

Programma Operativo Nazionale 2014-2020
Dottorati di ricerca su tematiche dell'innovazione e green
D.M. n. 1061 del 10 agosto 2021

Università degli Studi di Napoli Federico II
Dottorato di Ricerca in Information Technology and Electrical Engineering
XXXVII CICLO
TEMATICHE GREEN (AZIONE IV.5)

BORSA N. 1 - Dottorando VINCENZO MAISTO

Titolo del progetto:

“Architetture di calcolo innovative per Green Computing”

SETTORE SNSI: Industria intelligente e sostenibile, energia e ambiente

SETTORE PNR: Digitale, Industria, Aerospazio

a. Motivazioni

La sempre crescente mole di dati ed informazioni, ormai fondamento della odierna economia, offre opportunità di conoscenza e di sviluppo economico in numerosi campi, ma al contempo pone una serie di sfide tecnologiche decisive legate alla gestione e la processazioni dei cosiddetti *big data*. Come mostrato da numerosi scenari applicativi –dalle simulazioni per calcolo scientifico alla genomica, fino alle criptovalute– applicazioni di calcolo su larga scala possono avere una impronta ecologica molto pesante, in ragione del **costo energetico** che tali applicazioni impongono. La presente proposta prevede lo studio e la sintesi di soluzioni originali per il calcolo ad alte prestazioni (HPC) che realizzino a pieno il paradigma noto come **Green Computing**, sulla base di architetture a basso o bassissimo consumo energetico (*low-power* e *ultra-low-power* HPC).

In tal senso, un fattore abilitante fondamentale è legato alla disponibilità di metodologie per l'*hardware/software co-design*, che consentano di co-ottimizzare sistema ed applicazione, raggiungendo il più fine grado di ottimizzazione e, quindi, massimizzando l'utilizzo delle risorse di calcolo effettivamente disponibili. Evitando l'aggravio dovuto all'integrazione di sistemi progettati separatamente, l'impiego esteso di metodologie di hardware/software co-design ha il principale beneficio di aumentare l'efficienza energetica del sistema. Inoltre, inglobare requisiti di efficienza energetica già in fase di analisi e progetto, crea forti potenzialità di miglioramento rispetto a tecnologie tradizionali. In tal senso, microprocessori e System-on-Chip di nuova concezione, GPU ed acceleratori dedicati, quali dispositivi ad hardware programmabile (FPGA), nuovi paradigmi come il cloud computing e il rivoluzionario mondo del Quantum Computing, offrono dimensioni di ottimizzazione e gradi di libertà molto più ampi rispetto a sistemi centralizzati o alle architetture classiche. Tuttavia, soltanto se sostenuta da opportune metodologie di progetto, la possibilità di sfruttare questi scenari tecnologici permetterà di rispondere alla impellente necessità dell'efficienza e sostenibilità energetica, ben espressa dai desiderata del Green Computing.

b. Obiettivi

L'attività di ricerca proposta è orientata alle grandi sfide del Green Computing, ambito di forte interesse all'interno del *Green Deal* europeo, e fondamentale per la futura sostenibilità della

nostra società, ormai sempre più dipendente dai dati e dalle relative capacità di elaborazione. In tale contesto, l'**obiettivo generale** della proposta è il contenimento del consumo energetico delle grandi infrastrutture di calcolo e delle applicazioni basate su calcolo ad alte prestazioni. L'utilizzo di tecnologie, architetture e metodologie innovative è fondamentale ed indispensabile per raggiungere tale scopo. Il progetto proposto punta quindi ad estendere le conoscenze e le tecniche ingegneristiche per la realizzazione di infrastrutture di calcolo, ponendo come requisito chiave, assieme a quello delle prestazioni, quello dei bassi o bassissimi consumi energetici. Nello scenario ultimo cui guarda la proposta, la massimizzazione dell'utilizzo delle risorse di calcolo impiegate non è più visto come un mero parametro economico, ma come la chiave fondamentale per la sostenibilità ambientale delle applicazioni.

