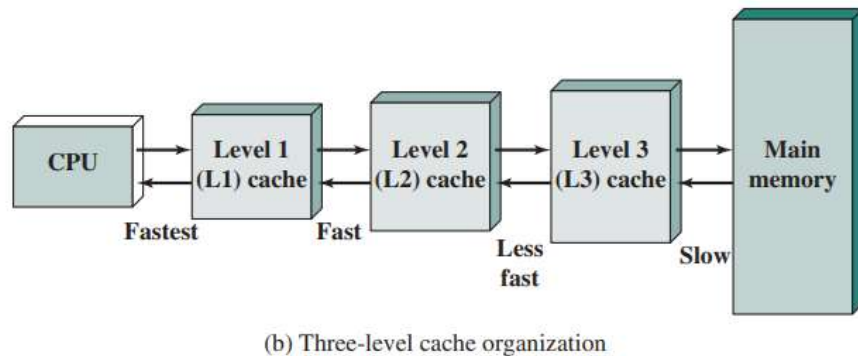
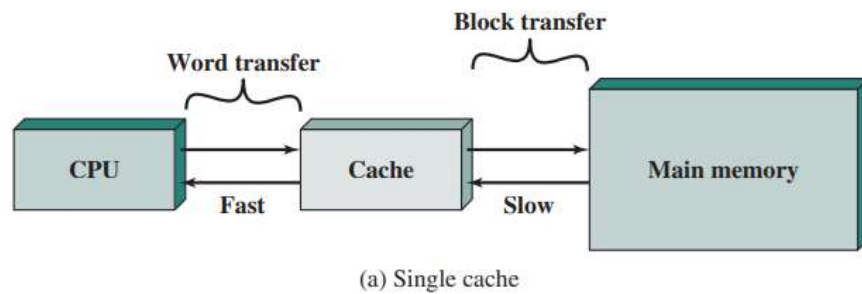


Cache Memory Principles

الـ Cache memory تم تصميمها لدمج وقت الوصول للذاكرة باهظة الثمن وعالية السرعة مع حجم ذاكرة الكبير للذاكرة وقل تكلفة والسرعة منخفضة.

مثال :



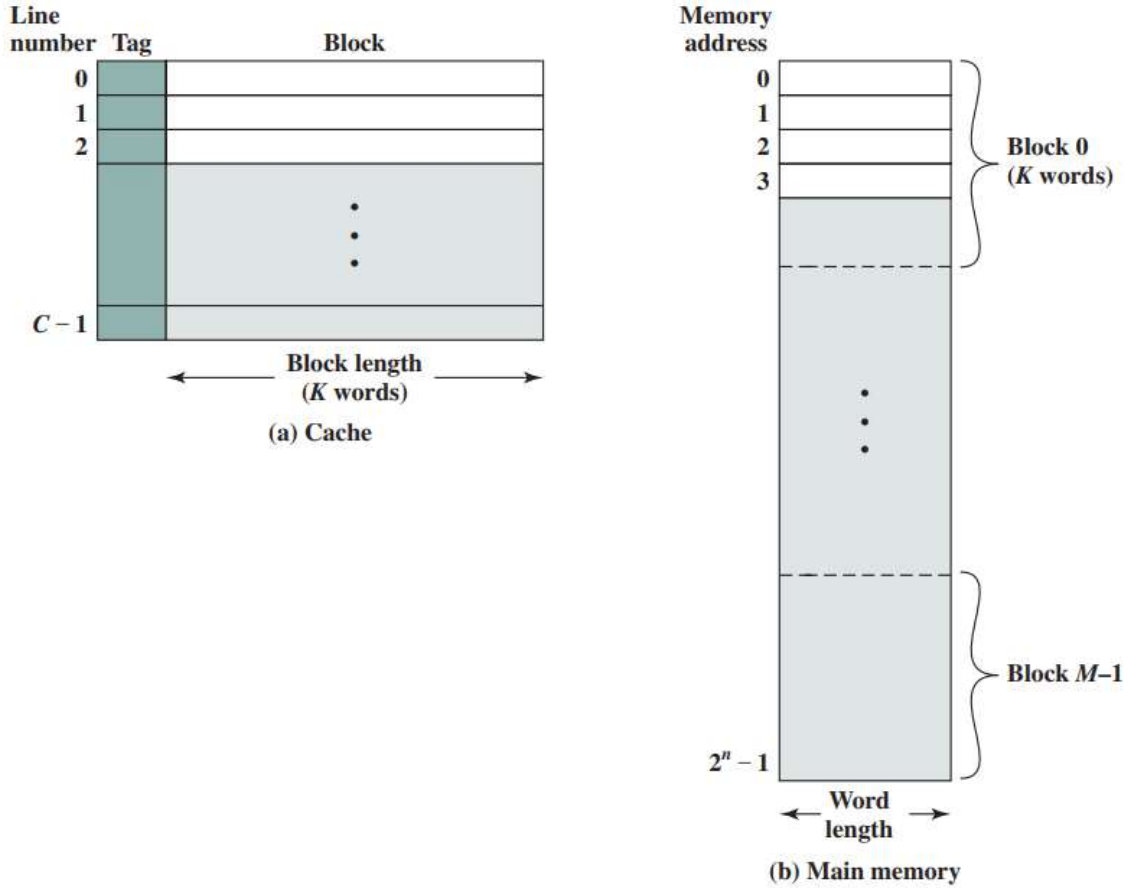
توجد ذاكرة رئيسية كبيرة وبطيئة نسبيا بالاضافة الى ذاكرة تخزين اصغر واسرع التي هي cache memory .

