



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES  
**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

## DESENVOLVIMENTO DO MODELO SUPIM-DAVS PARA A PLATAFORMA NEC SX-AURORA TSUBASA

Lucas Roges de Araujo

Relatório parcial de Iniciação Científica do programa PIBIC, orientada pelo Dr. Haroldo Fraga de Campos Velho (INPE) e Prof. Andrea Schwertner Charão (UFSM).

INPE  
São José dos Campos  
2021



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES  
**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

## DESENVOLVIMENTO DO MODELO SUPIM-DAVS PARA A PLATAFORMA NEC SX-AURORA TSUBASA

Lucas Roges de Araujo

Relatório parcial de Iniciação Científica do programa PIBIC, orientada pelo Dr. Haroldo Fraga de Campos Velho (INPE) e Prof. Andrea Schwertner Charão (UFSM).

INPE  
São José dos Campos  
2021



## RESUMO

Atualmente, diversas aplicações foram desenvolvidas e são capazes de simular situações reais e resolver diversos problemas através do cálculo de métodos matemáticos complexos, os quais utilizam grande poder computacional. Em paralelo, o desenvolvimento de novas tecnologias, para aceleração do desempenho deste tipo de aplicações, torna imprescindível o estudo acerca desses recursos, para encontrar quais destes melhor atendem as necessidades das aplicações em produção. No Brasil, o INPE utiliza algumas aplicações para simulação e estudos relacionados ao espaço, sendo parte deles com uma adaptação do modelo SUPIM, da Universidade de Sheffield. Arelado ao desenvolvimento de novas tecnologias, a empresa NEC apresentou há pouco tempo co-processadores vetoriais, de diversas escalas, para aceleração de aplicações de alto desempenho. Sendo assim, o objetivo do trabalho é estudar a arquitetura desenvolvida e adaptar o código do SUPIM para execução e análise do desempenho utilizando este recurso.

Palavras-chave: Processamento paralelo. Modelo de dinâmica da ionosfera. Módulo de processamento vetorial.



## LISTA DE FIGURAS

	<u>Pág.</u>
2.1 Diagrama da arquitetura SX-Aurora TSUBASA (Processador VE20). . .	7
2.2 Comparação de funcionamento entre o processador vetorial e GPGPUs. .	7



## LISTA DE ABREVIATURAS E SIGLAS

CPU	–	<i>Central Processing Unit</i>
DAVS	–	Data Assimilation and Visualization
Embrace	–	Estudo e Monitoramento Brasileiro do Clima Espacial
GPGPU	–	<i>General Purpose Graphics Processing Unit</i>
GPU	–	<i>Graphics Processing Unit</i>
HPC	–	<i>High Performance Computing</i>
INPE	–	Instituto Nacional de Pesquisas Espaciais
MPI	–	Message Passing Interface
SIMD	–	Single Instruction Multiple Data
SUPIM	–	Sheffield University Plasmasphere-Ionosphere Model
VE	–	<i>Vector Engine</i>



## SUMÁRIO

	<u>Pág.</u>
<b>1 INTRODUÇÃO</b> . . . . .	<b>1</b>
1.1 Objetivos . . . . .	2
1.1.1 Objetivo geral . . . . .	2
1.1.2 Objetivos específicos . . . . .	2
<b>2 DESENVOLVIMENTO</b> . . . . .	<b>5</b>
2.1 Revisão de Literatura . . . . .	5
2.1.1 Processamento paralelo e vetorial . . . . .	5
2.1.2 Arquitetura NEC SX-Aurora TSUBASA . . . . .	6
2.1.3 SUPIM . . . . .	8
<b>3 CONCLUSÃO</b> . . . . .	<b>11</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> . . . . .	<b>13</b>



# 1 INTRODUÇÃO

A difusão de diferentes técnicas de programação paralela, muitas delas atreladas as novas tecnologias e arquiteturas de computador criadas, para uso por aplicações de alto custo computacional, fazem com que o campo de High Performance Computing (HPC) se encontre em constante evolução. Essa evolução tem como principal objetivo a otimização do tempo de execução destas aplicações, através da exploração destes recursos desenvolvidos.

As principais maneiras de acelerar a execução de uma aplicação estavam, por muito tempo, ligadas a frequência do processador e/ou a quantidade de *threads* que o processador consegue executar simultaneamente. Também em nível de *hardware*, recursos como a memória cache e sistemas de interconexão de alta velocidade, para sistemas distribuídos, passaram a ser recursos importantes para reduzir o tempo de execução de uma aplicação.