L'approccio fondamentale che permette di operare in tale direzione è l'hardware/software co-design. Tecniche convenzionali di integrazione, infatti, non possono affrontare due requisiti così contrastanti come la sostenibilità ambientale del calcolo effettuato e l'incremento spinto delle sue prestazioni. Il progetto punta a superare tale limitazione e ribaltare il paradigma e l'accezione energivora dell'HPC. In tal senso, la definizione di metodologie di hardware/software co-design sarà orientata verso diversi **obiettivi tecnologici** e declinata su opportunità che vengono sia dal fronte delle tecnologie di calcolo convenzionali, nelle loro recenti evoluzioni, sia dall'emergente paradigma del Quantum Computing. Tra le diverse, tali opportunità includono:

- Architetture di calcolo eterogenee (direttamente richiamate dal PNR nell'ambito 5.4.1 *Transizione digitale-i4.0*, Articolazione 4: *Dispositivi e sistemi eterogenei*), con riferimento a tecnologie hardware riconfigurabili di nuova concezione, strettamente integrate con memorie ad alta banda, come HBMv2, tramite tecniche di 3D stacking. Assieme ad esse, si annoverano dispositivi di storage di nuova generazione estesi con supporto FPGA per l'accelerazione dedicata, che permettono l'esplorazione di innovativi approcci di *processing in-memory* (PIM). Tali tecniche, efficacemente sfruttate tramite opportune metodologie di hardware/software co-design, possono giocare un ruolo fondamentale nel miglioramento dei livelli di efficienza energetica.
- Supporto alla cosiddetta disaggregazione (*disaggregation*) delle risorse, un paradigma fondamentale per sostenere la scalabilità di infrastrutture cloud e datacenter, principalmente applicato alle risorse di storage, da estendere alla gestione di acceleratori eterogenei per il supporto all'efficienza energetica in applicazione di larga scala (applicazioni datacenter/HPC) in ambito Green Computing. Le attività saranno in linea con i risultati conseguiti dal gruppo di ricerca proponente, anche nell'ambito di specifici progetti europei H2020 basati su tecnologie emergenti nel comparto HPC.
- Metodologie innovative per lo sviluppo, re-ingegnerizzazione ed ottimizzazione di algoritmi classici in paradigma quantistico, che permettano la semplificazione della progettazione di algoritmi tramite tecnologie e metodologie di sviluppo semi-automatiche verso un modello unificato e semplificato di un calcolatore quantistico. In parallelo, si valuteranno approcci per ampliare e semplificare l'utilizzo di calcolatori basati su tecnologie quantistiche, integrandoli in architetture di calcolo eterogenee che integrino microprocessori con componenti ottimizzati, basati su tecnologie di hardware riprogrammabile, quali FPGA. Un secondo obiettivo in questo contesto sarà anche dimostrare e utilizzare il *quantum advantage* nel regime NISQ (Noisy, Intermediate-Scale Quantum) con le macchine quantistiche oggi disponibili.

Il raggiungimento dei precedenti obiettivi sarà misurato nella capacità di contribuire allo sfruttamento energeticamente sostenibile di risorse di calcolo per applicazioni di larga scala. Nei suoi diversi risvolti, la proposta è direttamente allineata con molteplici ambiti ed articolazioni del PNR, tra cui in particolare:

- Ambito 5.4.1 Transizione digitale – i4.0
 - Articolazione 4. Dispositivi e sistemi eterogenei
- Ambito 5.4.2 High performance computing e big data
 - Articolazione 1. Hardware e software a supporto della realizzazione ed evoluzione dei grandi hub HPC e big data europei e nazionali per il calcolo scientifico, la ricerca e la scienza aperta
 - Articolazione 2. Ricerca di base e fondamentale in ingegneria, scienze e tecnologie informatiche per HPC e big data
 - Articolazione 4. Architettura, ingegneria, scienze e tecnologie informatiche per la evoluzione dei dati della PA verso sistemi aperti, big data e servizi cloud
 - Articolazione 5. Applicazioni HPC, big data e sistemi di servizi cloud per la società, per la sua resilienza, per lo sviluppo sostenibile, per gli spazi dati comuni locali, nazionali ed europee
- Ambito 5.4.3 Intelligenza artificiale
 - Articolazione 5. Intelligenza artificiale per l'ambiente e le infrastrutture critiche
 - Articolazione 6. Intelligenza artificiale per la produzione industriale
- Ambito 5.4.5 Tecnologie quantistiche
 - Articolazione 1. Tecnologie quantistiche per computer e simulatori
 - Articolazione 2. Tecnologie quantistiche per la comunicazione
 - Articolazione 4. Tecnologie quantistiche per l'efficienza e la sostenibilità energetica
 - Articolazione 5. Infrastrutture di ricerca per le tecnologie quantistiche

Il contenimento del consumo di risorse necessarie per l'elaborazione dei dati ed il valore estratto, in tempi vantaggiosi, dai dati stessi, sosterrà inoltre l'industria nel compimento di alcuni obiettivi tematici del *Green Deal*, quali il raggiungimento di un'economia industriale pulita e circolare (Obiettivi Tematici del Green Deal europeo OT3: *Mobilitare l'industria per un'economia pulita e circolare*).

c. Cronoprogramma

Formazione (Mese 1 – 14):

- Studio ed analisi delle tecniche di progettazione e ottimizzazione hardware/software per l'HPC.
- Studio ed analisi delle tecniche per il trattamento di dati su larga scala (big data). A tal fine, si considereranno i corsi di Big Data Analytics and Business Intelligence, Hardware and Software Architectures for Big Data offerti dall'Università degli Studi di Napoli Federico II.
- Studio ed analisi delle tecniche di progettazione e tecnologie di hardware a basso e bassissimo consumo energetico (low-power, ultra-low-power).
- Studio ed analisi delle tecnologie di integrazione di sistemi eterogenei e di tecnologie eterogenee.

