

มหาวิทยาลัยสงขลานครินทร์

คณะวิศวกรรมศาสตร์

สอบปลายภาค : ภาคการศึกษาที่ 2

ปีการศึกษา : 2558

วันที่สอบ : 30 เมษายน 2559

เวลาสอบ : 9.00-12.00 น.

รหัสวิชา : 242-441

ห้องสอบ : A201

ชื่อวิชา : สถาปัตยกรรมและการจัดองค์ประกอบพีวเตอร์ขั้นสูง

(Advanced Computer Architecture and Organization)

ผู้ออกข้อสอบ : พิชญ์ ตันชัย

คำสั่ง: อ่านรายละเอียดของข้อสอบ และคำแนะนำให้เข้าใจก่อนเริ่มทำข้อสอบ

อนุญาต: เครื่องเขียนต่างๆ เช่น ปากกา หรือ ดินสอ, กระดาษโน้ตขนาด A4 เขียนด้วยลายมือตนเอง พร้อมชื่อนามสกุลและรหัสนักศึกษา จำนวน 1 แผ่น (2 หน้า)

ห้าม: ห้ามนำข้อสอบทั้งหมดหรือบางส่วนออกนอกห้องสอบ และห้ามหยิบยืมหนังสือ เอกสาร เครื่องคำนวณ หรืออุปกรณ์อื่นๆซึ่งกันและกัน

เวลา: 3 ชั่วโมง (180 นาที)

คำแนะนำ:

- * ข้อสอบมี 16 หน้า 3 ข้อใหญ่ คะแนนรวม 160 คะแนน (25%) ให้ทำทุกข้อ และตอบเป็นภาษาไทย
- * อ่านคำสั่งในแต่ละข้อให้เข้าใจก่อนลงมือทำ
- * เขียนคำตอบลงในข้อสอบ คำตอบส่วนใดอ่านไม่ออก จะถือว่าคำตอบนั้นผิด
- * อย่าลืม เขียน ชื่อ-นามสกุล และรหัสนักศึกษา ลงในข้อสอบทุกแผ่น
- * แนบกระดาษโน้ตขนาด A4 ในข้อสอบ เมื่อส่งข้อสอบ

ทุจริตในการสอบ: มีโทษขั้นต่ำปรับตกในรายวิชานั้นและพักการเรียน 1 ภาคการศึกษา โทษสูงสุดให้ออก

ข้อที่	เวลา (นาที)	คะแนน	ได้	คะแนนดิบ	ร้อยละ	สะสม
1	65	65		160	100	25%
2	95	95				
3	20	20				
รวม	180	160				

คำถามข้อที่ 1 Multiprocessors and Thread-Level Parallelism (65 คะแนน/65 นาที)

- 1.1 จงระบุว่าประโยคต่อไปนี้เป็นจริง (T) หรือเท็จ (F) (25 คะแนน)
- _____ ใน Uniformed Memory Access (UMA) ทุกโพรเซสเซอร์มีหน่วยความจำท้องถิ่นเป็นของตนเอง (local memory)
 - _____ ใน Non-Uniformed Memory Access (NUMA), ทุกโพรเซสเซอร์สามารถเข้าถึงหน่วยความจำทุกส่วนได้ในเวลาเท่าเทียมกัน
 - _____ ตัวอย่างของสถาปัตยกรรมแบบ Non-Uniformed Memory Access (NUMA) คือ Parallel Vector Processor (PVP) และ Symmetric Multiprocessors (SMP)
 - _____ ใน Distributed Shared-Memory (DSM) ตัวควบคุมแคชสามารถสอดส่องดู (snoop) ข้อมูลที่วิ่งอยู่บน shared memory bus ได้
 - _____ Centralized Memory Multiprocessors เรียกอีกชื่อหนึ่งว่า symmetric multiprocessors (SMPs) เพราะมีหน่วยความจำหลักหน่วยเดียวและมีความสัมพันธ์แบบ symmetric กับทุกโพรเซสเซอร์
 - _____ Centralized Memory Multiprocessors มีหน่วยความจำขนาดใหญ่เพื่อให้หน่วยความจำหลักสามารถรองรับความต้องการของโพรเซสเซอร์ทั้งหมดได้
 - _____ การสื่อสารข้อมูลระหว่างโพรเซสเซอร์ใน Distributed Memory Multiprocessors มีความซับซ้อนมากกว่าใน Centralized Memory Multiprocessors
 - _____ ในโพรเซสเซอร์แบบเวกเตอร์จะสามารถประมวลผลชุดคำสั่งได้ทั้งแบบเวกเตอร์และสเกลาร์ (scalar)
 - _____ ใน Processing Array ความเร็วของซีพียูจะเพิ่มขึ้นเมื่อประมวลผลคำสั่งแบบมีเงื่อนไข
 - _____ โดยธรรมชาติแล้ว Processor Arrays ออกแบบเพื่อให้รองรับผู้ใช้ได้หลายคน
 - _____ ตัวอย่าง interconnections ของ Massively Parallel Processors (MPPs) คือโครงสร้างแบบ hypercube และ mesh
 - _____ Massively Parallel Processors (MPPs) สื่อสารข้อมูลซึ่งกันและกันโดยใช้ shared memory
 - _____ ใน Symmetric Multiprocessors ตำแหน่งที่อยู่เดียวกันที่อ้างถึงโดยซีพียูที่แตกต่างกันหมายถึงตำแหน่งที่แตกต่างกันในหน่วยความจำ
 - _____ ใน Symmetric Multiprocessors โพรเซสเซอร์สื่อสารกันผ่านทาง shared data
- ชื่อ _____ รหัสนักศึกษา _____