Além das diversas soluções em nível de *hardware*, técnicas de programação foram desenvolvidas para facilitar a paralelização de aplicações, principalmente ao tratar-se de aplicações com cálculos complexos, que por si só já apresentam dificuldades no processo de desenvolvimento do código. Essas técnicas e bibliotecas criam interfaces de abstração para facilitar a paralelização durante a execução, a nível de instrução. Trabalhos na área (DIAZ et al., 2012; CZARNUL et al., 2020) trazem uma revisão acerca desses diversos métodos, os quais serão abordados na Subseção 2.1.1.

Nesse contexto, uma das técnicas que envolve tanto adaptação em *hardware*, quanto em código, é a de processamento vetorial. O processamento vetorial é uma abordagem de paralelismo em nível de dados, visto que intruções em código são realizadas simultaneamente sobre diferentes pontos de um conjuntos de dados, comumente um vetor, seja esse vetor unidimensional ou multidimensional. Empresas consolidadas no mercado de processamento, como a Intel, já desenvolveram tecnologias voltadas para o processamento vetorial. Mais recentemente, a empresa NEC também lançou dispositivos especializados, com um módulo de vetorização aplicável em várias escalas (desde o *desktop* até supercomputadores).

A introdução dessas novas técnicas de aceleração do processamento atraiu outros tipos de aplicações para o campo de HPC, como inteligência artificial, aprendizado de máquina e análise de dados. Ainda assim, as aplicações clássicas, como as voltadas para genética, modelagem financeira, previsão climática, pesquisas espaciais e outras áreas, ainda são alvo de estudo para adaptação em relação as novas tecnologias

emergentes, dado que sua estrutura também pode tirar proveito destes recursos.

Dentre os escopos de aplicação citados, o foco do Instituto Nacional de Pesquisas Espaciais (INPE) evidentemente é na parte de pesquisas espaciais, dentro dos diversos escopos internos que essa área possui. Nesse contexto, a aplicação Sheffield University Plasmasphere-Ionosphere Model (SUPIM) (BAILEY; SELLEK, 1990; BAILEY et al., 1993; BAILEY et al., 1997) é um modelo de previsão ionosférica da Universidade de Sheffield, que vem sendo utilizado e modificado por grupos de pesquisa do INPE, de maneira a otimizar a aplicação como um todo.

Dessa forma, levando em consideração os estudos realizados com a aplicação SUPIM e sua demanda de recursos para aceleração, atrelada ao advento de novas tecnologias, como os processadores vetoriais recentemente lançados pela NEC, tem-se uma combinação de fatores que instigam a possibilidade de adaptar essa aplicação para a arquitetura em questão e analisar seu desempenho em relação aos locais atuais de execução.

## 1.1 Objetivos

Na Subseção 1.1.1 é descrito o objetivo geral do trabalho, enquanto na Subseção 1.1.2 são listados os objetivos específicos.

### 1.1.1 Objetivo geral

O objetivo geral do trabalho é a adaptação do modelo de dinâmica da ionosfera SUPIM, com a integração de Data Assimilation and Visualization (DAVS), para execução na arquitetura NEC SX-Aurora TSUBASA, de maneira a investigar as partes mais propícias de código a serem alteradas para obtenção de melhoria no desempenho da aplicação utilizando o *hardware* em questão.

Para tal, é necessário fazer um estudo da arquitetura e de seus benefícios para processamento, acesso a memória e entrada/saída. Com o estudo da arquitetura, a aplicação também deve ser explorada para decidir quais pontos de código podem ser beneficiados pelos recursos oferecidos.

### 1.1.2 Objetivos específicos

Segue na lista abaixo os objetivos específicos do projeto de pesquisa:

- Implementar o modelo de dinâmica da ionosfera SUPIM-DAVS, sistema

utilizado operacionalmente no programa Estudo e Monitoramento Brasileiro do Clima Espacial (Embrace) do INPE no novo *hardware/software* de co-processamento vetorial NEC SX-Aurora TSUBASA.

- Instalação, compilação e testes preliminares de desempenho para o modelo SUPIM-DAVS na nova plataforma de *hardware* da NEC.
- Avaliação de estratégias de paralelismo no *hardware* de co-processamento vetorial.



