Project Description:

Background

Intel and ICHEC Partnership:

ICHEC is the national high-performance computing (HPC) centre and works on parallelisation and performance optimisation of different types of application codes on multiple hardware platforms (multi-core CPUs, accelerators such as GPUs, FPGAs and many-core CPUs, extreme-scale supercomputing). ICHEC also applies a number of high-end computing techniques (such as data analytics, machine learning, deep learning and blockchain) in its research and technology transfer activities when engaging with public and private sector organisations. Over the years, ICHEC and Intel have a long-standing collaboration. Particularly in Ireland, ICHEC regularly engages with Brian Quinn (Director of European Innovation at Intel Labs Europe), and also other Intel teams at the Exascale Labs in Paris (France) and the HPC code modernisation group at Portland (USA). ICHEC was globally the sixth HPC centre to be designated by Intel as an "Intel Parallel Computing Centre (IPCC)" and has since continued working on multiple Intel platforms to optimise multiple applications (such as Wave Watch 3, real-time migration for Oil and Gas, etc.).

Quantum Initiatives:

Recently, Ireland decided to join the EuroHPC initiative, which in future is believed to include quantum-computing platforms as accelerators in extreme-scale supercomputers. Being the national HPC centre, ICHEC identified a huge potential in developing programming expertise on quantum platforms and has been defining its strategy to address this. Intel has a long history in developing both a physical quantum platform and a quantum simulator (qHiPSTER) that can be deployed on classic HPC systems. The quantum simulator is aimed at providing a test-bed for programmers to create a software ecosystem and develop proof-of-concepts in parallel with the development of quantum computers.

Synergy between Intel and ICHEC:

During a recent discussion with ICHEC, Brian Quinn (Intel, Ireland) put forward the idea of Intel and ICHEC collaborating to pioneer quantum computing programming efforts in Ireland. For this, Intel proposed to leverage the quantum simulator (qHiPSTER) and that ICHEC implement a proof-of-concept application on this simulator. Following this, ICHEC investigated and prepared a short proposal to implement a quantum version of an existing natural language processing (NLP) algorithm on qHiPSTER. Intel Ireland and the Intel Quantum Computing team in the USA internally reviewed this proposal and gave it full support. Following this, Intel deemed the proposed project to be of high-value to Intel in Ireland and worldwide, and of interest to the quantum computing community at large. Thus, Intel are keen to extend their existing collaboration with ICHEC by leveraging state funding to develop the proposed solution and pioneer the quantum programming efforts in Ireland with global significance.

Problem / Opportunity

Natural Language Processing:

Natural language processing (NLP) is often used to perform tasks like machine translation, sentiment analysis, relationship extraction, word sense disambiguation and automatic summary generation. Most traditional NLP algorithms for these problems are defined to operate over strings of words, and are commonly referred to as the "bag of words" approach. The challenge, and thus limitation, of this approach is that the algorithms analyse sentences in a corpus based on meanings of the component words and lack information from the grammatical rules and nuances of the language. Consequently, the qualities of results of these traditional algorithms are often unsatisfactory when the complexity of the problem increases. On the other hand, an alternate approach called "compositional semantics" incorporates the grammatical structure of sentences in a language into the analysis algorithms. Compositional semantics algorithms include the information flows between words in a sentence to determine the meaning of the whole sentence. One such model is "distributional compositional semantics" (DisCo), which is based on tensor product composition to give a grammatically informed algorithm that computes the meaning of sentences and phrases. This algorithm has been noted to offer significant improvements to the quality of results, particularly for more complex sentences. However, the main challenge in its implementation is the need for large classical computational resources.

