

## Le Latenze nelle Memorie RAM

Ora vedremo come avvengono le operazioni di lettura o scrittura all'interno di un **Chip di Memoria**. All'interno di un Chip di Memoria, le celle sono organizzate sotto forma di tabella, ossia sono disposte per **Righe** e per **Colonne**.

L'indirizzo richiesto dal Processore viene tradotto dal Controller di Memoria in **Indirizzo di Riga (RAS)** e **Indirizzo di Colonna (CAS)**.

Per poter accedere al dato desiderato è necessaria anzitutto l'**Attivazione della Riga** (operazione **ACTIVE**) indicata dall'Indirizzo di Riga. Solo dopo che la riga viene "attivata" è possibile accedere alle celle in essa contenute. E' possibile attivare solo una riga per volta.

Quando una riga è attiva, è possibile effettuare sulle sue celle, tutte le **Letture** e le **Scritture** (operazioni **READ** e **WRITE**) che si desiderano, specificando semplicemente l'Indirizzo di Colonna.

Nel momento in cui viene richiesta una cella che si trova su un'altra riga, è necessario la **Disattivazione della Riga** (operazione **PRECHARGE**). Solo dopo aver "disattivato" la riga, è possibile procedere all'attivazione di un'altra riga e il ciclo di accesso si ripete.

Le operazioni **ACTIVE**, **READ/WRITE** e **PRECHARGE** non sono istantanee bensì il chip di memoria impiega un certo tempo a realizzare ciascuna di esse. Durante tali tempi, chiamati **Tempi di Latenza (DELAY)**, il processore deve attendere.

Le Latenze generalmente sono espresse in **Cicli di Clock**, anziché in unità di tempo.

☞ Esempio: una RAM ha una Frequenza di 1000 MHz quindi ha un Ciclo di Clock di  $1 / 1.000.000.000 \text{ sec} = 1 \text{ nanosecondo (ns)}$ . Un Tempo di Latenza di 3 (ossia 3 cicli), in questo caso, significherebbe un tempo di attesa pari a 3 ns.

I principali Tempi di Latenza delle RAM sono i seguenti:

**tRCD (RAS to CAS delay)**: il numero di cicli che è necessario attendere dopo il comando **ACTIVE**, prima di poter dare un comando **READ** o **WRITE** (è il tempo che impiega il chip ad "attivare" la riga).

**tCL (CAS latency)**: il numero di cicli che è necessario attendere fra un comando **READ** e il momento in cui il dato è disponibile sul bus. Analogamente dopo il comando **WRITE**, è il tempo in cui si può considerare il dato effettivamente memorizzato. In ogni caso, solo dopo tCL è possibile dare un nuovo comando **READ** o **WRITE**.

**tRP (RAS Precharge)**: il numero di cicli che è necessario attendere fra il comando **PRECHARGE** e il successivo comando **ACTIVE** (è il tempo che impiega il chip a "disattivare" la riga).

Il **Tempo di Latenza più importante** è il **tCL (CAS latency)**, perché tCL è la latenza che si ripete più volte durante i normali cicli di accesso alla RAM: spessissimo infatti capita di leggere/scrivere più volte sulla riga attiva, prima di dover accedere ad una riga diversa.

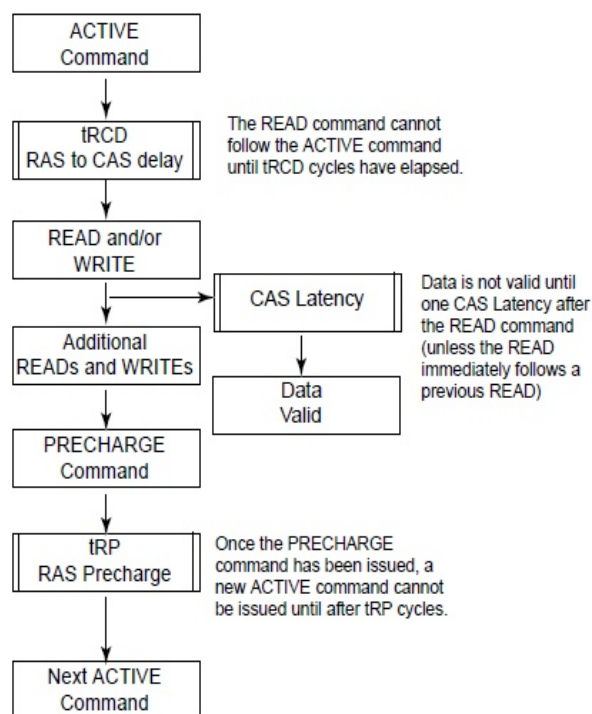
I **Tempi di Latenza**, assieme alla Frequenza, sono uno dei parametri per valutare le prestazioni di una RAM.

