A White Paper advocating a *Heliophysics Theory Mission*

Principle Authors: J. R. Jokipii(1) and Gary P. Zank(2)

(1)Distinguished Professor of Astrophysics, University of Arizona

(2)Pei-Ling Chan Chair of Physics –and- Director, Center of Space Plasma and Aeronomic Research –and- Chair, Physics Department, University of Alabama in Huntsville

1. Introduction

Theory and modeling, together with observations are the foundations of a sound, balanced scientific program in any discipline. The past few decades have witnessed a shift from a strongly exploratory and discovery-driven science to a more mature explanatory science. With its maturing, solar and space physics places increasing demands on theory. Theory provides a meaningful context for basic space physics observations, often revealing that seemingly disparate observed phenomena correspond to the same physical processes in a system(s). Critically, besides organizing and understanding observations, theory drives the prediction of new or unexpected important phenomena that in turn drive new missions.

Some 34 years ago it was realized that, within what was then called the Space Physics Division of NASA, theory and modeling was not adequately funded. At that time the Colgate Committee was formed. It found that, compared to peer disciplines such as fusion plasmas, space physics theory and modeling was significantly under-supported. The committee recommended that theory funding be increased and attention be given to the encouragement of critical-size theory groups. This recommendation led directly to the creation of the Space Physics Theory Program and indirectly raised the general awareness of the importance of a balanced program. This program still exists as the Heliophysics Theory Program, but at a significantly reduced fraction of the overall budget. This program has been an important factor in the significant progress made in the area of Heliophysics science of the past few decades, but the number of proposals supported now numbers 9 with a typical grant size of some $400k or less per year. This has decreased from 11 or 12 funded proposals just 20 years ago, and these were funded at levels that were comparable to today's (i.e., ~$400k in 1990 dollars)! There has therefore been a considerable erosion in both the number of theory groups supported in the US and in the level of individual support. With the reduction in the theory program, most theory is now being funded through guest investigator programs associated with missions or through the SR&T program. Unfortunately, the level of these programs is also being reduced, to the point that the last GI program was completely cancelled. The deleterious effect of the erosion in theory support can scarcely be over-estimated, especially in academic environments where it is now almost impossible to develop a major spacecraft or experimental program that can sometimes indirectly support the development of theory.

In significant measure, this erosion of theory and modeling funding appears to be the result of significant budget constraints combined with Mission cost growth. With respect to the latter, there appear to be several factors at work. Proposed budgets for a mission may not always reflect accurately the real costs – this may be due to overly aggressive or optimistic budgeting to ensure the selection of an instrument or mission, or it may be due to the shifting of the mission schedule and dates, or unexpected development costs, etc. NASA has frequently responded by adding further support to the mission (or perhaps de-scoping the mission, but often still with an increased budget), and the increase in mission funding has all too often resulted in a corresponding decrease in the level of support for theory and modeling related programs.

The purpose of this white paper is two-fold. 1) We recommend that theory and modeling be funded at a significantly enhanced level that would allow sizable teams of researchers to pursue major science projects using a combination of theory, modeling, and data analysis. We suggest that an individual grant could total as much as $1M/year, and that as many as 10 groups should be supported. 2) We recommend that the theory program be given the status of a Mission and be similarly protected against the vagaries of NASA funding, particularly in maintaining the level of funding and adjusting for inflationary increases.

2. Current state of Theory and Modeling

The Colgate report listed 6 specific problem areas in basic theory as important for the progress of Space Physics:

(1) Magnetic-field reconnection
(2) Interaction of turbulence with magnetic fields
(3) Behavior of large-scale flows and their interactions
(4) Acceleration of energetic particles
(5) Particle confinement and transport
(6) Collisionless shocks.

Although this list was compiled some 34 years ago, it is remarkably applicable to the present time. Significant progress has been made.

We now have a significantly improved understanding of each of the problems listed, due to more and better observations allied with theoretical and modeling advances, but in none of them can we claim to have achieved a more-or-less definitive understanding of the problem. Of course, the richness of each of these problems has spawned both increased understanding in related fields and new fields of inquiry.

Some case studies provide an interesting perspective on the development of theory, the associated modeling and simulations, and the role of data analysis. One example, a subset of the collisionless shock example, illustrates both our increasing understanding and the further fundamental questions that are opened up by theory.