Quantum Computing:
Quantum computers have the ability to solve complex problems that are beyond the capabilities of classical computers and will enforce the next genuine disruption to technical computing. The impacts of Quantum Computing will be significantly greater than those brought by many-core architectures and accelerators such as GPUs. While many enterprises (including Intel, IBM, Google, etc.) have been developing physical quantum computing devices, another line of developing quantum computing platforms is through the creation of simulators that are deployed on classical HPC (High Performance Computing) systems. It is widely acknowledged that quantum devices and simulators of size ~50 qubits allow for implementation of proof-of-concept algorithms and have computation power that exceeds many of currently available Peta-scale supercomputers. With the availability of such quantum computing platforms, it is essential that we develop the software ecosystem and programming expertise to target the quantum platforms.

Opportunity:
In this context, both Intel and ICHEC have identified the opportunity to leverage the Intel Quantum Simulator (that can simulate up to ~42 qubits) to port and implement a quantum version of an existing compositional semantics NLP algorithm to analyse the meaning sentences in a corpus. This project aims to

1. Leverage and evaluate the computing power that quantum devices can offer to computation intensive NLP algorithms such as in the distributional compositional semantics model.
2. Develop the ecosystem of proof-of-concept applications ported to the emerging quantum computing domain, particularly using a highly relevant application domain such as NLP.
3. Pioneer a collaborative innovation environment in Ireland between industry and research organisations to develop expertise to program quantum computers.

Solution / Product
The distributional compositional semantics (DisCo) model explained above was originally developed by its authors with direct inspiration from quantum theory. Following this, the authors of DisCo have since co-developed a quantum version of the algorithm that can be implemented on quantum computers [1], which is based on the DisCo model that was introduced in [2] and [3].
The quantum version of the DisCo model presents two algorithms:

1. A quantum algorithm for the "closest vector problem"
   An algorithm for the "closest vector problem" is used to determine the word/phrase out of a set of words/phrases that has the closest relation (for instance, meaning) to a given word/phrase. This finds application in many computational linguistic tasks such as text classification, word/phrase similarity, test classification and sentiment analysis. A quantum version of this algorithm in [1] enables its implementation on a quantum computing platform.
2. A quantum algorithm for the "CSC sentence similarity" model
   This algorithm is an adaptation of the "closest vector problem" quantum algorithm to perform sentence similarity calculations in the distributional compositional framework. This algorithm is based on tensor product composition that gives a grammatically informed algorithm to compute meaning of sentences/phrases and stores the meanings in quantum systems. Based on this, a quantum implementation of this algorithm has lower storage and compute requirements compared to a classic HPC implementation [1].

In this project, we will implement the two quantum algorithms ("closest vector problem" and "CSC sentence similarity") of the DisCo model on the Intel Quantum Simulator (qHiPSTER). Given a corpus, the implemented solution will be able to compute the meanings of two sentences (built from words in the corpus) and decide if their meanings match.

In this project, we estimate to target corpuses with ~2000 most common words using the Intel Quantum Simulator (qHiPSTER) that can simulate ~42 qubits. qHiPSTER will be installed on the Irish national supercomputer (Kay) that will become operational in September 2018 and is operated by ICHEC.

[1]. William Zeng and Bob Coecke, "Quantum Algorithms for Compositional Natural Language Processing", Proceedings of SLPCS, 2016.
[2]. Stephen Clark, Bob Coecke and Mehrnoosh Sadrzadeh, "A Compositional Distributional Model of Meaning", Proceedings of 2nd Quantum Interaction Symposium, 2008.
[3]. Bob Coecke, Mehrnoosh Sadrzadeh and Stephen Clark, "Mathematical Foundations of a Compositional Distributional Model of Meaning", Special issue of Linguistic Analysis, 2010.

Novelty

A survey of problems that have been ported/implemented on quantum computers points to fundamental mathematical and scientific algorithms [4]. Even though compositional NLP algorithms have been researched and theoretically proposed, their implementations have been limited on classical HPC platforms for want of computational resources. Furthermore, in our literature survey, we do not find references to previous or on-going work to implement compositional NLP algorithms on gate model quantum computers, which are the universal type quantum computers. On the other hand, there has been limited work on formulating the problem to classify text documents (fragments) as a binary optimisation problem for implementation on the annealing-based D-Wave systems [5]. Annealing-based systems target a narrow set of optimisation and sampling problems in areas such as financial services, healthcare and manufacturing [6].

