

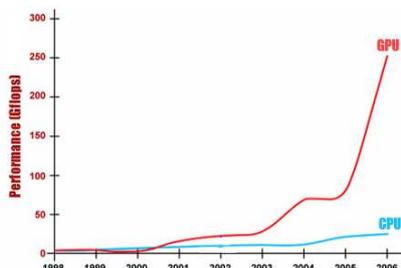
บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของงานวิจัย

ในยุคปัจจุบันเทคโนโลยีทางด้านสถาปัตยกรรมกราฟิกหรือจีพียู (GPU: Graphics Processing Unit) ได้มีความก้าวหน้าเป็นอย่างมาก อีกทั้งมีการพัฒนาความสามารถในด้านการประมวลผลอย่างต่อเนื่องไม่หยุดนิ่ง จึงทำให้ในระยะเวลาไม่กี่ปีที่ผ่านมาประสิทธิภาพในการประมวลผลของจีพียูเพิ่มขึ้นมากอย่างต่อเนื่องเมื่อเปรียบเทียบกับซีพียูดังแสดงในภาพที่ 1.1 ซึ่งแสดงให้เห็นถึงประสิทธิภาพของจีพียู (เส้นบน) ที่เพิ่มขึ้นอย่างมากเมื่อเปรียบเทียบกับซีพียู (เส้นล่าง) ในการประมวลผลแบบ GFLOPS โดยแท้ที่จริงแล้วหน้าที่หลักของจีพียูนั้นเริ่มต้นจากการเป็นตัวช่วยในการประมวลผลการทำงานในด้านภาพกราฟิกบนหน้าจอคอมพิวเตอร์ให้มีประสิทธิภาพสูงสุด ซึ่งได้แก่งานประเภทให้ความบันเทิง เช่น เกมส์ หรือวิดีโอ และด้วยโครงสร้างของจีพียูที่มีลักษณะเป็นแบบหลายแกน (Many Core) ที่ทำให้จีพียูมีประสิทธิภาพในการประมวลผลสูง จึงทำให้ในปัจจุบันได้มีการนำความสามารถในการประมวลผลทางด้านการคำนวณของจีพียู มาใช้กับงานที่ไม่ใช่ประเภทการประมวลผลทางด้านกราฟิกเพียงอย่างเดียวแต่นำจีพียูมาประมวลผลแทน หรือช่วยซีพียูในการทำงานที่ซับซ้อนหลาย ๆ อย่างสำหรับงานที่ต้องการความสามารถในการคำนวณสูง (High Performance Computing) เช่น งานในด้านการคำนวณ และงานทางด้านวิทยาศาสตร์ เป็นต้น โดยใช้คูดำเอพีไอ (CUDA API) และเมื่อเทียบกับซีพียูแล้วการใช้งานจีพียูสามารถเพิ่มประสิทธิภาพของแอปพลิเคชันได้กว่าสิบหรือร้อยเท่า (M. Garland et al, M. Boyer et al, B. Burke, A. Humber)

ภาพที่ 1.1 ความแตกต่างของประสิทธิภาพการประมวลผลระหว่าง GPU และ CPU



ที่มา : “Different cores for different chores” โดย Tony Smith, 2006, จาก

http://www.reghardware.com/2006/10/26/the_story_of_amds_fusion/

เทคโนโลยีเวอร์ชวลแมชชีน (Virtual Machine หรือ VM) เป็นเทคโนโลยีการจำลองเครื่องคอมพิวเตอร์โดยซอฟต์แวร์บนเครื่องคอมพิวเตอร์จริง ทำให้สามารถสร้างเวอร์ชวลแมชชีนหรือคอมพิวเตอร์เสมือนหลาย ๆ เครื่องให้ทำงานขนานกันไปบนคอมพิวเตอร์จริงเพียงเครื่องเดียว เครื่องคอมพิวเตอร์เสมือนแต่ละเครื่องจะรันระบบปฏิบัติการของตัวเอง เรียกว่าระบบปฏิบัติการเกส (Guest Operating System) การใช้เวอร์ชวลแมชชีนนั้นได้รับความนิยมมากในปัจจุบัน เพราะทำให้เกิดการเพิ่มการใช้ประโยชน์ (Utilization) ของเครื่องคอมพิวเตอร์จริงมากขึ้น