## 2 DESENVOLVIMENTO

O desenvolvimento parcial do projeto de iniciação científica teve foco na etapa de revisão de literatura, descrita na Seção 2.1. Uma etapa experimental também está prevista no projeto, a partir do acesso remoto a uma máquina de testes oferecida pela empresa NEC. Tal acesso é gerenciado pela empresa, levando em conta uma fila de instituições e pesquisadores interessados. Nos preparamos para o início das atividades em agosto ou início do mês de setembro de 2020, seguindo a recomendação de representantes da NEC do Brasil. Esse acesso só se concretizou na segunda quinzena de janeiro de 2021, de modo que a etapa experimental não faz parte deste relatório e será realizada por um bolsista substituto.

### 2.1 Revisão de Literatura

A revisão de literatura foi realizada acerca dos seguintes temas: processamento paralelo e vetorial, a arquitetura NEC SX-Aurora TSUBASA e o modelo SUPIM. Os temas encontram-se, respectivamente, nas Subseções 2.1.3, 2.1.1 e 2.1.2.

#### 2.1.1 Processamento paralelo e vetorial

O processamento paralelo de aplicações computacionais é consequência da combinação da produção de máquinas com múltiplas unidades de processamento, com suporte para execução de tarefas simultâneas e o desenvolvimento de técnicas em nível de *software*, para tirar proveito destas arquiteturas produzidas.

Nesse contexto, Diaz e co-autores (2012) classificam e descrevem alguns tipos de modelos de programação paralela, considerando seu nível de abstração em relação ao sistema computacional. Para a combinação processamento paralelo e HPC, se destacam os seguintes modelos: puros, heterogêneos e híbridos.

Dentre os modelos puros, destaque para os padrões da indústria: OpenMP e Message Passing Interface (MPI), comumente aplicados no escopo de HPC. O OpenMP (OpenMP Architecture Review Board, 2018) é um modelo *multithread* com foco em memória compartilhada. Destaca-se a facilidade de programação pelo nível de abstração que é capaz de oferecer através de suas diretivas de paralelismo. Para o escopo de computação de memória distribuída, o MPI é o mais utilizado e sua comunicação funciona através do compartilhamento de mensagens, dado que o sistema distribuído não é capaz de compartilhar variáveis em memória. Esse modelo se destaca pela escalabilidade, em relação a modelos de memória compartilhada, pois os dados podem ser compartilhados através da mensagens pela rede ao invés

do acesso direto em memória.

Nos modelos heterogêneos, conhecidos por complementarem o uso da *Central Processing Unit* (CPU), tornando o ambiente de execução heterogêneo. Nele, houve a introdução da *General Purpose Graphics Processing Unit* (GPGPU), a qual é descrita como um ponto de inflexão na história da computação moderna (MACEDONIA, 2003). De fato, estes dispositivos tem sido cada vez mais acopladas aos grandes centros de processamento computacional, como um recurso para aceleração de aplicações. Com a criação da linguagem CUDA (NVIDIA et al., 2020), a NVIDIA ajudou a consolidar ainda mais as GPUs como recursos para processamento de aplicações não-gráficas.

A partir da *Graphics Processing Unit* (GPU), já existe um modelo próximo ao do coprocessador vetorial da NEC, com as mesmas características de ter uma frequência menor e uma quantidade maior de unidades de processamento por dispositivo. A ideia base por trás do processamento vetorial é o *Single Instruction Multiple Data* (SIMD), onde uma mesma operação (instrução) é aplicada para um conjunto de dados.

Modelos de processamento vetorial costumam aplicar essas operações de maneira paralela para estruturas de dados com uma ou mais dimensões. Um exemplo comum de código é alguma operação matemática, dentro de um laço, aplicada para elementos de um mesmo conjunto de dados.

