmasphere, etc.. The coupling efforts are still ongoing.
- The role of GEM has changed: GGCM development occurs largely outside of GEM but GEM provides a focal point for guidance, metrics, basic science questions, verification etc.

GGCM basics

- All GGCMs solve the MHD equations of a variation thereof. Newer models with a global scope use a hybrid approach (see Omidi tutorial).
- Although there are 20+ approximations used to derive MHD from the fundamental equations (Maxwell & Newton) they are a surprisingly good approximation of the real world as long as the scales are large enough and the plasmas are close enough to thermal equilibrium. The reason is simple: MHD describes the conservation of mass, momentum, energy, and magnetic flux.
- All GGCMs use discrete numerical schemes to approximate the solution of the MHD equations for given boundary conditions. As a consequence the solutions are not solutions of the MHD equations but the solution of an modified system of equations.

Which equations are really solved?

Some models, for example BATS`R'US use explicitly equations different from MHD (Powell et al., 1999):

$$\mathbf{F} = \begin{pmatrix} \rho\mathbf{u} \\ \rho\mathbf{u}\mathbf{u} + \left(p + \frac{\mathbf{B}\cdot\mathbf{B}}{2}\right)\mathbf{I} - \mathbf{B}\mathbf{B} \\ \mathbf{u}\mathbf{B} - \mathbf{B}\mathbf{u} \\ \mathbf{u}\left(E + p + \frac{\mathbf{B}\cdot\mathbf{B}}{2}\right) - (\mathbf{u}\cdot\mathbf{B})\mathbf{B} \end{pmatrix}^T,$$

$$\frac{\partial \mathbf{U}}{\partial t} + (\nabla \cdot \mathbf{F})^T = \mathbf{S},$$

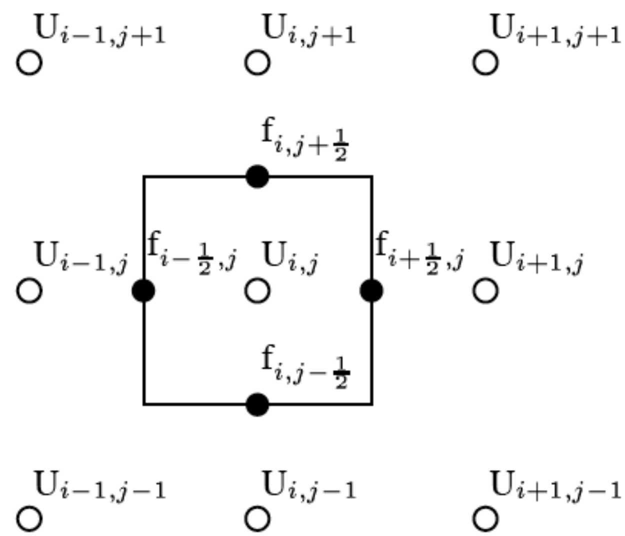
$$\mathbf{S} = -\nabla\cdot\mathbf{B}\begin{pmatrix} 0 \\ \mathbf{B} \\ \mathbf{u} \\ \mathbf{u}\cdot\mathbf{B} \end{pmatrix}.$$

$$\mathbf{U} = (\rho, \rho u, \rho v, \rho w, B_x, B_y, B_z, E)^T$$

Which equations are really solved?

- All models incur extra terms from numerical (discretization) errors:

- Numerical dispersion (odd derivatives)
- Numerical diffusion (even derivatives)



$$\frac{\partial U}{\partial t} = -\nabla \cdot \mathbf{F}(U)$$

$$\begin{aligned}\frac{\partial U}{\partial t} = &- (f_{i+\frac{1}{2},j}(U) - f_{i-\frac{1}{2},j}(U))/\Delta x \\ &- (f_{i,j+\frac{1}{2}}(U) - f_{i,j-\frac{1}{2}}(U))/\Delta y\end{aligned}$$

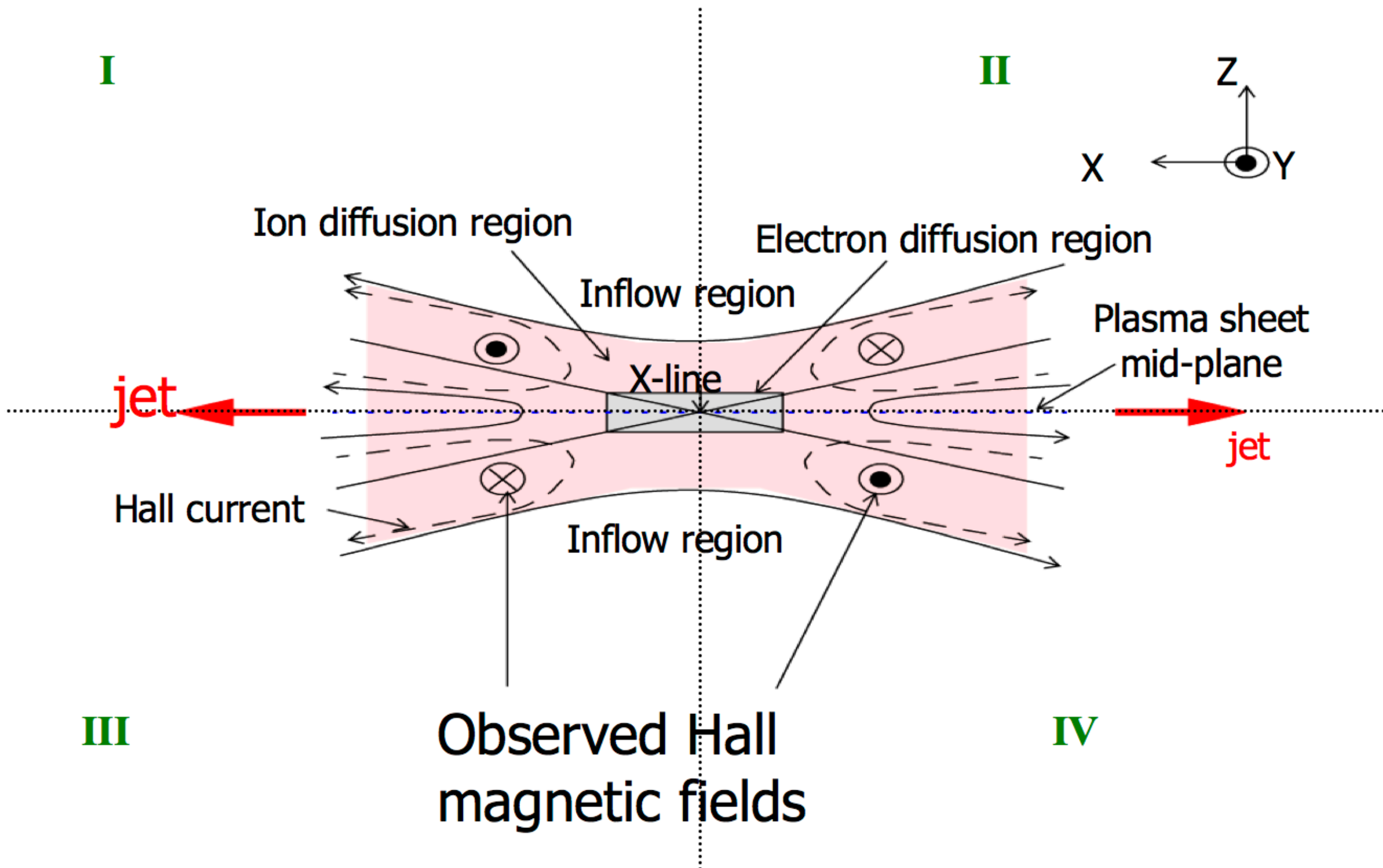
$$\begin{aligned}\Delta x \frac{\partial U}{\partial t} &= -(f_{i+\frac{1}{2}} - f_{i-\frac{1}{2}}) \\ &+ a_1(\Delta x)^2 \frac{\partial^2}{\partial x^2} F(U) + b_1(\Delta x)^3 \frac{\partial^3}{\partial x^3} F(U) \\ &+ a_2(\Delta x)^4 \frac{\partial^4}{\partial x^4} F(U) + b_2(\Delta x)^5 \frac{\partial^5}{\partial x^5} F(U) + \ldots\end{aligned}$$

Which equations are really solved?

- Minimizing these numerical errors is the “holy grail” of numerical modeling (Not just in space science).
- Error terms scale with some power (p>1) of the cell size, thus high resolution is desirable or even necessary for useful modeling. Consequently, computer power is an important ingredient for GGCM modeling.
- High numerical dispersion leads to “wiggles” that can make a code unstable. There is no known numerical scheme that is dispersion free. Numerical dispersion mostly affects discontinuities (shocks, RDs, TDs).
- Numerical diffusion counter-acts numerical dispersion by smoothing the grid oscillations.
- Numerical diffusion tends to make solutions often “look better”.
- Numerical diffusion “miraculously” lead to proper shocks (which exist in ideal MHD only as a singularity, or a “weak solution”!).
- Numerical diffusion also leads “miraculously” to magnetic reconnection.
- However, numerical diffusion can also lead to unphysical solutions because the fluid and magnetic Reynolds numbers in space plasmas are very high, much higher than the numerical ones -- except maybe in very limited regions.
- “Modern” schemes (FCT, TVD, ENO, ..) attempt to be “intelligent” and try to introduce only as much diffusion as necessary to counter-act dispersion.

Reconnection

- Magnetic reconnection drives the dynamics of the magnetosphere.
- Reconnection is usually envisioned as a X-line process (with or without a guide field). In this picture the fundamental property is the reconnection rate, which appears to be determined by the properties of the diffusion region:



Reconnection

- "GEM reconnection challenge" questioned MHD reconnection rates. It has been argued that only models that include the Hall term (full particle, hybrid, Hall-MHD) produce proper reconnection rates.