ระบบเวอร์ชวลแมชชีนที่มีอยู่ในขณะนี้ไม่สนับสนุนการคำนวณสมรรถนะสูง ในสภาพแวดล้อมจำลองของเวอร์ชวลแมชชีนนั้น เป็นเรื่องยากที่จะรันแอปพลิเคชันสมรรถนะสูงที่ต้องการสั่งงาน หรือเรียกใช้งานจีพียูโดยใช้คูด้าเอพีไอผ่านเวอร์ชวลแมชชีนเนื่องจากเกสโอเปอเรติงซิสเต็มในเวอร์ชวลแมชชีนไม่สามารถเรียกใช้งานอุปกรณ์จีพียูจากไดร์เวอร์ของอุปกรณ์จีพียูที่มีอยู่ในระบบปฏิบัติการโฮส (Host Operating System) บนเครื่องคอมพิวเตอร์จริงได้โดยตรง ยิ่งไปกว่านั้นระบบเวอร์ชวลแมชชีนที่มีอยู่ยังไม่สนับสนุนการใช้งานจีพียูร่วมกันจากเวอร์ชวลแมชชีนหลายเวอร์ชวลแมชชีน (ที่อาจอยู่บนเครื่องคอมพิวเตอร์จริงเครื่องเดียวกันกับอุปกรณ์จีพียูหรืออยู่ต่างเครื่องก็ได้) อีกด้วย

วิทยานิพนธ์ฉบับนี้จึงนำเสนอ**ระบบเวอร์ชวลคูด้า** (Virtual CUDA) สำหรับให้บริการ การเข้าถึงทรัพยากรจีพียูได้จากเวอร์ชวลแมชชีนโดยผ่านคูด้าเอพีไอที่จำลองขึ้นมา ซึ่งพัฒนาซอฟต์แวร์ต้นแบบ (Prototype) ของระบบเวอร์ชวลคูด้าประกอบด้วย ฟรอนท์เอนด์ไลบรารี (Front-End Library) ซึ่งจำลองคูด้าเอพีไอ และแบ็คเอนด์ซอฟต์แวร์ (Back-End Software) โดยการออกแบบระบบเวอร์ชวลคูด้าขึ้นเพื่อให้มีความสามารถในการรองรับการทำงาน ดังนี้ คือ 1) เพื่อให้คูด้าแอปพลิเคชันบนเวอร์ชวลแมชชีนหรือเครื่องคอมพิวเตอร์เครื่องอื่นที่ไม่มีอุปกรณ์จีพียูเป็นของตนเอง แต่เชื่อมต่อกับเครื่องคอมพิวเตอร์ที่มีอุปกรณ์จีพียูผ่านเน็ตเวิร์คให้สามารถเรียกใช้งานทรัพยากรจีพียูโดยใช้คูด้าเอพีไอได้เสมือนมีทรัพยากรจีพียูเป็นของตนเอง 2) เพื่อเพิ่มการใช้ประโยชน์ของจีพียูโดยอนุญาตให้เวอร์ชวลแมชชีนมากกว่าหนึ่งเครื่องสามารถเข้าใช้งานจีพียูร่วมกันได้ โดยที่ระบบแบ็คเอนด์ของเวอร์ชวลคูด้าจะจัดสรรการเข้าถึงจีพียูของแต่ละผู้ขอรับบริการ เพื่อไม่ให้เกิดปัญหาการเรียกใช้งานถูกยกเลิก กรณีที่ทรัพยากรจีพียูถูกเรียกใช้จนเต็มหรือการรอใช้ทรัพยากรจีพียูจนเกิดปัญหาการติดตาย (Deadlock) และ 3) เพื่อให้สามารถนำไปใช้งานบนเวอร์ชวลแมชชีนได้หลายชนิดไม่ขึ้นอยู่กับชนิดใดชนิดหนึ่ง ซอฟต์แวร์ต้นแบบนี้ได้รับการพัฒนาขึ้นบนเควีเอ็ม (KVM) ซึ่งเป็นเวอร์ชวลแมชชีนบนระบบลินุกซ์และสามารถนำไปพัฒนาใช้งานบนเวอร์ชวลแมชชีนแบบอื่นได้โดยง่าย

