

บทที่ 4

ระบบหน่วยความจำ

หน่วยความจำคืออุปกรณ์ที่สามารถจำค่าลอจิกซึ่งเขียนลงไปไว้ได้ และสามารถถูกอ่านค่าที่มันจำค่าไว้ขึ้นมาใช้ใหม่ได้เมื่อใดก็ได้ตามที่ต้องการ โดยปกติหน่วยความจำของระบบคอมพิวเตอร์แบ่งออกเป็นสองส่วน คือหน่วยความจำของระบบหรือ Main memory และ Storage memory โดยหน่วยความจำหลักของระบบมักจะสร้างขึ้นจากวัสดุสารกึ่งตัวนำ ส่วนหน่วยความจำที่เป็น Storage memory มักสร้างขึ้นจากเทคโนโลยีของแผ่นจานแม่เหล็ก เช่น ฮาร์ดดิสก์ เป็นต้น

จากที่ได้กล่าวมาแล้วในบทที่ 1 ว่าเครื่องคอมพิวเตอร์ทั่วไปทำงานบนหลักการ Stored program concept นั่นคือเก็บโปรแกรมและข้อมูลไว้ในหน่วยความจำและทำการเข้าถึงข้อมูลในโปรแกรมและหน่วยความจำโดยการอ้างถึงตำแหน่ง ในทางหลักการแล้วหน่วยความจำที่เก็บโปรแกรมและข้อมูลนี้จะเป็นหน่วยความจำหลักหรือดิสก์ก็ได้ แต่เนื่องจากดิสก์เป็นอุปกรณ์ทางแมคคาณิกซึ่งมีการทำงานที่ช้ามากเมื่อเทียบกับหน่วยความจำที่เป็นสารกึ่งตัวนำ ดังนั้นในทางปฏิบัติแล้วก่อนที่จะรันโปรแกรมใดๆ ซีพียูจะอ่านโปรแกรมเข้ามาเก็บในหน่วยความจำหลักเสียก่อน ในสไลด์หน้าที่ 4 แสดงให้เห็นถึงโครงสร้างของการต่อหน่วยความจำหลักเข้ากับโพรเซสเซอร์ ซึ่งจะเห็นว่าโพรเซสเซอร์ระบุค่าตำแหน่งของหน่วยความจำที่ต้องการที่รีจิสเตอร์ MAR (Memory address register) ซึ่งจะส่งข้อมูลออกมาที่แอดเดรสบัส นอกจากนี้ซีพียูยังรับ/ส่งสัญญาณควบคุมระหว่าง Control unit กับหน่วยความจำผ่านทางคอนโทรลบัส ซึ่งสัญญาณนี้ได้แก่สัญญาณ Read, Write, MFC เป็นต้น หน่วยความจำจะส่งข้อมูลตำแหน่งที่ระบุให้กับซีพียูผ่านทางดาต้าบัส โดยปกติความกว้างของดาต้าบัสจะมีขนาดเท่ากับขนาดของเวิร์ดของซีพียูนั้นๆ แต่ก็ไม่จำเป็นต้องเป็นแบบนี้เสมอไป ยกตัวอย่างเช่น ในซีพียู Pentium ซึ่งเป็นซีพียูขนาด 32 บิต แต่มีการอ่านข้อมูลจากหน่วยความจำครั้งละ 64 บิต เป็นต้น ขนาดของแอดเดรสบัสจะเป็นตัวบ่งบอกถึงปริมาณของหน่วยความจำสูงสุดที่ซีพียูจะสามารถอ้างถึงได้ ยกตัวอย่างเช่น หากแอดเดรสบัสมีขนาด 32 บิต ก็จะสามารถ อ้างถึงหน่วยความจำได้สูงสุดเท่ากับ $2^{32} = 4$ กิกะไบต์ เป็นต้น