Generalized Ohm's law (dimensionless form)

$$\mathbf{E} + \mathbf{v} \times \mathbf{B} = \frac{1}{S}\mathbf{J} + d_e{}^2 \frac{d\mathbf{J}}{dt} + \frac{d_i}{n}\left(\mathbf{J} \times \mathbf{B} - \beta_e \nabla p_e\right)$$

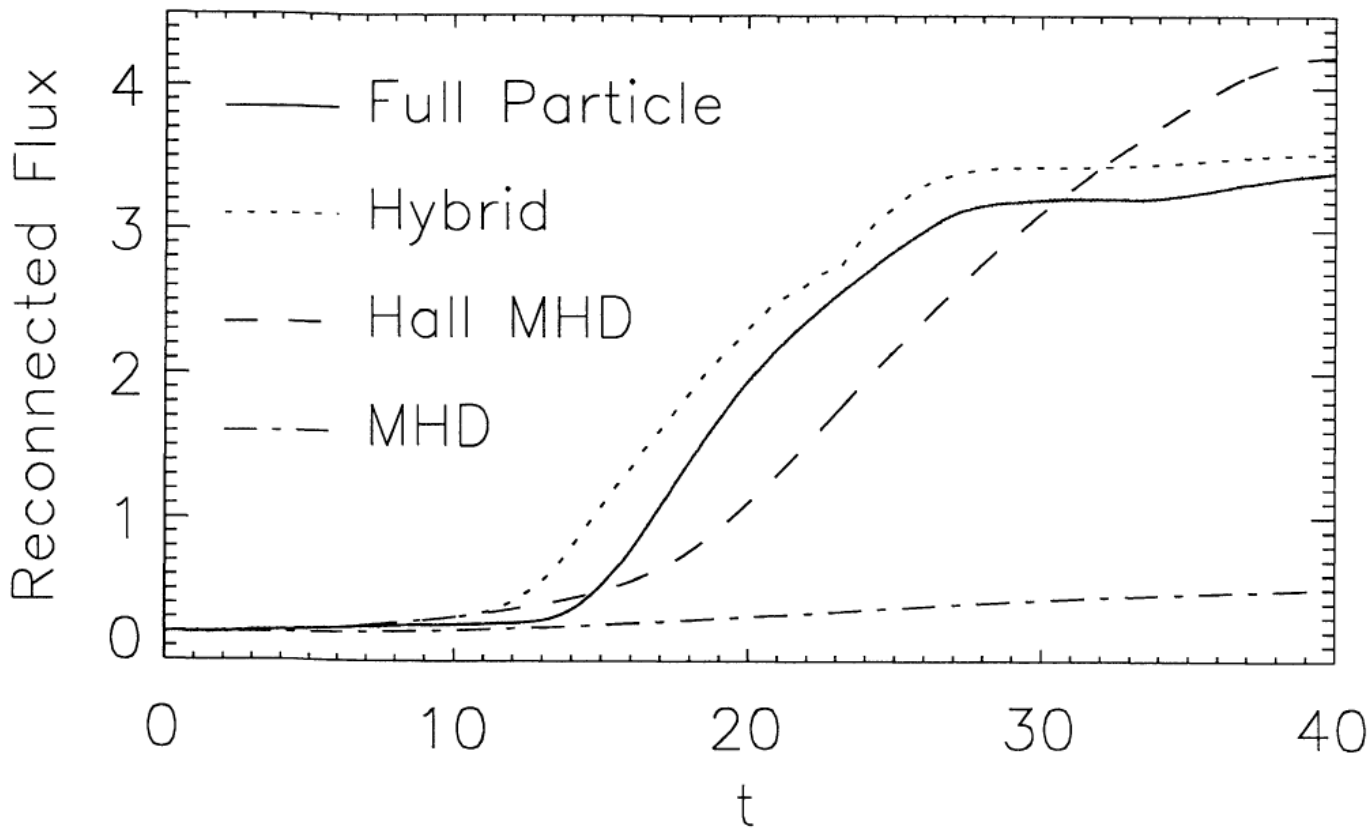
Electron skin depth $\quad d_e \equiv L^{-1}\left(c/\omega_{pe}\right)$

Ion skin depth $\quad d_i \equiv L^{-1}\left(c/\omega_{pi}\right)$

Electron beta $\quad \beta_e$

Reconnection

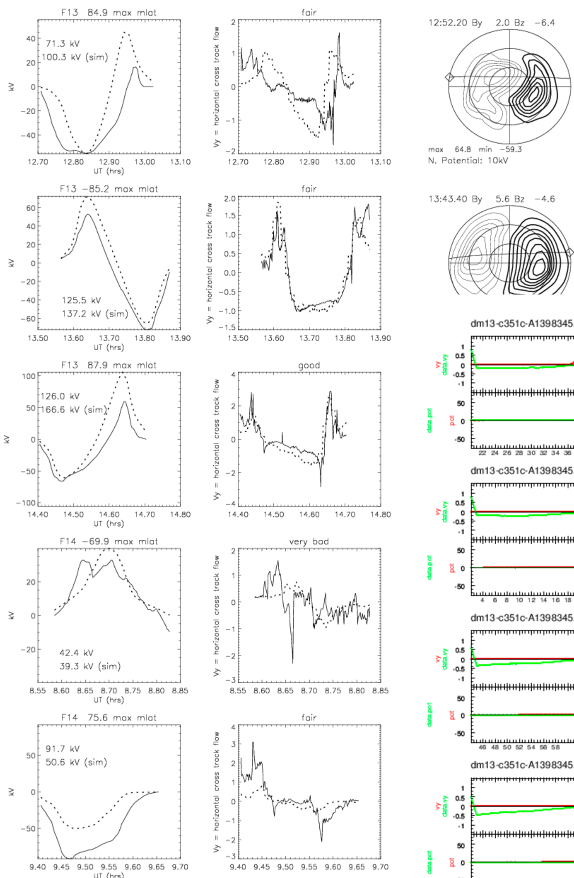
- The principal “GEM reconnection challenge” result: only models that include the Hall term (full particle, hybrid, Hall-MHD) produce proper reconnection rates, resistive MHD does not.



Birn et al., 2001

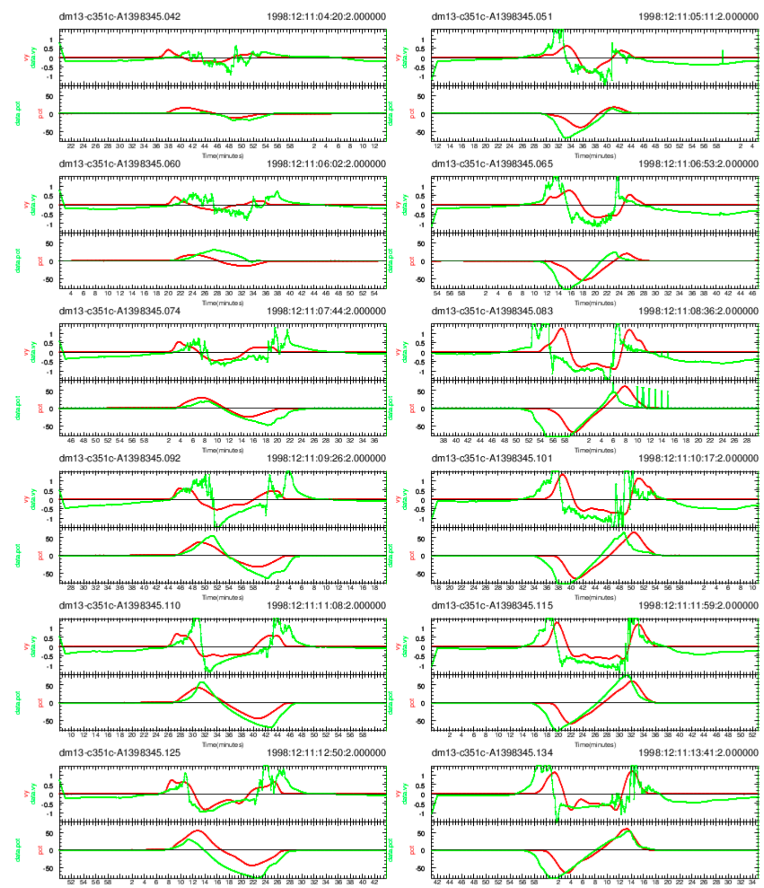
GGCM Dayside Reconnection I

- Judging from the reconnection challenge results the reconnection rates in the GGCMs should be way too low.
- However, in GGCMs we can measure reconnection rates and compare with observations.
- The magnetopause reconnection rate in GGCMs is quite accurate, often within the error bars of measurements, but rarely worse than a factor of 2 off. If the GGCM rate is off it is usually to high.
- Example comparisons:



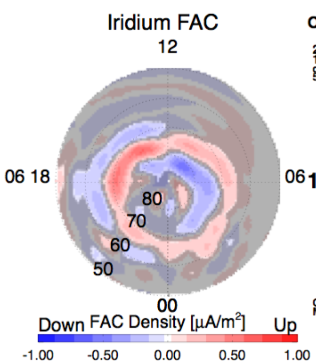
<== LFM, Slinker, 2003

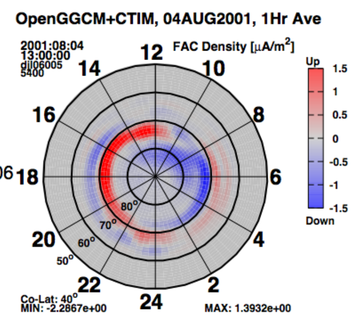
OpenGGCM, 2003, GEM Electrojet Challenge ==>

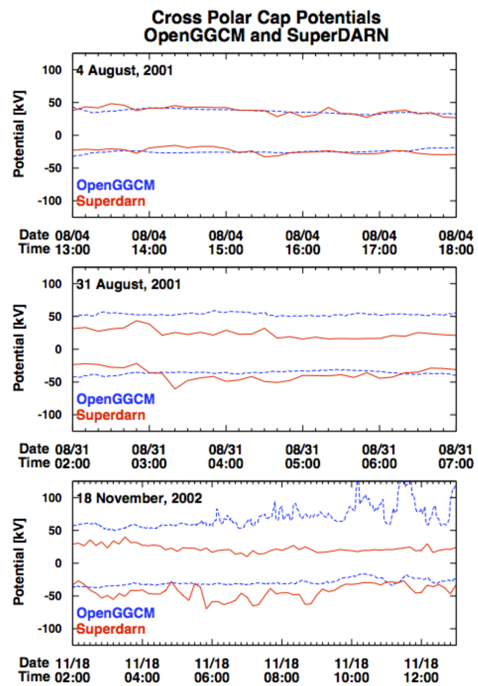


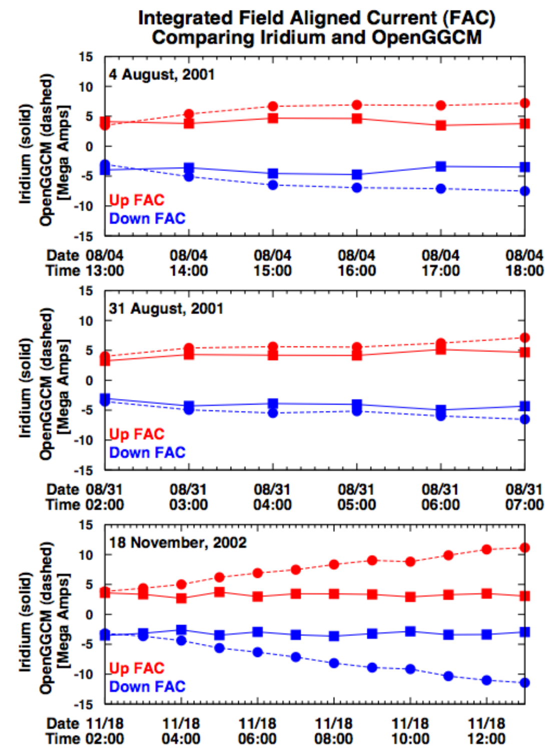
Dayside Reconnection I

- Similar relations hold for the FAC, which is generated by reconnection.
- Example comparisons IRIDIUM:



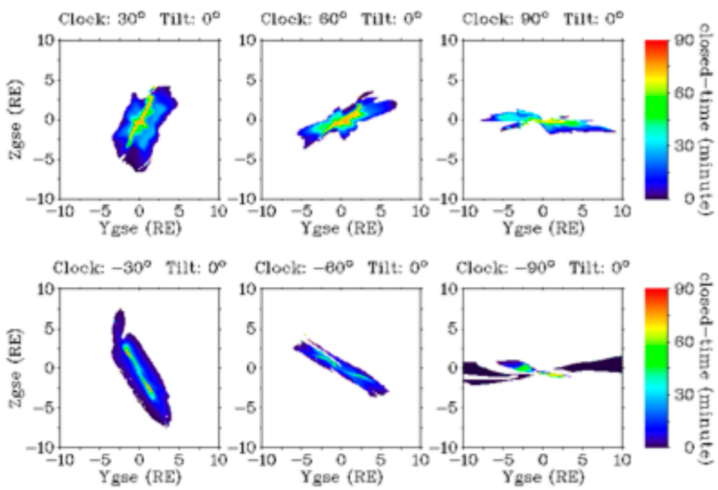




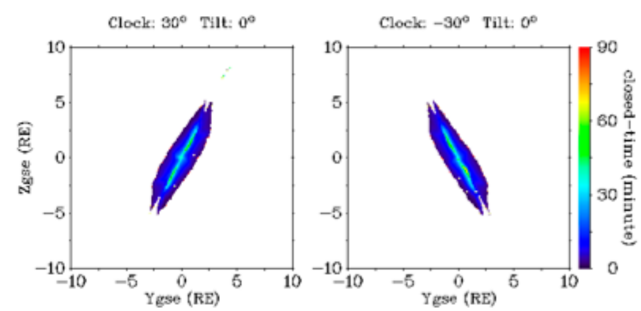


What gives?

