Title: **Design Space Exploration for Accuracy-Aware Computing**

*Keywords :* Approximate Computing, Energy Efficiency, Error Resilience, Input-Aware Design, Arithmetic Circuits.

The increasing complexity of large-scale applications, such as scientific computing and data analytics, has driven growing demands for computational resources. Traditional methods are becoming insufficient to handle the surging data volume. This thesis addresses the gap between data growth and processor advancements by introducing strategies for improving computational efficiency without sacrificing performance. A key focus is Approximate Computing (AxC), which leverages the inherent error tolerance in many applications—such as multimedia processing and machine learning—to balance performance with acceptable error levels.

We conducted a comprehensive review of circuit-level functional approximations in various multipliers, aiming to deepen our understanding of the techniques used. We also discussed key accuracy metrics for evaluating the performance and reliability of these operators at the application and system levels. In addition, we reviewed automatic methods from the literature that support the broader adoption of approximate circuits, highlighting the need for a systematic approach to their design.

The key contribution of the thesis is the introduction of input-aware approximate computing, which advances the design of efficient systems by optimizing performance and energy consumption while maintaining acceptable accuracy. The goal was to develop methods that are broadly applicable but specifically optimized for certain workloads. Initially, manual identification of circuit sections for approximation highlighted scalability challenges and the need for specialized expertise. To address this, an automated design approach was implemented, optimizing circuits based on specific input distributions. Tested with the FIT filter, this approach consistently improved accuracy and energy efficiency compared to traditional methods. Input-aware approximation is especially valuable for applications with predictable input patterns, such as embedded systems, machine learning inference, and real-time signal processing.

This thesis also explores the relationship between approximate computing and the resilience of deep neural networks (DNNs), which are crucial in many applications. It analyzes how DNNs, especially those implemented on custom systolic arrays, respond to faults from approximate multipliers, using a Fault Injector Framework for efficient simulation. The study found that while DNNs are inherently fault-tolerant, approximations can significantly impact accuracy. A per-bit analysis revealed that critical bits, like the most significant bit (MSB), are particularly affected. Additionally, the trade-off between energy efficiency and resilience was highlighted, showing that energy savings from approximate computing must be balanced with the potential for performance degradation.

Titre: **Exploration de l'espace de conception pour le calcul conscient de la précision**

*Mots-clés :* Calcul approximatif, Efficacité énergétique, Résilience aux erreurs, Calcul conscient de la précision, Circuits arithmétiques.

L'augmentation de la complexité des applications à grande échelle, telles que le calcul scientifique et l'analyse de données, a entraîné une demande croissante de ressources informatiques. Les méthodes traditionnelles deviennent insuffisantes pour gérer le volume de données en constante augmentation. Cette thèse traite de l'écart entre la croissance des données et les avancées des processeurs en introduisant des stratégies visant à améliorer l'efficacité computationnelle sans sacrifier les performances. Un accent particulier est mis sur le calcul approximatif (AxC), qui exploite la tolérance aux erreurs inhérente à de nombreuses applications, telles que le traitement multimédia et l'apprentissage automatique, pour équilibrer les performances avec des niveaux d'erreur acceptables.

Nous avons réalisé une revue complète des approximations fonctionnelles au niveau des circuits dans divers multiplicateurs, afin d'approfondir notre compréhension des techniques utilisées. Nous avons également discuté des principaux critères de précision pour évaluer la performance et la fiabilité de ces opérateurs aux niveaux des applications et des systèmes. En outre, nous avons passé en revue les méthodes automatiques issues de la littérature qui soutiennent une adoption plus large des circuits approximatifs, en soulignant la nécessité d'une approche systématique pour leur conception.

La contribution principale de cette thèse est l'introduction du calcul approximatif sensible aux entrées, qui fait progresser la conception de systèmes efficaces en optimisant la performance et la consommation d'énergie tout en maintenant une précision acceptable. L'objectif était de développer des méthodes largement applicables, mais spécifiquement optimisées pour certaines charges de travail. Initialement, l'identification manuelle des sections de circuits à approximations a mis en évidence des défis de scalabilité et la nécessité d'une expertise spécialisée. Pour y remédier, une approche de conception automatisée a été mise en œuvre, optimisant les circuits en fonction de distributions d'entrées spécifiques. Testée avec le filtre FIT, cette approche a constamment amélioré la précision et l'efficacité énergétique par rapport aux méthodes traditionnelles. L'approximation sensible aux entrées est particulièrement utile pour les applications avec des schémas d'entrées prévisibles, telles que les systèmes embarqués, l'inférence d'apprentissage automatique et le traitement de signaux en temps réel.

Cette thèse explore également la relation entre le calcul approximatif et la résilience des réseaux neuronaux profonds (DNNs), cruciaux dans de nombreuses applications. Elle analyse comment les DNNs, en particulier ceux implémentés sur des matrices systoliques personnalisées, réagissent aux fautes provenant des multiplicateurs approximatifs, en utilisant un cadre d'injection de fautes pour une simulation efficace. L'étude a révélé que bien que les DNNs soient intrinsèquement tolérants aux fautes, les approximations peuvent avoir un impact significatif sur la précision. Une analyse au niveau des bits a révélé que les bits critiques, comme le bit le plus significatif (MSB), sont particulièrement affectés. De plus, le compromis entre l'efficacité énergétique et la résilience a été souligné, montrant que les économies d'énergie issues du calcul approximatif doivent être équilibrées avec le potentiel de dégradation des performances.