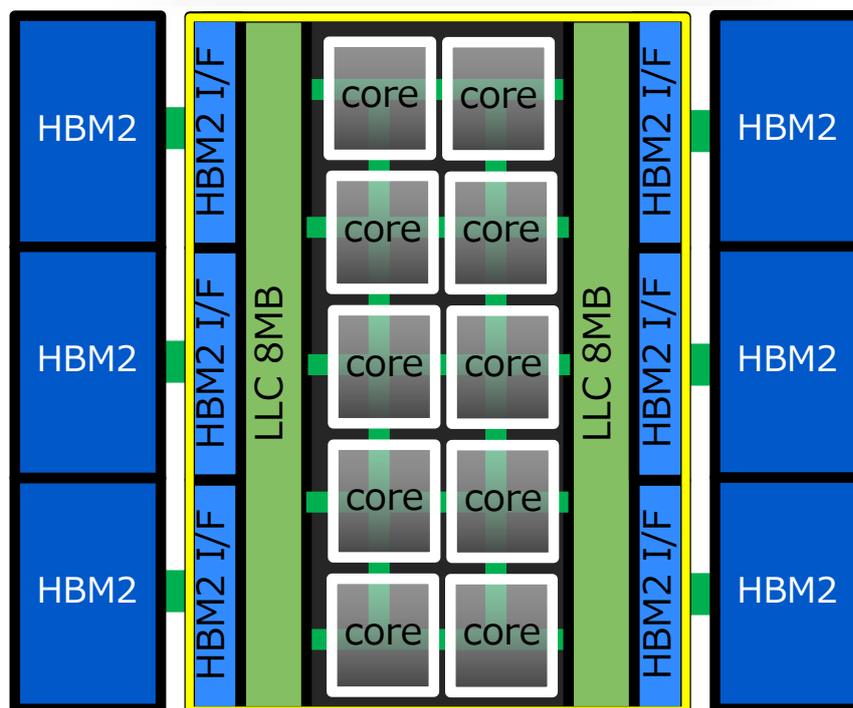
### 2.1.2 Arquitetura NEC SX-Aurora TSUBASA

A arquitetura NEC SX-Aurora TSUBASA foi apresentada em 2018 (YAMADA; MOMOSE, 2018) e traz como diferencial a utilização de Vector Engines (VEs) para processamento das aplicações. Além das especificações, a apresentação inclui a comparação de desempenho com CPU e GPU através da execução de alguns *benchmarks*. Sua eficiência apresentou melhores resultados em relação a CPU, além de resultados competitivos com a GPU.

Em detalhes mais técnicos e recentes apresentados através de uma palestra (NEC; HOSHI, 2020) durante o XXI Simpósio em Sistemas Computacionais de Alto Desempenho (WSCAD 2020), são mostrados detalhes da segunda geração de VEs. Na Figura 2.1, é possível visualizar a estrutura do dispositivo, que conta com os núcleos de processamento, memória cache e módulos de memória HBM2.

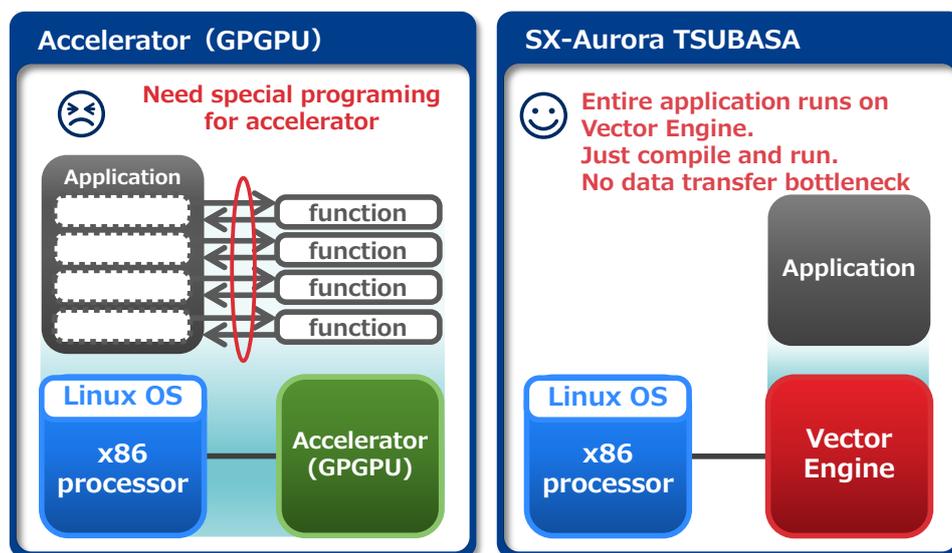
Além dos detalhes próprios da máquina, a apresentação indicou as diferenças entre

Figura 2.1 - Diagrama da arquitetura SX-Aurora TSUBASA (Processador VE20).



Fonte: (NEC; HOSHI, 2020)

Figura 2.2 - Comparação de funcionamento entre o processador vetorial e GPGPUs.



Fonte: (NEC; HOSHI, 2020).

seu modo de executar a aplicação e outros aceleradores comumente usados (GP-GPUs). Como indicado na Figura 2.2, a aplicação é executada por inteiro no dispositivo, enquanto em GPGPUs costuma haver uma passagem dos valores para processamento do acelerador e o retorno dos cálculos realizados, o que causa um determinado *overhead* no tempo de execução, devido ao tempo de transferência de dados entre a CPU e a GPU.