- GGCM results and 2d reconnection-in-a-box results are at odds.
- Implication: reconnection rates are principally determined by boundary conditions, not the microphysics.
- Other evidence:
  - Siscoe-Hill model of CPCP saturation depends explicitly on ionosphere conductance, but not on microphysics.
  - Recent plasma entry simulations show striking dependence on ionosphere parameters (next slide).



CTIM conductance: Plasma entry windows are not symmetric wrt clock angle

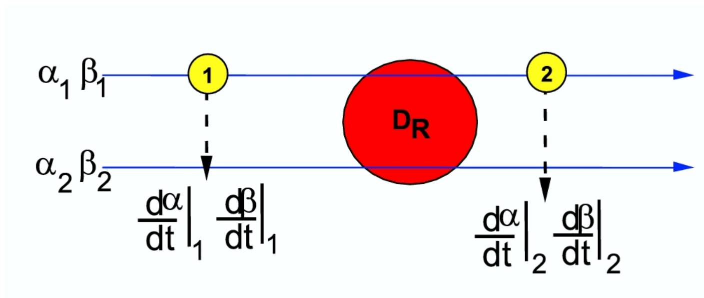


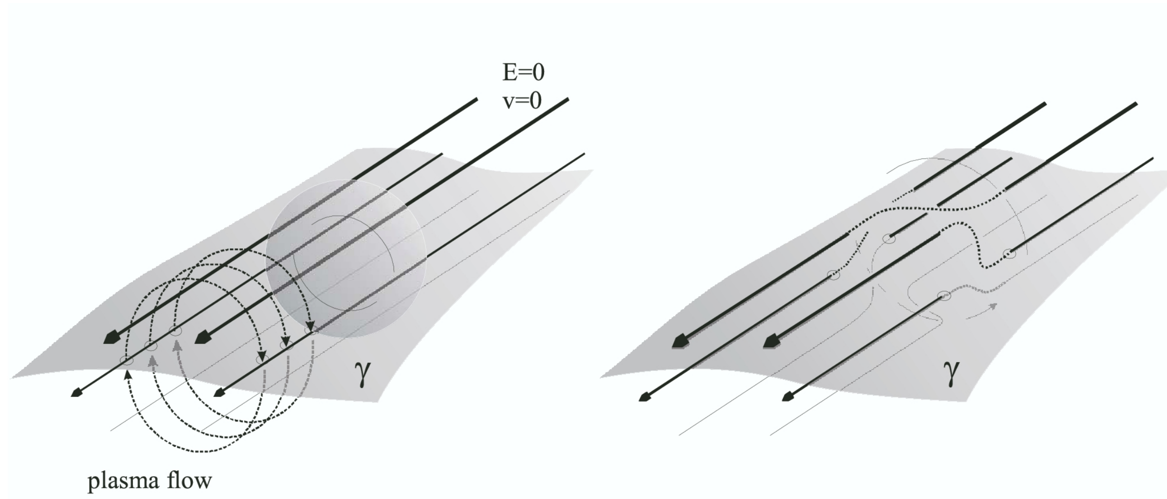
Uniform conductance: Plasma entry windows are symmetric (W.Li thesis, 2007)

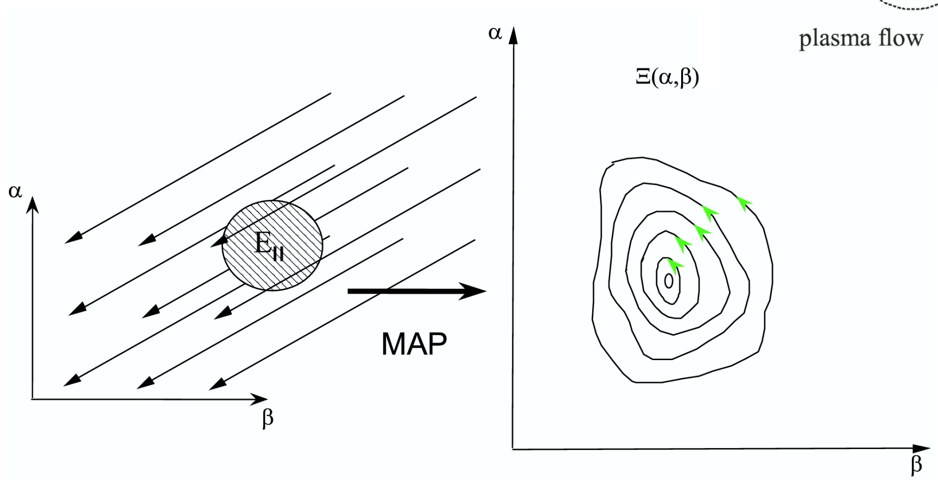
Dayside Reconnection II

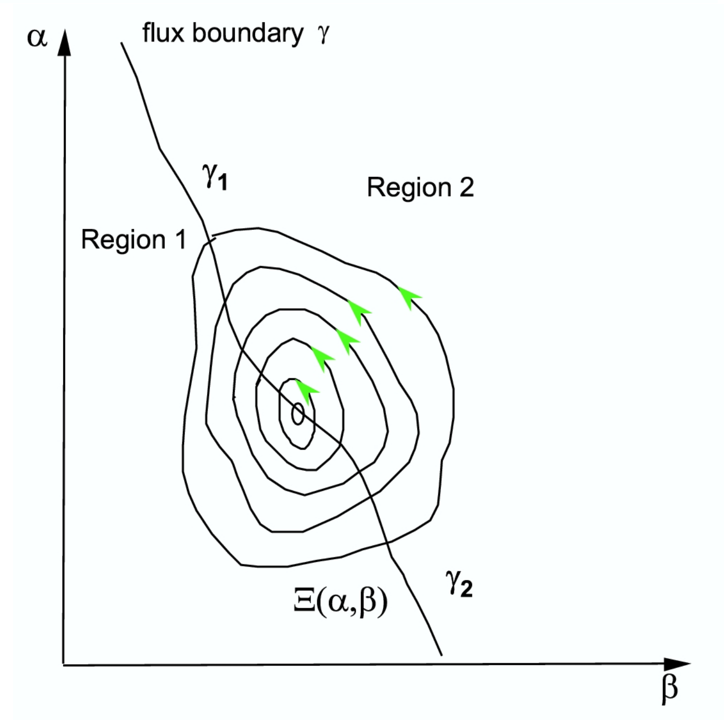
- What are the questions?
  - Anti-parallel or component reconnection? (Is this even a good question?)
  - Better: where does reconnection occur? --> need to define what reconnection is in 3D!
  - How does it occur?
  - What does control the rate if it is not the micro-physics?
- How can GGCMs address these questions?
- Staying close to the data always helps: FTE example.

What is Reconnection and how to Measure the Rate





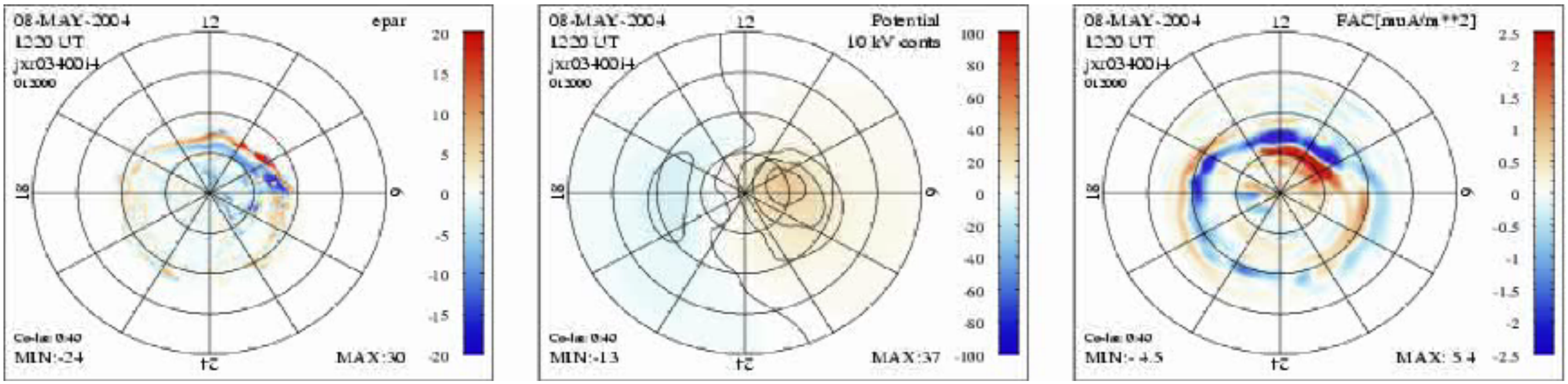




Hesse, Forbes & Birn, (ApJ, 2006): reconnection rate ~ maximum integrated parallel electric field Ξ(α,β).