- Analisi della teoria della Quantum Information e degli algoritmi quantistici esistenti. Studio del nuovo campo di progettazione del Quantum Engineering e delle tecniche di ingegnerizzazione di hardware e software quantistico. Allo scopo, si considereranno corsi che spaziano dalla progettazione di circuiti, con *Introduction to Quantum Circuits*, alla teoria dell'informazione con *Quantum Information*, offerti dall'Università degli Studi di Napoli Federico II.

Sviluppo di metodologie (Mese 12 – 28):

- Sviluppo di metodologie di integrazione efficiente in architetture di calcolo eterogenee di tecnologie quali:
 - hardware riconfigurabile (field programmable gate array, o FPGA)
 - memorie ad alta banda (HBMv2)
 - 3D stacking
 - dispositivi di storage estesi con supporto per l'accelerazione dedicata (*processing in-memory*, o PMI)
- Sviluppo di una metodologia a supporto della *resource disaggregation* per sistemi di supercalcolo, fondamentale per:
 - Scalabilità
 - Elasticità, dinamismo e resilienza (anche di sistemi cloud)
 - Integrazione efficiente di sistemi e tecnologie eterogenei
 - Massimizzazione dell'efficienza energetica delle risorse di memoria
- Sviluppo di metodologie avanzate di utilizzo del paradigma quantistico nei livelli di progettazione di:
 - Algoritmi quantistici
 - Sistemi di integrazione con architetture classiche
 - Ottimizzazione di algoritmi classici in paradigma quantistico
- Sviluppo di un modello unificato e semplificato di un calcolatore quantistico, al fine di massimizzarne:
 - L'usabilità dal punto di vista applicativo
 - L'integrabilità con sistemi convenzionali
 - L'efficienza di utilizzo di risorse energetiche
- Dimostrazione della possibilità di utilizzare il *quantum advantage* nel regime NISQ (Noisy, Intermediate-Scale Quantum) con le macchine quantistiche oggi disponibili.

Sperimentazione (Mese 22 –36):

Il programma di ricerca prevede la collaborazione con un'azienda operante nel segmento HPC/datacenter per un periodo di **sei mesi**, divisi in più fasi, prevalentemente collocate nel periodo di sperimentazione del progetto. L'azienda fornirà dei casi applicativi di interesse industriale e l'accesso a piattaforme di calcolo avanzate, offrendo quindi un riscontro diretto dell'impatto e dell'interesse suscitato nel mondo industriale da parte delle tecnologie sperimentali e delle architetture sviluppate nel progetto. Il segmento delle applicazioni su big data in ambienti HPC/datacenter rappresenta bene le sfide tese alla sostenibilità energetica. La collaborazione consentirà quindi reciproci spunti per ricalibrare in maniera fine gli obiettivi di ricerca, sul versante accademico, e migliorare i processi di sviluppo industriali, sul fronte dell'azienda.

Parallelamente alla collaborazione diretta con l'azienda, la parte sperimentale dell'attività di ricerca avrà a disposizione diverse piattaforme moderne ad alte prestazioni come FPGA, dotate di infrastrutture avanzate di memoria basate su HBMv2 (ad esempio, schede Xilinx Alveo per HPC), consentendo la continua sintesi, integrazione e collaudo dei sistemi progettati. Assieme alle architetture classiche, l'attività sperimentale intende servirsi anche di piattaforme per il calcolo quantistico. Oltre a simulatori come quello offerto nel Microsoft Quantum Development Kit, è di primario interesse per la proposta accedere all'hardware quantistico ad oggi reso pubblicamente disponibile da società come IBM (Quantum Experience) e Microsoft (Azure Quantum platform). Si considererà inoltre la possibilità di stabilire rapporti diretti con gruppi di ricerca che non offrono pubblicamente l'accesso alle proprie piattaforme quantistiche, negoziando la possibilità di utilizzare l'hardware a fini sperimentali.

