

Avaliação de Desempenho utilizando biblioteca MPI sobre Infraestrutura de Comunicação de Baixo Nível no Nanvix

João Fellipe Uller

Universidade Federal de Santa Catarina
Depto. de Informática e Estatística (INE)

INE410129-41000025DO/ME (20201) - Computação Paralela
Prof. Dr. Márcio Bastos Castro

joao.f.uller@grad.ufsc.br

29 de Setembro de 2020

- 1 Motivação
 - Contexto
 - *Lightweight manycores*
- 2 LWMPI
 - Visão geral
 - Nanvix
 - Estrutura
 - Suporte Atual
- 3 Proposta
 - Objetivo geral
 - Metodologia
 - Arquitetura alvo
- 4 Referências

- Poder computacional vs Consumo energético
- Preocupação crescente com a eficiência energética

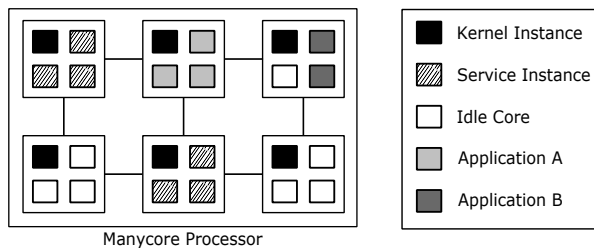
- Milhares de núcleos de baixa potência num mesmo *chip*, agrupados em *clusters*
 - Alto grau de paralelismo entre *threads*!
- Comunicação baseada em *Message Passing* através de uma *Network-on-Chip* (NoC) de alta vazão
- Subsistemas de memória restritivos

- Porte de aplicações é uma tarefa complexa
 - Precisa-se levar em conta características arquiteturas
- Necessidade de ambientes de desenvolvimento que equilibrem desempenho, portabilidade e programabilidade
 - **Implementação de uma biblioteca de comunicação segundo o padrão Message Passing Interface (MPI), compatível com as restrições de Lightweight Manycores**

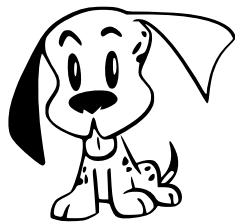
- **Lightweight Message Passing Interface**
- Biblioteca de comunicação compatível com o padrão MPI (versão 3.1)
 - Implementação "*from scratch*" para ser compatível com as restrições de *lightweight manycores*
 - Implementa um subconjunto das funções MPI
- Desenvolvida sobre a infraestrutura de Comunicação entre Processos (IPC) do Nanvix

Nanvix

- Sistema operacional (SO) de código aberto, compatível com POSIX (<https://github.com/nanvix/>)
- Sistema distribuído para arquiteturas *manycore*
- Baseado em uma estrutura **multikernel**
 - Múltiplas instâncias de um *microkernel* assimétrico



(a) Nanvix *Multikernel*



(b) Nanvix *Logo*

- *Mailbox*
- *Portal*
- *Sync*

Estrutura LWMPI

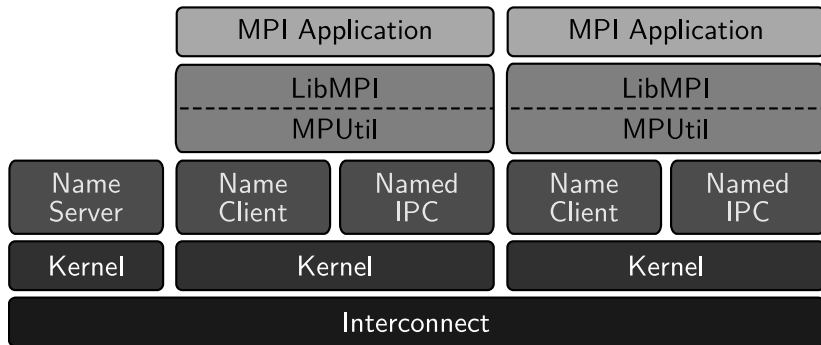


Figura: Estrutura LWMPI

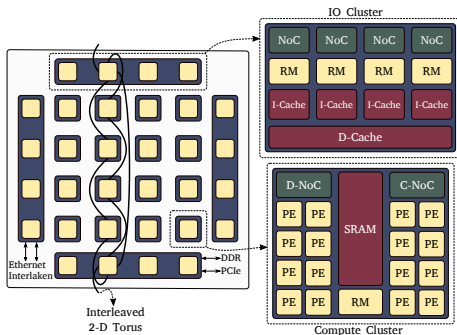
- Comunicadores
- Grupos de comunicação
- Tratadores de erros
- *Datatypes* padrão da linguagem C
- Comunicação ponto a ponto no modo síncrono (`MPI_Send` e `MPI_Recv`)


- Porte de algumas aplicações utilizadas em *benchmarking* para avaliar o desempenho da biblioteca desenvolvida (3 a 5 aplicações)
- Comparação com a infraestrutura de comunicação de baixo nível do Nanvix (Nanvix IPC)
 - Forma de avaliar o *overhead* introduzido pelo *runtime*
- Avaliação do tempo total, *speedup*, além de técnicas de *profiling* para avaliar possíveis gargalos e possibilidades de melhoria


- Suite de *benchmarking* utilizada para medir o desempenho de *lightweight manycores*
- Caracterizado por sete aplicações que empregam diferentes padrões de computação e de paralelismo para avaliar um conjunto maior de características
 - *Features from Accelerated Segment Test (FAST)*
 - **Friendly Numbers (FN)**
 - **Gaussian Filter (GF)**
 - *Integer Sort (IS)*
 - **K-Means (KM)**
 - *LU Factorization (LU)*
 - *Traveling-Salesman Problem (TSP)*
- Cenários variando de 1 a 15 *workers*, além do processo líder


Kalray MPPA-256

- 16 *Compute Clusters* (CCs)
 - 16 *Processing Elements* (PEs) + 1 *Resource Manager* (RM)
 - 2 MB de memória local compartilhada
- 4 *IO Clusters* (IOs) para comunicação com periféricos
- 2 *Networks-on-Chip* (NoCs)
 - *Command NoC* (C-NoC)
 - *Data NoC* (D-NoC)



 João V. Souto and Pedro H. Penna and Castro, Márcio. *An Inter Cluster Communication Facility for Lightweight Manycore Processors in the Nanvix OS*. UFSC, 2019.

 Message Passing Interface Forum. *MPI: A Message Passing Interface Standard Version 3.1* HLRS, 2015

 Pedro H. Penna and Souto, João and Lima, Davidson F. and Castro, Márcio and Broquedis, François and Freitas, Henrique H. and Mehaut, Jean François. *On the Performance and Isolation of Asymmetric Microkernel Design for Lightweight Manycores*. HAL-Archive-Ouvertes, 2019.

Dúvidas?
Obrigado!