สไลด์หน้าที่ 5 จะแสดงให้เห็นถึงโครงสร้างของหน่วยความจำประเภทสารกึ่งตัวนำทั่วๆ ไป โดยในรูปได้ยกตัวอย่างหน่วยความจำขนาด 16 ตำแหน่ง แต่ละตำแหน่งเก็บข้อมูลได้ 8 บิต หรือที่เรียกว่าหน่วยความจำขนาด 16×8 นั่นเอง เนื่องจากมีจำนวนแอดเดรสเท่ากับ 16 แอดเดรส ดังนั้นจึงต้องการความกว้างของแอดเดรสบัสเท่ากับ 4 เส้น วงจรแอดเดรสดีโคดีเตอร์จะรับค่าแอดเดรสมาถอดรหัสทำให้เอาต์พุต word line ที่ต้องการแอกตีฟ ส่งผลให้ในการอ่านนั้นหน่วยความจำเซลล์ที่อยู่ในแถวที่แอกตีฟส่งค่าออกมาให้กับวงจรตรรกะซึ่งต่อค้นเอาไว้ก่อนส่งข้อมูลลงไปที่ดาต้าบัส

หน่วยความจำไม่จำเป็นจะต้องเก็บข้อมูลตำแหน่งละ 8 บิตเสมอไป หน่วยความจำบางตัวอาจเก็บข้อมูลตำแหน่งแอดเดรสละ 1 บิตก็ได้ ซึ่งยกตัวอย่างให้ดูในสไลด์หน้าที่ 6 ซึ่งจะเห็นว่าเป็นโครงสร้างของหน่วยความจำขนาด 1024 แอดเดรส แต่ละแอดเดรสมีข้อมูลเก็บอยู่จำนวน 1 บิต หรือเรียกว่าหน่วยความจำขนาด 1024x1 โดยจะเห็นว่าโครงสร้างภายในของหน่วยความจำถูกจัดในรูปแบบของอะเรย์ขนาด 32*32 ซึ่งก็คือมีจำนวนแถวเท่ากับ 32 แถว แต่ละแถวเก็บข้อมูลได้ 32 บิต สัญญาณแอดเดรสขนาด 5 บิตจะถูกส่งเข้าวงจรถอดรหัสแถว ซึ่งก็จะทราบได้ว่าแถวที่เท่าใดแอกทีฟ และใช้สัญญาณแอดเดรสที่เหลืออีก 5 บิตในการป้อนเข้าสวิตช์ 32-to-1 multiplexer เพื่อเลือกที่จะเอาเฉพาะบิตที่ต้องการของหนึ่งใน 32 บิตของแถวที่อ่านได้ออกสู่ดาต้าบัสต่อไป

หน่วยความจำที่เป็นวัสดุสารกึ่งตัวนำสามารถแบ่งออกได้เป็นสองจำพวกคือหน่วยความจำแบบ nonvolatile และแบบ volatile โดยหน่วยความจำแบบ nonvolatile จะสามารถเก็บข้อมูลที่เขียนลงไปในตัวมันไว้ได้แม้จะถูกปลดไฟเลี้ยงให้มันก็ตาม ในขณะที่หน่วยความจำแบบ Volatile จะเก็บข้อมูลเอาไว้ได้ขณะที่ยังมีไฟเลี้ยงจ่ายให้เท่านั้น หากปลดแหล่งจ่ายไฟออกข้อมูลจะสูญหายไป