<u>Collisionless Shocks</u> Our current understanding of perpendicular shocks is rightfully regarded as one of the major achievements of modern collisionless nonlinear plasma physics. A perpendicular shock wave has the upstream magnetic field aligned perpendicular to the shock normal. Deviations up to nearly 45^0 for the magnetic field from the shock normal are described as quasi-perpendicular, and a clear physical picture can be drawn since reflected ions will gyrate back to the shock front in this case. Nearly perpendicular shocks have received considerable

attention because of their relatively clean, laminar appearance in the time series data. The particle gyromotion in the quasi-perpendicular magnetic field acts to prevent particles behaving excessively diffusively (i.e., the particles tend to be bound together), thereby simplifying the collisionless processes responsible for thermalization of the plasma, and the injection and acceleration of energetic non-thermal particles. Despite much effort, key questions remained unanswered or open to interpretation, and prior to Cluster, single, and dual, spacecraft studies were unable to place quantitative limits on the spatial scales, or address non-stationarity of the overall shock transition. By taking advantage of the sharp, quasi-perpendicular shock transitions and four spacecraft techniques, Cluster studies can probe internal shock scales and physics, including energetic particle origin, of the Earth's quasi-perpendicular bow shock. This has led to a resurgence in theoretical studies of the perpendicular shock, with many of the formerly accepted assumptions now being questioned. For example, the role of dimensionality in kinetic simulations and the importance of the electron/ion mass ratio are now recognized as introducing different behavior in the simulations depending on the assumptions. In all of this, the role of magnetic-field-line mixing must also be considered.

The capacity for the Voyager Interstellar Mission to surprise remains undiminished, with the recent crossing of the heliospheric termination shock (HTS) by the Voyager 2 spacecraft. With a working plasma instrument, the Voyager 2 observations revealed a quasi-perpendicular heliospheric termination shock that appears to be considerably different in character from quasi-perpendicular shocks observed in the inner heliosphere. The Voyager 2 observations suggest the possibility of a broad structure, or a velocity slowdown before the shock, possibly considerable fine-scale structure, many partial termination shock crossings, very little heating of the solar wind protons, to the extent that the downstream flow appears not to be subsonic, and a density spike appears to be present. These results suggest that the heliospheric termination shock is unlike any heliospheric shock observed elsewhere. Nonetheless, the observations were not entirely unexpected theoretically. It had been argued that the primary dissipation mechanism at a quasi-perpendicular heliospheric termination shock would be reflected pickup ions, and that the solar wind ions would be heated very little. This basic concept appears to have been borne out by the Voyager 2 observations.

Thus, theory appears to be making headway in our basic understanding of quasi-perpendicular shock physics but fundamental questions remain unresolved, and have not been clarified observationally, in part because without an accepted theoretical structure, it is difficult to know what we should be looking for. For example, even the question of the scale length over which changes in the electric field occur and its relation to the scale size over which changes in the magnetic field occur is unclear theoretically and observationally.

<u>Transport and acceleration of energetic particles</u> The acceleration of energetic particles and cosmic rays is thought by many to be caused by shock waves. Perpendicular shocks, such as much of the termination shock, are efficient and fast accelerators. With the crossings of the termination shock, it has become clear that the early theoretical models were too simple. The energetic particles did not behave as expected. In response, a number of approaches have been suggested. Non-shock acceleration mechanisms such as reconnection and statistical acceleration have been resurrected and the effects of pre-existing turbulence have been examined. Thus, theory drove certain expectations of particle acceleration at the heliospheric termination shock,

some of which were met and others not by the Voyager observations. Clearly, a major theory effort is now needed to address these recent developments.

All of these acceleration issues require an increased understanding of particle transport, both in formulating the theories and in interpreting spacecraft observations which are most-often remote from the acceleration site. These theoretical and modeling issues are of basic importance throughout physics, ranging from thermonuclear plasmas to supernova blast waves. Thus, increased knowledge here will have broad impact.