Relation of Ξ, Φ and J_par

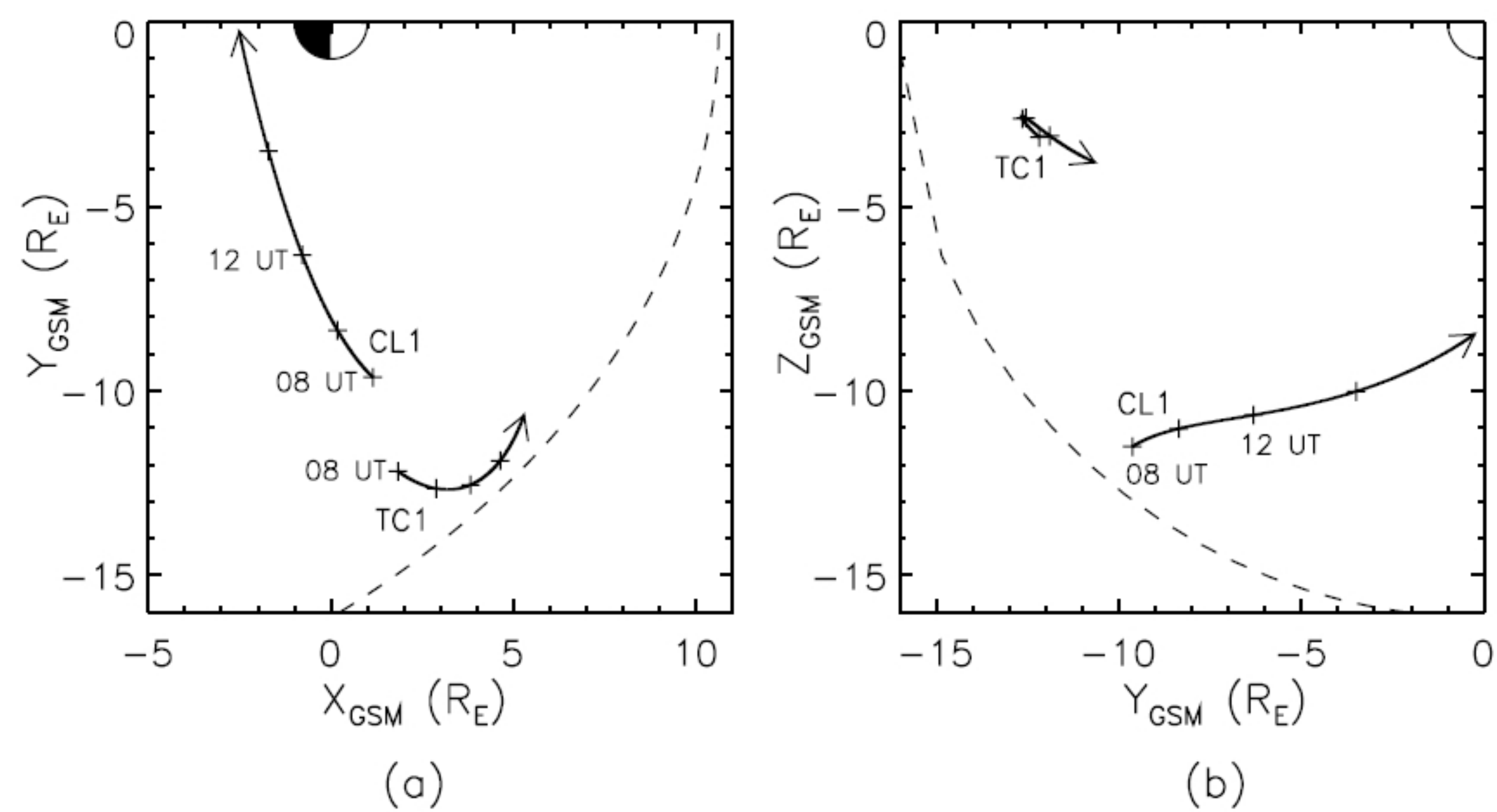
In the magnetosphere: Ionosphere provides natural Euler potentials: map epar = int_E_par into the ionosphere.



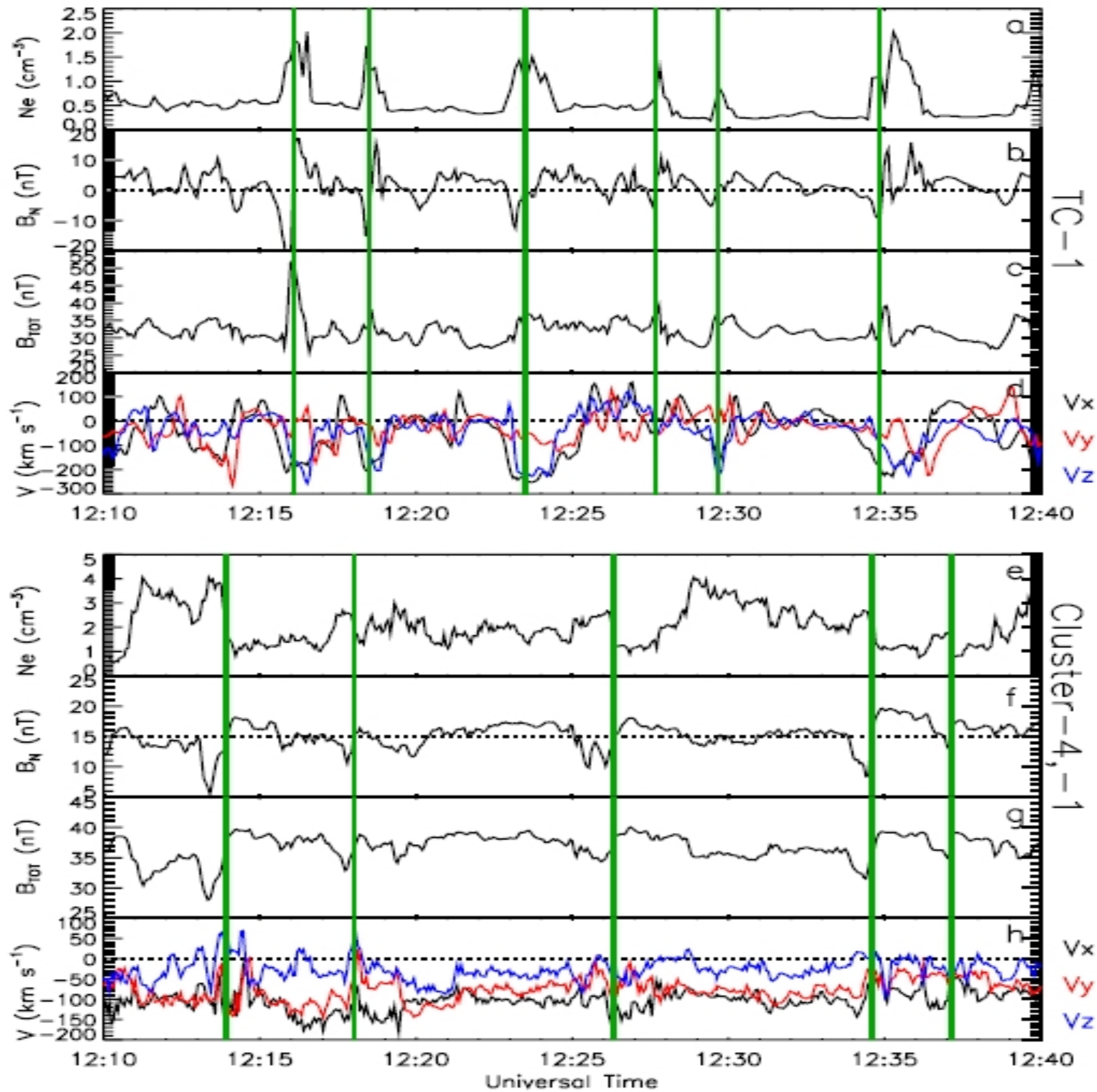
- Largest E_par occurs at pc boundary
- Also substantial E_par away from pc boundary
- The maximum in Ξ closely matches cross pc potential

Cluster-DS1 FTE Event May 8, 2004

Cluster-DS1 Locations:



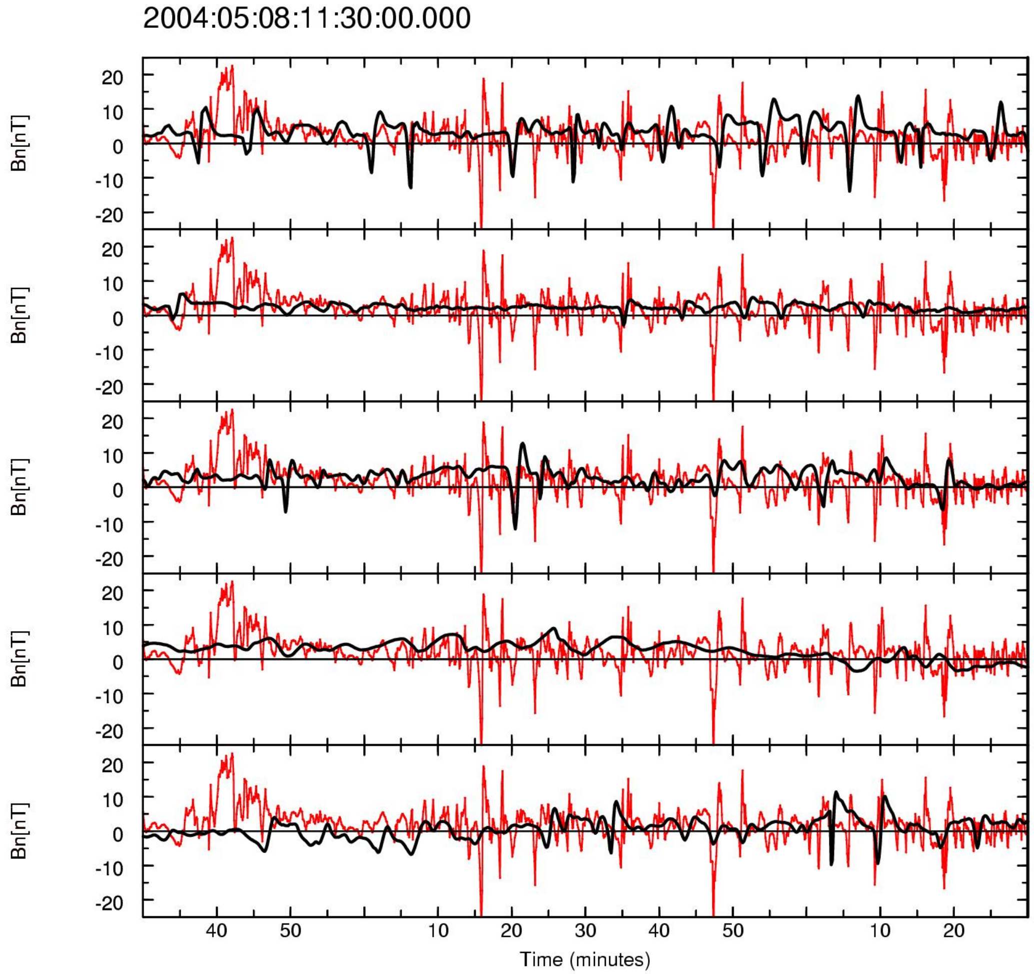
Doublestar (top) and Cluster mag field (boundary coordinates) and velocity:



FTE Simulation - data comparisons

Parameter dependence:

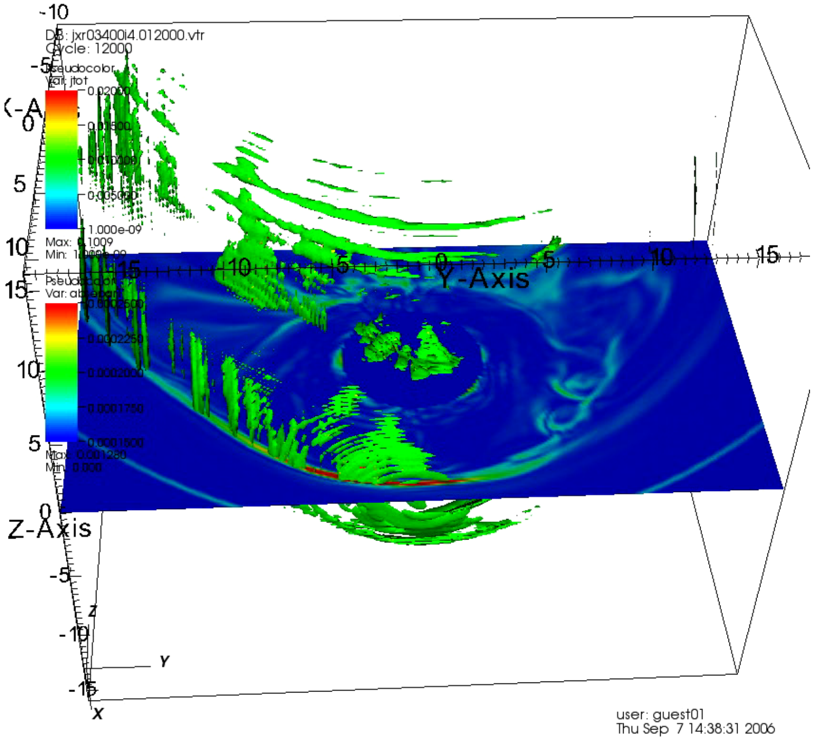
For lower resolution or larger values of resistivity FTEs become smaller, less frequent, or disappear (too diffusive).

