HISTORICO DO DESENVOLVIMENTO COMPUTACIONAL DO BRAMS

Jairo Panetta

ITA/IEC
PETROBRAS/E&P
BMFBovespa

CPTEC, 21/03/2016

Contribuições

- Alvaro Luiz Fazenda
- Celso Luiz Mendes
- Daniel Massaru Katsurayama
- Daniel Merli Lamosa
- Demerval Soares Moreira
- Eduardo Hidenori Enari
- Eduardo Rocha Rodrigues
- Eugenio Sper de Almeida
- Haroldo Fraga de Campos Velho
- Luiz Filipe Guedes Mota
- Luiz Flavio Rodrigues
- Manoel Baptista da Silva Junior
- Marcelo Saraiva Limeira
- Marco Dias Gubitoso
- Marcio Augusto de Moraes
- Paulo Yoshio Kubota
- Pedro Pais Lopes
- Rafael Mello da Fonseca
- Roberto Pinto Souto
- Saulo Rabelo Maciel de Barros
- Simone Shizue Tomita
- Stephan Stephany

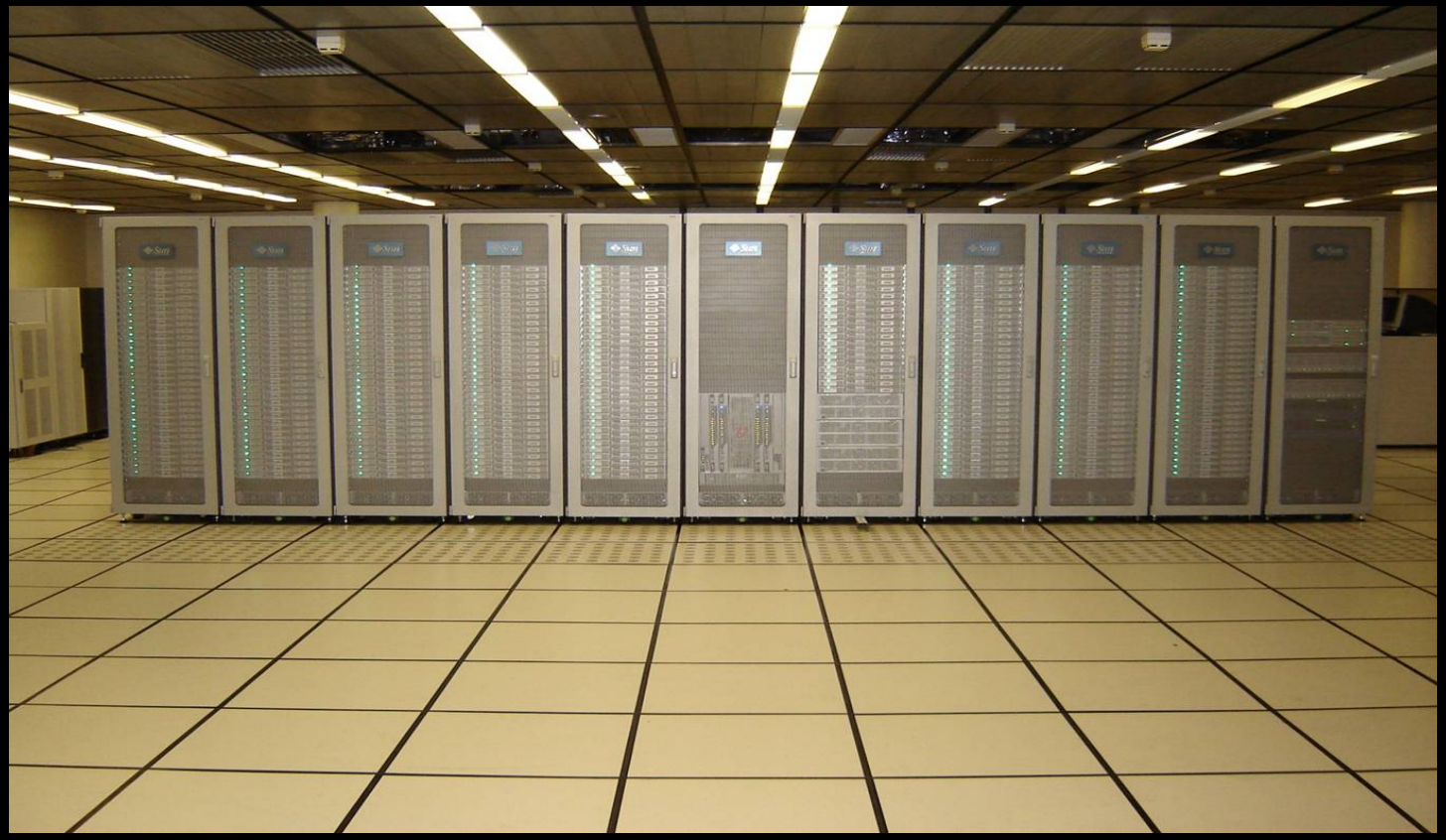
Necessidade de Paralelismo

2000 **1 proc** **355 s/dia**

Necessidade de Paralelismo

2004 **8 proc** **23 s/dia**

Necessidade de Paralelismo

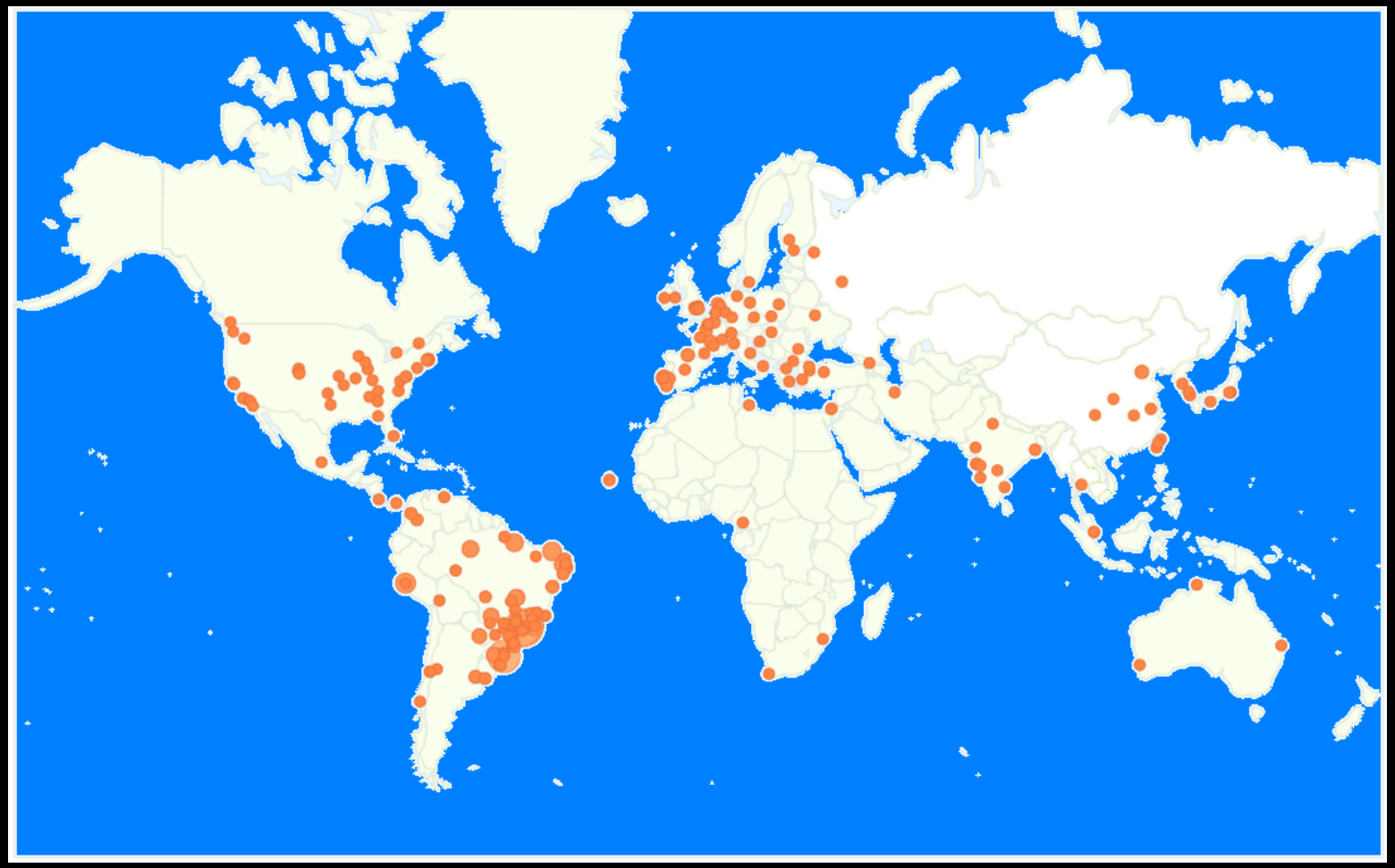


2007 **63 proc** **12 s/dia**

Necessidade de Paralelismo

2010 **1752 proc** **4 s/dia**

BRAMS: FINEP 2002 a 2006

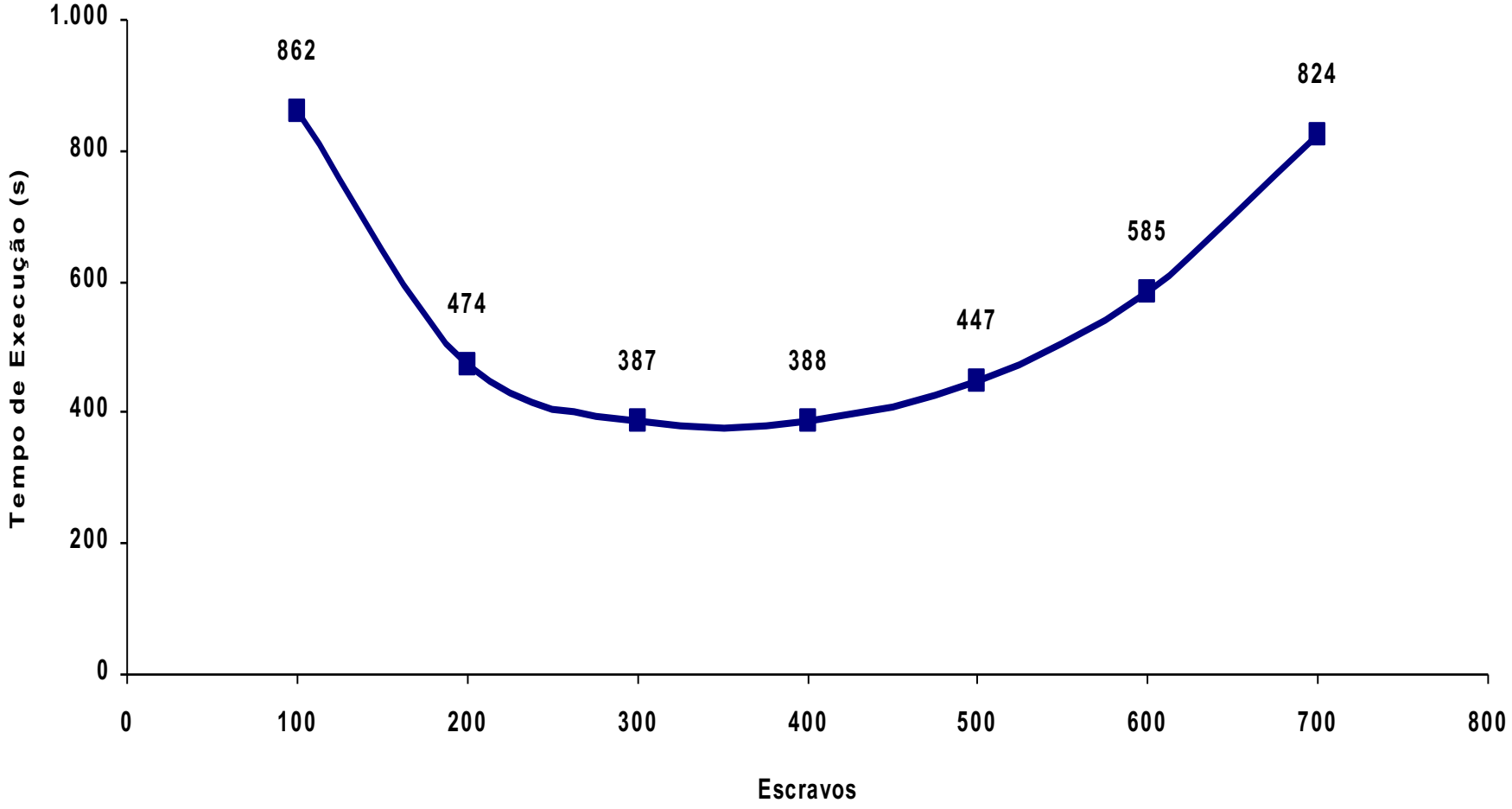


brams.cptec.inpe.br

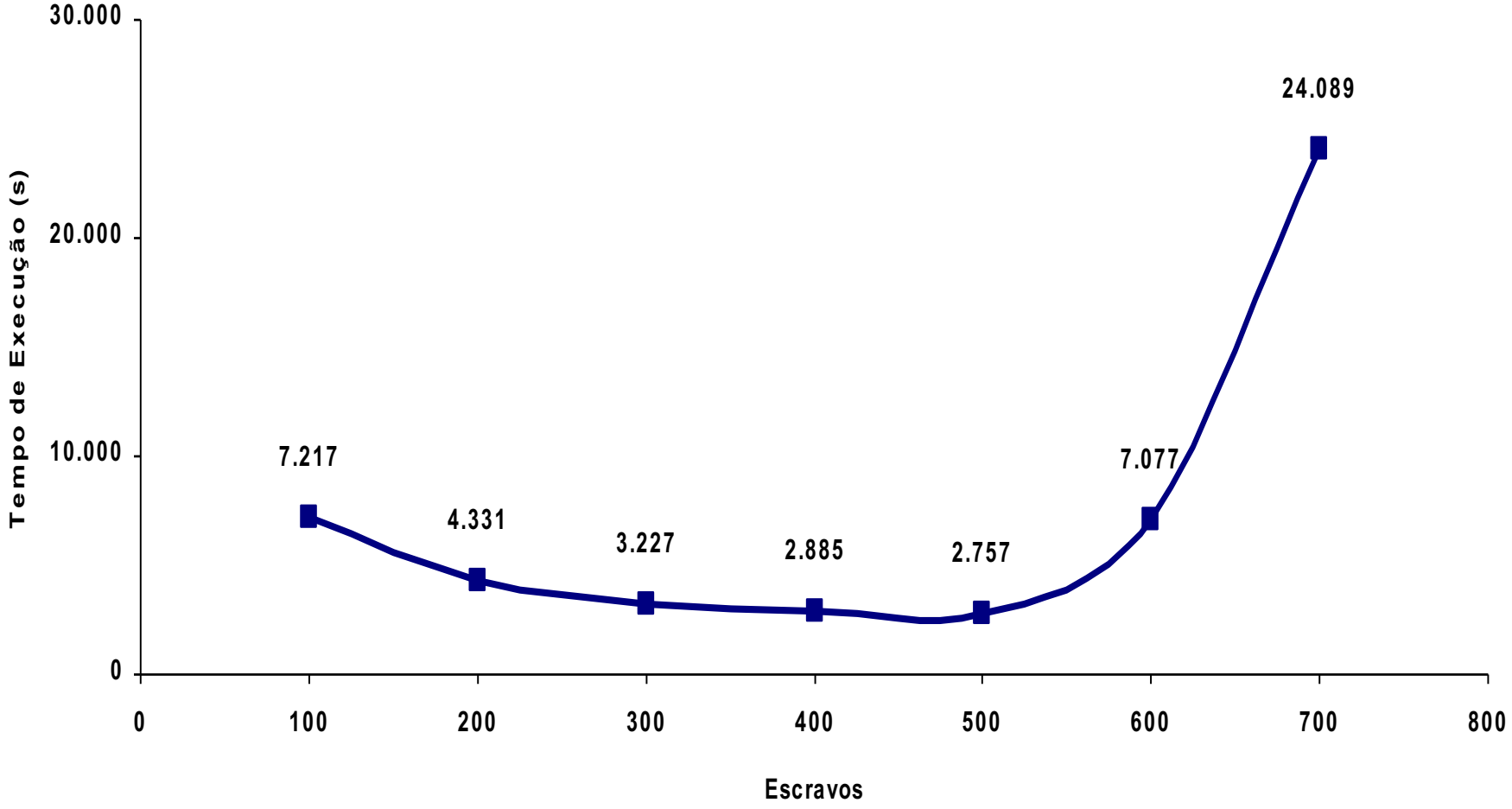
Objetivo 2007 - 2008

- Escalar de 100 núcleos para 1000 núcleos
  - Financiado CNPq, Grandes Desafios da Computação
- Objetivo CPTEC:
  - Aumentar a resolução do BRAMS de produção
  - Domínio: América Latina
  - Resolução: de 40km para 20km e talvez 10km