Magnetohydrodynamic turbulence in the solar wind Magnetohydrodynamic (MHD) turbulence is characterized by nonlinear interactions among fluctuations of the magnetic field and flow velocity over a range of spatial and temporal scales. It plays an important role in plasma heating, the transport of energetic particles, such as galactic and anomalous cosmic rays, and radiative transfer and is ubiquitous in the solar and interplanetary plasma. The solar wind exhibits turbulent behavior, as can be observed from in situ data. Solar wind fluctuations are observed over spatial scales that range from many AUs to electron kinetic scales. These observations have stimulated an ongoing effort to develop theoretical treatments of MHD turbulence. Large-scale fluctuations can be described by fluid models, but the smaller than proton scales requires Hall MHD or a kinetic description. The coupling of large-scale and kinetic-scale processes in the solar wind is mediated by the turbulence. Energy is transferred from large-scales to small-scales through turbulent fluctuations, including eddies and waves, that interact. The cascade is eventually halted by kinetic processes and heats the plasma. With an explicitly turbulence-based model that includes appropriate source terms, the radial evolution of turbulence intensity and the temperature of the plasma has been computed successfully from 1 to > 50 AU.

Solar wind turbulence remains incompletely understood despite considerable progress in the last 20 years. It is an important problem since turbulence mediates the complex dynamical couplings between large and small scales, slow fluid motions and fast kinetic processes, and low energy and high energy charged particles. Numerous important theoretical problems remain unresolved, ranging from compressible effects, passive scalar transport, the dimensionality and symmetries of interplanetary turbulence, etc.

While the Colgate Report list of fundamental plasma physics problems remains comprehensive, at least one additional area of fundamental research has assumed particular importance for solar and interplanetary plasmas. In the examples described above, a common element was the interaction of the heliosphere with interstellar material, especially the neutral gas (hydrogen). This took the form of a new shock dissipation mechanism based on interstellar pick-up ions, the acceleration of anomalous cosmic rays, and even the heating of the solar wind by pick-up ions providing the energy (through a resonant instability associated with pick-up) that cascades to kinetic scales and heats the solar wind. A more detailed example related to the formation of the "hydrogen wall" is given below. Thus, to the Colgate Report list, we would add the additional fundamental problem of ***Partially-Ionized Plasmas***.

Structure of the heliosphere The physics of the outer heliosphere and the large-scale structure of the heliosphere is determined fundamentally by its interaction with the partially ionized local interstellar medium (LISM). To illustrate, an important development in outer-heliospheric research was the prediction and discovery of the “hydrogen wall.” Theoretical models predict that the partially ionized LISM and the solar wind are separated by a complex set of plasma and neutral-atom boundaries, of which the termination shock has now been observed by the Voyager spacecraft. A specific theoretical prediction of the models is that a wall of interstellar neutral hydrogen should exist in the upstream region owing to the relative motion of the heliosphere and the LISM. This hydrogen wall is predicted to have a number density slightly more than twice the interstellar density, to be hotter than interstellar hydrogen, and to be some 100 AU wide. The physical reason for the hydrogen wall is the deceleration and diversion of the interstellar plasma flow about the heliosphere leading, through charge-exchange coupling, to a pile-up and heating of interstellar neutral hydrogen. The result is the formation of a giant wall, which acts to filter neutral hydrogen as it enters the heliosphere. Confirmation of the hydrogen wall’s existence was not expected for decades, but a serendipitous convergence of predictive theoretical modeling, observations to place limits on the cosmological deuterium/hydrogen ratio, and a multi-disciplinary investigation spanning space physics and astrophysics led to the detection of the hydrogen! This was the first of the boundaries separating the solar wind and the LISM to be discovered, and it offers a glimpse into the global structure of the three-dimensional heliosphere.

The research leading to the discovery of the hydrogen wall is an excellent example of theory driving the frontiers of space science and motivating the development of new observational techniques and methodology to complement traditional space physics tools.

3. *Theory and Modeling*

It is important to emphasize that the terms theory and modeling are not synonymous, and th must both be present in a balanced program. We view theory as the extension of our basic knowledge of phenomena. Theory attempts to find new physical descriptions of poorly understood phenomena. Modeling, on the other hand, involves the implementation of theory to complicated, realistic situations and scenarios. This often involves considerable investment in computer hardware and software. The end results are quantitative predications that can be directly compared to observations. An illustrative example is the creation of a synthesized time dependence of the predicted magnetic field or energetic particles along a real spacecraft trajectory to compare with an observed timeline.

