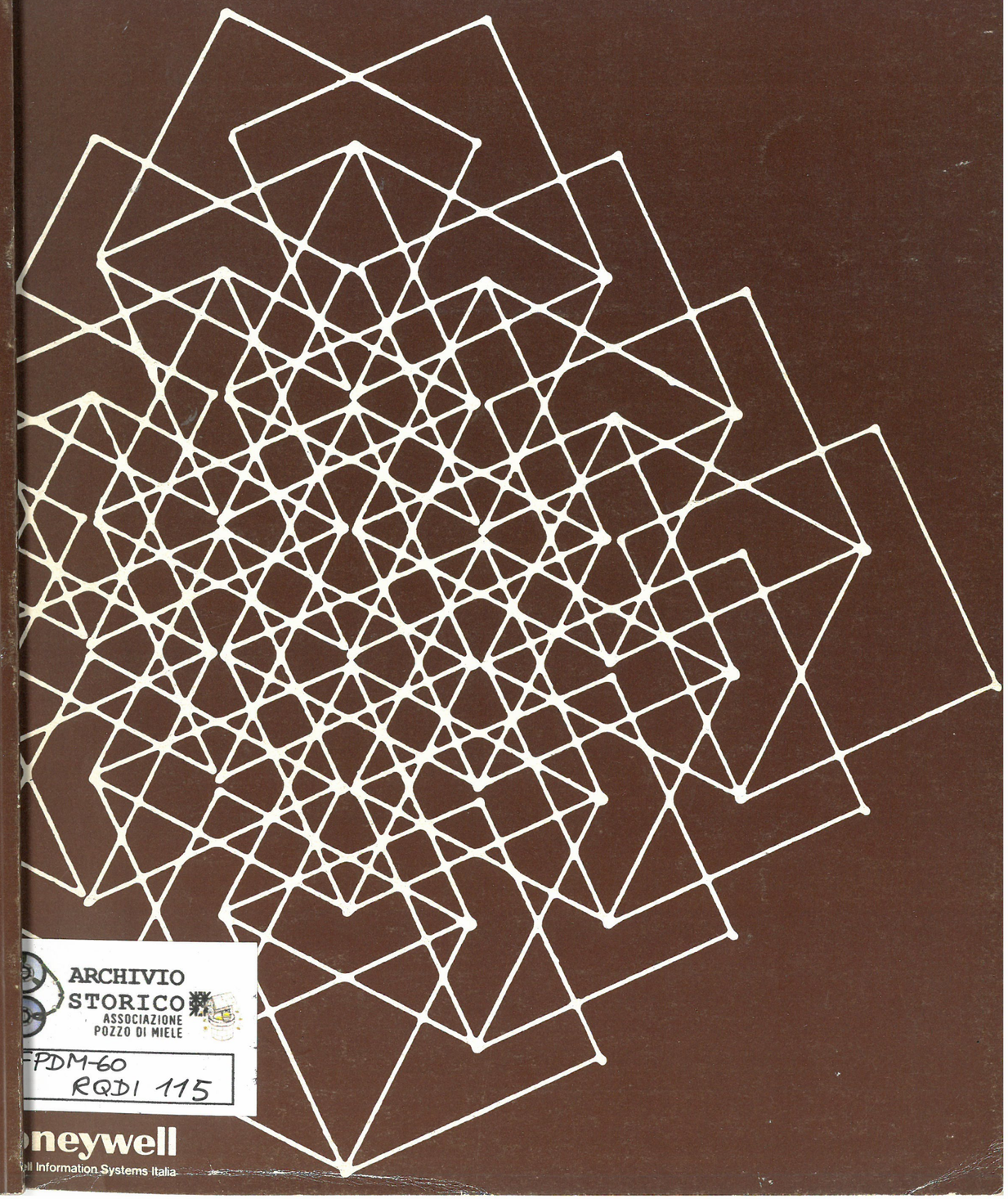
Quaderni di informatica 17

segna di tecnologia ed applicazioni degli elaboratori elettronici - Anno IX - Numero 1 - 1982



neywell
ell Information Systems Italia

FPDM-60

Honeywell

Honeywell Information Systems Italia

Quaderni di informatica

Rassegna di tecnologia ed applicazioni degli elaboratori elettronici
Anno IX - Numero 1 - 1982