Tempo de Execução Original, 20 km



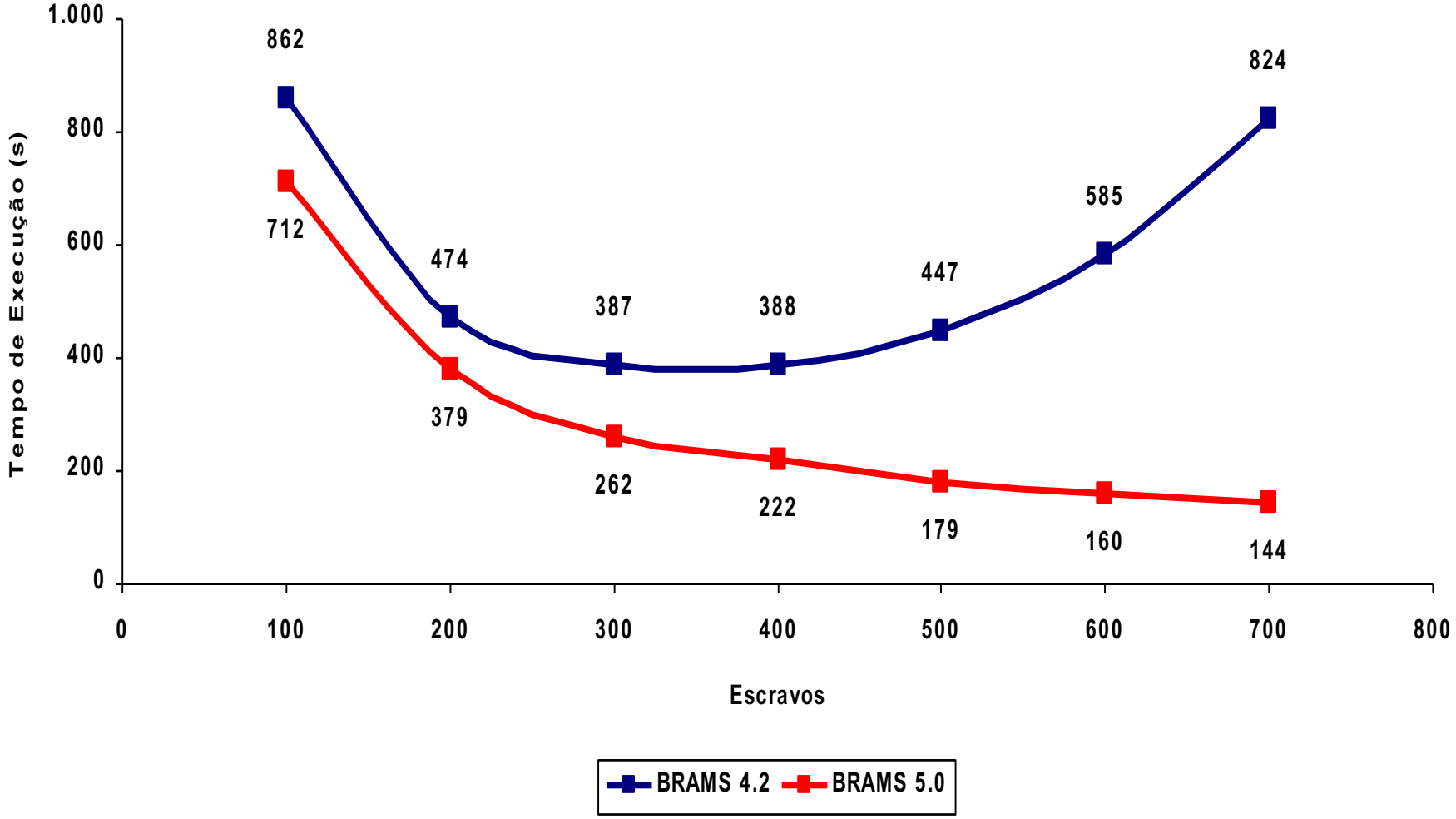
Tempo de Execução Original, 10 km



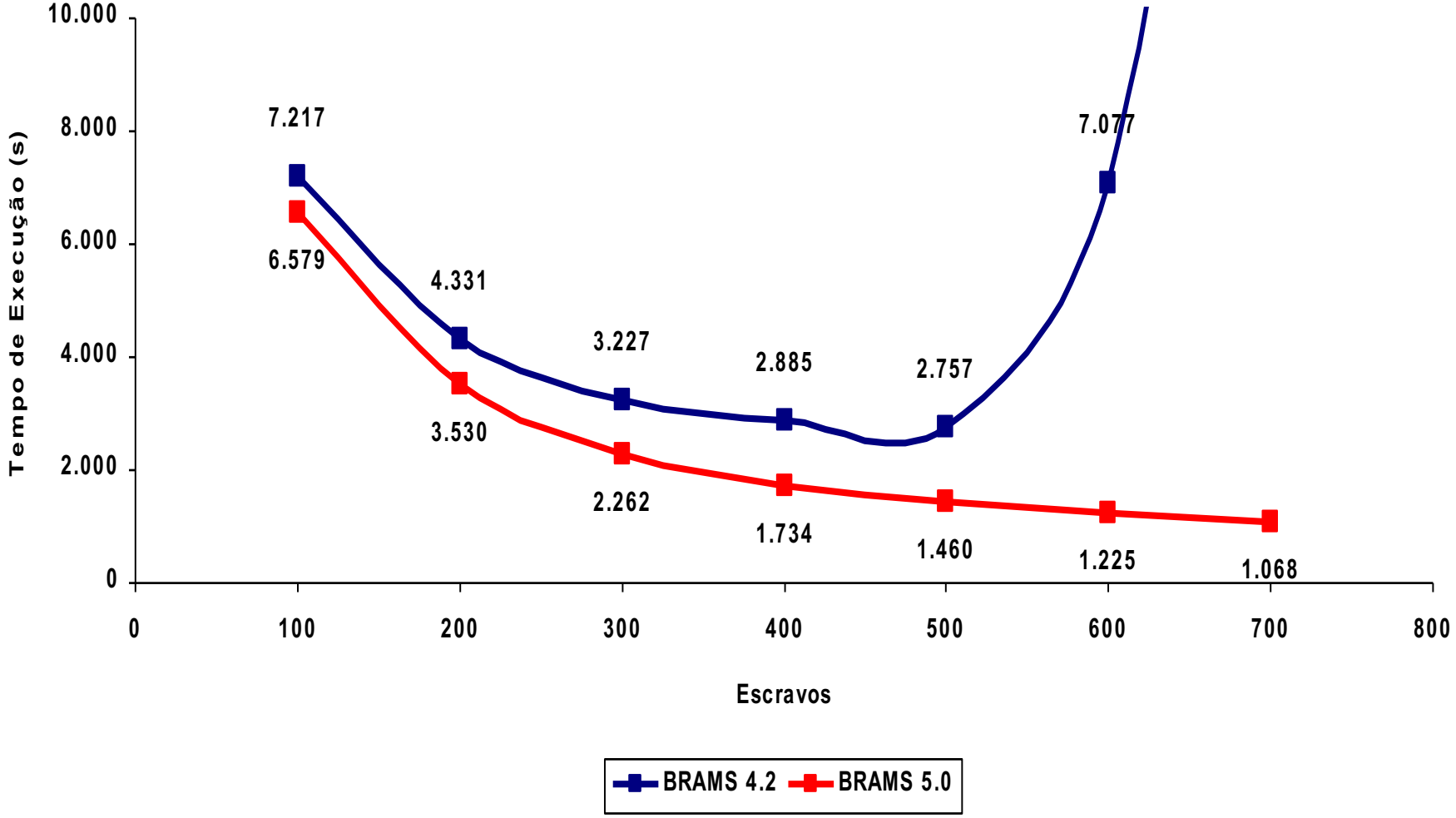
Diagnóstico

- Fases com tempo crescente:
  - Algoritmo possui componente sequencial
- Motivo: composição e decomposição do domínio
  - Um processo compõe e decompõe campos
- Exemplo: Saída
  - Todos os processos enviam seus sub-domínios para um único processo
