

# 동적 스케줄링으로 양자-HPC 클러스터 효율 개선

(2025.09.03., 양자정보연구지원센터)

## □ 슈퍼컴퓨터와 양자컴퓨터의 협업: 동적 스케줄링으로 양자-HPC 클러스터 효율 개선

### ○ 연구 개요

- (배경) 슈퍼컴퓨터(HPC)와 양자 프로세서(QPU)를 결합한 하이브리드 컴퓨팅 시스템 등장
- 기존 방식에서는 자원 비효율 문제 발생: HPC 노드와 QPU를 동시에 예약해도 실제로는 한쪽만 사용되는 경우 많음
- 목표: 낭비를 줄이고 작업 완료 시간을 단축하며 하이브리드 시스템 효율 개선
- (연구 목적) HPC와 QPU 간 동적 스케줄링(dynamic scheduling, malleability) 적용 가능성 평가
- 소수의 QPU를 효율적으로 활용하면서 HPC 자원의 활용 극대화

### ○ 하이브리드 시스템과 자원 문제

- (QPU 역할) 특정 문제(최적화, 재료 모델링, 양자 시뮬레이션 등)에서 가속기 역할 수행
- 대부분 계산은 클래식 HPC 시스템에서 처리
- (기존 문제) 한 하이브리드 작업이 전체 런타임 동안 HPC 노드와 QPU 모두 예약
- CPU 코어가 놀거나 QPU가 기다리는 비효율 발생
- QPU 수가 적을수록 병목 가능성 증가

### ○ 접근법

- (연구팀) E4 Computer Engineering, LINKS Foundation, Barcelona Supercomputing Center, CINECA, Università di Torino 참여
- 워크플로 기반 관리 (Workflow-based) : 작업을 독립적 단계로 나

누어 필요할 때만 자원 요청

- StreamFlow WMS 사용, HPC 단계 → 양자 단계 → 품질 평가 반복
- 양자 자원만 필요 시 예약, HPC 노드는 비활성 시간에 다른 작업에 활용 가능
- 유연성 기반 관리 (Malleability-based) : 실행 중인 작업이 사용 노드 수를 동적으로 조절
- DMR(Dynamic Management of Resources) 프레임워크 사용
- 양자 단계로 전환 시 HPC 노드를 최소화하고, 클래식 단계로 돌아오면 확대
- (테스트 사례) 기존 클러스터링 알고리즘을 하이브리드 HPC-QC로 변환
- HPC 단계: k-means, DBSCAN, 계층적 클러스터링 병렬 수행
- 양자 단계: QUBO 문제로 변환 후 양자/에뮬레이터에서 해결

○ 실험결과

- (실험 환경) SLURM 관리 클러스터, 3개의 CPU 노드 + 1 “양자” 노드
- QPU 부재로 양자 에뮬레이터 사용, 인위적 지연 포함
- (비교 시나리오) 기본(Baseline): 전체 런타임 동안 HPC+QPU 정적 할당
- 워크플로(Workflow): 필요 시만 양자 자원 요청
- 유연성(Malleability): 양자 단계에서 HPC 노드 최소화, 종료 후 확장
- 짧은 양자 단계: HPC 사용량 절약은 작지만, 병렬 작업 처리에서 malleability가 유리
- 긴 양자 단계: baseline은 자원 비효율, workflow는 자원 요청 지연, malleability는 완료 시간 단축 + 자원 활용 극대화
- 동시 작업 시 malleability가 가장 효율적

○ 의의 및 적용

- (시사점) 정적 스케줄링은 하이브리드 워크로드에 부적합
- 자원이 제한된 클러스터에서 동적 할당이 HPC 활용과 큐 대기 시간 감소에 중요
- 소프트웨어 · 스케줄링 개선은 양자 속도 향상만큼 중요
- (고려사항) 워크플로: 모듈화, 워크플로 언어 필요
- 유연성: 기존 코드 통합 용이, 상태 관리 복잡성 증가

○ 한계 및 향후 연구

- (실험 제한) 소규모 클러스터 및 시뮬레이션 양자 워크로드
- 실제 QPU, 다양한 워크로드, 경쟁 조건 테스트 필요
- (전망) 국가 슈퍼컴퓨터 센터에서 페타/엑사스케일 HPC + QPU 결합 시, 동적 스케줄링이 시스템 용량 결정에 핵심
- HPC-QC 통합에서 “성숙도 격차(maturity gap)” 존재, 오케스트레이션이 지속적 과제

○ 결론

- 이번 연구는 하이브리드 HPC-QC 시스템에서 동적 스케줄링 및 malleability 적용을 통해 자원 활용률 향상, 작업 완료 시간 단축, 동시에 실행되는 여러 작업 처리 효율 개선
- 향후 실제 양자 하드웨어와 다양한 워크로드에서 확장 가능성 있음
- 효율적 하이브리드 작업 관리가 실용적 성능 향상에 핵심적 역할을 함

(원문)

1. <https://thequantuminsider.com/2025/08/16/how-can-supercomputers-and-quantum-computers-work-together-scientists-report-dynamic-scheduling-boosts-quantum-hpc-cluster-efficiency/>