الـ cache تحتوي على اجزاء من الذاكرة الرئيسية .

عندما المعالج يحتاج الى قراءة كلمة اول شيء سيذهب الى الـ cache يبحث ان كانت الكلمة موجودة ام لا اذا كان كذلك سيتم تسليم الكلمة للمعالج.

اذا لم يكن كذلك الامر تتم قراءة كتلة من الذاكرة الرئيسية تتكون من عدد محدد من الكلمات في cache ثم يتم تسليمها الى المعالج.

في صورة عند (b) يتم استخدام مستويات متعددة من cache .
 الـ L2 هي ابطأ وعادة ما تكون اكبر من L1 و L3 ابطأ وعادة ما تكون اكبر من L2 .



هاذي الصورة توضح هيكل نظام Cache/main memory .
 في الذاكرة الرئيسية ما يصل الى $2^n - 1$ عنوان قابل للعنونة ولكل word عنوان فريد.

ولغرض العام في الرسم تعتبر هذه الذاكرة مكونة من عدد من الكتل ذات طول ثابت لكل منها K Words .

اي ان هناك $M = 2^n / K$ كتل في الذاكرة الرئيسية.

تتكون ذاكرة cache من كتل m تسمى lines .
 وكل line يحتوي على K Words بالاضافة الى tag مكونة من بضع بتات.

وكل line يتضمن ايضا بتات تحكم (control bits) (غير موجودة في الصورة) مثل bit to indicate يشير الى ما اذا كان الـ line تم تعديله منذ تحميله في cache .

الطول للخط بدون الـ tag و الـ control bit هو نفسه line size .

قد يصل حجم الخط الى 32 بت بحيث تكون الـ word عبارة عن بايت واحد في هذه الحالة حجم الخط هو 4 byte عدد الخطوط اقل بكثير من عدد كتل الذاكرة الرئيسية ($m \ll M$).