หน่วยความจำแบบ Nonvolatile จะมีหลายชนิด โดยในสไลด์หน้าที่ 7-14 แสดงหน่วยความจำแบบ nonvolatile ชนิดต่างๆ โดยเริ่มตั้งแต่หน่วยความจำ ROM ซึ่งมีโครงสร้างดังรูปในสไลด์หน้าที่ 9 ซึ่งแสดงตัวอย่างของรอมอย่างง่ายซึ่งมีขนาด 4x8 โดยวงจร 2-to-4 line decoder ทำหน้าที่ถอดรหัสสัญญาณแอดเดรสและทำให้สัญญาณ ในแถวที่เลือกมีค่าแอกทีฟที่ลอจิก 0 โครงสร้างของรอมจะเป็นการใช้ไดโอดต่อระหว่างเส้น bit line และ เส้น word line โดยหากตำแหน่งใดมีไดโอดต่ออยู่ตำแหน่งนั้นก็จะเก็บค่าลอจิกศูนย์ แต่ถ้าตำแหน่งใดไม่มีไดโอดอยู่จะถือว่าตำแหน่งนั้นเก็บค่าลอจิกหนึ่ง รูปในสไลด์หน้าที่ 10 เป็นการแสดงโครงสร้างของรอมอีกแบบหนึ่งซึ่งสร้างจากการใช้มอสทรานซิสเตอร์แทนไดโอด ซึ่งจะมีการทำงานที่ละม้ายคล้ายคลึงกัน หน่วยความจำ ROM จะไม่สามารถโปรแกรมโดยผู้ใช้ได้ ขั้นตอนการโปรแกรมจะถูกทำในระหว่างการผลิต ซึ่งทำให้รอมไม่สะดวกในการใช้งานโดยผู้ใช้โดยทั่วไปนักจึงมีการสร้างรอมแบบที่สามารถโปรแกรมโดยผู้ใช้ได้ ซึ่งเรียกว่า PROM : Programmable read only memory ซึ่งสามารถเขียนโปรแกรมได้โดยผู้ใช้ แต่สามารถทำได้เพียงครั้งเดียวเท่านั้น โดยในตัว PROM จะแทนที่ไดโอดด้วยฟิวส์ โดยการโปรแกรมจะทำการป้อนพัลส์กระแสสูงเข้าไปละลายฟิวส์กรณีที่ต้องการบันทึกลอจิก 1

หน่วยความจำ PROM แม้จะมีข้อดีเหนือกว่า ROM ตรงที่สามารถโปรแกรมได้โดยผู้ใช้แล้วแต่ก็ยังมีข้อจำกัดในแง่ของความยืดหยุ่นในการใช้งาน สาเหตุเนื่องจากว่ามันไม่สามารถโปรแกรมใหม่ได้ ดังนั้นจึงมีการพัฒนาหน่วยความจำ EPROM ขึ้นมาแทนที่ รูปในสไลด์หน้าที่ 12 แสดงให้เห็นถึงหน่วยความจำอีพรอมหรือ EPROM ซึ่งสามารถโปรแกรมใหม่ได้โดยใช้ไฟฟ้าและสามารถลบได้โดยการใช้รังสีอัลตราไวโอเลตความเข้มสูงฉายเข้าไปบนตัวชิปเป็นเวลาประมาณ 5-10 นาที โดยที่ตัวถังของอีพรอมจะทำเป็นช่องใสให้แสงผ่านได้ ดังนั้นเมื่อโปรแกรมเสร็จจะต้องหาวัสดุมาปิดช่องรับแสงนี้

ไว้เพื่อไม่ให้รังสียูวีจากภายนอกเข้าไปทำความเสียหายกับข้อมูลที่บันทึกไว้ในรูปสไลด์หน้าที่ 13 แสดงให้เห็นถึงตัวถังของอีพროม 3 เบอร์ด้วยกัน ซึ่งจะสังเกตเห็นว่าการบอกค่าความจุของหน่วยความจำจะบอกว่าหน่วยความจำตัวนั้นมีจำนวนกี่เซลล์ เช่นหน่วยความจำเบอร์ 27C256 จะเป็นหน่วยความจำที่มีขนาด 256 กิโลบิตเซลล์ โดยจัดโครงสร้างเป็นแบบ 32x8 หรือจะให้ความจุเท่ากับ 32 กิโลไบต์ หน่วยความจำทั้งสามในสไลด์หน้าที่ 13 จะเห็นว่ามีขาที่คอมแพททิเบิลกัน นั่นหมายถึงเราสามารถเปลี่ยนตัวหน่วยความจำไปใช้เบอร์ที่มีความจุสูงกว่าได้โดยไม่ต้องออกแบบลายวงจรพิมพ์ของแผงวงจรใหม่

