

## Projet : Communication et démonstration d'un travail de recherche fondé sur CUDA : étude de cas pratique

### I. Objectifs du projet

#### 1. Objectif général

L'objectif de ce projet est de communiquer efficacement un travail dans un cadre de recherche expérimentale, en s'appuyant sur les principes fondamentaux de la programmation CUDA appliquée à un cas concret.

#### 2. Objectifs spécifiques

- Analyser une stratégie de parallélisation sur GPU existante appliquée à un réseau de neurones peu profond (*shallow neural network*) et l'utiliser comme référence expérimentale (*baseline*).
- Proposer une stratégie de parallélisation alternative sur GPU pour ce réseau de neurones peu profond.
- Comparer la stratégie de référence et la stratégie alternative, en spécifiant les métriques de performance (temps d'exécution et accélération) et le protocole d'évaluation.
- Présenter l'ensemble du travail sous la forme d'un article scientifique structuré et rédigé selon les normes de la discipline.

### II. Description des étapes du projet

#### 1. Prise en main du code de la stratégie de parallélisation de référence

Afin de maîtriser la stratégie de référence comme un point de comparaison fiable, les étudiants devront :

- Étudier le code de la stratégie de référence fourni implémentant la propagation avant (*forward pass*) et la rétropropagation (*backward pass*) d'un réseau de neurones peu profond.
- Comprendre le mapping du calcul sur les threads, la granularité de parallélisme (par neurone, par échantillon, par couche, etc.), l'organisation des grilles et blocs, et l'utilisation des mémoires du GPU (globale, partagée, registres, etc.).

- Analyser les limitations potentielles de la stratégie de référence au regard des hypothèses de conception retenues (granularité, accès mémoire, synchronisations, etc.).
- Reproduire les résultats de performance du code de référence.

## 2. Proposition d'une stratégie de parallélisation alternative

Afin de concevoir une stratégie de parallélisation alternative pour la propagation avant et la rétropropagation d'un réseau de neurones peu profond, les étudiants devront :

- Proposer une stratégie de parallélisation différente de la stratégie de référence.
- Justifier leurs choix sur le plan algorithmique et sur le plan architectural (GPU).
- Implémenter cette stratégie proposée avec CUDA en lui appliquant toute optimisation pertinente (choix de granularité, organisation des accès mémoire, utilisations des mémoires partagées, réduction des synchronisations, etc.).
- Vérifier la correction fonctionnelle des résultats de cette stratégie alternative et récupérer les mesures de performance.

## 3. Évaluation expérimentale des performances de la stratégie alternative

Afin de valider expérimentalement la stratégie alternative, les étudiants devront :

- Comparer la stratégie de référence et la stratégie alternative selon : le temps d'exécution, l'accélération obtenue (ratio par rapport à la stratégie de référence), la scalabilité selon la taille du réseau.
- Étudier l'impact : de la taille du réseau, de la taille de l'échantillon, du nombre de neurones.
- Présenter les résultats sous forme de : tableaux, graphiques clairs et commentés.

## 4. Rédaction d'un article scientifique résumant l'étude sur les deux stratégies

Afin de développer les compétences de communication autour d'un travail de calcul haute performance, les étudiants devront rédiger un article scientifique comprenant :

- **Titre** : un intitulé informatif et représentatif de l'étude.
- **Résumé** : synthèse structurée des objectifs, des méthodes, des résultats principaux et des perspectives.
- **Introduction** : justification du problème, revue rapide du contexte et formulation des objectifs.

- **Description du réseau** : architecture du réseau de neurones peu profond utilisé, caractéristiques et hypothèses.
- **Analyse de la stratégie de référence** : approche, architecture, coûts et limites.
- **Stratégie alternative proposée** : description détaillée, motivations et points distinctifs par rapport à la stratégie de référence.
- **Section expérimentale** : protocole d'évaluation, jeux de données, matériel/software, métriques.
- **Discussion** : interprétation des résultats, limites.
- **Conclusion et perspectives futures** : synthèse des enseignements et avenues de recherche.
- **Références** : format APA.

### III. Contraintes du projet

- Rédaction en anglais de l'article, dont la longueur ne doit pas dépasser 12 pages.
- Possibilité d'appliquer le formatage typique d'un article de conférence donné.
- Travail collaboratif en groupes de 3 à 5 étudiants, dont les noms devront figurer dans l'article.
- Date de remise de l'article : 5 février 2026. Aucun livrable additionnel (par ex. code source, données brutes, etc.) n'est exigé.
- Un espace Drive sera communiqué pour le dépôt de l'article.