Thus, Intel and ICHEC observe that this project is

1. a novel effort to implement the more powerful distributional compositional semantics algorithm for NLP on the standard gate model platform, and
2. will produce early and original results in the NLP domain on the Intel Quantum Computing platform that (from Intel and ICHEC investigations) will be the first of its kind in Ireland, and potentially worldwide as well.

Quantum computing represents a significant research endeavour within Intel's R&D corporate portfolio. The Intel Quantum Simulator (qHiPSTER) is key part of that endeavour, providing feed-forward insights on the impact various quantum algorithms are likely to have on underlying quantum hardware – thereby helping to optimise Intel's research and development efforts. The quantum simulator is a single node or distributed high-performance implementation of a quantum simulator that can simulate general single-qubit gates and two-qubit controlled gates. The Intel Quantum Simulator, which has been used to simulate algorithms of more than 40 qubits, is targeted at algorithm developers who wish to test their quantum software in a simulation. On a laptop, one can simulate about 30 qubits. But quantum computing is exponential in its behaviour, so one needs a supercomputer for simulating 40 qubits and verifying algorithms. The novelty of this research comes from ICHEC developing and evaluating the performance of compositional NLP algorithms leveraging the full qubit potential of the Intel quantum simulator and providing feedback to Intel to guide its quantum hardware and quantum software research.

[4]. Quantum Algorithm Zoo, https://math.nist.gov/quantum/zoo/
[5]. Ramin Ayanzadeh, "Quantum Artificial Intelligence for Natural Language Processing Applications", Proceedings of SIGCSE, 2018.
[6]. Accenture Labs, "Innovation with Quantum Computing", 2017.

Project Objectives

1. Implement the quantum version of a natural language processing (NLP) algorithm (called distributed compositional semantics model) on the Intel Quantum Simulator (qHiPSTER).
2. Develop, demonstrate and evaluate the solution on the new Irish national supercomputer "Kay".
3. Establish a pioneering Irish initiative for programming quantum computers and simulators through a collaboration between ICHEC and Intel.

Potential Risks

Risks	Steps for management
Installation and performance issues of the Intel Quantum Simulator (qHiPSTER) on the ICHEC Kay supercomputer.	The first task in the project work plan is to install and test qHiPSTER on the ICHEC supercomputer. Reservations can be made for this project on the Kay supercomputer to achieve the desired performance. The simulator developers from Intel will also be available to assist with solving installation and configuration issues.
Issues in mapping and/or implementing the DisCo model algorithms on the programming model of qHiPSTER.	The developers of qHiPSTER will be available to provide user support during implementation of the algorithms. Also, the authors of the DisCo model at the University of Oxford may be consulted for a deeper understanding of their algorithms.
Performance of the algorithm implementations is lower than expected and thus evaluation is delayed.	The size of the corpuses targeted for the evaluations may be reduced accordingly to achieve reasonable performance to get evaluation results.

Execution Plan

The Irish Centre for High-End Computing - ICHEC's knowledge and expertise in the field of High Performance Computing would be consolidated into a unique solution that aims to address the raised technical requirements and challenges, allowing the Company to incorporate an innovative solution within its operations, which would ultimately enhance its productivity, and competitive advantage within the international ICT sector – mainly the emerging Quantum Computing domain. With regard to technology transfer, the execution plan calls for a significant degree of integrated and collaborative activity between ICHEC and the company in order to ensure the NLP models and algorithms are compatible with Intel's roadmap. Therefore, the Company shall be involved thought-out the PLC, facilitating sector knowledge and the required business logic through the input of their assigned project representative(s).