Sommario

Direttore Responsabile
F. Filippazzi

Comitato di Redazione
A. Cicu, C. Falcetti, P. Lupo,
M. Nobile, G. Occhini
G. Rapelli, F. Sala

Redazione
Honeywell Information Systems Italia
Centro di Ricerca e Progettazione
20010 Pregnana Milanese (Milano)

Stampa
tipolito Maggioni s.a.s.

Autorizzazione del Tribunale di Milano
n. 93 del 20 Marzo 1974

Limiti fisici e strutturali degli elaboratori elettronici (*)

IVO DE LOTTO,

VIRGINIO CANTONI

Istituto di Informatica e Sistemistica
Università degli Studi
Pavia

1. Introduzione

La ricerca dei limiti teorici intrinseci alla tecnologia dell'elaboratore non presenta solo un rilevante interesse concettuale, ma anche finalità pratiche.

La conoscenza di tali limiti e dei fattori che li determinano costituisce infatti la base sia per prevedere l'ulteriore evoluzione del settore sia per orientare gli sviluppi tecnologici nelle direzioni più opportune.

L'argomento è stato oggetto ormai di analisi dettagliate, tra cui recentemente quella di R.W. Keyes [1], che riprende studi precedenti [2,3,4] alla luce degli ultimi sviluppi tecnologici.

Questo articolo riassume lo stato delle conoscenze sul tema, collegandolo alle linee di tendenza della tecnologia e dell'architettura dell'elaboratore.

2. Energia dissipata per bit elaborato

Elaborare dell'informazione implica trasmettere informazione tra un blocco di elaborazione e l'altro (tra un circuito e l'altro) e in questa operazione esiste il ben noto limite fissato da Shannon [5,6]:

$$C = B \lg_2 \left(1 + \frac{P}{N}\right) \text{ bit/sec} \qquad (1)$$

(*) Questo articolo è stato tratto da una relazione invitata degli Autori al Convegno Annuale AICA 1981.

dove B è la banda del canale in herz, P è la potenza del segnale e N la potenza del rumore. Si osservi che esiste una capacità C non nulla di trasmettere senza errore informazione anche per $P \ll N$. L'energia per bit trasmesso è: (J/bit)

$$W_o = \frac{P}{C} = P/[B \lg_2 (1 + \frac{P}{N})] \qquad (2)$$

che per $P \ll N$ diventa, ponendo $N = BKT$:

$$W'_o = \lim_{P \to 0} W_o = KT \cdot \lg_e 2 \qquad (2')$$

Un limite simile si trova calcolando la dissipazione di energia necessaria per scrivere 1 in una cella di memoria, considerando questa operazione come un passo di elaborazione elementare dell'informazione; R.W. Landauer [7], dopo aver mostrato che ogni elaborazione dell'informazione è realizzata con componenti nonlineari, mostra che è irreversibile e quindi dissipativa e calcola la dissipazione minima necessaria per l'operazione citata. L'energia per bit data dalla (2') è molto piccola, dell'ordine di $4 \cdot 10^{-21}$ J, ma viene quasi raggiunta in alcuni casi di trasmissione dell'informazione (nei segnali mandati dal Voyager da Giove [8], alla temperatura del rumore del ricevitore di 28.5K, $W_o' \simeq 2 \cdot 10^{-21}$ J $\simeq 6 \cdot KT$) e di elaborazione in essere viventi (batteri che rivelano la direzione del campo magnetico terrestre con poche particelle di magnetite, con energie dell'ordine di KT [9] o elaborazioni elementari del sistema nervoso di esseri viventi che dissipano dieci o venti KT [10]). Nella realtà degli elaboratori di informazione che ci sono noti, nei quali le transizioni tra uno stato e l'altro avvengono molto velocemente e quindi lontano dalle condizioni di minima energia dissipata, si è assai lontani da questi limiti teorici (circa 10^8 KT per circuiti elettronici, qualche decina di KT per i componenti logici a effetto Josephson).