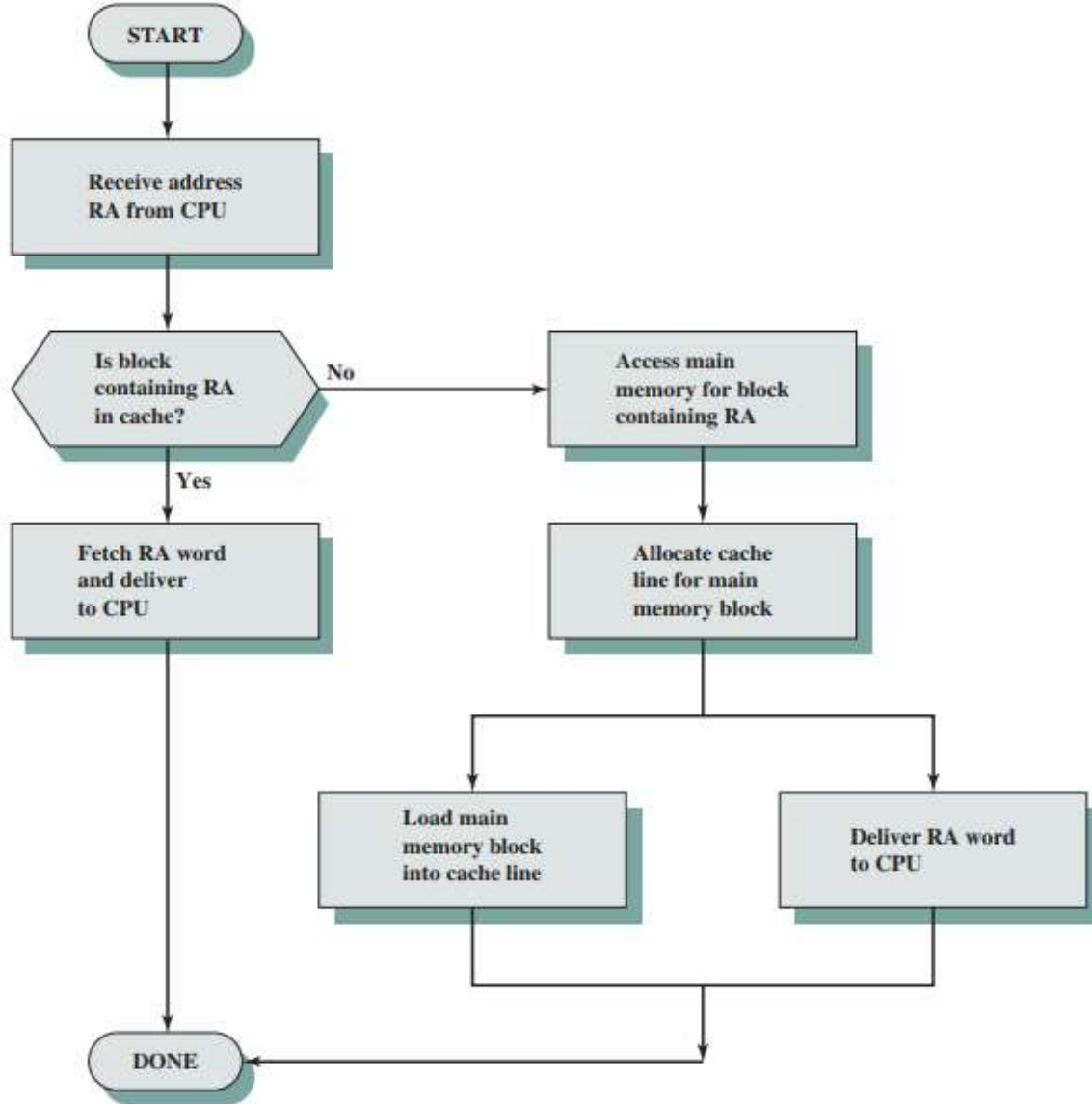
في اي وقت توجد مجموعة فرعية من كتل الذاكرة في Lines cache .

اذا تمت قراءة word من الـ block من الذاكرة فسيتم نقل تلك الكلمة الى احد الـ lines في cache .

نظرا لوجود blocks اكثر من الـ lines لا يمكن تخصيص خط فردي بشكل فريد ودائم لـ block معين.

لذلك كل Line يتضمن tag يعرف ويحدد الكتلة المعنية التي يتم تخزينها حاليا.

وعادة الـ tag تكون جزءا من عنوان الذاكرة الرئيسية كما هو موضح .



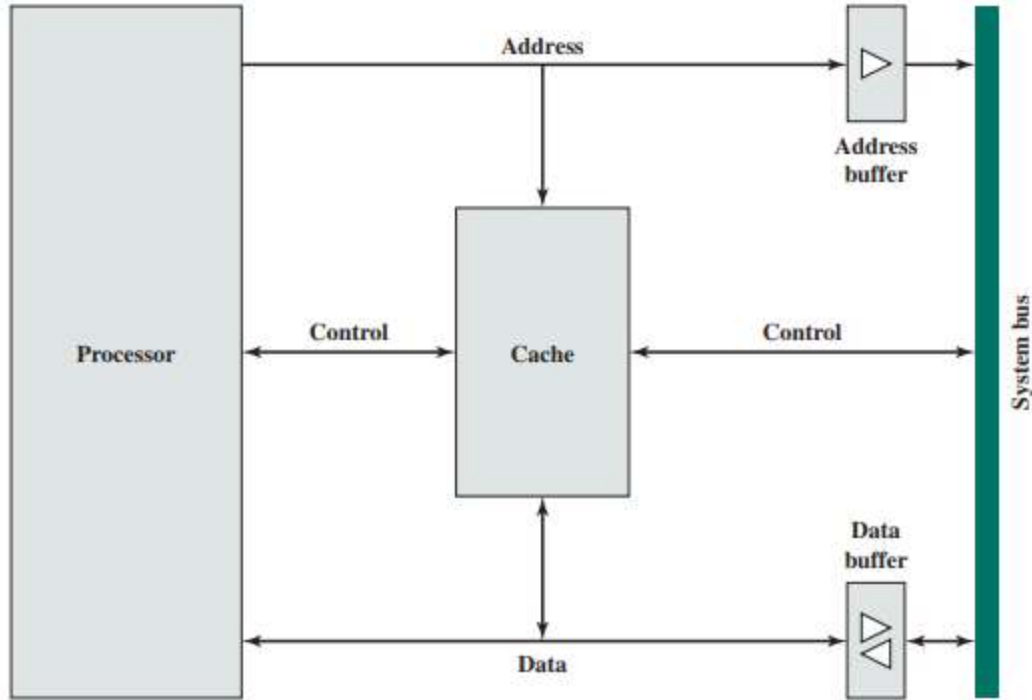
توضح هاذي الصورة عملية القراءة.