  - O único processo compõe o domínio completo e escreve
- Solução:
  - Evitar composição/decomposição, quando possível
  - Quando impossível, minimizar componente sequencial

Tempo de Execução Final, 20km



Tempo de Execução Final, 10km



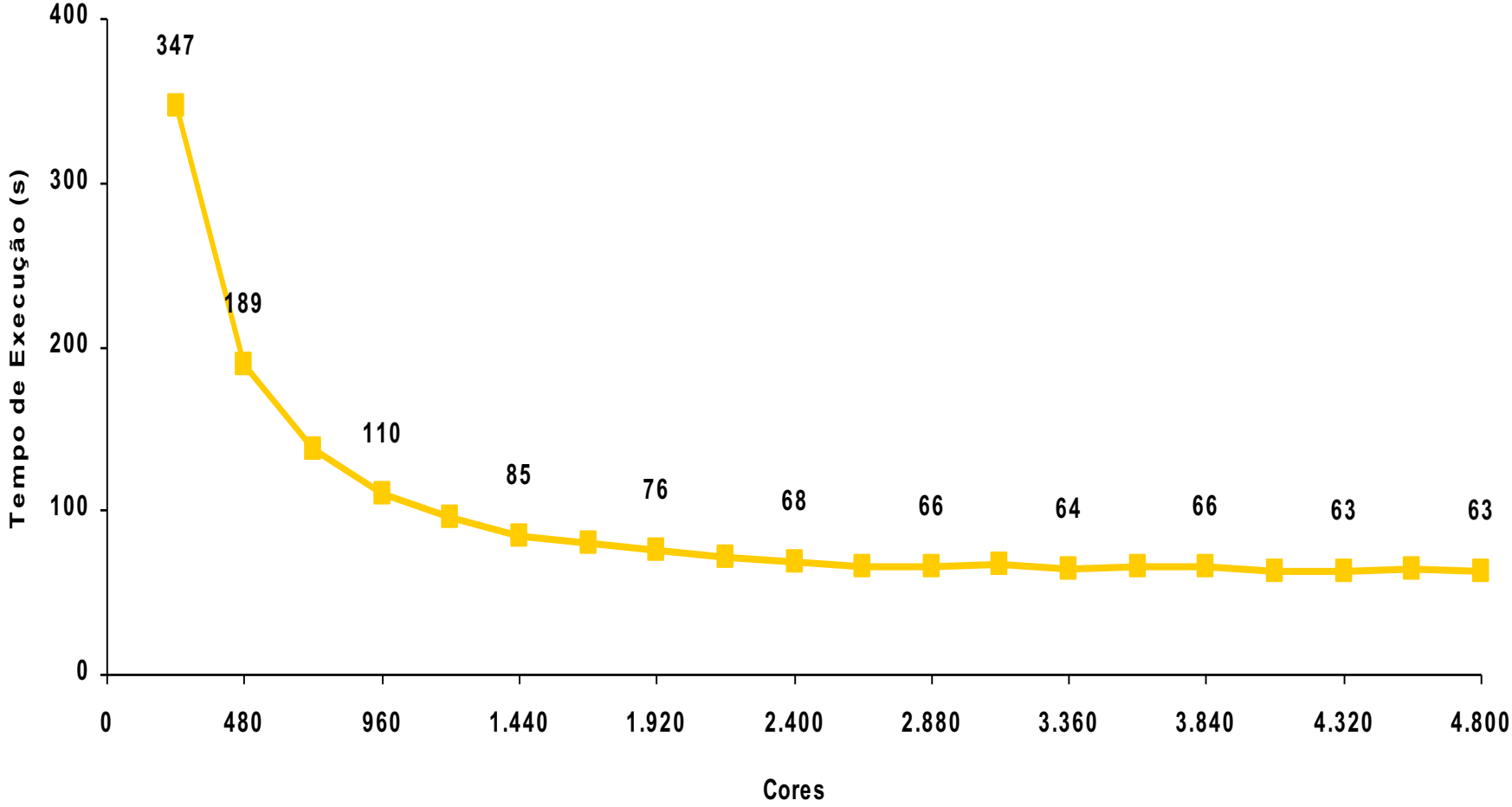
Objetivo 2010

- Escalar de 1.000 núcleos para 10.000 núcleos
  - Financiado CNPq, Grandes Desafios da Computação
- Visa aumentar a escalabilidade mantendo resolução