La ragione sta nel fatto che per evitare che l'informazione si degradi durante il processo di elaborazione i segnali elettrici che fungono da supporto devono essere continuamente riformati, il che richiede componenti con una caratteristica statica ingresso-uscita non lineare a S. Nei circuiti a semiconduttore ciò si ottiene con escursioni di tensione di almeno alcune decine di KT/q,

che, com'è noto, a 300 K vale 25 mV. Escursioni grandi rispetto a KT/q sono dovute anche al fatto che la produzione industriale dei componenti elettronici non può garantire una piccolissima tolleranza in dette caratteristiche statiche, non potendo controllare esattamente i molti parametri che la influenzano (spessori, livelli di drogatura, concentrazione di stati superficiali tra isolante e semiconduttore, concentrazione delle cariche intrappolate, ecc.); inoltre nella vita delle apparecchiature in cui i componenti sono inseriti la T può cambiare nel tempo e a seconda del luogo in cui il componente è collocato, e quindi cambia anche KT/q.

Oltre alla necessità di usare escursioni di tensione di alcune decine di KT/q, ogni operazione elementare coinvolge nei circuiti elettronici molti elettroni. Infatti una transizione di livello di tensione ai capi di un condensatore di capacità C, comporta il trasferimento di N=CV/q elettroni. Essendo $V \gg KT/q$ segue che $N \gg CKT/q^2$, dove C è la capacità del componente e delle relative interconnessioni, attualmente dell'ordine del $10^{-2} \div 10^{-1}$ pF. In conseguenza per T=300 K, $N \gg 10^4$ elettroni. Escursioni più basse sono possibili alla temperatura dell'azoto liquido (77K) dove KT/q=7mV, ma intervengono altri fenomeni a limitare la prestazione limite del componente elettronico. Restando quindi alla temperatura ambiente, gli elevati valori di energia per bit rispetto a W'$_0$, sono dovuti alle escursioni elevate di tensione necessarie per garantire la formazione del segnale, cioè l'immunità da rumore con elevata probabilità, e al numero elevato di elettroni in gioco, essendo relativamente grandi le C, cioè i volumi di spazio da riempire con campo elettromagnetico per trasferire l'informazione.

Si è più fortunati [11] nel caso di componenti ad effetto Josephson dove l'energia dissipata per bit deve essere:

$$W = Li^2 \gg L\frac{KT^2}{\Phi_o^2} \qquad (3)$$

dove L è l'induttanza del circuito che genera il campo magnetico (l'equivalente della capacità nei componenti a semiconduttore), i è l'escursione di corrente necessaria per garantire l'immunità da rumore (equivalente della V nella precedente analisi), $\Phi_o = h/2q = 2 \cdot 10^{-15}$ Vs è il quanto di flusso magnetico, misurato in Volt·secondo (equivalente della carica elementare q). Un'induttanza dell'ordine di $L_m = \Phi_o^2/KT \simeq 10^{-7}$h a 3,5 K è fisicamente fattibile, per cui pochi quanti di flusso sono sufficienti; in questa situazione l'Eq.3 mostra che in questi dispositivi si è assai prossimi al limite teorico dato dall'Eq.2'.