In tale contesto sarà anche pianificato un periodo di permanenza all'estero di almeno **tre mesi** presso un centro di ricerca individuato sulla base delle relazioni sviluppate in ambito europeo dal gruppo di ricerca proponente.

d. Risultati attesi

Il dottorato si articola in tre anni. Durante il primo anno, si punterà a formare una robusta conoscenza nel campo dell'HPC, big data, low-power (green) computing e quantum computing. L'insieme di queste conoscenze concorrerà a formare un profilo che spazi in tutti i campi dell'ottimizzazione e della progettazione dei sistemi di calcolo eterogenei ad alte prestazioni, in linea con le potenzialità delle tecnologie classiche come di quelle emergenti. Insieme a questo studio, una prima attività di sperimentazione sarà iniziata per mettere a frutto le conoscenze acquisite ed indirizzare le attività sperimentali degli anni successivi.

Il secondo anno mira ad uno studio più puntuale e verticale delle tecnologie e metodologie necessarie per l'attività di ricerca, assieme al raffinamento dell'attività sperimentale in programma. Questo raffinamento porterà a contribuire al mondo della ricerca con risultati e lavori scientifici, che abbiano un impatto significativo sulla comunità scientifica ed industriale.

Il terzo ed ultimo anno del corso di Dottorato sarà focalizzato sulla sperimentazione e sintesi delle nuove metodologie e tecnologie, basate sullo studio degli anni precedenti. Il risultato finale sarà rappresentato da sperimentazioni sistematiche e dal rilascio di strumenti software che consentano a terzi la riproducibilità delle esperienze condotte così come l'integrazione delle metodologie sviluppate in sistemi reali.

e. Impresa

E' previsto un periodo di mesi 6 presso la A3Cube, un'azienda molto attiva in ambito datacenter/HPC, nata negli US e trasferitasi in Italia, sita in Via Dogana Vecchia Centrale Idroelettrica Orlandi, 28066 Galliate (Novara).

f. Istituzione ospitante all'estero

E' previsto un periodo di mesi 3 presso il Politecnico di Valencia (Universitat Politècnica de València, UPV), Camí de Vera, s/n, 46022 València, Valencia, Spagna.

g. Prodotti misurabili della ricerca, comunicazione e disseminazione

L'attività di ricerca prevedrà la condivisione degli avanzamenti teorici, metodologici e tecnologici con il resto della comunità scientifica tramite pubblicazioni in riviste e presentazioni in congressi. Per il mondo dell'HPC, saranno considerate varie riviste e congressi di settore, tra cui (solo a titolo di esempio) IEEE HPC and Simulation, HPC and Architectures, IEEE Transactions on Computers, IEEE Transactions on Parallel and Distributed Systems, che trattano anche temi di interesse per questa proposta, quali architetture eterogenee, low-power computing, ed aspetti innovativi come il quantum computing. Inoltre, molte pubblicazioni di ingegneria trovano spazio nell'ambito del quantum computing e del quantum engineering in varie riviste di fisica quantistica, tra cui ad esempio l'archivio di *Quantum Physics* gestito dalla Cornell University oppure *Fortschritte der Physik*, *Physcal Review* o *IOPscience*. Si prevede di ottenere nel periodo coperto alla borsa di dottorato almeno tre presentazioni a congresso ed una rivista scientifica.

Oltre alle pubblicazioni scientifiche, il programma ambisce a stabilire rapporti con centri di ricerca internazionali e sviluppare collaborazioni in ambito europeo, anche sulla base delle precedenti esperienze del gruppo di ricerca proponente nell'ambito dei bandi H2020 FET-HPC. Particolare attenzione sarà posta ai bandi di ricerca ed innovazione come il bando europeo EuroHPC Joint Undertaking (JU), per la realizzazione in Europa di centri di supercalcolo, avente tra i propri obiettivi primari aspetti come le alte prestazioni ed i bassi consumi, di diretto interesse per la presente proposta. In questi bandi, in particolare, è sempre più forte e presente l'interesse per la sostenibilità ambientale. Si prevede di registrare nel periodo coperto alla borsa di dottorato la partecipazione ad almeno una proposta di finanziamento in un bando Europeo, su temi direttamente coperti dalla presente attività.

In aggiunta alla disseminazione scientifica, la proposta prevede anche di condurre attività di comunicazione di carattere divulgativo. Uno dei canali per tali attività sarà rappresentato dal Laboratorio nazionale CINI HPC Key Technologies and Tools, presso il quale il gruppo di ricerca proponente è particolarmente attivo. Tra le diverse azioni, si prevede la partecipazione alle attività del laboratorio incentrate sui temi del Quantum Computing e la regolare partecipazione al workshop annuale organizzato dal laboratorio. Inoltre, si cercherà di instaurare scambi con le comunità scientifiche interessate a temi di Big Data e di Intelligenza Artificiale, che potrebbero essere direttamente interessate alle metodologie oggetto della presente proposta.