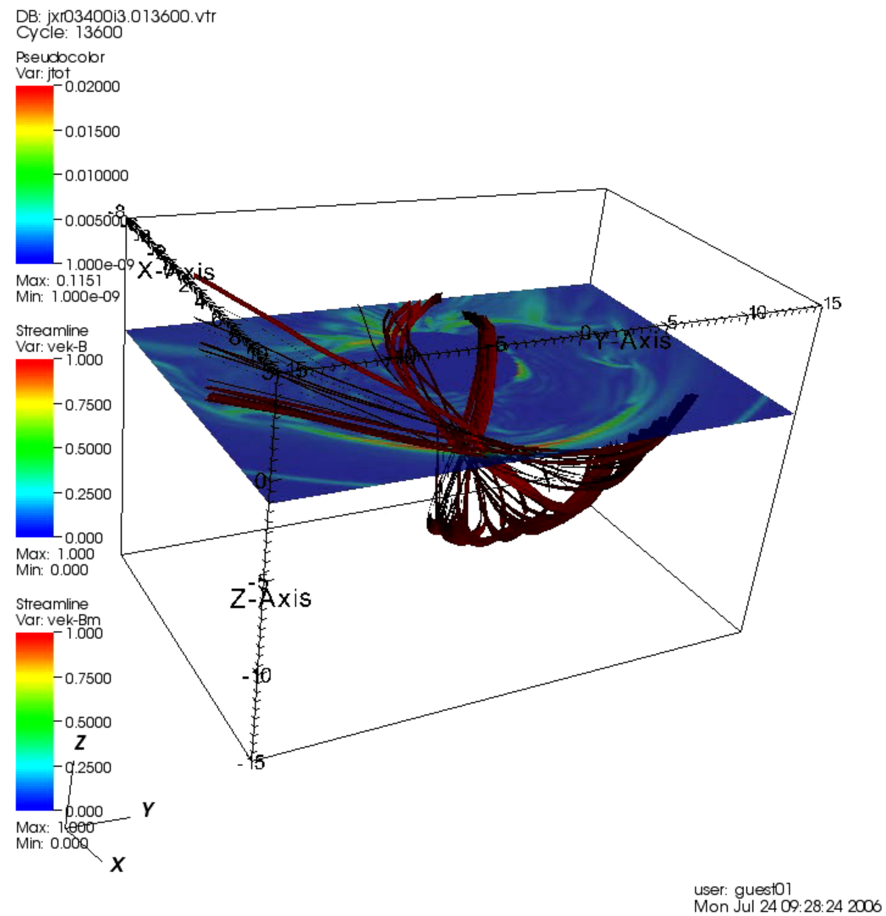


Key lessons:

- Diffusion error terms need to be kept down for an acceptable result
- Without a detailed data comparison one would never know whether the model bears any relation to reality!

FTE Parallel Electric Fields





Equatorial plane: current density

green isosurface: parallel E

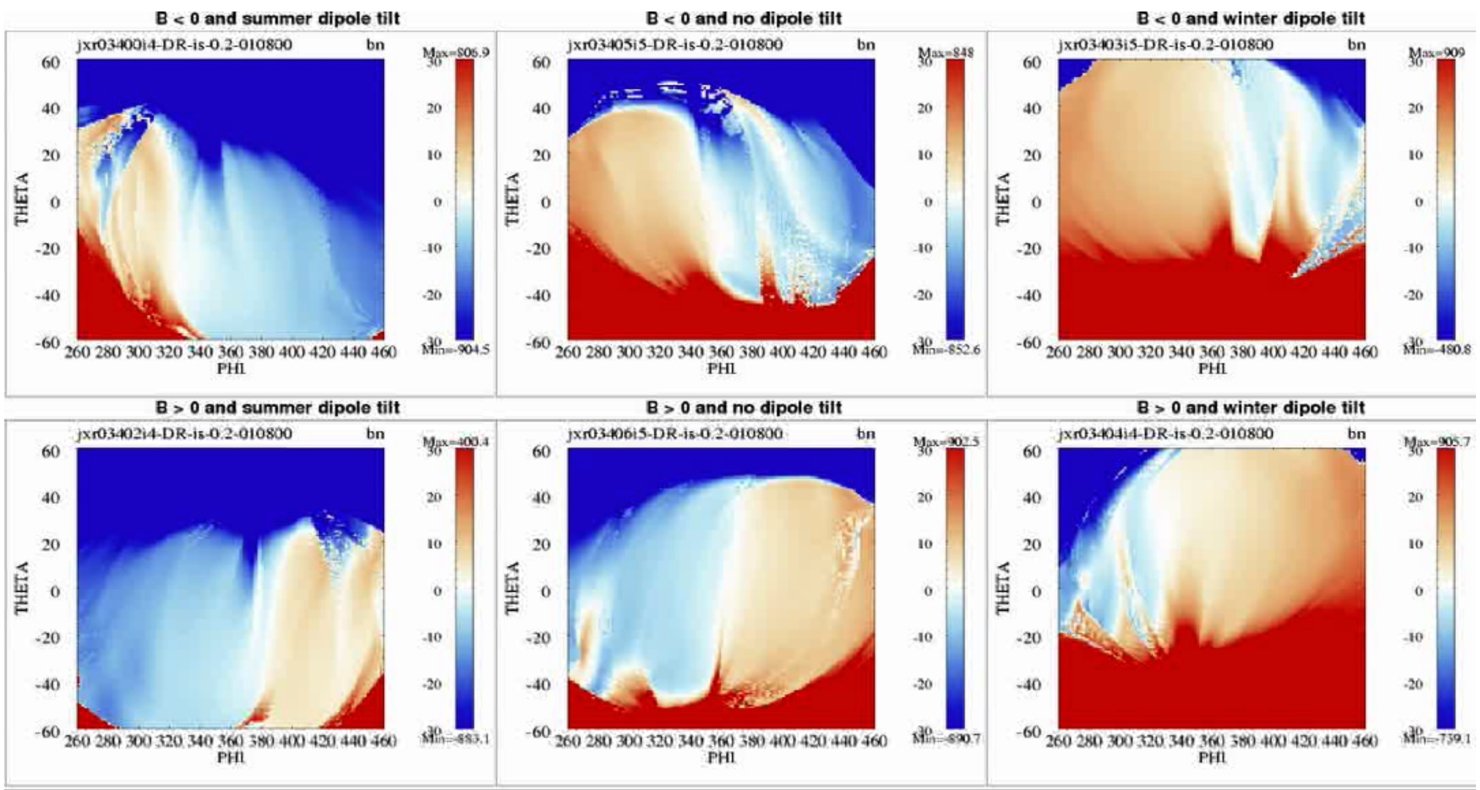
FTE generation near subsolar!

no FTEs on dusk side! (B_y - tilt effect, next slide)

E-par mostly trails FTEs.

FTE By and Tilt Dependence

- 6 runs: (+By,-By) and (summer,zero,winter) tilt.
- Color coded: normal magnetic field on the magnetopause (approximately).
- FTE hemisphere depends on tilt and sign of By:
  - dawn FTEs only for summer/-By and winter/+By.
  - dusk FTEs only for summer/+By and winter/-By.
  - FTEs in both hemispheres if tilt ~=0.
- Contrary to all existsing models!

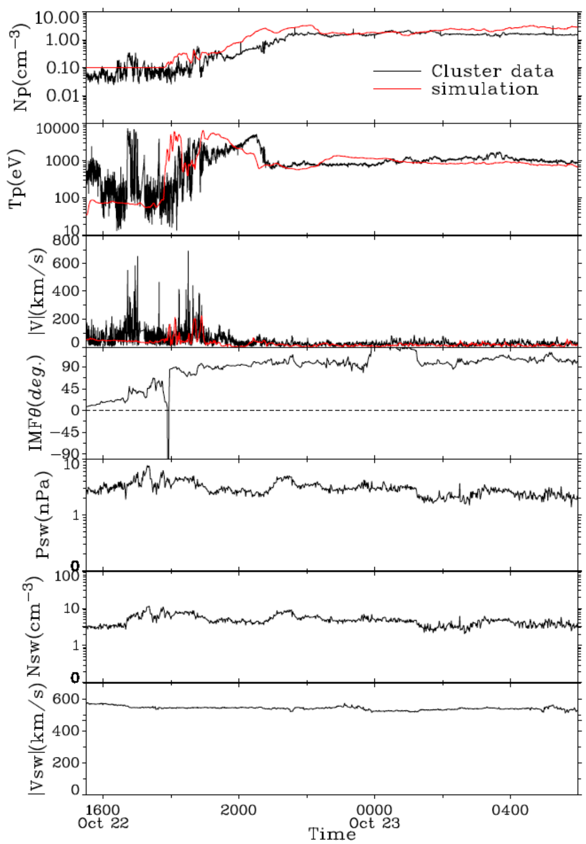


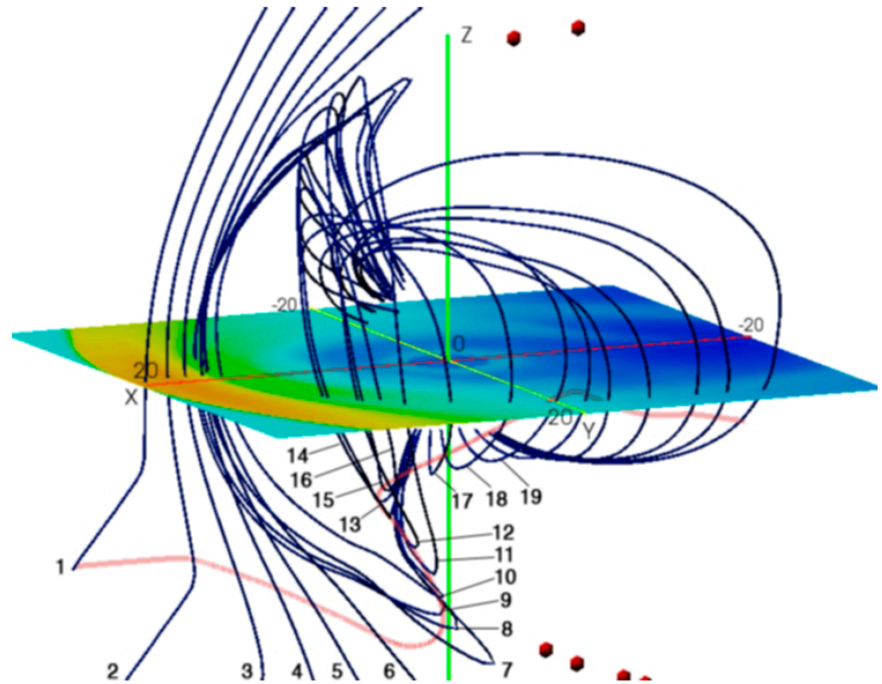
Dayside Reconnection: Where to go?