3. Potenze dissipate

I circuiti logici a semiconduttore abbiamo visto che dissipano per ogni elaborazione elementare più di 10^8KT Joule, cioè molto di più del limite teorico dato dalla Eq.2'. Nei moderni circuiti integrati la potenza dissipata dai componenti è circa $^1/_3 \div ^1/_4$ di quella dissipata lungo le connessioni interne per trasmettere i segnali da un componente all'altro; quindi una valutazione dei limiti imposti dalla dissipazione termica può essere fatta restringendo il calcolo alla potenza dissipata nelle connessioni. L'energia dissipata per caricare una linea di lunghezza L e capacità per unità di lunghezza C_1, alla tensione V e riportarla quindi alle condizioni iniziali è, per una lunghezza $L \ll$ lunghezza d'onda della massima frequenza del segnale trasmesso:

$$E_1 = p \cdot t = C_1 L V^2 \text{ Joule} \qquad (4)$$

dove p rappresenta la potenza media e t il tempo richiesto dall'operazione.

Se Q_c è la quantità di calore che si può estrarre per unità di tempo e di superficie (~100 W/cm^2) e A è la superficie media per circuito, connessioni comprese, ovviamente:

$$E_1 < Q_c A t \quad \text{e cioè} \quad p < Q_c A \qquad (5)$$

La lunghezza L media si può calcolare in funzione della dimensione [12] lineare media del singolo circuito ($A^{1/2}$) e del grado di integrazione n (numero di circuiti elementari per chip): $L = f(n) A^{1/2}$ e quindi dalle (4) e (5) si ottiene:

$$pt^2 > (C_i V^2)^2 f^2(n)/Q_c \qquad (6)$$

che può considerarsi un vincolo di base dovuto alla fisica e alla tecnologia, non facilmente superabile. Infatti, per quanto visto, V non può ridursi molto; C_i è sostanzialmente dipendente dalla costante dielettrica del materiale semiconduttore utilizzato e non è facile ridurlo a valori minori di 2pF/cm, per un'impedenza caratteristica della linea (stripline) di circa 100Ω; f(n) per ora non sembra potersi ridurre rispetto a quanto indicato in [12].

Un secondo vincolo nasce dalla necessità di asportare il calore prodotto dal circuito integrato, attraverso un raffreddamento del suo supporto. Se Q_s è la quantità di calore asportabile nell'unità di tempo e per unità di superficie (nel raffreddamento ad aria $\simeq$ 1W/cm^2) ed S è la superficie attiva del supporto, deve risultare:

$$np < Q_s S$$

Il tempo di un'operazione logica dipende anche dal tempo di propagazione di un segnale tra un integrato e l'altro e si può esprimere in funzione della velocità di propagazione del segnale sulla scheda v_s, della lunghezza media della linea di connessione tra un integrato e l'altro, espressa attraverso un fattore m della distanza media tra integrati $S^{1/2}$ e del peso relativo k tra le trasmissioni tra integrati rispetto al totale delle trasmissioni originate in un integrato (tra circuiti dell'integrato e tra circuiti di integrati diversi):

$$t > kmS^{1/2}/v_s$$

Dalle ultime due relazioni, eliminando S, si ricava:

$$t^2 > p(n\ k^2\ m^2/v_s^2\ Q_s) \qquad (7)$$

Questa equazione rappresenta pure un limite fisico, che potrebbe essere migliorato aumentando Q_s (raffreddamento dell'integrato con sistemi più efficenti del raffreddamento ad aria), diminuendo k attraverso un'allocazione ottimale delle funzioni logiche entro i singoli integrati.

I limiti imposti dalla (6) e dalla (7) sono riportati in Fig. 1 [1], ove si è assunto, per un circuito LSI al silicio: C_i = 2pF/cm, V = 1 Volt, Q_c=100 W/cm^2, n = 1000, f(n)$\simeq$2 e con raffreddamento ad aria: Q_s = 1W/cm^2 e $km/v_s \simeq 10^{-10}$ sec/cm.

