



POLITECHNIKA POZNAŃSKA

WYDZIAŁ INFORMATYKI I TELEKOMUNIKACJI

Instytut Informatyki

Steszczenie rozprawy doktorskiej

Energooszczędne zarządzanie zasobami
dla obliczeń stencilowych w systemach
superkomputerowych

Energy-aware resource management for
stencil computations in High
Performance Computing

mgr inż. Miłosz Ciżnicki

Promotor: prof. zw. dr hab. inż. Jan Węglarz

Promotor pomocniczy: dr hab. inż. Krzysztof Kurowski

Poznań, 2023

W obliczu osiągnięcia przez superkomputery wydajności na poziomie exascale dzięki dużej liczbie rdzeni obliczeniowych, niezwykle istotne staje się nie tylko dążenie do maksymalizacji wydajności, ale także skalowalności aplikacji równoległych, lokalności danych i świadomego zużycia energii. Obecnie istniejące aplikacje równoległe muszą ewoluować poprzez przeprojektowanie algorytmów i struktur danych, aby wykorzystać ostatnie udoskonalenia w wydajności energetycznej heterogenicznych architektur komputerowych, włączając wielordzeniowe procesory i akceleratory GPU. W kolejnych latach jednym z największych wyzwań dla obliczeń exascale będzie osiągnięcie zdolności aplikacji równoległych do pełnego wykorzystania lokalności danych, która będzie wymagana, aby uzyskać oczekiwaną wydajność obliczeniową i efektywność energetyczną. Opracowywane w przyszłości aplikacje równoległe będą musiały zmierzyć się z wielopoziomowymi hierarchiami pamięci, uwzględniając koszt energetyczny niezbędny do komunikowania się i przenoszenia danych poza układ scalony. W związku z tym będą musiały być opracowane i zastosowane nowe skoordynowane metody szeregowania, aby zrównoważyć zużycie energii zasobów oraz zminimalizować zagłócenie procesów podczas ich wykonywania.

Obliczenia stencilowe stanowią ważną klasę aplikacji, które pojawiają się w wielu kodach HPC wykonywanych na blokowych siatkach strukturalnych używanych do modelowania różnych zjawisk fizycznych, przykładowo obliczeniowa mechanika płynów, geometria obliczeniowa, rozwiązywanie równań różniczkowych cząstkowych lub modyfikacja zdjęć i wideo. Ze względu na to, że czas przetwarzania i zużycie pamięci rośnie liniowo wraz z liczbą elementów siatki, nasze badania skupiają się na równoległych implementacjach obliczeń stencilowych razem z szeregowaniem zadań i technikami optymalizacji biorącymi pod uwagę koszt energetyczny i lokalność danych. Podczas naszych badań udowodniliśmy, że ostatnie zmiany wprowadzone w heterogenicznych architekturach komputerowych doprowadziły do tego, że różne cechy wydajnościowe i energetyczne są krytyczne dla efektywnych i skalowalnych obliczeń stencilowych.

Według naszego stanu wiedzy, żadne z dotychczasowych badań nie uwzględniło tak szczegółowej analizy zużycia energii przy dystrybucji obliczeń sten-

ciowych na heterogenicznych zasobach z ograniczeniami czasu wykonania. Dla wybranych referencyjnych obliczeń stencilowych dotychczasowe badania nie próbowały minimalizować zużycia energii na komunikacje wewnątrz węzła, jak i również pomiędzy węzłami w odniesieniu do różnych topologii ich połączeń, co jak zostało wykazane w rozprawie znacząco może wpływać na oszczędność energii w szczególności w przypadku większych obliczeń.

W niniejszej rozprawie sformułowano i zaprezentowano model szeregowania dla efektywnych obliczeń stencilowych. Topologia uwzględnia różne połączenia komunikacyjne w ramach węzła, jak i pomiędzy węzłami dla heterogenicznych procesorów. Opisano innowacyjne, analityczne oraz metodologiczne podejście modelowania i przewidywania zużycia energii i czasu wykonania dla referencyjnych obliczeń stencilowych na pojedynczych, jak i wielu węzłach dla heterogenicznych architektur HPC. Przebadano i zaprezentowano nowe heurystyczne algorytmy dla równoczesnej minimalizacji zużycia energii i czasu wykonania dla małych oraz dużych zadań obliczeń stencilowych. Zaproponowano metodę zarządzania zadaniami, uwzględniającą zużycie energii, która bierze pod uwagę złożone sprzętowe architektury HPC, wliczając w to różne konfiguracje sprzętowe oraz topologie sieciowe na bazie rzeczywistych największych instalacji superkomputerowych. Opisano nowy algorytm przeszukiwania tabu w celu rozwiązania problemu i przedstawiono wyniki eksperymentów obliczeniowych, uwzględniając jego wydajność w porównaniu do prostych heurystyk. Opracowano elastyczny i generyczny model szeregowania oraz algorytmy heurystyczne, które mogą być łatwo rozszerzane albo zaadaptowane w istniejących platformach programistycznych dla obliczeń stencilowych.

Niniejsza rozprawa doktorska stanowi wkład w rozwój energooszczędnych metod obliczeniowych w kontekście heterogenicznych architektur HPC, proponując nowe podejścia modelowania, szeregowania zadań i optymalizacji, które mogą znacząco poprawić wydajność i efektywność energetyczną obliczeń stencilowych w złożonych konfiguracjach wielowęzłowych.