- GGCMs do well in reproducing dayside reconnection (even if the microphysics may be wrong).
- This opens the door to study processes in more detail:
  - Dependence on SW/IMF conditions.
  - Dependence on geophysical conditions (dipole tilt, conductance, winds, …).
  - Spacecraft signatures, for example FTE fringes.
  - Ionospheric signatures, for example PMAF, flow bursts, …
  - Conditions for steady versus time dependent reconnection.
  - …..
- Data comparison is essential for firm conclusions.

Reconnection - Convection

- Also for northward IMF, see formation of the cold dense plasma sheet by reconnection. GGCM does not just produce reconnection well but also associated convection.
- Example CDPS formation:

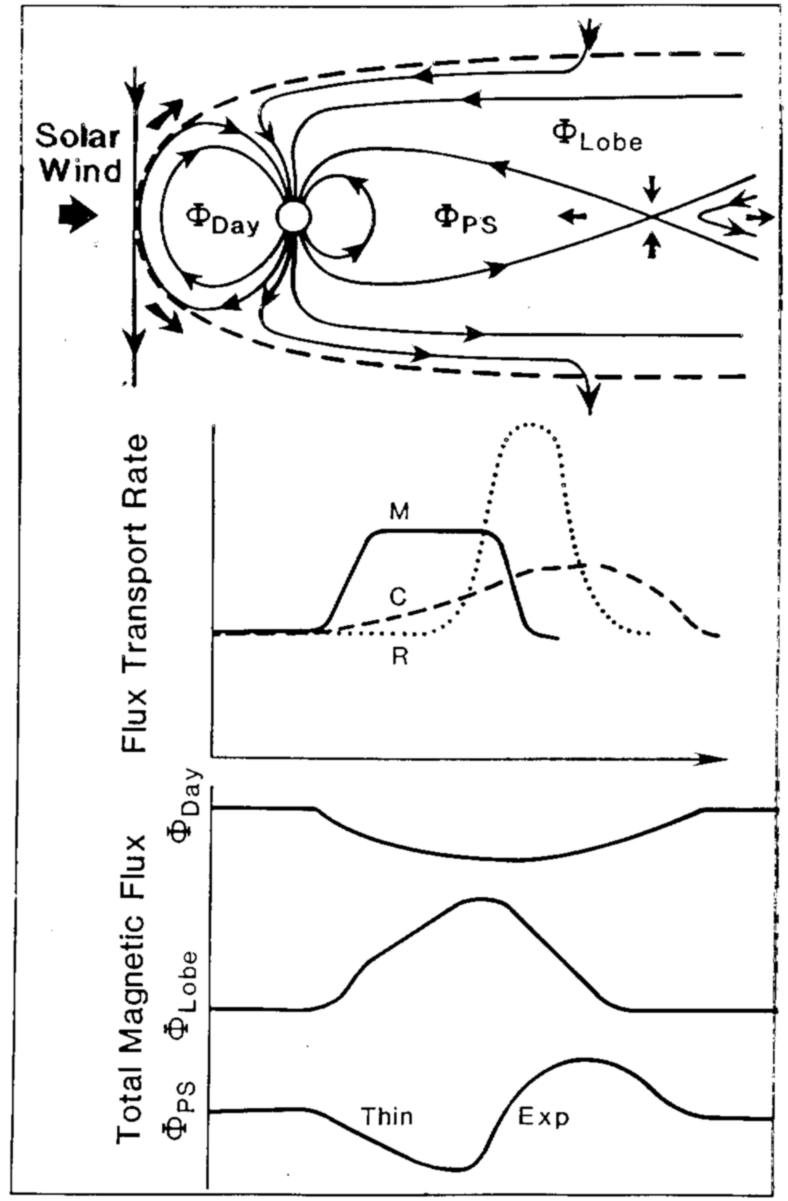




Oieroset et al.; Li et al., 2005

Convection and Substorms

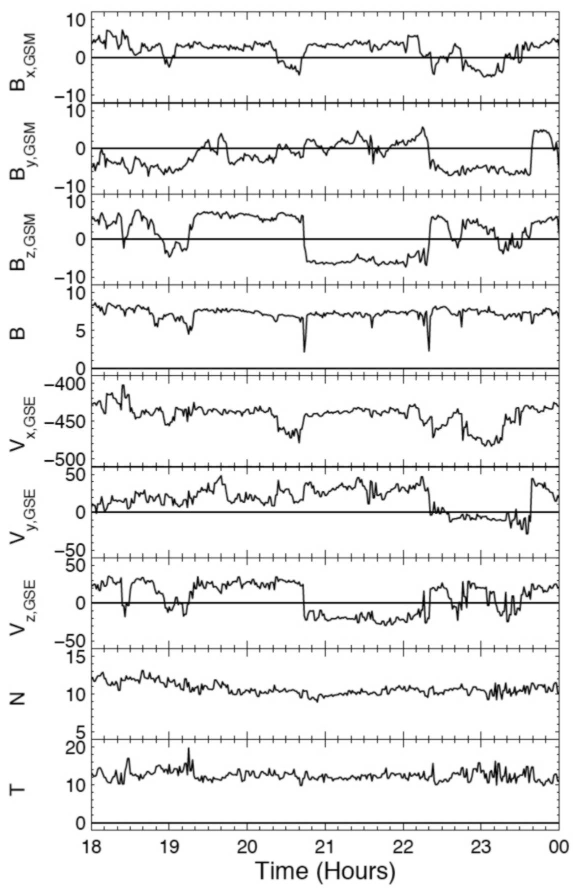
- Nightside reconnection rate more difficult to assess.
- However, the nightside rate must balance dayside rate, at least on long scales (>1h).
- Nightside reconnection rate in global models is quite accurate on average but we know much less about accuracy of short scales.
- Imbalances between dayside/ nightside rate lead substorms:
- If rates balance: steady magnetospheric convection (SMC).



Russell, 1993

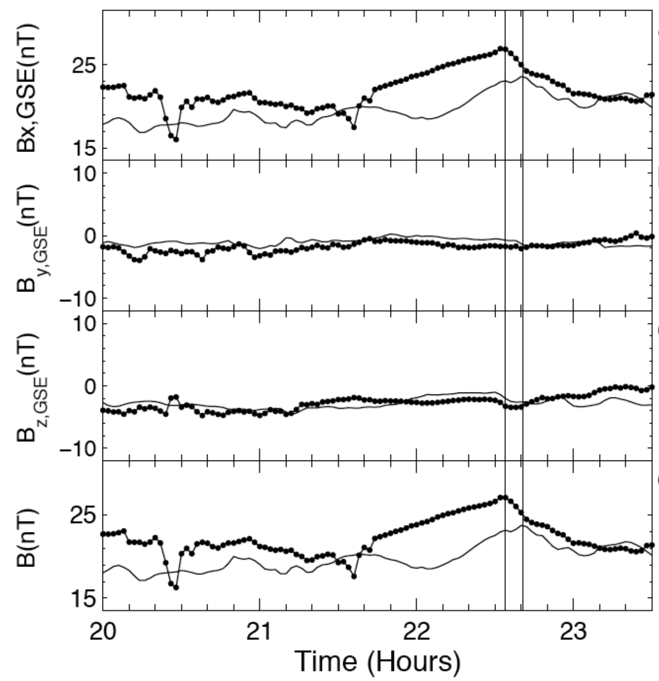
Substorms I

- Only Phi_lobe can be observed by monitoring lobe B.
- Example: GEM substorm challenge:



Wind SW/IMF

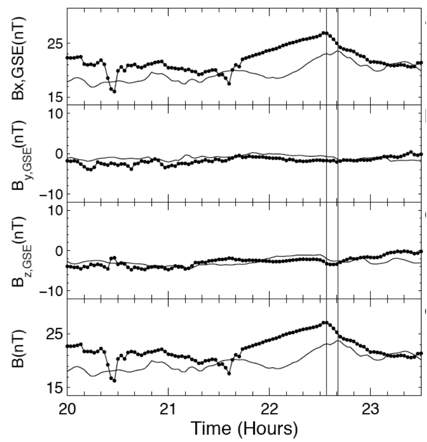
IMP8 in n-lobe shows loading with flux/energy and unloading -->



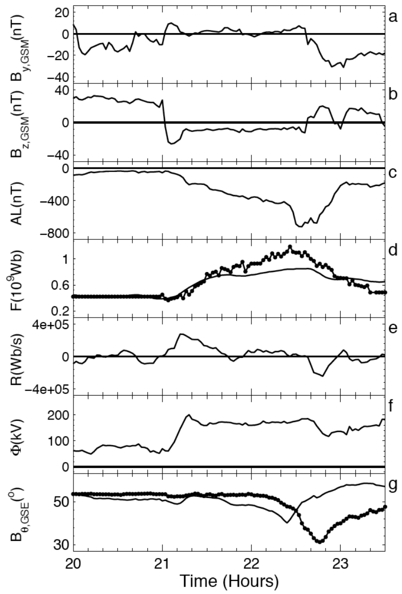
From Raeder et al., JGR, 2001

Substorms I

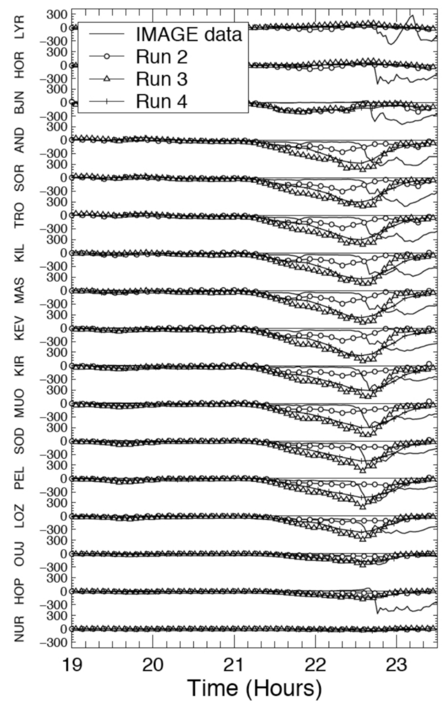
- Do GGCMs actually produce substorms?
- GEM substorm challenge (2001!): very mixed results:



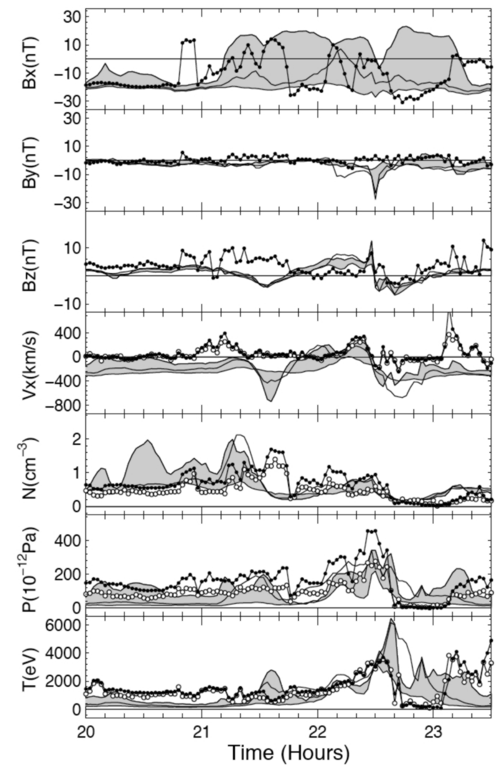
Lobe flux: so-so. Too little loading - unloading.