Com relação ao seu uso em código, a arquitetura apresenta um modo de programação semelhante ao visto com a ferramenta OpenMP, pois sua utilização em aplicações se dá através de diretivas que precedem o local das operações vetoriais em código, considerando um certo grau de abstração. Inclusive, as duas tecnologias podem ser usadas simultaneamente em um mesmo trecho de código (SERPA et al., 2020), caracterizando um modelo de programação híbrido.

O desempenho da arquitetura já foi testado em outros casos, desde a utilização com outros *benchmarks* e algumas aplicações práticas (KOMATSU et al., 2018), até aplicações voltadas para situações reais (MUSA et al., 2019). Para o caso com situação real, onde o equipamento foi utilizado para simulação de previsão de tsunamis, o desempenho da aplicação superou as outras arquiteturas da comparação (NEX SX-ACE e Xeon Gold), apesar de mostrar menor escalabilidade.

### 2.1.3 SUPIM

As primeiras publicações do modelo SUPIM surgiram no início da década de 90. Com alguns trabalhos já publicados acerca do tema, Bailey e co-autores (1997) trazem uma revisão dos avanços realizados na área através da utilização da aplicação. O autor descreve que o modelo tem o objetivo de entender os processos físicos e químicos que acontecem na parte de cima da atmosfera terrestre, com foco em regiões de baixa latitude (regiões próximas à linha do Equador).

Além do desenvolvimento por parte dos pesquisadores que criaram a aplicação, a utilização do SUPIM por grupos de pesquisa brasileiros fez com que diversos trabalhos fossem desenvolvidos nesse âmbito, trazendo benefícios para a melhoria e evolução do modelo.

Junto aos estudos acerca dos modelos matemáticos e a modificação nesse quesito para melhoria dos resultados da aplicação, a partir de certo ponto houve o interesse em otimizar o desempenho do SUPIM. (PEREIRA et al., 2010) utilizam técnicas como OpenMP, MPI, e até mesmo uma abordagem híbrida, para execução da aplicação.

O trabalho destaca a dificuldade de paralelização nas principais sub-rotinas, devido a dependência de dados e a utilização de variáveis globais, o que pode ter ocasionado na eficiência obtida (33%).

Além da paralelização, outros recursos foram desenvolvidos para inserção no modelo. Entre eles, a incorporação de um módulo de assimilação de dados (PETRY et al., 2014). Essa técnica pode ser descrita como um método de combinação de dados observados e dados modelados matematicamente, de maneira a obter as melhores condições iniciais para execução da aplicação e, conseqüentemente, obter melhores resultados de predição. É utilizando essa variante, chamada SUPIM-DAVS, que o trabalho atual está em desenvolvimento.



### 3 CONCLUSÃO

A partir da revisão de literatura realizada, foi possível observar tanto a evolução do processamento paralelo até o advento dos aceleradores e co-processadores modernos que existem no mercado. Estas técnicas oferecem, primeiramente, um paralelismo a nível de instrução, onde são criados fluxos paralelos de código que executam tarefas independentes e simultâneas. Algumas dessas técnicas englobam aplicações por inteiro, enquanto outras são aplicadas somente para trechos de código.

Além do paralelismo em nível de instrução, é possível paralelizar o tratamento de dados atrelados a alguma operação. Para tal, uma ou mais operações são aplicadas, ao mesmo tempo, para diferentes índices de um conjunto de dados. Nesta categoria, se encaixam os equipamentos de processamento vetorial. Como aplicações de HPC costumam apresentar estruturas de dados robustas, este tipo de equipamento costuma melhorar o desempenho da execução, trazendo benefícios para os grupos de interesse.

Junto a evolução dos modelos computacionais para execução de aplicação com alta demanda computacional, a aplicação SUPIM, que é alvo de utilização na arquitetura em questão, também sofreu diversas modificações ao longo dos anos, incluído o acréscimo de novas etapas, além de tentativas de paralelização com técnicas mais estabelecidas.

Dessa forma, com o surgimento dessa nova arquitetura vetorial lançada pela NEC, torna-se relevante estudar o código da aplicação e buscar por trechos com operações vetoriais que possam tirar proveito do que o *hardware* tem para oferecer, isto , avaliar o custo-benefício de eventual ganho de desempenho frente ao custo do equipamento. Além disso, aliar a técnica de vetorização do código com outros modelos paralelos, utilizando programação híbrida, pode auxiliar melhor na exploração de todos os recursos disponíveis.



