Complex Model Acceleration Using Neural Networks (CMAUNN)

*A **ReNew** research suggestion white paper*

Paul Cadaret
UNICON Inc, Rancho Santa Margarita, CA, 92688
Phone: 949-742-7538; Fax: 877-243-6431; Email: alias-renew-2009[at]unicon.ws

Background

UNICON's recent participation in the *Small Business Innovation Research* (**SBIR**) programs at the DOD and DOE has inspired us to envision how very large and fast artificial neural networks could be constructed. We believe that our technology [1] allows us to employ very large neural networks in a way that enables a concept we call "***exhaustive learning***" to be applied for the benefit of certain types of complex and slow computational problems [2]. In mid 2007 we developed a paper [3] for the DOD that described how an unusual neural network configuration could be utilized as a fast hyperdimensional computational engine. This ReNeW paper *focuses on* a variant of our previously described method [3] that we believe can be tailored to effectively perform ***real-time*** **tokamak instability prediction using theory trained neural networks** (**TIPUTTNN**) and therefore minimize the risk of damaging, costly, and dangerous disruption events. However, we believe that the computational approach described is generic and it can be used to provide the DOE with a variety of fast and effect computational tools to allow various DOE-related experiments to utilize complex computational models more effectively for enhanced performance and safety.

1 The Current Tokamak Stability Problem

Within the last few months we have come to better understand certain needs of the DOE fusion energy community. We understand that the community has been pursuing activities that are significantly focused on high-energy tokamak reactor science. Such scientific exploration demands experimentation. High-energy tokamak experimentation inherently comes with the significant risk that plasma disruption events will occur. Such disruption events can be extremely damaging to equipment, time-consuming/costly to repair, and extremely dangerous to operational personnel. We believe that the community needs mitigation methods to more effectively manage the risks inherent in high-energy tokamak experimentation. Although we understand that complex physics-based tokamak stability models are available, the traditional software implementation methods used to instantiate these models result in systems that are generally quite slow. Therefore, currently available tokamak instability prediction systems are generally too slow to be useful to meet the DOE's ***real-time*** diagnostic or predictive needs. In summary, we believe that there is an urgent need for **real-time tokamak reactor instability prediction/mitigation systems** for experiments such as DIII-D and ITER [4]. Other needs may also exist for computational systems that are based on similar pattern recognition concepts.

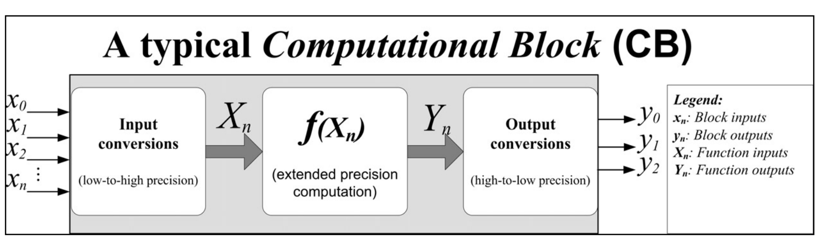
2 Technical Requirements For A Solution

We believe that any effective real-time tokamak reactor instability prediction system must be based on verified physics models. We understand that currently available tokamak stability models require a large number of operational parameters as inputs and they generate a number of outputs. To be effective, we understand that a tokamak instability prediction system must have a computational response time of less than 10msec (>100Hz). Therefore, what appears to be needed are **(a)** effective methods of gathering a wide variety of tokamak reactor operational parameters in real-time, **(b)** a means to compute plasma stability results in real-time based on the data collected, and **(c)** a means to direct operational parameters toward greater plasma stability (safety) in real-time when instability conditions are detected. We understand that the operational data required to satisfy item (a) is currently available (at least for DIII-D). This paper describes methods [3] that we believe can simultaneously address items (b) and (c) for DIII-D and ITER.