L'attuale componentistica con prodotto $p \cdot t \simeq 10^{-13}$ Joule che si pone all'estremo sinistro dell'area tratteggiata dei circuiti integrati è ancora lontana dal limite imposto dall'Eq. 6, ma non lontanissima da quello imposto dall'Eq. 7.

4. Interconnessioni interne ed esterne ad un circuito integrato

È noto come con l'aumentare del grado di integrazione n, aumentino notevolmente i problemi per le connessioni interne ad un circuito integrato e quelli relativi alle comunicazioni tra circuiti integrati. I secondi trovano una parziale soluzione nelle tecniche di multiplexing e nell'allocazione ottimale delle funzioni logiche nel circuito integrato, anche se spesso a scapito del tempo di comunicazione tra i circuiti integrati e quindi delle prestazioni globali del sistema; per i primi l'unica soluzione sembra essere quella delle interconnessioni su più strati.

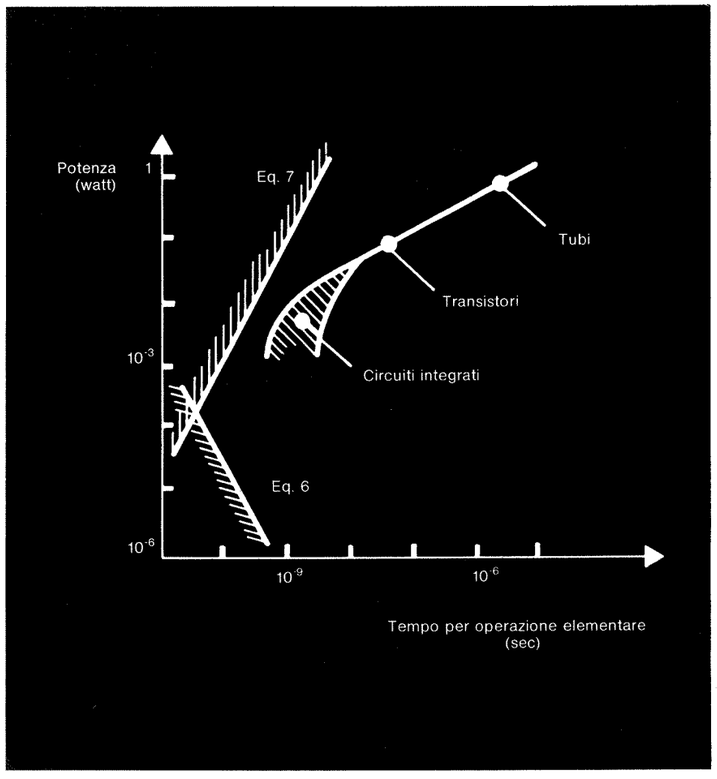


Fig. 1 – Velocità e potenza dissipata dei dispositivi elettronici per l'elaborazione digitale. Sono indicate le prestazioni delle varie tecnologie nonchè i limiti risultanti dall'analisi teorica (equazioni 6 e 7).