PC flux, dipolarization: so-so.



Electrojet: parameter dependent.



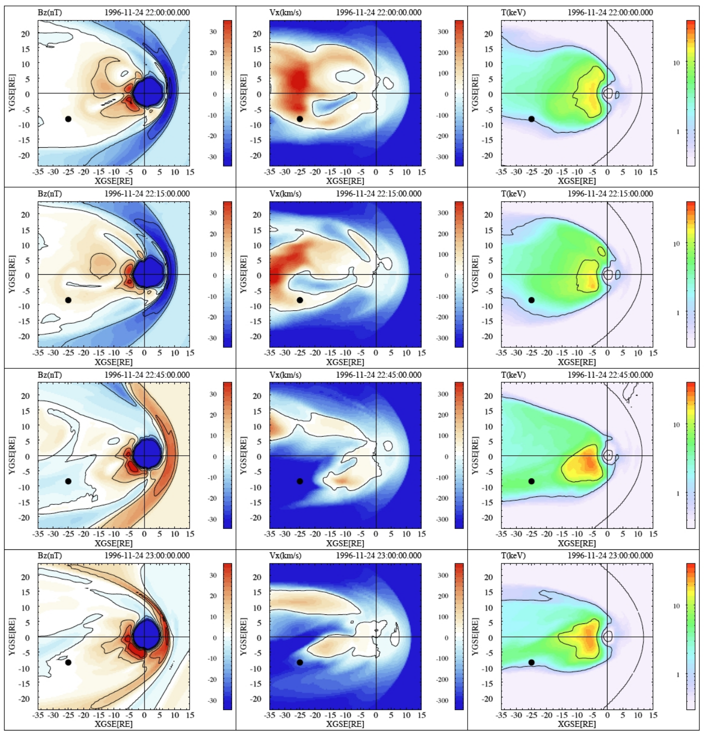
Plasma sheet flows, thinning, Bz: reasonable.

From Raeder et al., JGR, 2001

Substorms I

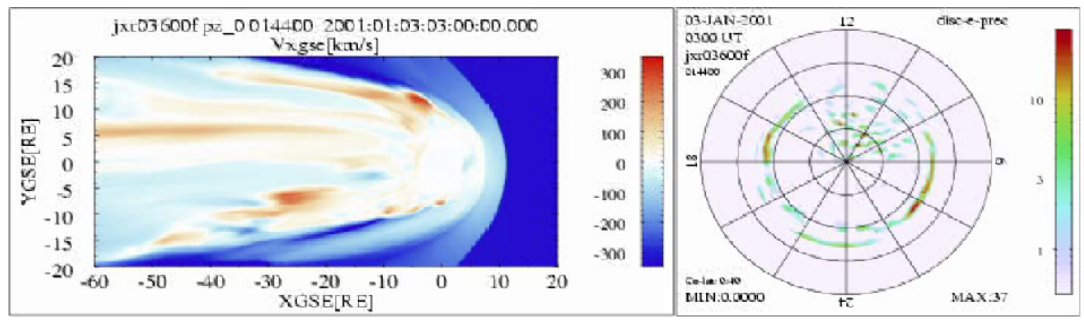
- What have we learned from GGCMs?
- Tail reconnection is much more “fragmented” than “old” models would suggest:

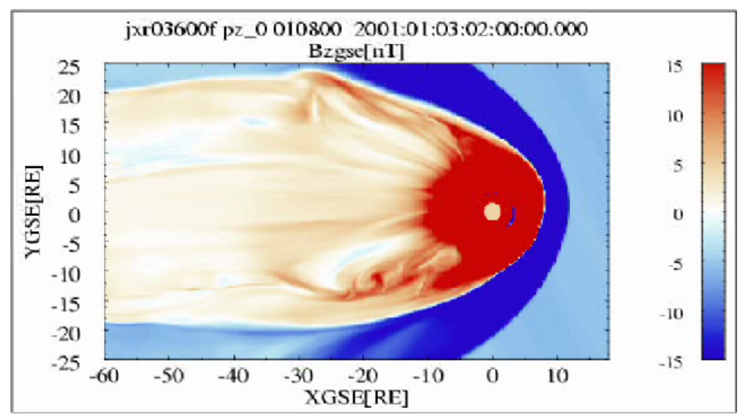
a)
b)
c)
d)



Bz, Vx, Temp

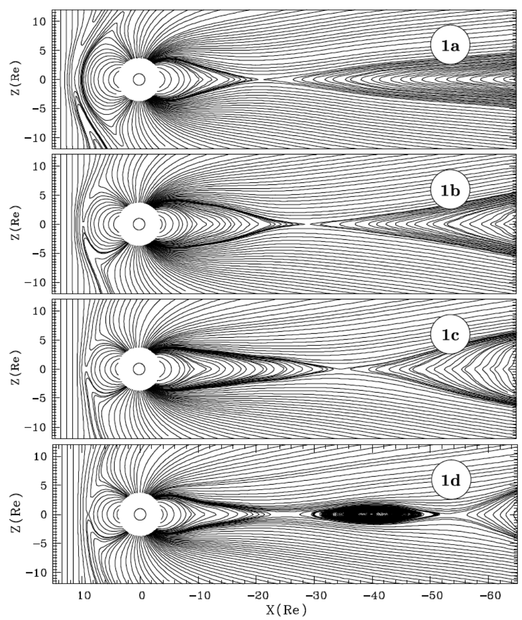
And not just during substorms: BBFs and PBIs:





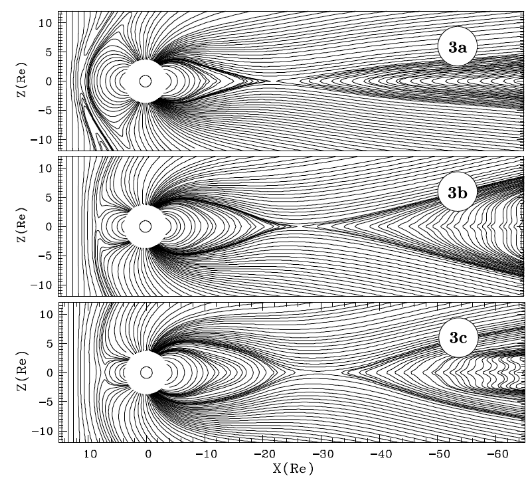
Substorms II

- Where can the GGCMs take us with substorms?
- Today, no GGCM reproduces all aspects of a substorm.
- A principal problem --> GGCMs reconnect too much in the tail, SMC is more likely than loading - unloading.
- A principal question: What controls reconnection in the tail:
  - If dayside is a guide, b.c. and ionosphere should have a major influence.
  - OpenGGCM results point to both E_par and ionosphere.
  - High ionosphere conductance can suppress convection entirely:



<== Normal conductance

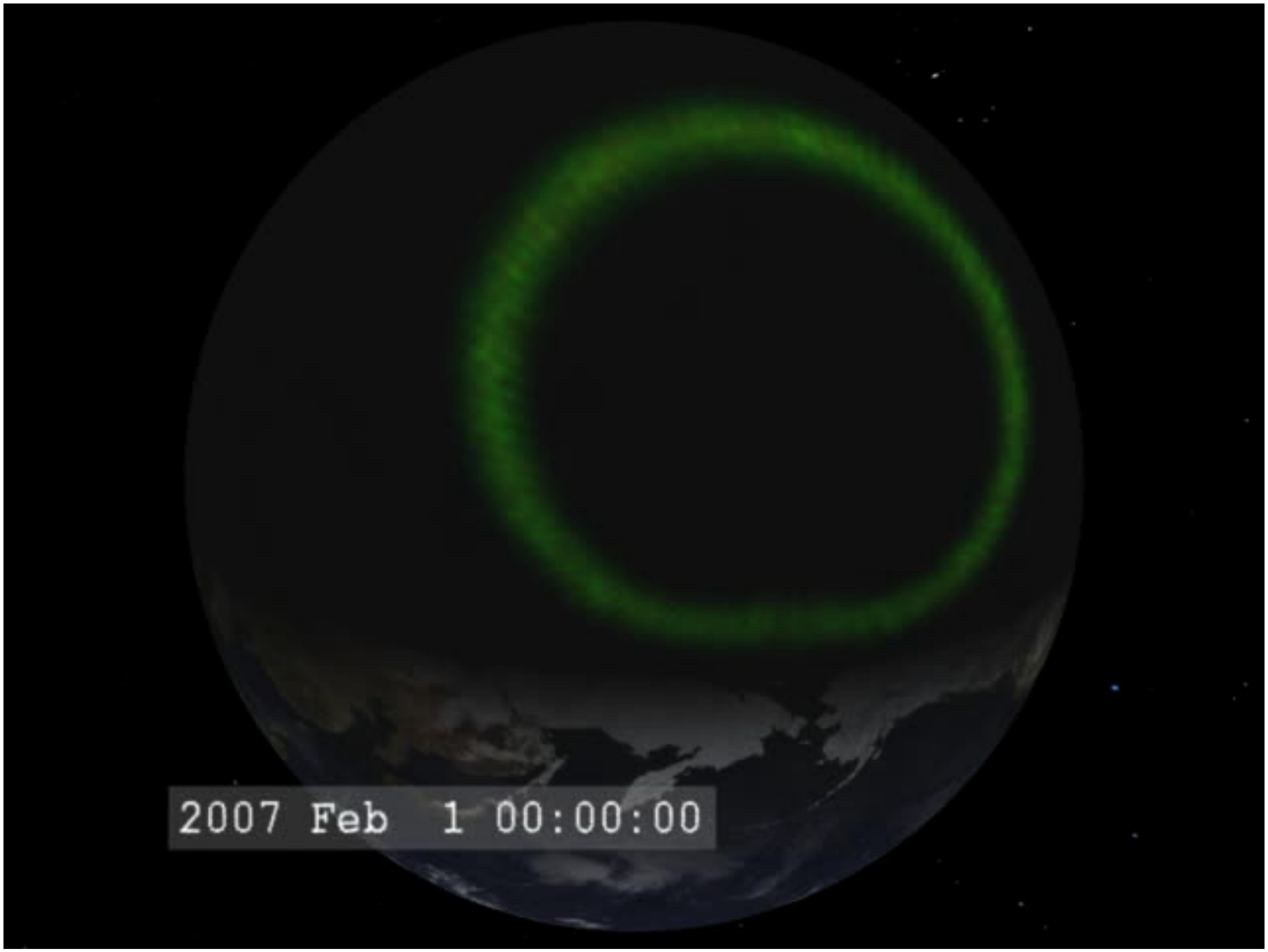
High conductance ==>



Substorms III

- What can/should GGCMs do with substorms?
  - Need to push resolution to suppress numerical diffusion.
  - Need to think about other terms in Ohm's law and parameterization.
  - Need to look closely at ionosphere coupling.
  - Need to include inner magnetosphere p/e-.
  - There is much room for numerical experiments.
- What do we need to look out for?
  - Other processes than reconnection: interchange, balloning, ….
  - The first substorm signature (possibly pointing to the trigger) is brightening of the equatormost arc!?
  - An IMF northward turning triggers a substorm (~50%)!
  - A blob in the tail does not make a substorm! Model must produce: electrojet, dipolarization, aurora/WTS, northward expansion, injection, energy/flux storage-release, fast earthward/tailward flows.
  - The model alone can not be trusted, it must be corroborated with data, but that is what GGCMs are for! And THEMIS, of course:

THEMIS



Computing: The Old Days...