3 Elements Of Research Thrusts Needed To Provide The Solution

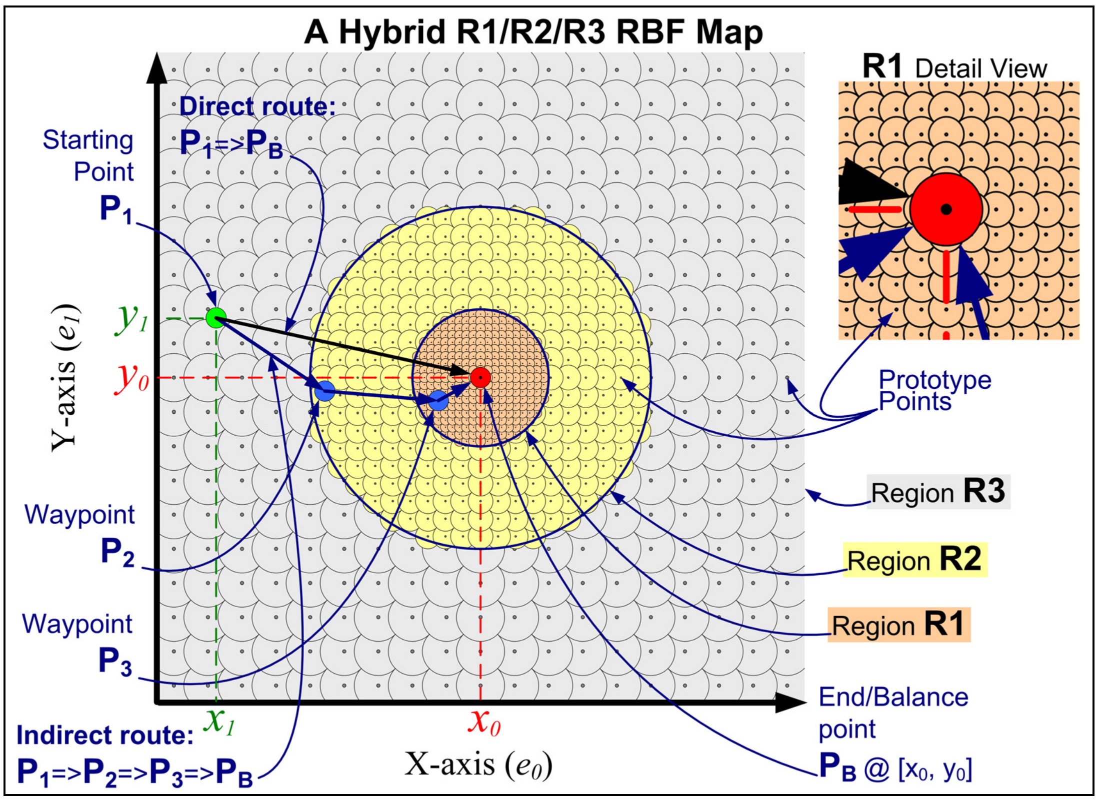
3.1 A Brief Overview Of The Unusual Computational Method Proposed

When one thinks of speeding up a computational system the traditional view is often to (a) get a faster computer, (b) decompose the computational problem into subsets and apply more computers to the problem (if possible), (c) develop custom hardware to perform the required operations, or (d) develop a computational system that is some combination of the above methods. Our approach [3] takes a fundamental look at the nature of the computational process itself. Instead of focusing on the steps involved in a computational process our approach focuses only on the ***inputs-to*** and ***outputs-from*** a *computational block* (**CB**) as shown in the figure below.



The inputs to the CB can be thought of as defining a computational *feature space* (**FS**) of some size (perhaps quite large). Typically, when processing a real-world scientific problem we take a series of finite precision inputs $\boldsymbol{x_n}$, convert them to an extended precision representation $\boldsymbol{X_n}$, process these extended precision values with some algorithm $f(X_n)$, and then convert the extended precision results Y_n back to some lower resolution form y_n where they can be utilized. Often, these computations are performed inefficiently because the complex computations $f(X_n)$ are performed at a substantially higher resolution than that required by the resolution of the results y_n ultimately delivered. Unfortunately, the typical methods used to model our physical world generally rely upon various combinations of complex mathematical functions (i.e.: exponential, trigonometric, hyperbolic, etc.) whose typical implementations require high-precision inputs and generate high-precision outputs. With such a computational approach it is generally difficult to *scale down* precision and *scale up* computational performance to meet real-world needs.

Our approach considers the CB inputs to define a FS of some size (again, possibly quite large). Although the size of the CB input FS might be extremely large, this is generally unimportant. What is important is the precision of the results delivered by the CB (y_n) to meet a specific computational requirement. From the required precision of these results (y_n) we can derive our need for computational precision. Depending on the type of problem the application *may* also benefit from independent CB regions of computational precision. Our approach advocates the use of a large/fast (generally pre-trained) *Radial Basis Function* (**RBF**) *Artificial Neural Network* (**ANN**) as the basis for extremely fast computational systems. This approach then enables the concept of *exhaustive-learning* to be exploited [2]. The figure below (extracted from our earlier paper [3]) illustrates this approach.



This figure shows 3 discrete computational regions (R1, R2, and R3). Each region shown in this example is fully covered by an array of regularly spaced RBF neurons using traditional hyperspherical *influence fields* (IF). As shown the influence fields are of a uniform size (or precision) within each region. The RBF neurons used would be modified such that in addition to delivering an indication that an input pattern (x_n) match was found, each neuron would also deliver the computational output values (y_n) that must be "computed" (delivered) by the CB. Any input pattern presented (x_n) that falls within a neuron's IF will be declared a "match" by that neuron and the associated CB output values (y_n) would be delivered as our computational system results. *The size of the RBF influence fields shown in each region directly correlates to the amount of error that we are willing to accept within each region.* Numerous such regions may exist.