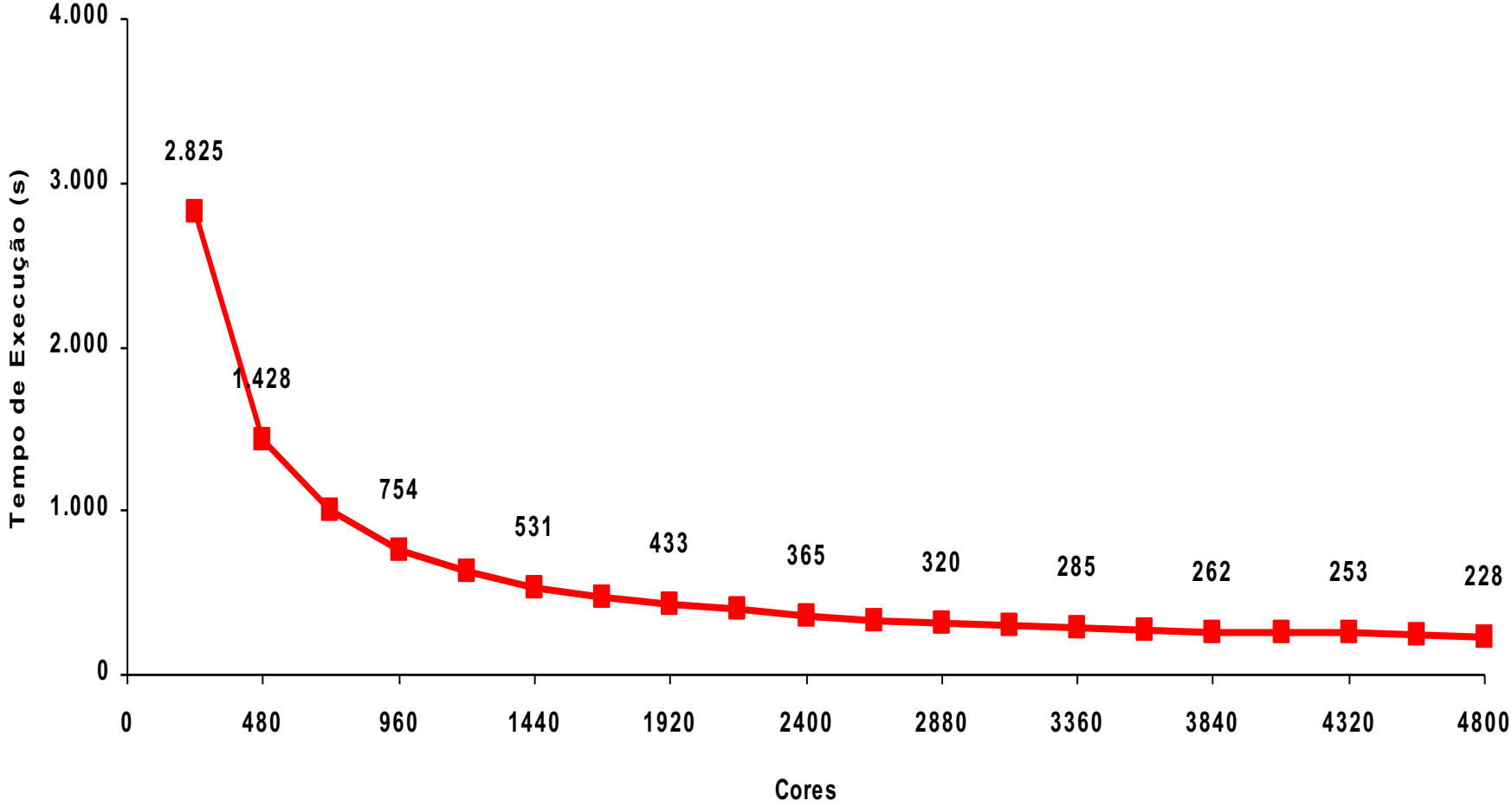
Diagnóstico

- Mestre utiliza tempo em excesso
- Estrutura mestre-escravo impede escalabilidade
- Solução: eliminar mestre-escravo; todos os processos computam