Generalmente tali caratteristiche vengono espresse dai costruttori con la notazione seguente:

**PCxxxx – tCL – tRCD – tRP – altre latenze...**

ove PCxxxx indica la tipologia di memoria (e, quindi, il suo clock).

☞ Ad esempio, la dicitura **PC2-6400-5-5-5-9-1T** indica una RAM di tipo **PC2-6400** (ossia **DDR2** a **800 MHz**) con **tCL** pari a **5**, **tRCD** pari a **5** e **tRP** pari a **5** e altre latenze meno importanti.



## Il Prefetch Buffer

Il **Prefetch Buffer** è un "Buffer" (ossia una "Memoria di Transito") presente nei moderni chip DDR, DDR2, DDR3 e permette un veloce e facile accesso a gruppi di dati adiacenti locati in una stessa riga fisica della memoria.

Quando viene richiesto un dato su una riga, oltre ad esso, viene letto nel buffer anche **un insieme di dati subito successivi**. Essi sono "catturati" nel buffer e vengono resi disponibili sul bus di uscita in rapida sequenza. In tal modo si evita di dover fare più singoli accessi e si rende la lettura più rapida.

✎ Ovviamente questo presuppone che la CPU richieda dati che sono "adiacenti" in memoria, ma questo avviene molto spesso, basti pensare ad una CPU a 64 bit che legge da una memoria organizzata con celle di 16 bit: ogni dato richiesto (64 bit) sono in effetti 4 celle di memoria adiacenti (4 x 16 bit)

Il numero di dati che vengono letti contemporaneamente è detto **Profondità del Prefetch Buffer** e in genere i valori tipici sono **2n, 4n, o 8n**, a seconda del tipo di memoria.

✎ Ad esempio, se il Prefetch Buffer è di tipo "4n" (come nelle memorie DDR2) e la memoria ha locazioni di 16 bit, al momento della lettura di un dato vengono in effetti lette 4 locazioni da 16 bit e riposte nel buffer.

Inoltre se in un certo tempo "internamente" vengono letti più dati, nello stesso tempo è necessario trasmetterli "esternamente", quindi **il Bus Esterno ha una Frequenza Multipla di quella del Bus Interno**.

✎ Ad esempio, se il Prefetch Buffer è di profondità 8n (come nelle memorie DDR3) e l'interno della memoria opera a 200 MHz, si ha un bus esterno che funzionerà a  $8 \times 200 = 1600$  MHz.

Più precisamente si ha che le memorie DDR hanno buffer 2n, le DDR2 hanno buffer 4n e le DDR3 hanno buffer 8n e la Frequenza di Lavoro fra DDR, DDR2 e DDR3 sale soprattutto in virtù dell'aumento della Profondità del Prefetch Buffer.

## Architettura e Standard Attuali per il sottosistema RAM

In un PC la memoria RAM è suddivisa fisicamente in uno o più moduli denominati **Moduli DIMM (Dual In-Line Memory Module)** che vengono montati in appositi slot presenti sulla scheda madre.

Su ogni modulo DIMM sono presenti **più Chip di Memoria** opportunamente saldati e organizzati in modo da gestire ciascuno una parte del dato: non dimentichiamo che il Memory Bus (Bus Dati) è a 64 bit, ossia 8 byte.

