Ze-yao MO, 2018. Extreme scale parallel computing: bottlenecks and strategies. *Frontiers of Information Technology & Electronic Engineering*, 19(10):1251-1260. https://doi.org/10.1631/FITEE.1800421

Extreme-scale parallel computing: bottlenecks and strategies

Key words: Extreme scale; Numerical simulation; Parallel computing; Supercomputers

Corresponding author: Ze-yao MO
E-mail: zeyao_mo@iapcm.ac.cn
ORCID: Ze-yao MO, http://orcid.org/0000-0003-3280-5682

Motivations

1. Parallel computing research bridges numerical simulation and computer architectures, and supports the high accuracy of numerical simulations by continuously improving the computing capabilities of numerical software.

2. With increases of supercomputer performance, parallel computing research has exhibited three major bottlenecks: limited computational scale, inadequate computing efficiency, and low programming productivity.

3. These bottlenecks collectively inhibit the acquisition of computing capability of extreme-scale numerical applications, and keep widening the gap of software development productivity and supercomputer advancements.

Main idea

We systematically analyze these bottlenecks and propose a series of research-demanding key technology issues and effective coping strategies.

Major results

1. The computational scale bottleneck lays in the inability for numerical algorithms to achieve linear complexity

$t_{N,P} = O(N \log(NP)),$ which hurts the weak scalability of these algorithms and leads to limited gains for parallel computing on large-scale problems.

2. Common algorithm frameworks and application-feature-driven preconditioners can help to achieve linear complexity. The former simplifies the integration of fast algorithms and the latter solve the complexity problem by application-specific algorithms.

Major results (Cont.d)

1. The computing efficiency bottleneck lays in the inability for high performance implementations to adapt to ever-changing complex computer architecture, which keeps widening the gap of real-world application performance and theoretical peaks.

2. By abstracting parallelisms in numerical algorithms as parallel computing patterns, the work required to efficiently implement numerical algorithms goes from $O(M\ N)$ to $O(\alpha N)$.

3. Parallel computing patterns can be implemented in parallel programming frameworks to support a wide range of applications.

Major results (Cont.d)