- o) _____ ใน Symmetric Multiprocessors ระยะเวลาในการเข้าถึงหน่วยความจำจะเท่ากันในทุกซีพียู
- p) _____ การสำเนาข้อมูลไปใช้ในหน่วยความจำแคชของแต่ละซีพียูจะช่วยลดการแข่งขันหรือการช่วงชิงข้อมูลกันระหว่างโพรเซสเซอร์ แต่ข้อมูลดังกล่าวในแต่ละซีพียูอาจไม่เหมือนกันหรือไม่ถูกต้องได้ ต้องปรับข้อมูลให้ทันสมัยอยู่เสมอด้วยกลไกของ Cache Directory
- q) _____ ใน Cache Coherence การเขียนค่าใดกลงในแคชของโพรเซสเซอร์ใดๆต้องได้รับ exclusive access ที่จะเขียนข้อมูล ณ ตำแหน่งนั้นก่อน
- r) _____ ใน Cache Coherence การเขียนค่าใดกลงในแคช จะต้องตรวจสอบ (validate) ค่าข้อมูลนั้นในแคชอื่นๆก่อน
- s) _____ เราสามารถลดความถี่ของการเข้าถึงข้อมูลจากระยะไกลที่มีผลต่อประสิทธิภาพของการประมวลผลแบบขนานได้โดยการแคช shared data ในฮาร์ดแวร์หรือการสร้าง data layout ในซอฟต์แวร์รุ่นใหม่ เพื่อให้สามารถเข้าถึงข้อมูลนั้นได้แบบท้องถิ่น (local)
- t) _____ ใน Snooping Protocol การดึงข้อมูลมาจาก cache block ของโพรเซสเซอร์อาจมีความซับซ้อน และใช้เวลานานกว่าการดึงข้อมูลโดยตรงจากหน่วยความจำได้
- u) _____ ใน Asymmetrical Cluster เครื่อง front end อาจเป็น single point of failure ได้
- v) _____ In Asymmetrical Cluster, ประสิทธิภาพของเครื่อง front end เป็นตัวจำกัด scalability ของระบบ
- w) _____ ใน Asymmetrical Cluster คอมพิวเตอร์แต่ละตัวใช้ระบบปฏิบัติการตัวเดียวกันและสามารถทำงานได้เหมือนกัน ผู้ใช้สามารถ login เข้าสู่คอมพิวเตอร์ใดๆเพื่อที่จะแก้ไขหรือ compile โปรแกรมต่างๆได้
- x) _____ ใน Symmetrical Cluster นั้น CPU cycle จะถูกออกแบบมาเพื่อการประมวลผลแบบขนาน
- y) _____ ใน Cluster of Workstations (COW) แต่ละโหนด (node) เป็นเครื่องเสมือน (virtual machine)

