

## الملخص

وحدة معالجة الرسومات GPU أصبح توجه كثير من الباحثين في مجال تطبيقات الحوسبة عالية الأداء لما يملكه من قدرات هائلة تدعم البرمجة المتوازية. الطرق التكرارية لأنظمة المعادلات الخطية الكبيرة تشكل لبنة أساسية في كثير من التطبيقات العلمية والهندسية هي واحدة من الأمثلة المهمة لتطبيقات الحوسبة عالية الأداء. ولكن الوصول الى أداء عالي يوازي القدرات الهائلة لمعالج GPU لعملية ضرب المصفوفات المتناثرة بالمتجهات (SpMV) والتي تعتبر النواة الأساسية لحل هذه المعادلات الخطية الكبيرة لا يزال تحدي للباحثين بسبب تأثير اختلاف أنماط وخصائص هذه المصفوفات على عملية الأداء. تم تقديم العديد من الطرق من قبل الباحثين لتخزين هذه المصفوفات المتناثرة ومحاولة تحسين تخزينها بغرض معالجة تأثير انماطها على أداء عملية SpMV. ولكن معظم هذه الطرق تتسم بمحدودية تقييم الأداء والتي تعني تقييم الأداء من جوانب قليلة (كقياس إنتاجية المعالج فقط) لا تعطي نظرة عميقة لأداء عملية SpMV. لذلك الهدف من هذه الرسالة هو تحليل وفهم وتحسين الأداء لطريقة جاكوبي التكرارية وحسابات SpMV في معالج GPU. ولتحقيق هذا الهدف قمنا بدراسة مفصلة للطرق الشائعة لإجراء عملية SpMV وهي (CSR, ELL, HYB, HYB). يشمل ذلك تقييم الأداء لكل طريقة بشكل مفرد وأيضاً مقارنة أداء جميع هذه الطرق باستخدام ثمانية مقاييس مختلفة وهي الوقت المستغرق للتنفيذ (execution time)، إنتاجية المعالج (GPU throughput)، إنتاجية الذاكرة (Memory throughput)، نسبة شغل المعالج (achieved occupancy)، عدد العمليات لكل نواة (instructions per warp)، نسبة شغل الأنوية (warp execution efficiency)، فعالية الذاكرة (memory efficiency)، عدد العمليات للذاكرة (memory transactions). باستخدام هذه النظرة المفصلة لأداء كل طريقة تم تحديد مكان ضعف الأداء والفجوات لهذه الطرق بالإضافة الى تقديم طريقة مقترحة (HCGHYB) لإجراء SpMV والتي تم فيها حل لبعض مكان الضعف الحالية. هذه الطريقة المقترحة قدمت أداء أفضل مقارنة بنظيرتها HYB والتي تعتبر من أكثر الطرق الشائعة في حسابات SpMV. وأخيراً تم اقتراح عدد من الحلول والتي يمكن أن تسهم في تحسين حسابات SpMV في معالج GPU.

## ABSTRACT

Graphics processing units (GPUs) have delivered remarkable performance for many high performance computing (HPC) applications through massive parallelism that they provide. Iterative solution of large linear equation systems that forms an essential building block for numerous scientific and engineering applications is one such application. However, obtaining performance for Sparse Matrix-Vector (SpMV) multiplication that is the core kernel of iterative solutions of linear equation systems has remained a challenge due to the varying sparsity patterns of the nonzero entries in the linear equation systems. A number of schemes have been proposed to improve SpMV storage and computation performance. However, these schemes are mainly evaluated and compared in terms of the SpMV throughput in FLOPS, which alone does not provide a deep insight into the SpMV storage and computations.

The aim of this thesis is to analyze, understand, and improve the performance of Jacobi iterative method and SpMV computations on GPU architectures. Towards this aim, we provide a detailed study of four notable schemes (CSR, ELL, HYB, and CSR5). The schemes are evaluated individually and compared using eight different performance metrics including execution time, GFLOPS, achieved occupancy, instructions per warp, warp execution efficiency, global memory throughput, global memory efficiency, and global memory transactions. Subsequently, using the deeper insights into the performance gaps of the current schemes, gained through the detailed performance analysis of the schemes, we propose a novel SpMV computations scheme called the heterogeneous CPU-GPU Hybrid (HCGHYB) scheme. HCGHYB provides improved performance over the HYB scheme, which is a popular choice in

many SpMV open source and commercial iterative solvers. Finally, we give directions for further improvements in SpMV computations schemes.