Potential Company Benefits

- Intel considers quantum computing as the ultimate in parallel computing, with the potential to tackle problems conventional computers can't handle. For example, quantum computers may simulate nature to advance research in chemistry, materials science and molecular modelling.
- The benefit of this research comes from ICHEC developing compositional NLP algorithms, testing them using the full qubit potential of the Intel quantum simulator - running on ICHEC's high-performance infrastructure, and providing feedback to Intel to guide its quantum hardware and quantum software research. Intel believes that currently there is significant learning to be achieved by going through this simulator approach with ICHEC.
- Additionally, this research partnership will give opportunity for Intel Ireland to grow its corporate mandate for researching emerging technology areas, and positions Ireland favourably to pitch for future related research projects.
- For further information about Intel's quantum research efforts – see Intel's press centre https://newsroom.intel.com/press-kits/quantum-computing.

Detailed Description of Work Plan:

TASK IDENTIFIER WORK PACKAGE 1:

TASK IDENTIFIER 1:

- **TASK TITLE:** RESOURCE SETUP AND PROBLEM MAPPING
- **TASK OWNER:** ICHEC 100%
- **TASK TEAM MEMBERS / % OF TIME ON TASK:** 1 RESEARCH FELLOW (20%), 2 POST-DOCS (EACH 100%)
- **TASK OBJECTIVE:** INSTALL AND TEST THE INTEL QUANTUM SIMULATOR (QHIPSTER) ON THE ICHEC SUPERCOMPUTER (KAY), AND DEFINE THE STRATEGY TO MAP THE QUANTUM VERSION OF THE DISCO MODEL ALGORITHM ON THE INTEL QUANTUM SIMULATOR.
- **DESCRIBE IN DETAIL HOW THE TASK WILL BE COMPLETED:**
  This work package will be implemented through the following four tasks:
  - T1.1. Install and test the Intel Quantum Simulator (qHiPSTER) on the ICHEC supercomputer (Kay).
  - T1.2. Investigate the quantum version of the distributed compositional semantics (DisCo) model algorithms, and define their mapping strategy on qHiPSTER.
  - T1.3. Specify the methodology to test and evaluate the implementation.
  - T1.4. Prepare a representative corpus (10x – 100x words) for preliminary testing and evaluation.
- **DURATION:** 12 WEEKS
- **DELIVERABLES:**
  - D1.1. A report describing the mapping of quantum DisCo model algorithms on qHiPSTER, testing and evaluation methodologies, and the representative corpora. (M03)
- **MILESTONES:**
  - M1.1. qHiPSTER is installed and tested on the ICHEC Kay supercomputer. (M01)
  - M1.2. Mapping of quantum DisCo algorithms on qHiPSTER is complete. (M03)
  - M1.3. Testing and evaluation methodologies is defined. (M03)
  - M1.4. Representative corpora for preliminary testing is ready. (M03)

TASK IDENTIFIER 2:

- **TASK TITLE:** SOLUTION DEVELOPMENT
- **TASK OWNER:** ICHEC 100%
- **TASK TEAM MEMBERS / % OF TIME ON TASK:** 1 RESEARCH FELLOW (20%), 2 POST-DOCS (EACH 100%)
- **TASK OBJECTIVE:** IMPLEMENT THE DISCO MODULE ALGORITHMS ON QHIPSTER AND PERFORM PRELIMINARY EVALUATIONS.
- **DESCRIBE IN DETAIL HOW THE TASK WILL BE COMPLETED:**
  This work package will be implemented through the following four tasks:
  - T2.1. Implement and test the quantum algorithm for the "closest vector problem" in the DisCo model on qHiPSTER based on the mapping strategy defined in T1.2.
  - T2.2. Evaluate the "closest vector problem" implementation using the representative corpora based on the evaluation methodology specified in T1.3.
  - T2.3. Implement and test the quantum algorithm for the "CSC sentence similarity" in the DisCo model on qHiPSTER based on the mapping strategy defined in T1.2.
  - T2.4. Evaluate the "CSC sentence similarity" implementation using the representative corpora based on the evaluation methodology specified in T1.3.
- **DURATION:** 32 WEEKS
- **DELIVERABLES:**
  - D2.1. A report summarising the results of tasks T2.1 and T2.2 that implement and evaluate the "closest vector problem" algorithm on qHiPSTER using the representative corpora. (M07)
  - D2.2. A technical presentation at Intel event at SC 2019. (M11)
  - D2.3. A report summarising the results of tasks T2.3 and T2.4 that implement and evaluate the "CSC sentence similarity" algorithm on qHiPSTER using the representative corpora. (M11)
- **MILESTONES:**
  - M2.1. Implementation and preliminary evaluation of the "closest vector problem" algorithm on qHiPSTER is complete. (M07)
  - M2.2. Implementation and preliminary evaluation of the "CSC sentence similarity" algorithm on qHiPSTER is complete. (M11)