1.2 จงเปรียบเทียบ Centralized Memory Multiprocessor และ Distributed-Memory multiprocessor (4 คะแนน)

ชื่อ _____ รหัสนักศึกษา _____

1.3 จงเติมคำในช่องว่าง

(12 คะแนน)

- a) _____ คือสถานการณ์ซึ่งมีเพียงโพรเซสเดียวที่จะสามารถกระทำกิจกรรมเฉพาะ ณ เวลาหนึ่งๆ ได้
- b) _____ จะการันตีว่าไม่มีโพรเซสใดจะสามารถก้าวข้ามหรือเดินหน้าต่อไปจากจุดหนึ่งในโปรแกรมได้ จนกว่าทุกโพรเซสจะถึงจุดนั้น
- c) _____ คือระบบคอมพิวเตอร์ที่มีหน่วยความจำแบบกระจายขนาดใหญ่โดยมีโหนดแยกหลายโหนด
- d) ใน _____ ตัวควบคุมแคชของซีพียูจะตรวจสอบบัสเพื่อที่จะหาว่า cache block ไหนถูกร้องขอข้อมูลจากซีพียูอื่น
- e) ใน _____ ฮาร์ดแวร์และซอฟต์แวร์ของระบบสร้างภาพของพื้นที่หน่วยความจำที่มีที่อยู่เดียว (single address space) ให้กับผู้ใช้งาน
- f) ใน Distributed Shared-Memory (DSM) ใช้ _____ สำหรับรองรับ distributed coherent caches
- g) ใน Distributed Shared-Memory (DSM) ใช้ _____ directory entry สำหรับแต่ละ cache block
- h) ใน Asymmetrical Cluster เครื่อง _____ ได้ตอบกับผู้ใช้ และอุปกรณ์อินพุตเอาต์พุต ในขณะที่ _____ อุทิศให้กับการประมวลผลโปรแกรมแบบขนาน
- i) ใน Symmetric Shared-Memory Multiprocessors มี cache miss ประเภทที่สี่ นอกเหนือจาก Compulsory Miss, Capacity Miss, and Conflict Miss นั่นคือ _____ ซึ่งเกิดจาก _____ ในขณะที่พยายามจะเขียนข้อมูลซึ่งได้ถูกสำเนาไปยัง cache block อื่นๆ แล้ว
- j) ใน _____ ทุก memory block สัมพันธ์กับ directory information ซึ่งมีหน้าที่เก็บข้อมูลการสำเนา cached block ต่างๆ และสถานะเพื่อใช้ในการทำ Cache Coherence

1.4 จงคำนวณหาค่าตอบต่อไปนี้

(6 คะแนน)

- a) หากการใช้จำนวนโพรเซสเซอร์จำนวน 256 คอร์ ทำให้ได้ speedup เป็น 100 จงระบุว่า สัดส่วนของโปรแกรมที่สามารถประมวลผลเป็นแบบขนานควรเป็นเท่าใด?

(3 คะแนน)

- b) สมมติว่าซีพียูแบบหลายโพรเซสเซอร์ มีค่า base CPI เป็น 0.4 และมีระยะเวลาในการเข้าถึง remote memory เป็น 100 ns ($\text{remote access} = 100/0.5 = 200$ clock cycles) จะมีผลกระทบต่อประสิทธิภาพเป็นอย่างไร หาก 0.4% ของคำสั่งทั้งหมดในโปรแกรมเกี่ยวข้องกับการเข้าถึงข้อมูลระยะไกล (remote access)

(3 คะแนน)

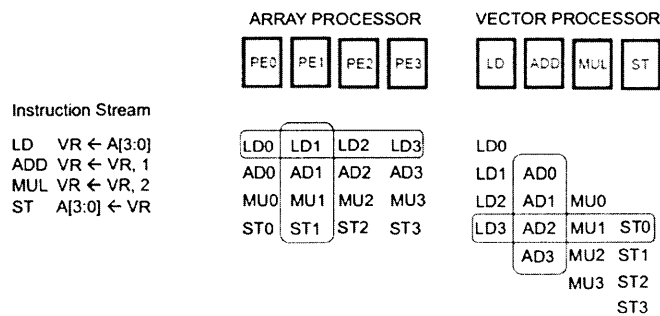
1.5 จงตอบคำถามต่อไปนี้

(18 คะแนน)