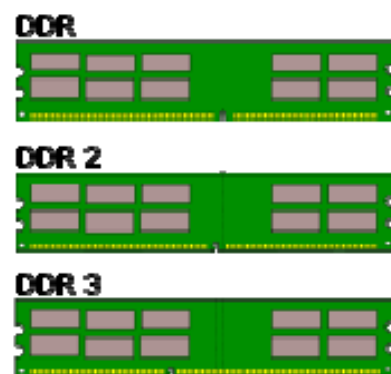
I Moduli DIMM delle diverse tipologie di RAM oggi presenti sul mercato (DDR, DDR2, DDR3) **non sono né fisicamente né elettricamente compatibili** e per questo presentano delle "scanalature" in posizioni fisiche diverse per prevenire montaggi errati.

I Moduli di tipo **DDR hanno 186 pin** mentre i **DDR2 e DDR3 hanno 240 pin**.

I comandi vengono impartiti ai moduli da un **Controller di Memoria**: esso è generalmente locato o nel **NorthBridge** della scheda madre o (nelle architetture più moderne) direttamente **Integrato all'Interno del Processore**.

Per velocizzare l'accesso alle RAM i Controller moderni supportano tecnologie quali il **Dual-Channel** o il **Triple-Channel** che distribuiscono la lettura/scrittura di un dato su 2 o addirittura 3 moduli contemporaneamente, riuscendo così a "parallelizzare" le operazioni ed aumentare la velocità complessiva della RAM.

Il **Jedec** è un organo che standardizza le caratteristiche dei moduli RAM e attribuisce un "nome ufficiale" alle varie tipologie di RAM presenti sul mercato. Ecco alcuni esempi:



Nome Standard	Tipo di RAM	Profondità Buffer	Freq. Clock Interno (MHz)	Frequenza Bus Esterno (MHz)	Trasferimenti al Secondo (MT/s)	Transfer Rate (Mb/s) (Single-channel)
PC-2700	DDR-333	2n	166	166	333	2667
PC-3200	DDR-400	2n	200	200	400	3200
PC2-3200	DDR2-400	4n	100	200	400	3200
PC2-6400	DDR2-800	4n	200	400	800	6400
PC3-8500	DDR3-1066	8n	133	533	1066	8533
PC3-10600	DDR3-1333	8n	166	667	1333	10667
PC3-12800	DDR3-1600	8n	200	800	1600	12800

Avrai notato come la **Frequenza Esterna** si può ottenere moltiplicando la *Frequenza Interna* per la *Profondità di Buffer*.

I **Trasferimenti al Secondo (MT/s)** sono il doppio della *Frequenza Esterna* perché sono tutte DDR e trasferiscono su entrambi i fronti di salita e discesa.

Infine il **Transfer-Rate (Mbyte/s)** si ottiene moltiplicando il numero di trasferimenti per 8, in quanto in ogni trasferimento viaggiano sul Memory-Bus 8 byte. Questo Transfer-Rate, in teoria, potrebbe raddoppiare o triplicare con il *Dual-Channel* o il *Triple-Channel*.

Queste sono le velocità massime teoriche, ma le velocità reali sono molto più basse a causa delle latenze e altri fenomeni.

### *Confronto fra Memorie DDR2 e DDR3*

Le **memorie DDR3** sono lo standard più recente e sono una evoluzione delle memorie DDR2. Sono sul mercato dall'anno 2007. Rispetto alle DDR2, le memorie DDR3 si differenziano per i seguenti motivi:

- Frequenza di Funzionamento **fra 800 e 1600 MHz** (le DDR2 variano fra 400 e 800 MHz)
- **Prefetch Buffer da 8 bit** (contro i 4 bit delle DDR2)
- Latenze più alte con un **CAS da 5 a salire** (contro i 3 in su delle DDR2)
- Tensione di **Alimentazione pari a 1,5V** (contro i 1,8V delle DDR2)
- **Maggiore Densità** di Memoria rispetto alla DDR2, ossia più celle di memoria nello stesso spazio
- Piedinatura da 240 pin **non compatibile** con la DDR2 e "tacca" in posizione diversa