แม้ว่าหน่วยความจำแบบอีพროมจะสามารถลบใหม่ได้โดยผู้ใช้แต่ก็จะมีข้อจำกัดหลงเหลืออยู่ นั่นคือการโปรแกรมจะต้องถอดตัวไอซีไปโปรแกรมด้วยเครื่องโปรแกรมที่ออกแบบมาพิเศษโดยเฉพาะ และยังเสียเวลาในการลบโปรแกรมประมาณ 5-10 นาที ดังนั้นจึงมีหน่วยความจำชนิดใหม่ออกมาเพื่อแก้ไขปัญหาดังกล่าว นั่นคือหน่วยความจำ EEPROM ซึ่งสามารถลบได้ด้วยไฟฟ้า และยังสามารถคงค่าข้อมูลเก็บเอาไว้ได้แม้ไม่มีไฟเลี้ยงจ่ายให้กับตัวมัน อย่างไรก็ตาม EEPROM ก็ยังมีข้อเสียอยู่ก็คือการลบข้อมูลหรือเขียนข้อมูลจะต้องใช้ค่าแรงดันไฟฟ้าที่ต่างจากค่าแรงดันปกติ

หน่วยความจำแบบนอนโวลตาไทล์แบบสุดท้ายที่จะพูดถึงก็คือหน่วยความจำแบบแฟลช ซึ่งมีคุณสมบัติคล้ายกับหน่วยความจำอีอีพโรมมาก กล่าวคือสามารถลบได้ด้วยไฟฟ้า แต่สิ่งที่แตกต่างกันก็คือหน่วยความจำแฟลชไม่มีความสามารถในการเขียนค่าลงตำแหน่งใดๆ เดี่ยวๆ ได้ตามที่ต้องการ เหมือนกับอีอีพโรม การเขียนค่าลงในหน่วยความจำแฟลชจะต้องทำครั้งละอย่างน้อย 1 บล็อกข้อมูล โดยก่อนเขียนข้อมูลในบล็อกนั้นจะถูกลบออกทั้งหมดก่อนแล้วจึงเขียนข้อมูลลงไปทั้งบล็อก ดังนั้นหน่วยความจำแบบแฟลชจะมีความยุ่งยากในการใช้งานมากกว่า แต่อย่างไรก็ตามเนื่องจากคุณสมบัติที่มีราคาถูกและสามารถลบได้โดยใช้แรงดันเดียวกับค่าแรงดันใช้งานปกติ จึงทำให้หน่วยความจำแฟลชมีความนิยมอย่างมากในปัจจุบัน ยกตัวอย่างเช่นในหน่วยความจำของกล้องถ่ายรูปหรือซิลิคอนสเตทิสก์ เช่น USB flash drive เป็นต้น ทั้งนี้เนื่องจากการอ่านเขียนข้อมูลของดิสก์โดยปกติจะทำได้ทีละอย่างน้อย 1 บล็อกข้อมูลหรือทีละ 1 คลัสเตอร์อยู่แล้ว

หน่วยความจำแบบนอนโวลตาไทล์จะใช้เก็บข้อมูลที่ต้องการให้คงค่าของข้อมูลอยู่แม้ว่าปิดไฟฟ้าให้กับระบบ ตัวอย่างการประยุกต์ใช้หน่วยความจำแบบนอนโวลตาไทล์ที่เห็นได้ชัดเจนที่สุด ได้แก่ การใช้ในการเก็บโปรแกรมไบออส (BIOS) ในเครื่องไมโครคอมพิวเตอร์ ซึ่งผู้ผลิตเมนบอร์ดมักจะใช้หน่วยความจำแฟลช หรืออีพโรมในการจัดเก็บไบออสเป็นส่วนใหญ่ สาเหตุสำคัญที่ต้องเก็บโปรแกรมไบออสไว้ในหน่วยความจำแบบนอนโวลตาไทล์เพราะทุกครั้งที่เปิดเครื่องจะต้องมีการรันโปรแกรมไบออสเพื่อตรวจเช็คทรัพยากรที่ต่ออยู่บนเมนบอร์ดเสียก่อนที่จะบูตเพื่อนำเอาระบบปฏิบัติการเข้ามาไว้ในหน่วยความจำหลักต่อไป