- a) จงอธิบายปัญหาต่างๆที่เกี่ยวข้องกับ shared data

(4 คะแนน)

- b) จากภาพต่อไปนี้ของแนวทางการประมวลผลแบบ Vector Processor และ Vector Array จงแสดงแกนเวลา (time) และ space ลงในภาพดังกล่าว พร้อมทั้งอธิบายความแตกต่าง (4 คะแนน)



c) จงแสดงภาพการเชื่อมต่อโหนดต่างๆของ Asymmetrical Cluster และ Symmetrical Cluster (4 คะแนน)

d) จงอธิบายสถานะต่างๆของ shared cache block sharing ใน Directory Based Cache Coherence Protocol. (6 คะแนน)

ชื่อ _____ รหัสนักศึกษา _____

คำถามข้อที่ 2 **Parallel Computing, Performance Analysis และ Load Balancing (95 คะแนน; 95 นาที)**

2.1 จงตอบคำถามต่อไปนี้ (40 คะแนน, 40 นาที)

a) จงแสดงปัจจัยทางฮาร์ดแวร์มา 4 ปัจจัยที่มีบทบาทสำคัญต่อ scalability (4 คะแนน)

b) จงแสดง 4 ปัจจัยที่มีผลต่อ parallel overhead (4 คะแนน)

c) จงเปรียบเทียบ *loop independent data dependence* และ *loop carried data dependence* โดยการยกตัวอย่าง code fragment สำหรับแต่ละแบบ (4 คะแนน)

Loop independent data dependence	Loop carried data dependence

d) จงแสดงตัวอย่างของปัญหาที่ส่งผลให้เกิดภาวะภาระงานไม่สมดุล หากข้อมูลไม่ได้กระจายตัวอย่างสม่ำเสมอ (2 คะแนน)

ชื่อ _____ รหัสนักศึกษา _____

e) จงระบุปัจจัยอย่างน้อย 4 ปัจจัยที่ต้องคำนึงถึงในการออกแบบ โปรแกรมที่เกี่ยวข้องกับ inter-task communications (4 คะแนน)

f) จงระบุและอธิบาย 3 ปัจจัยที่มีผลต่อ *Dynamic Load Balancing* (6 คะแนน)

g) จงอธิบายแนวทางการดำเนินนโยบาย *System Information exchange* ในการจัดการระบบให้สมดุลอย่างน้อย 3 แนวทาง (6 คะแนน)

h) จงอธิบาย 7 คุณสมบัติ (Property) ของระบบที่มีการจัดการระบบให้สมดุล (Load Balancing Systems) (7 คะแนน)

i) เราสามารถเพิ่ม throughput ในการประมวลผลแบบขนานได้อย่างไรบ้าง? (2 คะแนน)

j) จงระบุตัวแปรที่มีผลต่อประสิทธิภาพของหน่วยความจำและแนวทางการวัดค่าประสิทธิภาพดังกล่าว (3 คะแนน)

2.2 จงระบุว่าประโยคต่อไปนี้เป็นจริง (T) หรือเท็จ (F) (25 คะแนน)

- _____ Fine-grain Parallelism มีสัดส่วนการคำนวณต่ำกว่าการสื่อสาร
- _____ Fine-grain Parallelism มี overhead ในการสื่อสารต่ำ ซึ่งทำให้มีโอกาสมากกว่าในการปรับปรุงประสิทธิภาพของระบบ
- _____ ใน Fine-grain Parallelism การทำภาระงานให้สมดุลทำได้ยากกว่า Coarse-grain Parallelism
- _____ Coarse-grain Parallelism มีอัตราส่วนการคำนวณต่อการสื่อสาร (computation to communication ratio) ที่สนับสนุนการจัดการภาระงานให้สมดุล (load balancing)
- _____ การจัดการภาระงานให้สมดุล หมายถึงการกระจายงานระหว่างโพรเซสต่างๆเพื่อให้แต่ละโพรเซสทำงานอยู่ตลอดเวลา
- _____ กลยุทธ์ในการโปรแกรมแบบขนานคือ เน้นการทำงานแบบขนานเฉพาะบริเวณ hotspot ที่สิ้นเปลืองเวลาซีพียูมาก และไม่ต้องสนใจส่วนอื่นๆของโปรแกรมที่ใช้เวลา