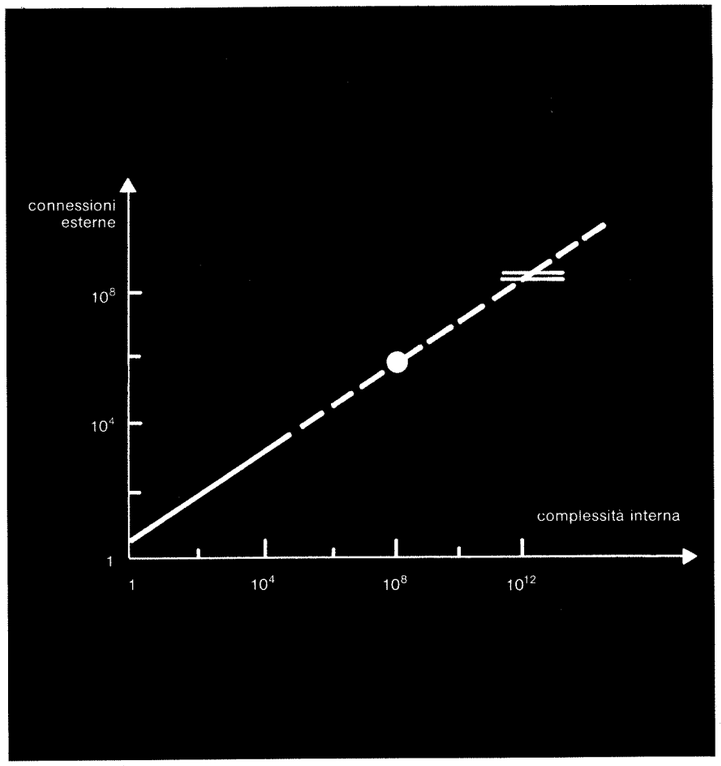


Fig. 2 - Correlazione tra complessità interna e numero di connessioni esterne. Il tratto continuo si riferisce a blocchi funzionali generici (regola di Rent); il tondino al numero di ricettori dell'occhio umano e alle fibre del nervo ottico; la zona a tratti al numero di fibre del corpo calloso ed al numero di neuroni di un emisfero del cervello umano.

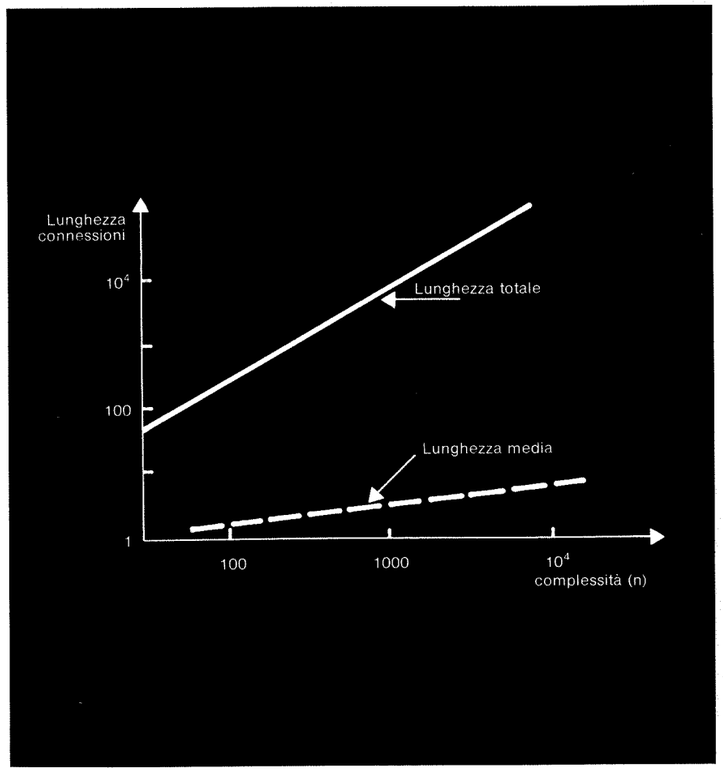


Fig. 3 - Lunghezza media della connessione tra due circuiti elementari e lunghezza totale delle interconnessioni in un circuito integrato in funzione della complessità di quest'ultimo.

A titolo informativo si ricorda la regola di Rent [13, 14], Fig. 2, che dà un'indicazione del numero di piedini di un circuito integrato in funzione del numero di circuiti elementari contenuti nell'integrato stesso.

Il tratto continuo descrive la regola di Rent quale risulta sperimentalmente per blocchi funzionali generici, fino a poco più di 10^4 circuiti elementari per blocco; il tondino il numero di ricettori dell'occhio umano e le fibre del nervo ottico; la zona a tratti in alto il numero di fibre del corpo calloso che unisce le due parti del cervello umano e il numero di neuroni di una di queste parti.