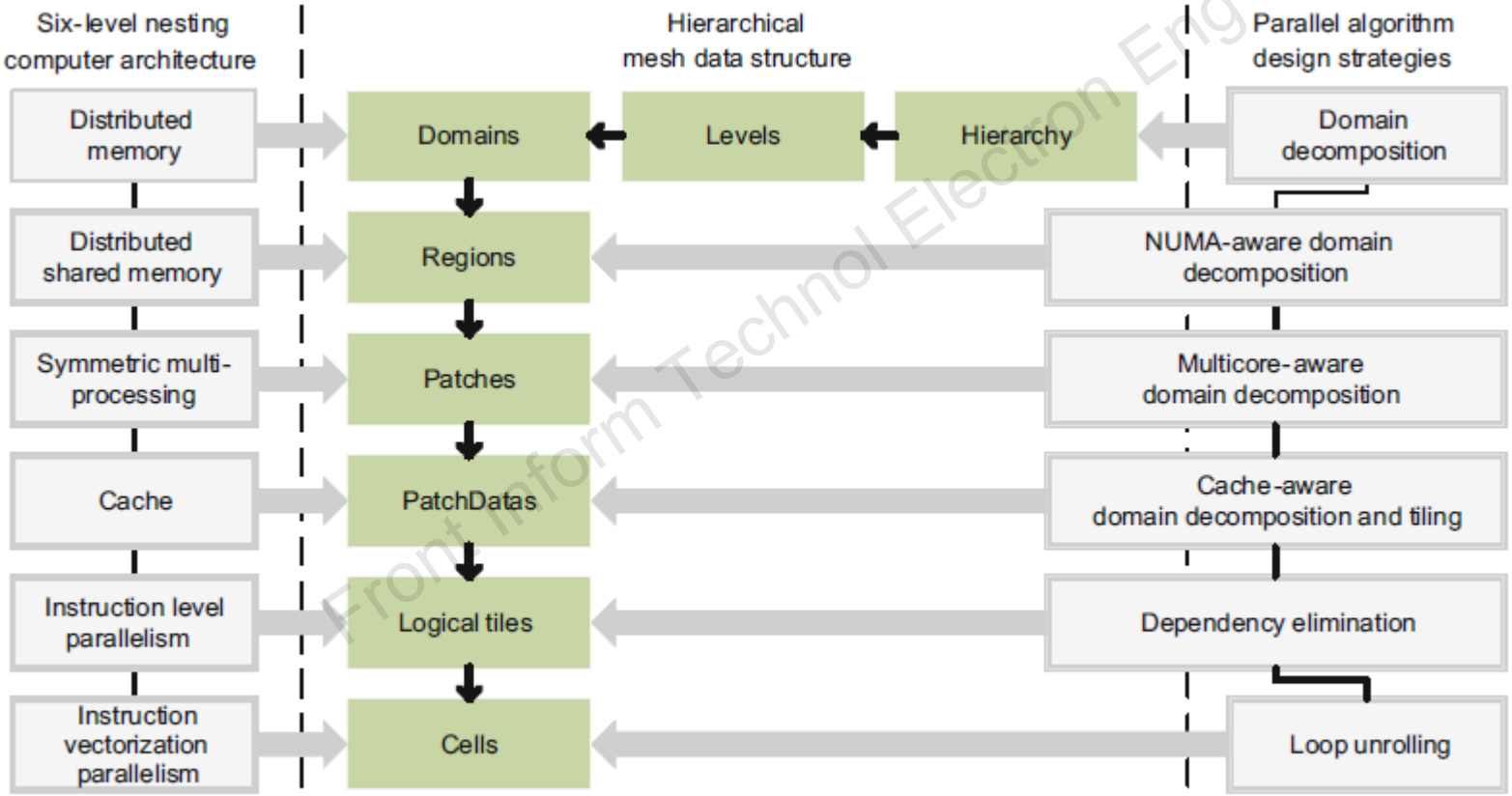

Fig. 1 Six-level nesting hierarchical supercomputer architecture and corresponding parallel algorithm design strategies

Major results (Cont.d)

1. The programming productivity bottleneck lays in the inability for parallel programming models to hide parallel computing complexity.

2. Reuse of parallel algorithm and programming is an effective way to improve productivity.

3. The key technologies comprise two aspects: the reuse of parallel computing patterns and reuse of numerical algorithm frameworks.

4. Based on a component-based programming framework like JASMIN, developers can achieve an ‘automatic parallelization and high scalability’ goal without knowledge of explicit parallel programming models.

Major results (Cont.d)

Supply-side technology in parallel computing research is essential for assessing weaknesses in computing capability.

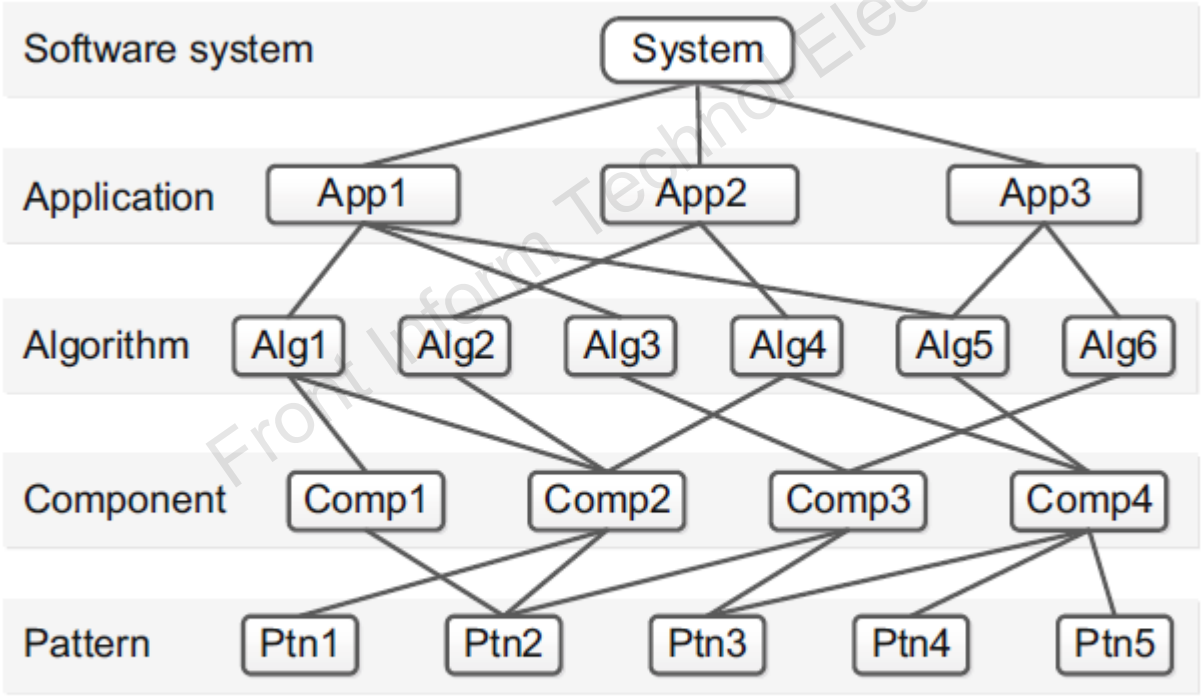

Fig. 2 Architecture of computing-capability evaluation models

Conclusions

1. To address the exascale computing challenge, we have systematically analyzed the parallel computing bottlenecks from three perspectives: computational scale, computing efficiency, and programming productivity.

2. For the computational scale bottleneck, the keys are numerical algorithm frameworks, application-driven pre-conditioners, and their coupling bottlenecks.

3. For the computing efficiency bottleneck, the keys would be data communication, load balancing, and floating-point performance optimizations.

4. For the programming productivity bottlenecks, domain-specific parallel programming frameworks would be a viable and effective solution.

5. To tackle these bottlenecks with minimal investment, supply-side supporting technology is essential.