ชื่อ _____ รหัสนักศึกษา _____

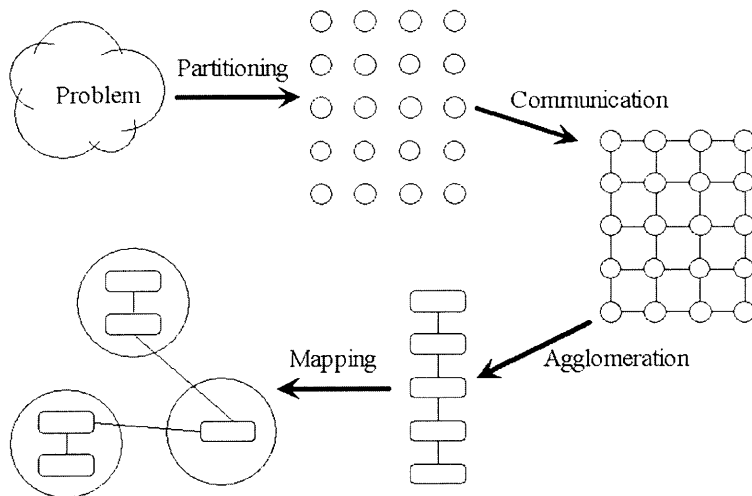
จีพียูไม่มาก

- g) ____ อุปสรรคพื้นฐาน (inhibitor) ในการประมวลผลแบบขนานคือการขึ้นต่อกันของข้อมูล (data dependence)
- h) ____ ปัญหาคอขวด (Bottleneck) ในโปรแกรมอยู่ในบริเวณที่ซ้ำผิปกติ หรือยับยั้งหยุดหรือช่วงเวลาการทำงานของส่วนที่เป็นโปรแกรมแบบขนาน
- i) ____ เมื่อพบปัญหาคอขวด ควรจัดรูปแบบการทำงานของโปรแกรมใหม่หรือใช้อัลกอริทึมใหม่ เพื่อลดหรือกำจัดบริเวณที่ทำงานซ้ำโดยไม่จำเป็น
- j) ____ หากทุกงานต้องรออยู่ที่จุดกั้นสำหรับเข้าจังหวะ (barrier synchronization point) งานที่เร็วที่สุดจะเป็นตัวกำหนดประสิทธิภาพรวมของระบบ
- k) ____ การจัดสมดุลภาระงานเกี่ยวกับ scheduling หรือการจัดสรรและจัดการทรัพยากร (resource allocation and management)
- l) ____ กลยุทธ์ในการจัดสมดุลภาระงานคือ พยายามย้ายงานจากเครื่องที่ไม่ค่อยไม่ภาระงานไปให้กับเครื่องที่มีภาระงานมาก.
- m) ____ การย้ายภาระงานมีเป้าหมายเพื่อให้ได้ response time สูงสุด เพื่อให้ได้ประสิทธิภาพโดยรวมสูงสุด
- n) ____ การย้ายภาระงานของโพรเซสที่ถูกบล็อก (blocked process) จะเป็นประโยชน์ในการลดภาระงานของโพรเซสเซอร์
- o) ____ โพรเซสขนาดเล็กจะทำให้เกิดภาระงานในเครือข่ายการสื่อสารมากกว่าโพรเซสขนาดใหญ่
- p) ____ ในการจัดสมดุลภาระงาน ควรพิจารณาย้ายโพรเซสที่ยังคงมีเวลาในการให้บริการเหลืออยู่สูงสุด
- q) ____ ในการจัดสมดุลภาระงาน ไม่ควรย้ายโพรเซสที่สื่อสารกันบ่อยให้ไปอยู่ที่ไหนคนเดียวกัน
- r) ____ ในการจัดสมดุลภาระงาน ไม่ควรย้ายโพรเซสที่กินทรัพยากรมากออกไปจากเครื่องที่ให้บริการ
- s) ____ ปัจจัยในการวัดความคุ้มค่าต่อราคา (cost-effectiveness) ของระบบคอมพิวเตอร์คือ utilization
- t) ____ Isoefficiency เป็นวิธีการวัดหรือระบุ scalability