หน่วยความจำแบบอนโวลไทล์ จะไม่สามารถเก็บข้อมูลไว้ในตัวมันได้หากปลดกระแสไฟฟ้าออกจากตัวมัน รูปในสไลด์หน้า 15 แสดงให้เห็นถึงโครงสร้างการทำงานของหน่วยความจำแบบ SRAM หรือหน่วยความจำแบบสแตติก ซึ่งประกอบไปด้วยโครงสร้างพื้นฐานจำนวน 6 ทรานซิสเตอร์ด้วยกัน ในการอ่านหรือเขียนข้อมูลจะทำได้โดยการสั่งให้เส้น word line มีค่าเป็นลอจิก 1 และทำการส่งหรือรับข้อมูลออกหรือเข้าสู่หน่วยความจำผ่านทางเส้น bit line หน่วยความจำแบบสแตติกจะมีข้อดีคือมีความเร็วในการทำงานที่สูงมาก แต่ข้อเสียของมันคือต้องการทรานซิสเตอร์ถึง 6 ตัวต่อหน่วยความจำ 1 เซลล์ ดังนั้นจึงทำให้มีราคาที่สูงมาก ตัวอย่างการประยุกต์ใช้หน่วยความจำแบบสแตติกได้แก่การสร้างเป็นรีจิสเตอร์หรือหน่วยความจำแคชในซีพียูซึ่งต้องการความเร็วในการทำงานที่สูงมาก

หน่วยความจำแบบโวลไทล์อีกชนิดหนึ่งที่มีราคาถูกกว่าสแตติกแรมคือหน่วยความจำที่เราเรียกว่า ไดนามิกแรม ซึ่งโครงสร้างพื้นฐานประกอบด้วยทรานซิสเตอร์ 1 ตัวทำงานร่วมกับคาปาซิเตอร์ ดังแสดงให้เห็นในสไลด์หน้า 16 การทำงานของมันคือจะมีการชาร์จประจุเก็บไว้ในคาปาซิเตอร์กรณีที่มีการเก็บลอจิก 1 และดิสชาร์จคาปาซิเตอร์ทิ้งไปกรณีเก็บลอจิกศูนย์ แต่เนื่องจากคุณสมบัติของตัวเก็บประจุจะมีการรั่วไหลของประจุหลังจากหยุดการชาร์จประจุลงเรื่อยๆ จนกระทั่งประจุหมดไป ดังนั้นก่อนที่ประจุจะรั่วออกไปจนต่ำกว่าค่า threshold นั้นจะต้องมีการรีชาร์จประจุเข้าไปใหม่เสียก่อน ซึ่งเราเรียกขั้นตอนนี้ว่าการรีเฟรช (Refresh) หน่วยความจำแบบไดนามิกแรมมีข้อดีตรงที่มีราคาถูกเนื่องจากต้องการจำนวนทรานซิสเตอร์น้อย แต่ก็มีข้อเสียคือมีความเร็วในการทำงานที่ต่ำกว่าสแตติกแรม แต่อย่างไรก็ตาม ในเครื่องคอมพิวเตอร์โดยส่วนใหญ่แล้วมักใช้ไดนามิกแรมในการสร้างเป็นหน่วยความจำหลักของระบบ สาเหตุหลักเนื่องจากราคาที่ถูกลงของมันนั่นเอง

สไลด์หน้า 18-19 แสดงให้เห็นถึงการจัดโครงสร้างการอ้างแอดเดรสของหน่วยความจำแบบไดนามิก โดยทั่วไปแล้วไดนามิกแรมมีความจุสูง จึงต้องการจำนวนขาสัญญาณแอดเดรสมากเป็นเงาตามตัวด้วย เพื่อที่จะลดจำนวนขาลงจึงมีการลดจำนวนขาแอดเดรสลงเหลือแค่ครึ่งเดียวแล้วใช้วิธีการมัลติเพล็กซ์ขา ร่วมกับการใช้ขา RAS (Row Address Strobe) และขา CAS (Column address strobe) สไลด์หน้า 20-21 แสดงให้เห็นถึงตัวอย่างหน่วยความจำแบบสแตติกและแบบไดนามิกในรูปแบบต่างๆ ทั้งแบบเดี่ยวๆ หรือแบบแผงหน่วยความจำ