## REFERÊNCIAS BIBLIOGRÁFICAS

BAILEY, G.; BALAN, N.; SU, Y. The sheffield university plasmasphere ionosphere model—a review. **Journal of Atmospheric and Solar-Terrestrial Physics**, v. 59, n. 13, p. 1541–1552, 1997. ISSN 1364-6826. The Ninth International Symposium on Equatorial Aeronomy. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1364682696001551>>. 2

BAILEY, G. J.; SELLEK, R. A mathematical model of the Earth's plasmasphere and its application in a study of He<sup>+</sup> at L = 3. **Annales Geophysicae**, v. 8, p. 171–189, mar. 1990. 2

BAILEY, G. J.; SELLEK, R.; RIPPETH, Y. A modelling study of the equatorial topside ionosphere. **Annales Geophysicae**, v. 11, n. 4, p. 263–272, abr. 1993. 2

CZARNUL, P.; PROFICZ, J.; DRYPCZEWSKI, K. Survey of methodologies, approaches, and challenges in parallel programming using high-performance computing systems. **Scientific Programming**, Hindawi, v. 2020, 2020. 1

DIAZ, J.; MUÑOZ-CARO, C.; NIÑO, A. A survey of parallel programming models and tools in the multi and many-core era. **IEEE Transactions on Parallel and Distributed Systems**, v. 23, n. 8, p. 1369–1386, 2012. 1

KOMATSU, K.; MOMOSE, S.; ISOBE, Y.; WATANABE, O.; MUSA, A.; YOKOKAWA, M.; AOYAMA, T.; SATO, M.; KOBAYASHI, H. Performance evaluation of a vector supercomputer sx-aurora tsubasa. **SC18: International Conference for High Performance Computing, Networking, Storage and Analysis**, p. 685–696, 2018. 8

MACEDONIA, M. The gpu enters computing's mainstream. **Computer**, v. 36, n. 10, p. 106–108, 2003. 6

MUSA, A.; ABE, T.; KISHITANI, T.; INOUE, T.; SATO, M.; KOMATSU, K.; MURASHIMA, Y.; KOSHIMURA, S.; KOBAYASHI, H. Performance evaluation of tsunami inundation simulation on sx-aurora tsubasa. In: **ICCS**. [S.l.: s.n.], 2019. 8

NEC; HOSHI, N. **An Evolved and Brand New Vector Technology SX-Aurora TSUBASA-Present Future**. 2020. Disponível em: <<http://wscad.sbc.org.br/2020/artigos/palestras/slides-nec.pdf>>. 6, 7

NVIDIA; VINGELMANN, P.; FITZEK, F. H. **CUDA Toolkit | NVIDIA Developer**. 2020. Disponível em:

<<https://developer.nvidia.com/cuda-toolkit>>. 6

OpenMP Architecture Review Board. **OpenMP Application Programming Interface**. 2018. 5

PEREIRA, A.; PETRY, A.; SOUZA, J.; CHARÃO, A.; VELHO, H. C. Aplicação de técnicas de paralelismo em modelo para previsão da dinâmica da ionosfera terrestre. **International Symposium on Computer Architecture and High performance**, 01 2010. 8

PETRY, A.; DE SOUZA, J. R.; VELHO, H. F. de C.; PEREIRA, A. G.; BAILEY, G. J. First results of operational ionospheric dynamics prediction for the brazilian space weather program. **Advances in Space Research**, v. 54, n. 1, p. 22–36, 2014. ISSN 0273-1177. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0273117714001793>>. 9

SERPA, M. d. S.; NAVAUX, P. O. A.; PANETTA, J. **Programação com Aceleradores Vetoriais**. 2020. Disponível em: <<http://wscad.sbc.org.br/2020/artigos/minicursos/minicurso1-Prog-Vetorial-texto.pdf>>. 8

YAMADA, Y.; MOMOSE, S. **Vector Engine Processor of NEC's Brand-New Supercomputer SX-Aurora TSUBASA**. 2018. Disponível em: <[https://old.hotchips.org/hc30/2conf/2.14\\_NEC\\_vector\\_NEC\\_SXAurora\\_TSUBASA\\_HotChips30\\_finalb.pdf](https://old.hotchips.org/hc30/2conf/2.14_NEC_vector_NEC_SXAurora_TSUBASA_HotChips30_finalb.pdf)>. 6