15 years ago one would have needed something like a Cray Y/MP:

- 8 processors
- ~ 1 GB memory
- ~ 8 x 0.2 Gflops
- ~ $10M
- ~ 200 kW power/cooling
- ~ $1M/y operations.

First global models ran on such machines.

Photograph courtesy of Charles Babbage Institute, University of Minnesota, Minneapolis

Computing:

1992 Y/MP block diagram:

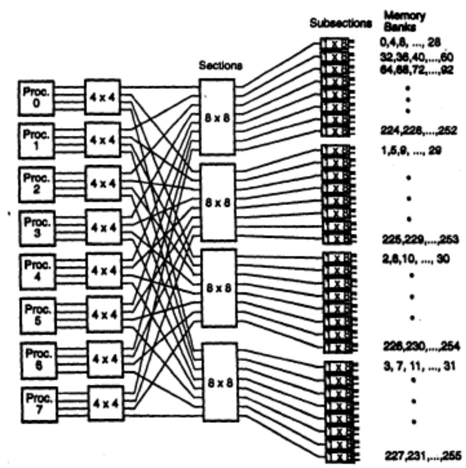


Figure 8.10 Schematic logic diagram of the "crossbar" network between 8 processors and 256 memory banks in the Cray Y-MP 816.

Hwang

Computing: Enter the Cell Chip

1992 Y/MP block diagram:

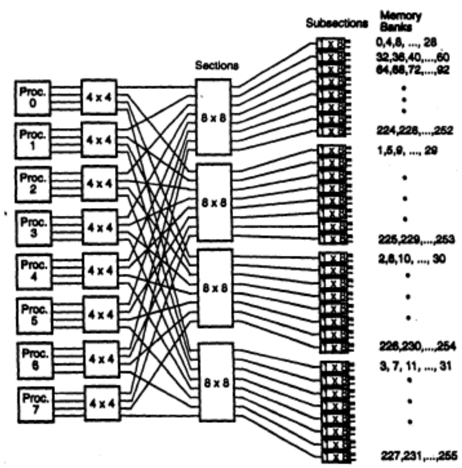
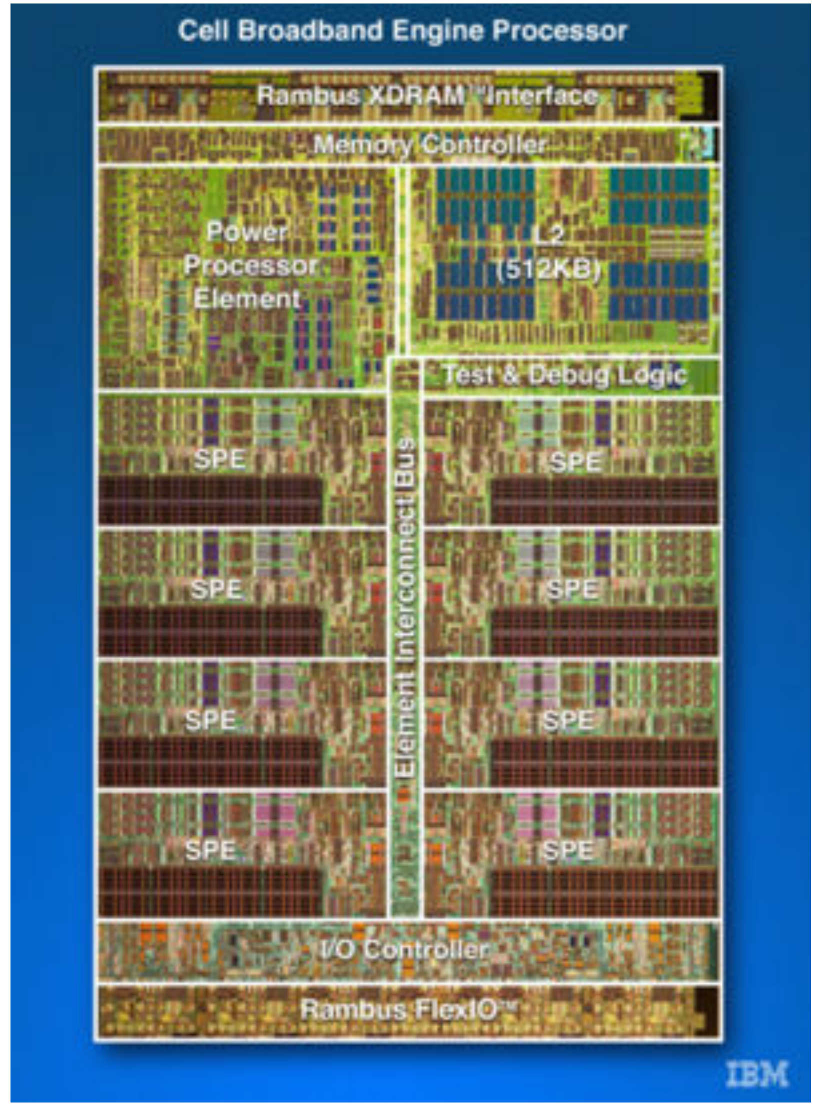


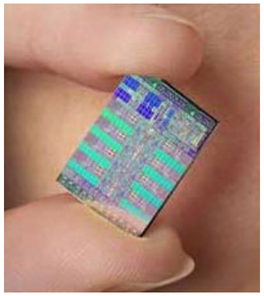
Figure 8.10 Schematic logic diagram of the crossbar network between 8 processors and 256 memory banks in the Cray Y-MP 816.

Hwang

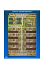
2006 Cell Broadband Engine (CBE) chip (IBM/Sony/Toshiba):

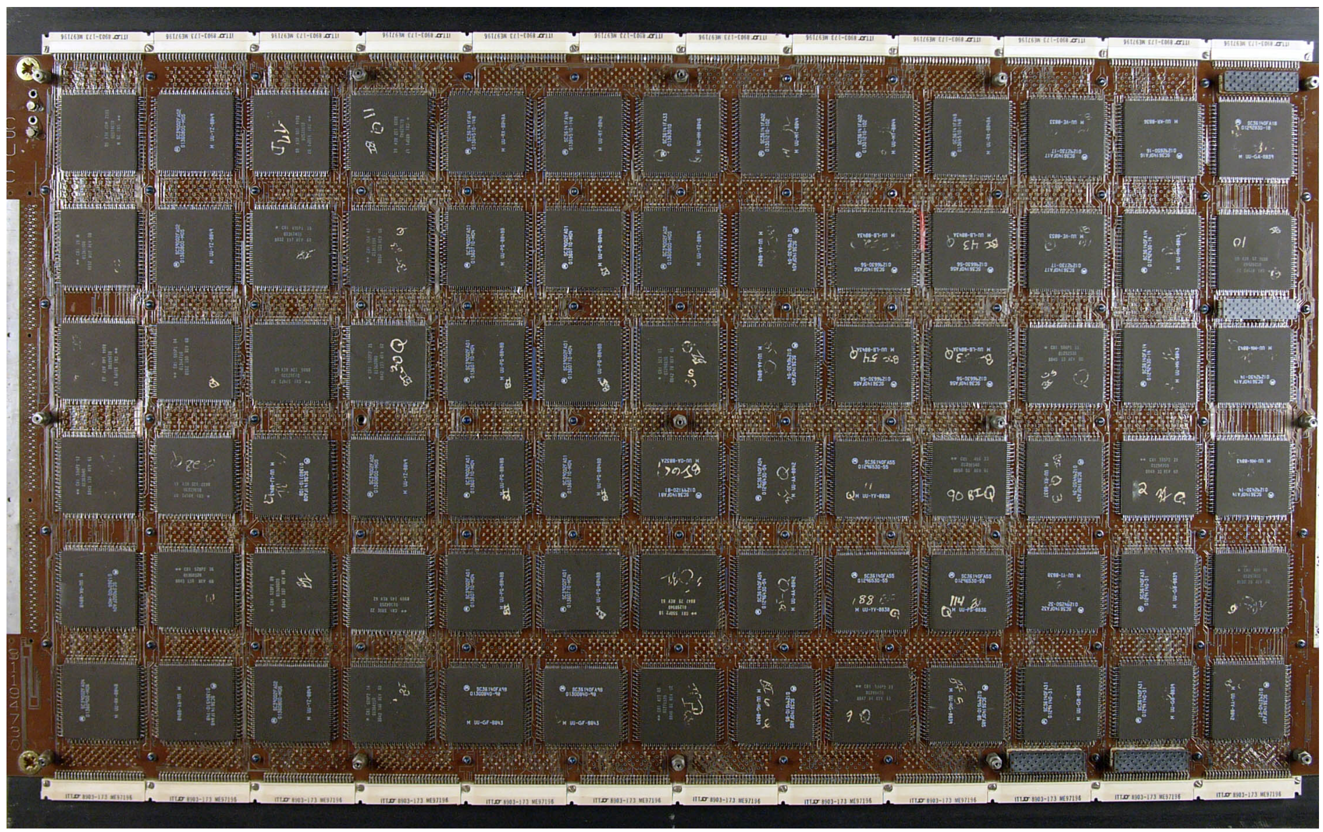


Computing: Cray on a Chip



Cell chip versus Y/MP processor board
(2 ft wide)





Computing: Comparing the Vitals

	Y/MP	Cell chip
Processors:	8	1+8(6)
Memory:	0.25-1.0 GB	0.25 GB
Peak speed:	~ 2 Gflops	~ 256 Gflops
Memory bw:	2 GB/s	25.6 GB/s
Clock:	0.333 GHz	3.2 GHz
Power:	100 kW	200 W
OS:	UNICOS	Linux
Price:	$10M	599.00 @Walmart
Compilers:	Crap back then	Crap now

Computing: The WalMart Supercomputer

Why NOT run the OpenGGCM on a PS3?

Computing: The WalMart Supercomputer

Why NOT run the OpenGGCM on a PS3?

Because it is a pain to program!

Computing: The WalMart Supercomputer

Why run the OpenGGCM on a PS3?

- Extremely low cost.
- Portable.
- Virtually no maintenance.
- Little expertise required: A 5th grader can do it!
- Virtually no infrastructure needed.
- Large ensemble predictions possible.
- Low “entry barrier.”
- Basically a “Space Weather Appliance.”

Computing: OpenGGCM on PS3

- Faster than real time with 128x64x64 (~0.5M cells) grid.
- Main part is still F77 (unchanged), ~ 500 lines of F77 code are recoded in C to use SPEs.
- Code is still MPI based: Multiple PS3 can be tied together via Ethernet for better resolved model.
- Current performance ~ 5-10 Gflops, still far from peak. New code is also 2 times faster on Opterons (use of SSE instr.).
- Small memory (~200MB shared, ~150kB in SPEs) is main bottleneck.
- It took ~ 3 months.
- All credit to Kai Germaschewski & Doug Larson.

Let's see it running!