Noting this distinction between theory and modeling allows us to consider several essential elements that a theory program needs to address.

<u>Formulation of theoretical models</u> In recognizing that multiple scales, regions, processes, and plasma populations are intrinsic to the challenging space physics problems of today, the correct mathematical and physical formulation is critical. In this, there is no substitute for time-honored analytical approaches to theoretical developments in plasma physics, fluid dynamics, and applied mathematics. Progress on highly nonlinear, coupled plasma problems may be made using techniques that range from the relatively standard to nonlinear, low-order reductive approaches and statistical methods. Current agency funding programs are not adequate to even support these basic theoretical efforts, and unfortunately innovative and bold ideas, approaches, and techniques in proposed research are often not encouraged and rewarded. Computation is no substitute for the development of rigorous theories and well-conceived models. Basic theory must be regarded as a

critical component of the funding profile for either NASA or the NSF and a Mission-level perspective must be encouraged

Invariably, many of the problems listed in the "extended" Colgate Report list impose significant computational demands in terms of CPU power and the concomitant development of sophisticated and efficient codes. Two challenges face the community. The first is to further develop existing codes and algorithms, such as three-dimensional MHD codes that incorporate adaptive mesh refinement, for example, or three-dimensional hybrid codes with improved electron/ion mass ratios or improved codes for data exploration. These problems do not demand the inclusion of new physics but demand instead substantial progress in current research areas. The second challenge lies in developing and implementing new computational approaches for both model solving/simulation and data exploration that exploit advances made by numerical mathematicians, statisticians, and computer scientists.

The coupling of different physical processes, scales, and regimes and the self-consistent incorporation of multiple scales, physical processes, and distinct regions into models will be the main challenge to theorists and modelers in the coming decade, demanding the formulation and development of sophisticated models and theory, the development of new and innovative algorithms, access to sophisticated computational resources, and the opportunity to test model predictions and validate theories against existing and future observations. We know that theory will demand sophisticated new measurements, which will in turn drive and define new space and ground-based missions (in situ, multipoint, remote, etc.). A theory program must focus on the investigation of well-chosen, theoretical problems and the development of coupled global models. For major advances to be made in space physics, fundamental theoretical analysis, sophisticated computational tools, and state-of-the-art data analysis must all be coupled intimately under a single umbrella program. Theoreticians working with pen and paper, computational space physicists, and data analysts will be needed collectively to achieve the major advances expected of space physics. Only by creating and maintaining major groups of this sort can a strong and vital connection between basic science, computation, and observations be achieved.

A well-balanced discipline would be in a position to leverage more fully the remarkable observational possibilities to increase our understanding of our home in space.

4. *Recommendation and Implementation*

The above specific examples and issues document the real decrease in support of theory, both absolutely and relative to the missions. We therefore recommend that theory and modeling be protected from further erosion and be restored to the relative position it had in the aftermath of the Colgate report.

We propose the creation of a New Theory MISSION. What is needed for a successful Theory Program? The answers are rather simple:

- Long-term, stable funding;

• Synergistically interacting groups of students, postdoctoral associates, research scientists, and several university or institutional faculty who are able to integrate research and education at some level.

To guard theory and modeling from continuing erosion, we propose that a theory mission be created with a specific goal or goals and a specified time line, in analogy with a hardware mission. One possible scenario would involve selecting one of the issues discussed above as the target of the Mission. For example, one might select a 'Mission to Understand the Interaction of the Heliosphere with the Local Interstellar Medium", or perhaps 'The Sources of the Interplanetary Magnetic Field as a focus for a 5-year effort. Within such a mission concept, one would fund separate groups (in analogy to separate instruments), some to work on aspects of basic theory, others to develop models and simulations and perhaps others to work on the connections with observations.

Such a mission should NOT be funded at the expense of the usual SRT or GI programs.

In summary,
1) we recommend that theory and modeling be funded at a significantly enhanced level that would allow sizable teams of researchers to pursue major science projects using a combination of theory, modeling, and data analysis. We suggest that an individual grant could total as much as $1M/year, and that as many as 5-10 groups should be supported at any one time, for up to 5 years; and 2) we recommend that the theory program be given the status of a Mission and be similarly protected against the vagaries of NASA funding, particularly in maintaining the level of funding and adjusting for inflationary increases.