หน่วยความจำแบบไดนามิกจะแบ่งออกเป็นสองชนิดคือแบบซิงโครนัสและแบบอซิงโครนัสโดยไดนามิกแรมที่เราใช้เป็นหน่วยความจำหลักในปัจจุบันจะเป็นแบบซิงโครนัสหรือที่เราเรียกว่า SDRAM (Synchronous DRAM) หน่วยความจำแบบนี้จะมีโครงสร้างของลอจิกเซลล์ที่คล้ายกับไดนามิกแรมทั่วไปแต่ตรงเอาต์พุตจะคั่นด้วยรีจิสเตอร์เอาไว้ การอ่านหน่วยความจำทำได้โดยการส่งค่าแอดเดรสของแถวให้ก่อนจะทำให้ข้อมูลของแถวที่ต้องการถูกส่งให้กับรีจิสเตอร์ จากนั้นจึงส่งค่าคอลัมน์ให้กับตัวชิป จึงจะได้ข้อมูลในแอดเดรสที่ต้องการออกมา ใน SDRAM นี้ไม่จำเป็นที่จะต้องส่งค่าคอลัมน์ใหม่ทุกๆ

ครั้ง แต่จะมีวงจรคอลัมน์เคาน์เตอร์ทำหน้าที่เพิ่มค่าของคอลัมน์แอดเดรสทุกๆ สัญญาณนาฬิกาแสดงในไทม์มิงไดอะแกรมในสไลด์หน้าที27

จะเห็นได้ว่าการอ่านหน่วยความจำไดนามิกแรมนั้นจะต้องมีการส่งค่าแอดเดรสของแถวก่อน แล้วตามด้วยแอดเดรสของคอลัมน์ แต่เนื่องจากไมโครโปรเซสเซอร์โดยทั่วไปไม่ต่อหน่วยความจำในลักษณะนี้แต่ทำการส่งแอดเดรสของหน่วยความจำออกมาพร้อมกันในครั้งเดียว ดังนั้นเราจึงไม่ต่อหน่วยความจำแบบไดนามิกเข้ากับโปรเซสเซอร์โดยตรงเหมือนอย่างหน่วยความจำแบบสแตติก แต่จะต่อผ่านอุปกรณ์ที่เรียกว่าตัวควบคุมหน่วยความจำหรือ Memory controller ซึ่งจะทำหน้าที่คอยนำสัญญาณแอดเดรสจากโปรเซสเซอร์มาแปลงให้เป็นแอดเดรสของแถวและคอลัมน์พร้อมทั้งจัดการเรื่องของไทม์มิงของสัญญาณ CAS และ RAS ให้ ซึ่งตัวควบคุมหน่วยความจำนี้ก็คือไอซี North bridge บนเมนบอร์ดนั่นเอง รูปในสไลด์ 28-31 จะยกตัวอย่างของไอซี North bridge เบอร์ต่างๆ ซึ่งแต่ละเบอร์ก็จะสนับสนุนการเชื่อมต่อหน่วยความจำที่ไม่เหมือนกัน

สไลด์หน้าที 32 แสดงให้เห็นถึง memory hierarchy ของระบบหน่วยความจำของคอมพิวเตอร์ โดยทั่วไป โดยจะเห็นได้ว่ารีจิสเตอร์มีความเร็วสูงสุด รองลงมาจะเป็นหน่วยความจำแคช และตามด้วยหน่วยความจำหลักของระบบ และหน่วยความจำสำรองเช่น ฮาร์ดดิสก์ถือเป็นหน่วยความจำที่ช้าที่สุด แต่อย่างไรก็ตามหน่วยความจำที่เร็วที่สุดก็จะมีราคาสูงที่สุดด้วย เราจึงเห็นว่ารีจิสเตอร์มีให้ใช้งานน้อยมาก หน่วยความจำแคช L1 ก็มีให้ใช้อย่างมากแค่ 128 กิโลไบต์(ในซีพียู Athlon XP) และหน่วยความจำแคช L2 ก็มีประมาณ 128K-1Mbyte ส่วนหน่วยความจำหลักมีใช้ประมาณ 128-512 เมกะไบต์และฮาร์ดดิสก์มีขนาดตั้งแต่ 20 กิกะไบต์ขึ้นไป