- u) ____ ระบบที่สามารถขยายตัวได้คือ ระบบที่ยังคง efficiency ได้เท่าเดิม เมื่อเพิ่มจำนวนโพรเซสเซอร์หรือขนาดของปัญหาให้ใหญ่ขึ้น
- v) ____ ระบบที่มีค่า Isoefficiency function มากกว่าจะขยายตัวได้ดีกว่าระบบที่มี Isoefficiency function น้อยกว่า
- w) ____ เราสามารถคงค่า speedup ได้โดยการเพิ่มขนาดของปัญหาและจำนวนของโพรเซสเซอร์ไปพร้อมๆกัน
- x) ____ ค่า throughput ของระบบคืออัตราส่วนของการสื่อสารต่อการคำนวณ
- y) ____ Utilization หมายถึง จำนวนของงานที่ประมวลผลต่อหน่วยเวลา

2.3 จงระบุเป้าหมายในการทำ mapping ที่ขัดแย้งกันสองเป้าหมายพร้อมอธิบายเหตุผล (2 คะแนน)

2.4 ภาพต่อไปนี้แสดงขั้นตอนของ design methodology ในการออกแบบการประมวลผลแบบขนาน จงจับคู่รายการ design methodology ต่อไปนี้ (13 คะแนน)



- A. การจัดสรรงานแบบ static และ dynamic
- B. ลดการคำนวณและการจัดเก็บข้อมูลที่ซ้ำซ้อน
- C. จัดการสื่อสารระหว่างงานต่างๆให้สมดุล
- D. เพิ่ม locality ของอัลกอริทึมแบบขนาน

ชื่อ _____ รหัสนักศึกษา _____

- E. แบ่งงานย่อยให้มีขนาดเท่าๆกัน
 - F. แต่ละงานสื่อสารกับกลุ่มเพื่อนบ้านจำนวนไม่มาก
 - G. แทนที่การสื่อสารด้วยการคำนวณที่ซ้ำซ้อนที่ใช้เวลาซีพียูน้อยกว่าเวลาที่ใช้ในการสื่อสารนั้น
 - H. การทำซ้ำข้อมูลไม่มีผลต่อ scalability
 - I. การสื่อสารของงานต่างๆสามารถเกิดขึ้นได้พร้อมๆกัน
 - J. จำนวนของงานเพิ่มขึ้นเมื่อขนาดของปัญหาเพิ่มขึ้น
 - K. การออกแบบให้มีจำนวนงาน 1 งานต่อโพรเซสเซอร์ และหลายงานต่อโพรเซสเซอร์
 - L. กำหนดจำนวนงานที่เหมาะสมให้กับระบบเป้าหมาย
- a) Partitioning

b) Communication

c) Agglomeration

d) Mapping

2.5 จงวาดภาพที่เกี่ยวข้องกับ Amdahl Laws ในรายการต่อไปนี้ พร้อมทั้งแสดงตัวแปรและข้อมูลของแกนต่างๆให้ชัดเจน (10 คะแนน)

- a) จงแสดงกราฟ scalability ของระบบแบบขนานที่แสดงให้เห็นถึงความสัมพันธ์ของการเปลี่ยนแปลงของ speedup จำนวนโพรเซสเซอร์ และขนาดของปัญหา (4 คะแนน)

- b) กราฟแสดงความสัมพันธ์ของ efficiency กับจำนวนโพรเซสเซอร์ เมื่อขนาดของปัญหาคงที่ (3 คะแนน)

ชื่อ _____ รหัสนักศึกษา _____

- c) กราฟแสดงความสัมพันธ์ของ efficiency กับขนาดของปัญหา เมื่อจำนวนโพรเซสเซอร์ คงที่ (3 คะแนน)

2.6 งานประยุกต์หนึ่งมีอัตราการเติบโตเป็น g และให้เวลาที่ใช้ในการประมวลผลแบบขนานเป็น T_p ซึ่งเป็นฟังก์ชันของขนาดของปัญหา overhead และจำนวนของโพรเซสเซอร์ จงแสดงการคำนวณหาค่า Isoefficiency Function ของงานประยุกต์นี้ (5 คะแนน)