TASK IDENTIFIER 3:

- **TASK TITLE:** SOLUTION EVALUATION AND PACKAGING
- **TASK OWNER:** ICHEC 100%
- **TASK TEAM MEMBERS / % OF TIME ON TASK:** 1 RESEARCH FELLOW (20%), 2 POST-DOCS (EACH 100%)
- **TASK OBJECTIVE:** PERFORM EXTENDED EVALUATIONS AND PACKAGE THE SOFTWARE SOLUTION ALONG WITH RESULTS AND DOCUMENTATION.
- **DESCRIBE IN DETAIL HOW THE TASK WILL BE COMPLETED:**
  This work package will be implemented through the following two tasks:
  - T3.1. Prepare a larger corpus (1000x words) for final testing and evaluation. Evaluate the final implementation of the DisCo model algorithms using the larger corpus.
  - T3.2. Package the final version of the implementation with documentation and evaluation results.
- **DURATION:** 10 WEEKS
- **DELIVERABLES:**
  - D3.1. A report describing the larger corpora and summarising the final evaluations using the larger corpora. (M14)
  - D3.2. Final version of the DisCo algorithm implemented on qHiPSTER with documentation. (M14)
  - D3.3. A technical poster and presentation for ISC 2020. (M14)
- **MILESTONES:**
  - M3.1. Larger corpora for final testing and evaluation is ready. (M14)
  - M3.2. Final version of the DisCo algorithm implementation and evaluation on qHiPSTER is complete. (M14)

TASK IDENTIFIER 4:

- **TASK TITLE:** TECHNICAL REPORT AND HANDOVER
- **TASK OWNER:** ICHEC 100%
- **TASK TEAM MEMBERS / % OF TIME ON TASK:** 1 RESEARCH FELLOW (20%), 2 POST-DOCS (EACH 100%)
- **TASK OBJECTIVE:** DELIVERY OF THE SOFTWARE SOLUTIONS AND RESULTS TO INTEL.
- **DESCRIBE IN DETAIL HOW THE TASK WILL BE COMPLETED:**
  In this work package, ICHEC will deliver the following components to Intel:
  - The final documented version of the DisCo algorithm implemented on qHiPSTER.
  - All deliverables prepared in Tasks 1, 2 and 3.
- **DURATION:** 2 WEEKS
- **DELIVERABLE:**
  - D4.1. A consolidation of reports from deliverables D1.1, D2.1, D2.2, D2.3, D3.1, D3.3 and final software from deliverable D3.2. (M14)
- **MILESTONES:**
  - M4.1. Handover of consolidated reports and final software to Intel is complete. (M14)

Gantt Chart:

Task No.	Month Number													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1.1														
1.2														
1.3														
1.4														
2.1														
2.2														
2.3														
2.4														
3.1														
3.2														
4														