1.2 วัตถุประสงค์ของงานวิจัย

ในงานวิทยานิพนธ์ฉบับนี้ มีวัตถุประสงค์ในการค้นคว้าวิจัยดังนี้

1. เพื่อศึกษาและพัฒนาวิธีการเข้าถึงอุปกรณ์จีพียูสำหรับคู้ด้าแอปพลิเคชันที่อยู่ในเวอร์ชวลแมชชีน
2. เพื่อศึกษาและพัฒนาวิธีการจัดการสรรกรเข้าถึงทรัพยากรจีพียูจากการเรียกใช้งานร่วมกันของเวอร์ชวลแมชชีนและหลีกเลี่ยงการเกิดปัญหาการติดตาย (Deadlock)
3. เพื่อให้ผู้ที่ใช้งานเวอร์ชวลแมชชีนสามารถเรียกใช้งานจีพียูจากเวอร์ชวลแมชชีนได้
4. เพื่อศึกษาประสิทธิภาพของระบบเวอร์ชวลคู้ด้า (Virtual CUDA)
5. เพื่อศึกษาลักษณะของแอปพลิเคชันที่เหมาะสมสำหรับการใช้งานระบบเวอร์ชวลคู้ด้า

1.3 ขอบเขตของงานวิจัย

งานวิทยานิพนธ์ฉบับนี้นำเสนอไดร์เวอร์จำลองสำหรับเครื่องเวอร์ชวลแมชชีนที่ต้องการเรียกใช้งานจีพียูจากเครื่องหลักของผู้ให้บริการอุปกรณ์จีพียู โดยมีขอบเขตการวิจัยดังนี้

1. ออกแบบและพัฒนาระบบเวอร์ชวลคู้ด้า สำหรับการเรียกใช้งานอุปกรณ์จีพียูจากเครื่องเวอร์ชวลแมชชีน
2. ออกแบบและพัฒนาแคปซูลที่เครื่องคอมพิวเตอร์จริงของผู้ให้บริการจีพียูเพื่อเข้าถึงจีพียูจากการร้องขอของเวอร์ชวลแมชชีน
3. ออกแบบและพัฒนาการจัดการสรรกรเข้าถึงจีพียูจากการร้องขอของเวอร์ชวลแมชชีน กรณีมีการขอใช้จีพียูและเข้าถึงจีพียูพร้อมกัน
4. การออกแบบระบบเวอร์ชวลคู้ด้าสำหรับเรียกใช้งานจีพียูกำหนดให้คู้ด้าแอปพลิเคชัน สามารถเรียกใช้งาน CUDA API Programming Interface ดังต่อไปนี้

- | | |
|-----------------------|------------------------------|
| ➤ cuInit | ➤ cuMemAlloc |
| ➤ cuDeviceGet | ➤ cuMemcpyHtoD/ cuMemcpyDtoH |
| ➤ cuCtxCreate | ➤ cuParamSetv |
| ➤ cuModuleLoad | ➤ cuParamSetSize |
| ➤ cuModuleGetFunction | ➤ cuFuncSetBlockShape |

➤ cuLaunchGrid

➤ cuMemFree

➤ cuCtxDetach

5. การทดลองของงานวิจัยนี้จะใช้การอิมพลีเมนต์บนเควีเอ็มด้วยระบบไฮสโอสที่เป็นระบบปฏิบัติการลินุกซ์ และมีระบบปฏิบัติการในเวอร์ชวลแมชชีนเป็นระบบปฏิบัติการลินุกซ์เช่นกัน

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. ระบบเวอร์ชวลคูด้าทำให้สามารถรันคูด้าแอปพลิเคชันจากเวอร์ชวลแมชชีนได้
2. ระบบเวอร์ชวลคูด้าทำให้สามารถเรียกใช้งานจีพียูจากเวอร์ชวลแมชชีนได้
3. ระบบเวอร์ชวลคูด้าทำให้เกิดการใช้งานจีพียูร่วมกันจากหลายเวอร์ชวลแมชชีนพร้อม ๆ กันได้ โดยไม่เกิดการติดตาย
4. ระบบเวอร์ชวลคูด้าทำให้การใช้ประโยชน์ (Utilization) ของจีพียูเพิ่มขึ้น อันเนื่องมาจากการอนุญาตให้มีการใช้งานจีพียูพร้อม ๆ กันจากหลายเวอร์ชวลแมชชีน
5. ระบบเวอร์ชวลคูด้าทำให้สามารถรันคูด้าแอปพลิเคชันจากเวอร์ชวลแมชชีนที่รันอยู่บนคอมพิวเตอร์ที่ไม่มีอุปกรณ์จีพียูเป็นของตนเอง สามารถเรียกใช้งานจีพียูจากคอมพิวเตอร์เครื่องอื่นที่มีอุปกรณ์จีพียูได้
6. ความรู้จากการศึกษาประสิทธิภาพของระบบเวอร์ชวลคูด้า และลักษณะของคูด้าแอปพลิเคชันที่เหมาะสมกับระบบเวอร์ชวลคูด้า