Tempo de Execução Final, 20km



Tempo de Execução Final, 10km

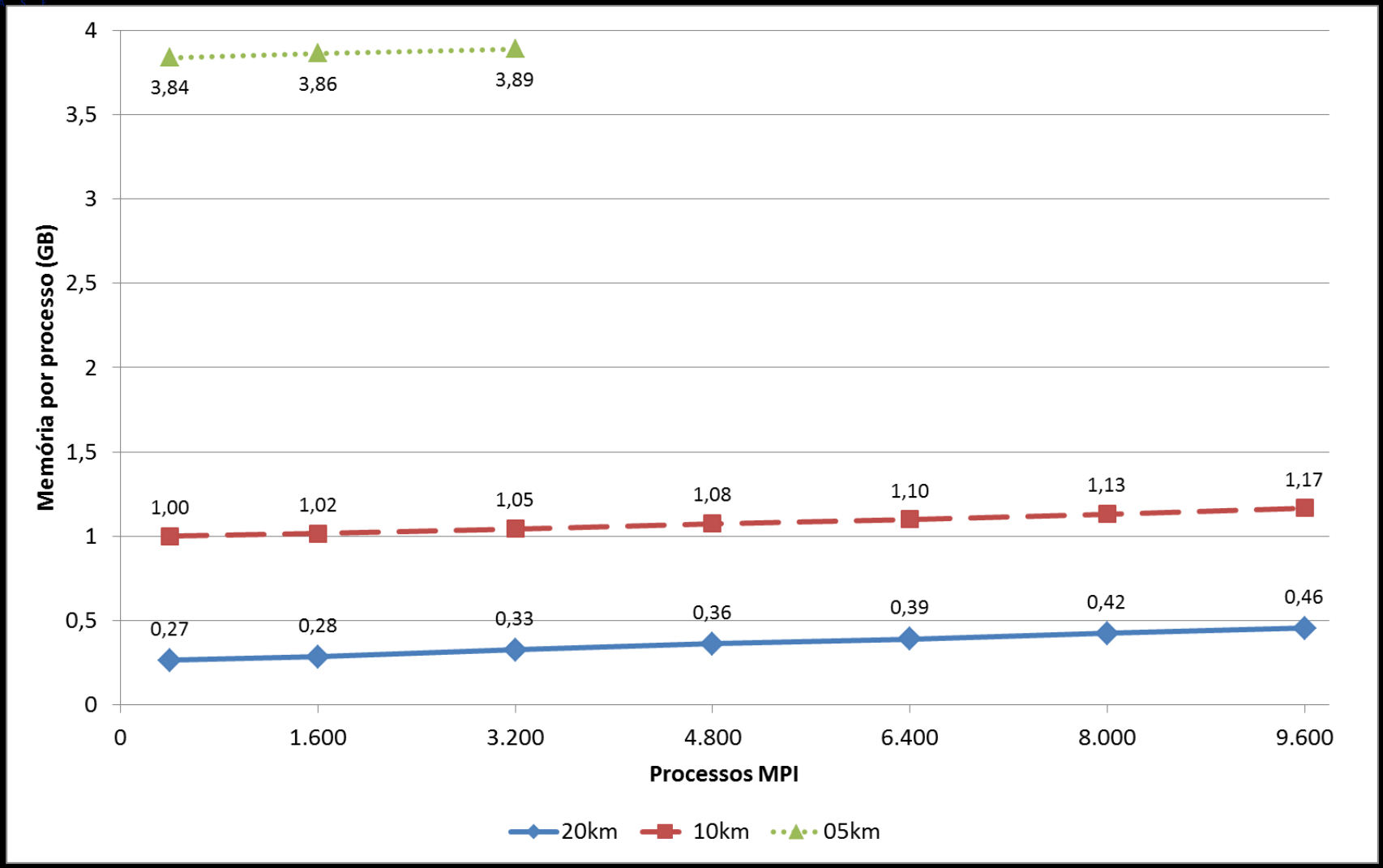


Objetivo 2011

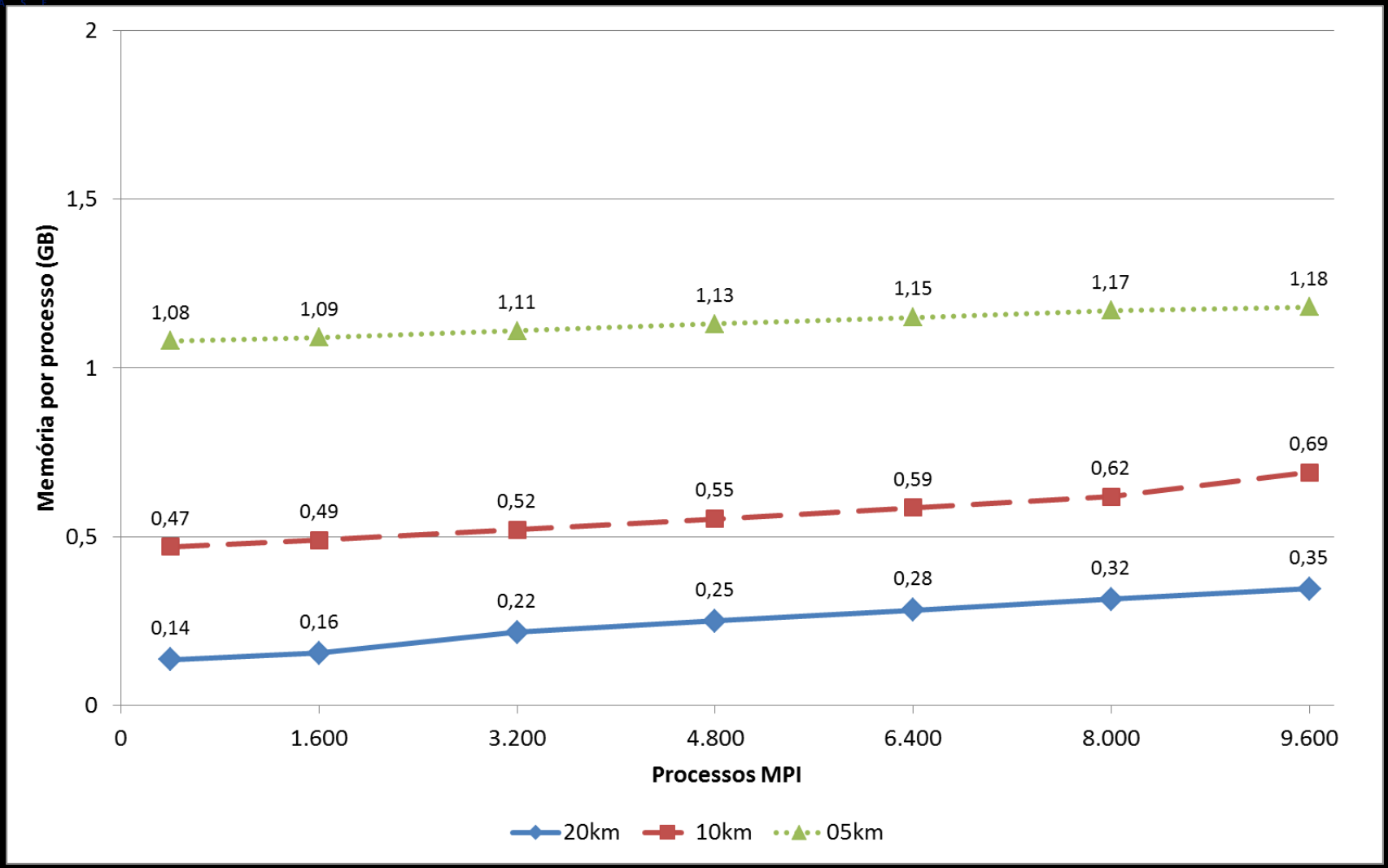
- Objetivo CPTEC:
  - Aumentar a resolução do BRAMS de produção
  - Resolução 5km

- Problema:
  - BRAMS em 5 km usa memória em excesso
  - 32GB por nó limitam execuções a 8 núcleos por nó, dos 24 núcleos disponíveis