Schedule 1
Project Plan

Project Title	Quantum Natural Language Processing using Intel Quantum Simulator (qHiPSTER)				
NUIG Principal Investigator	Dr. Venkatesh Kannan				
Industry Party Project Manager	Mr. Brian Quinn				
TASK #	**TASK TITLE**	**MILESTONE / DELIVERABLE**	**DELIVERABLE DESCRIPTION**	**TASK LEADER**	**TASK PARTNERS**
1	Resource Setup and Problem Mapping	D1.1 (M03)	A report describing the abstraction of the quantum DisCo model algorithms for implementation on any quantum platform, and their mapping on qHiPSTER, along with testing and evaluation methodologies, and the representative corpora.	ICHEC	Intel (external)
		M1.1 (M01)	qHiPSTER is installed and tested on ICHEC's Kay.		
		M1.2 (M03)	Abstraction and mapping of DisCo algorithms on qHiPSTER is complete.		
		M1.3 (M03)	Testing and evaluation methodologies are defined.		
		M1.4 (M03)	Representative corpora for testing are ready.		
2	Solution Development	D2.1 (M07)	A report summarising the implementation, testing and evaluation of the "closest vector problem" algorithm using the representative corpora.	ICHEC	Intel (external)
		D2.2 (M11)	A technical presentation at an Intel event at SC 2019.		
		D2.3 (M11)	A report summarising the implementation, testing and evaluation of the "CSC sentence similarity" algorithm using the representative corpora.		
		M2.1 (M07)	Implementation of the abstract and qHiPSTER-specific code,		

<table>
<tr><td rowspan="2"></td><td rowspan="2"></td><td></td><td>and evaluation for the "closest vector problem" is complete.</td><td rowspan="2"></td><td rowspan="2"></td></tr>
<tr><td>M2.2 (M11)</td><td>Implementation of the abstract and qHiPSTER-specific code, and evaluation for the "CSC sentence similarity" algorithm is complete.</td></tr>
<tr><td rowspan="6">3</td><td rowspan="6">Solution Evaluation and Packaging</td><td>D3.1 (M14)</td><td>A report describing the larger corpora, and summarising the final evaluations using the larger corpora.</td><td rowspan="6">ICHEC</td><td rowspan="6">Intel (external)</td></tr>
<tr><td>D3.2 (M14)</td><td>Final version of the abstract interface code for the DisCo algorithms, with documentation.</td></tr>
<tr><td>D3.3 (M14)</td><td>Final version of the qHiPSTER-specific code for the DisCo algorithms, with documentation.</td></tr>
<tr><td>D3.4 (M14)</td><td>A technical poster and presentation for ISC 2020.</td></tr>
<tr><td>M3.1 (M14)</td><td>Larger corpora for final testing and evaluation is ready.</td></tr>
<tr><td>M3.2 (M14)</td><td>Final version of the DisCo algorithms implementations and evaluation is complete.</td></tr>
<tr><td rowspan="3">4</td><td rowspan="3">Technical Report and Handover</td><td>D4.1 (M14)</td><td>A consolidation of reports from deliverables D1.1, D2.1, D2.2, D2.3, D3.1, D3.3 and final software from deliverables D3.2 and D3.3.</td><td rowspan="3">ICHEC</td><td rowspan="3">Intel (external)</td></tr>
<tr><td>M4.1 (M14)</td><td>Consolidated reports and final software are prepared.</td></tr>
<tr><td>M4.2 (M14)</td><td>Deliverable D3.3 is published on a GitHub page under the Apache License version 2.0.</td></tr>
</table>

Commencement Date	14 January, 2019
Completion Date	13 March, 2020
Reporting requirements	• Project kick-off meeting in January 2019 at ICHEC, Dublin. • One meeting every 3 weeks between ICHEC and Intel (video conference and at Intel Ireland Campus, Leixlip as needed) • Face-to-face meeting between ICHEC and Intel Quantum Computing team at SC 2019

	• Project closing meeting in March 2020 at Intel Ireland Campus, Leixlip.