يقوم المعالج بإنشاء عنوان قراءة (RA) read address للكلمة المراد قراءتها.

إذا كانت الكلمة موجودة في الـ Cache على عيني يتم تسليمها للمعالج.

أما إذا لا؟ يتم تحميل الكتلة التي تحتوي على هاذي الكلمة في ذاكرة cache وبعدها يتم تسليمها الى المعالج.

توضح الصورة العمليتين الأخيرتين اللتين تحدثان بتوازي ويعكس التنظيم الموضح في صورة هاذي :



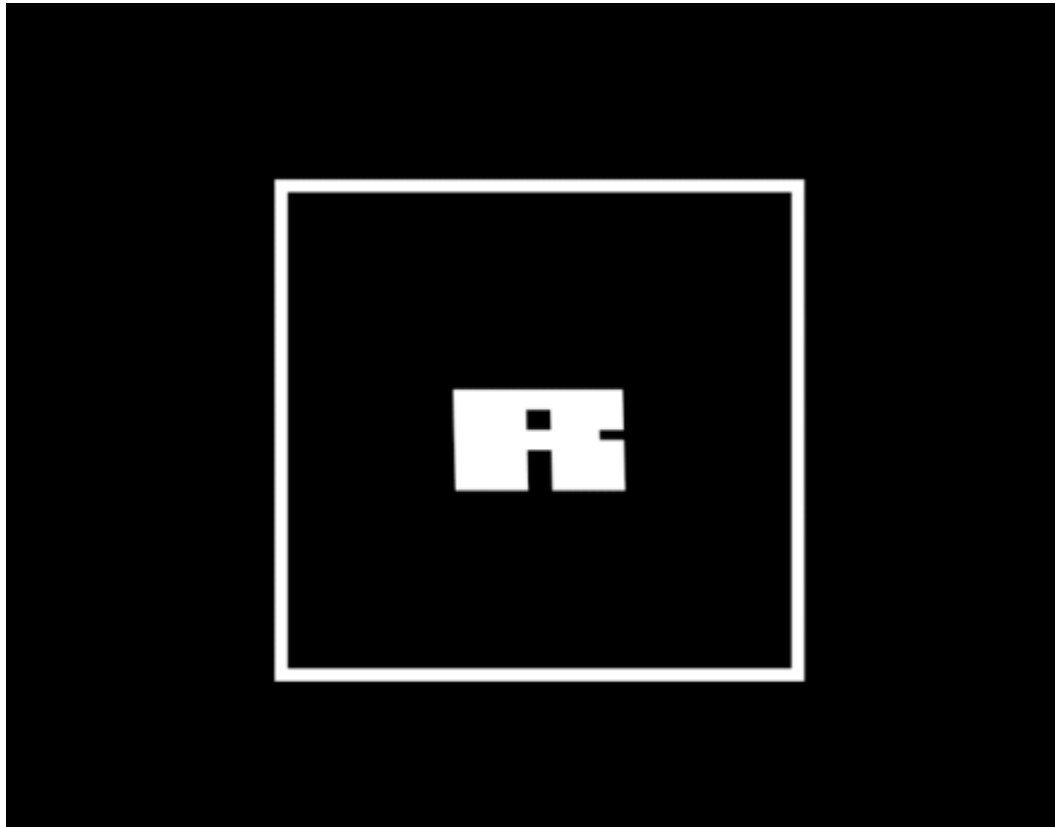
هو امر نموذجي لتنظيم التخزين cache .

في هذا التنظيم تتصل cache بالمعالج عبر data , control & address line .

والـ Address و Data خطوط تتصل ايضا بـ Data Buffer و الـ Data Buffer يتصل بـ System Bus الذي يتم منه الوصول الى الذاكرة الرئيسية منه.

عندما الـ hit cache يتم تعطيل data & address buffer ويتم الاتصال فقط بين المعالج والـ cache دون اي حركة من system bus .

وعند حدوث miss في cache يتم تحميل العنوان المطلوب على ناقل النظام ويتم ارجاع البيانات من خلال data buffer الى الـ cache & processor .



Twitter : https://twitter.com/dr_retkit

YouTube : <https://www.youtube.com/@retkit1823>