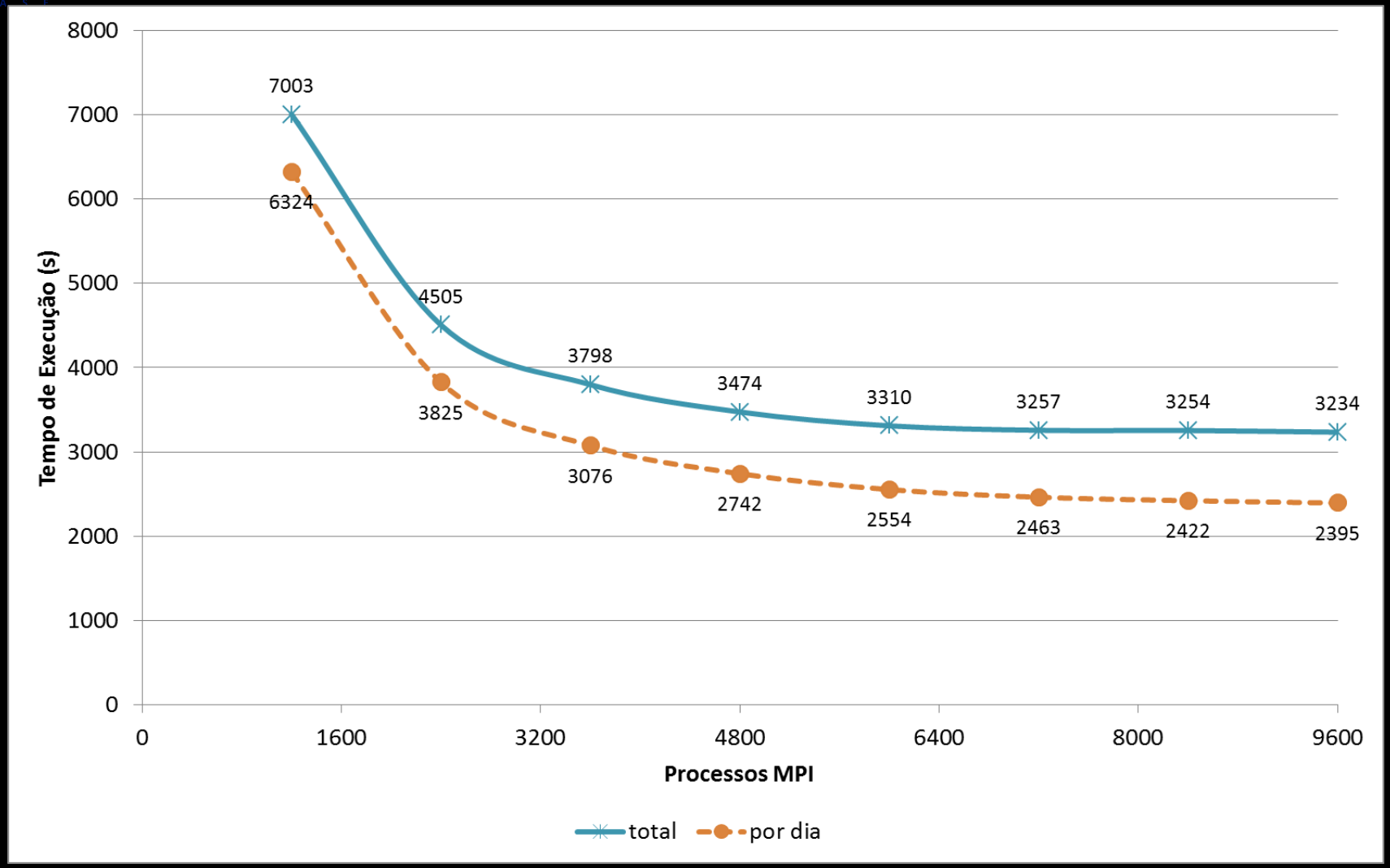
Uso de Memória 5 km



Após 3 meses de trabalho



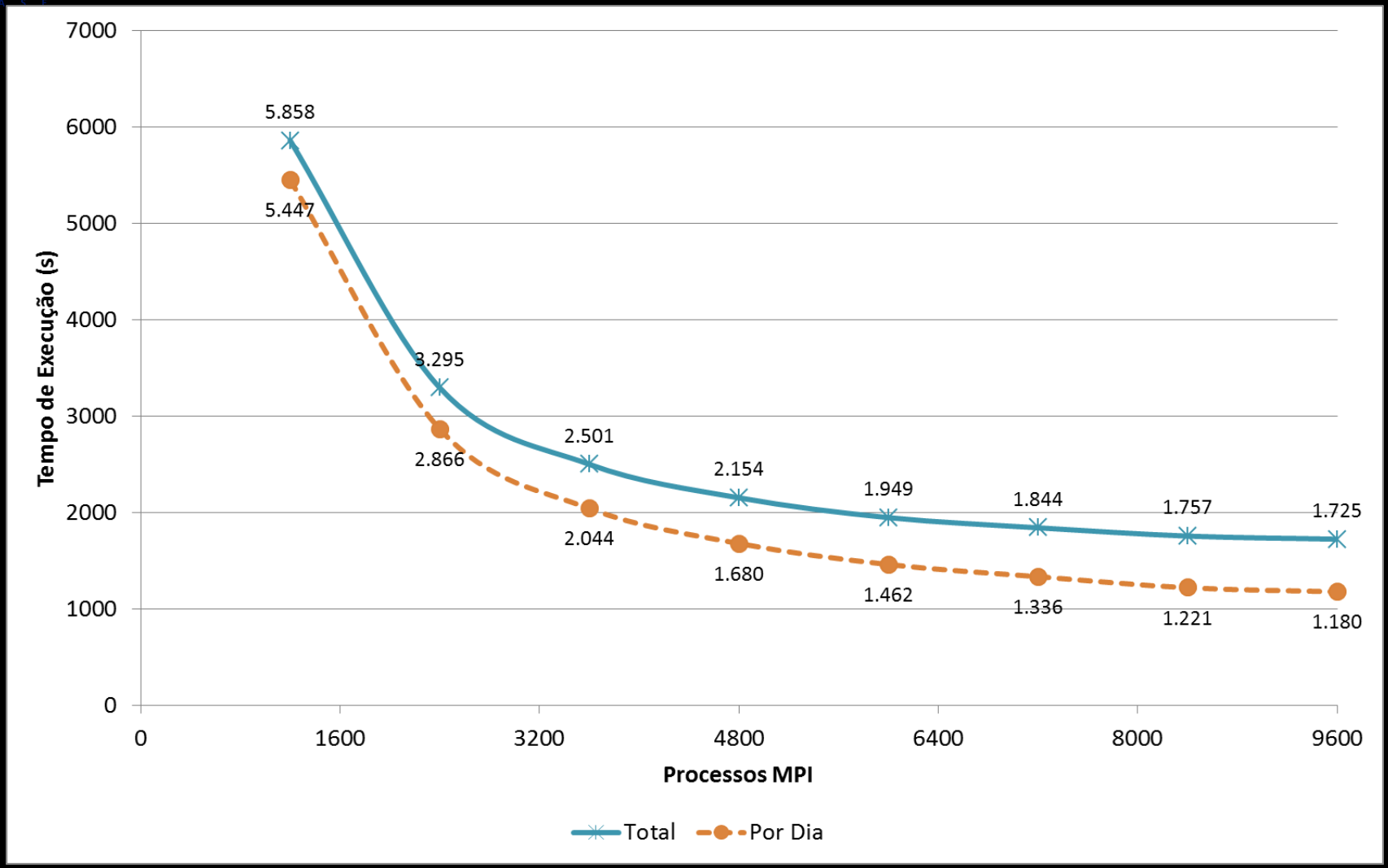
Tempo de Execução Após Redução de Memória



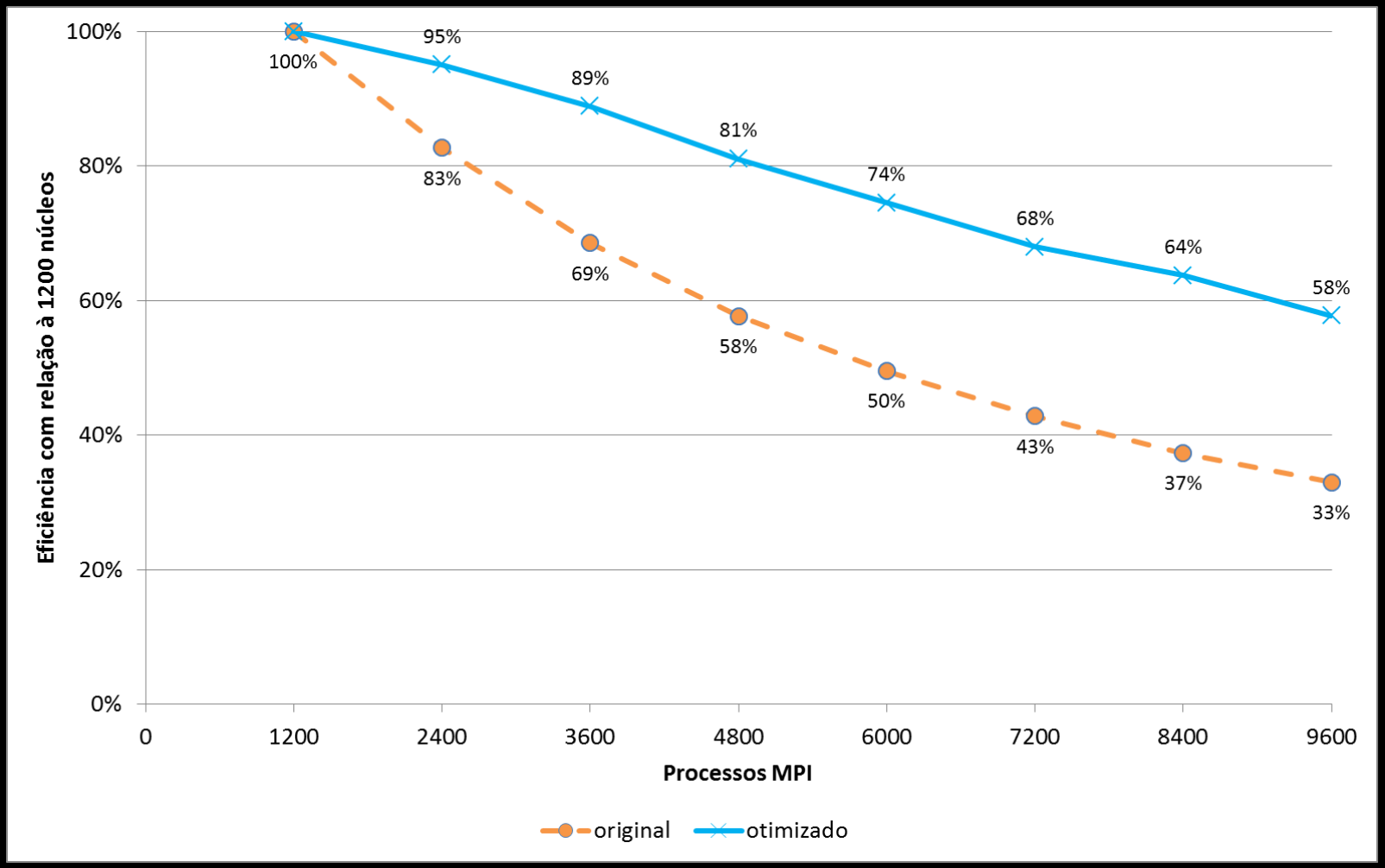
Diagnóstico

- Tempo da saída domina a computação:
  - Constante com o número de núcleos
- Solução: I/O Paralelo
  - MPI-I/O e HDF5