Using an ANN as the basis for a computational system would normally make little sense because large ANNs (including RBF ANNs) are generally known to be computationally burdensome in the extreme. However, we believe that our technology [1] can provide a robust computational solution whose performance can be greatly scaled. Our earlier paper [3] described a useful computational system configuration that was capable of unusually fast computational performance (>1.32MHz) despite significant model complexity. We should also point out that our approach is generally applicable to any problem domain where input feature space precision can be *scaled down* in order to *scale up* effective computational speed. The advantage of our approach is that it is completely invariant to the complexity of the model computational algorithm $f(X_n)$ used. This means that the more computationally burdensome the computational algorithm is, the more significant our achievable speed improvement will likely be.

3.2 Research Thrusts Needed To Develop Effective Tools

Looking forward across the spectrum of potential needs for ITER and the larger DOE fusion energy community we believe that various opportunities exist for the application of advanced pattern recognition methods as the basis for unusual computational solutions to address important technical challenges. The following list of research thrusts is provided for consideration:

A. **Tokamak instability prediction (TIP):** One research thrust suggested (the focus of this paper) is the study of pattern recognition methods as the basis for advanced computational systems. Complex fusion energy experiments will likely require complex physics based models to adequately describe them. The instantiation and application of such physics based models using traditional computational methods may sometimes result in complex models that are computationally burdensome to implement and therefore result in slow computational systems. Various needs exist for fast real-time diagnostics, prognostics, machine control applications, user interfaces, and other systems that rely on complex computational models. We understand that one important need in this area is for the implementation of real-time methods of detecting off-normal events; particularly the detection of tokamak operating parameter proximity to plasma instability boundaries. Because existing stability models are complex and slow, we suggest that this important application be used as a test-case for the computational system implementation methods proposed. As a starting point the methods described in our paper [3] are offered for consideration. Supporting research will likely be needed to explore effective methods of applying exhaustive-learning [2], effective model conversion and system training methods, effective validation methods, identifying computational speed capabilities and variations, and identifying computational robustness via “error” analysis.

B. **Other computational systems (OCS)**: We believe that ITER and the larger fusion energy community likely have a variety of needs similar in general form to the TIP application described above. Such system needs may require existing complex and slow computational models to be accelerated so that they can be utilized in some real-time context. Therefore, another research thrust suggested is the identification of additional opportunities for the application of fast/fuzzy computational systems that are based on the methods described in our papers [2] and [3]. Different applications will likely exhibit different computational requirements and hence result in somewhat different system solution configurations and training methods. Supporting research will likely be needed regarding effective methods of applying exhaustive-learning [2] in each specific context and rapid conversion/training methods for such systems appropriate to the specific computational problem to be solved.

Organizational issues: Given the fact that the solution enabling methods described in this section originate within the private sector we suggest that the DOE draw upon and partner with private sector resources in a significant way to accelerate the research and development of advanced systems to meet the urgent needs of the fusion energy community.

4 Anticipated Research Outcome

Should the DOE OFES look favorably on the research thrust topics listed above this section presents a forward looking ***and admittedly optimistic*** view of what types of tools might be developed as a result of the research and development efforts proposed.

(a) **GENERAL:** Given a significant and successful research and development effort we believe that unusually fast computational systems can be developed that are based on advanced fuzzy pattern recognition methods. We believe that such tools will allow complex and slow physics-based models to be employed to support real-time experimental needs. We believe that such computational systems can be made highly scalable in speed and invariant to the complexity of the underlying physics-based models used to train such systems. We believe that the approach proposed will provide the DOE with new and effective methods to adapt existing complex models for use in a wide variety of real-time system environments.

(b) **TIP:** Specifically, we believe that the proposed effort will enable the DOE to develop computational systems capable of processing large numbers of experimental data parameters at high rates so that effective real-time TIP systems can be developed. Such a capability would allow the DOE to utilize TIP systems to significantly enhance the safe operation of high-energy tokamak reactor experiments. The utilization of such TIP systems as the potential to greatly improve the safety of high-energy tokamak experimentation and reduce overall costs.

(c) **OCS:** Given a significant and successful research and development effort we believe that a variety of other complex computational models likely exist that could benefit the fusion energy community if these models could be instantiated in ways that provide real-time computational behavior. Such capabilities would allow the fundamental knowledge of science embodied within these models to be applied in a practical way to support experiments that are fundamentally real-time in nature.

5 Additional References

[1] *CogniMax® pattern recognition technology;* COGNIMAX is a trademark of UNICON Inc.

[2] CADARET, P., "*What Is Exhaustive-Learning?*" (WIEL). Available within the DTIC IR&D collection (www.dtic.mil) with document accession number is 08241207

[3] CADARET, P., "*Computational Acceleration Using Neural Networks*" (CAUNN), Proceedings Of The SPIE, Defense and Security Conference 2008; See the SPIE online library at http://spie.org/x399.xml; Also available within the DTIC IR&D collection (www.dtic.mil) with document accession number is 08241202

[4] **Theme I Panel priorities** recently observed:

Theme I Panel	Topics for Burning Plasma Theme	EPAct Report	ITER Research Plan v1	ITER/ITPA High Priority
	Diagnostics			Urgent
Off-normal Events	Disruption/VDE/runaway mitigation	II.B	2.2.3	Urgent