Tempo de Execução Otimizado



Eficiência Paralela

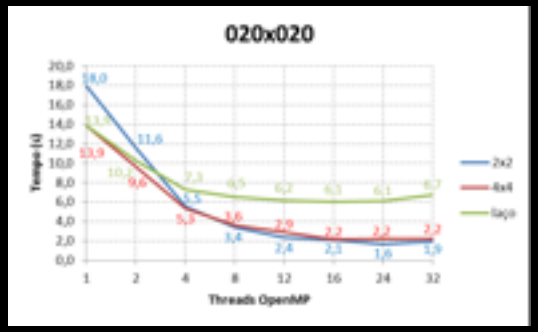


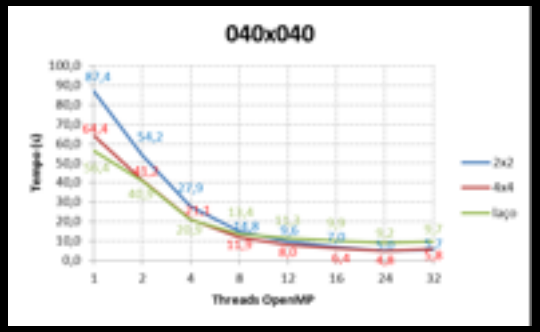
Escalabilidade ratificada no Santos Dumont até 13.000 núcleos

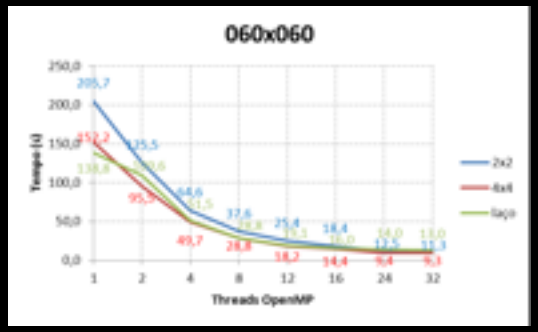
Desde então:

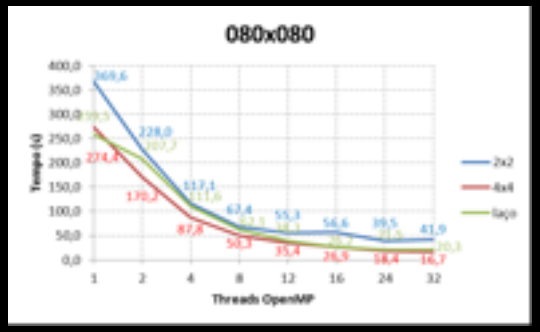
- Escalar de 10.000 núcleos para 100.000 núcleos
  - Financiamentos IBM, INTEL, FAPESP, CNPq
- Mecanismo: Paralelismo Híbrido
  - MPI + OpenMP
  - MPI + Aceleradores

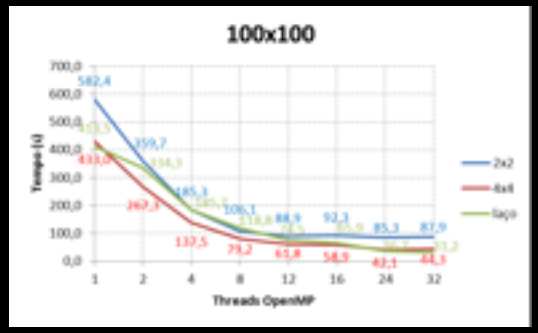
Paralelismo Hibrido: Advecção de Escalares em OpenMP



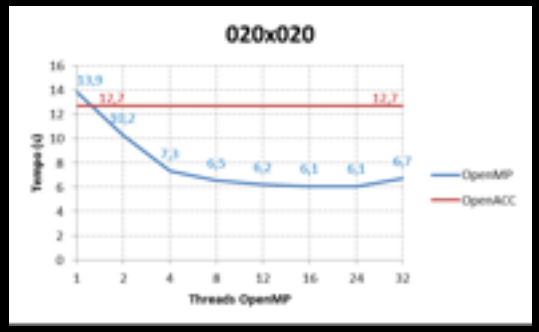


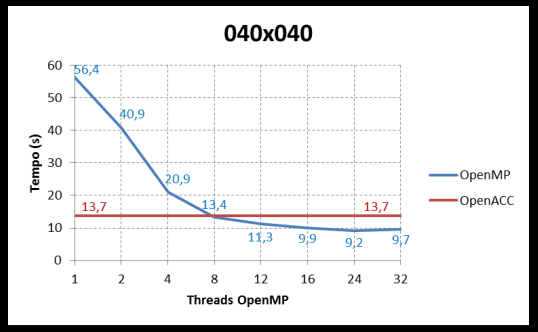


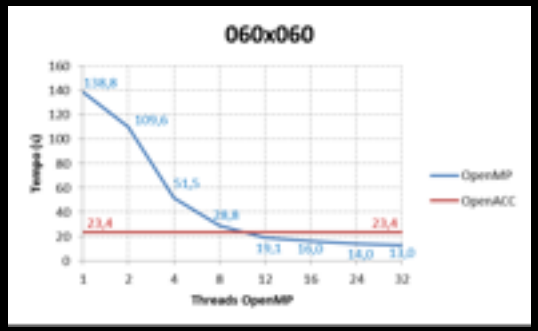


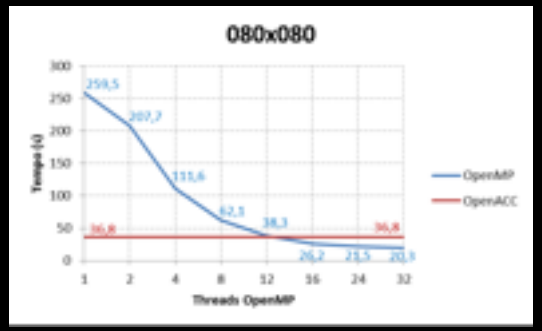


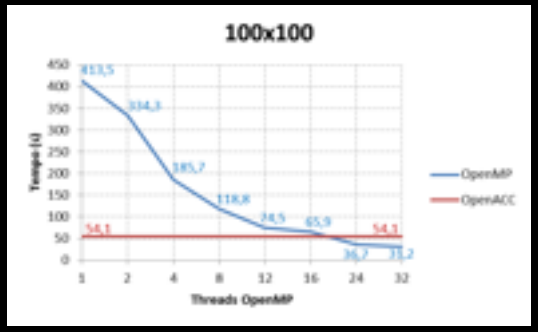
Paralelismo Hibrido: Advecção de Escalares na GPU











Conclusões

- Paralelismo para uma escala de processadores e resolução não atende a próxima escala de processadores e resolução
- Ineficiências em locais inesperados, sem “glamour” acadêmico
- Enorme esforço de desenvolvimento ao longo dos anos