Nei limiti di validità di questa regola, si verifica facilmente che essa rappresenta una reale difficoltà al progresso dell'integrazione e, indirettamente, alle prestazioni degli elaboratori. Il problema è assai meno drammatico per quegli elaboratori che possono essere collocati in un solo integrato e per quelle funzioni omogenee e complete che richiedono un ridotto flusso di informazione con l'esterno.

Per le connessioni interne ad un circuito integrato si ricordano i dati forniti da Heller [12], Fig. 3, che riguardano la lunghezza media della connessione tra due circuiti elementari espressa come multiplo della distanza media tra circuiti elementari, e la lunghezza totale delle interconnessioni, espressa nella stessa unità di misura, in funzione della complessità n del circuito integrato (numero di circuiti elementari per circuito integrato). Si trova che per un circuito integrato [16] di area di 0.57 x 0.57 cm^2 e $n \simeq 1.500$, la lunghezza totale delle connessioni è di 400 cm pari a $6 \cdot 10^5$ volte la loro larghezza (la larghezza media è di 6.5μ), cioè quasi tutta la superficie del circuito è occupata da connessioni, anche se situazioni meno drammatiche si hanno per strutture logiche più regolari [19].

La regola sperimentale data in Fig. 3 è tuttavia meno vincolante della regola di Rent, potendosi prevedere soluzioni tecnologicamente possibili, quali connessioni a più livelli (polisilicio, doppio strato metallico, siliciuri metallici, ecc.), dimensioni trasversali contenute, ecc.

5. Dimensioni e altri fattori limite nell'integrazione

Un'analisi esaustiva di questi argomenti richiederebbe molto più spazio di quello disponibile. Si rimanda quindi alla varia letteratura sull'argomento [19-22]. Nei circuiti di laboratorio si raggiungono già dimensioni di 1.5μ con registrazioni di 0.3μ con metodi ottici, di 1μ e di 0.2μ con un fascio di elettroni. Dimensioni più ridotte possono ottenersi con un fascio di raggi X. In realtà queste dimensioni non sembrano essere un limite immediato per la produzione di circuiti logici integrati, essendo più importanti quelli posti dal rendimento della produzione, legato alla densità dei difetti e ad altri fattori. Le dimensioni previste per una memoria a semiconduttore di 256 Kbit sono di circa 40 mm^2 di area attiva, corrispondente a celle di circa $160\mu^2$ contro gli attuali $1000\mu^2$ e più delle memorie di 16 Kbit.

La miniaturizzazione dei componenti, cercata per ridurre i tempi di commutazione (riduzione di C_i dell'equazione 4) e l'energia per elaborazione elementare, pone però alcuni problemi, che possono diventare dei limiti fisici, oltre a quelli sopra ricordati. Lo spessore della giunzione, come è noto, è dato da:

$$d = (2\varepsilon\ V/nq)^{1/2}$$

con ε la costante dielettrica del mezzo, V la tensione ai capi, q la carica dell'elettrone e n la densità degli atomi droganti (donatori o accettori). Ridurre d significa sostanzialmente aumentare n e controllare le dimensioni del volume drogato con precisione. Il numero globale di droganti in un volume pari a d^3 è allora:

$$M = nd^3 = (2\varepsilon\ V/q)\ d$$

cioè diminuisce con d, potendo rendere importanti le fluttuazioni poissoniane di n. Le impurezze infatti sono introdotte a caso nel reticolo del semiconduttore e il loro numero M fluttua con una distribuzione poissoniana, con deviazione standard $M^{1/2}$. L'errore relativo diventa importante quando M diventa piccolo, o quando i componenti su un integrato sono migliaia e quindi anche le code della distribuzione possono farsi sentire. Fluttuazioni di M incidono sulla tensione di soglia dei transistori ad effetto di campo, sulla tensione di rottura delle giunzioni, ecc. Esse quindi contribuiscono alla dispersione delle caratteristiche dei componenti, che si riflettono ad esempio sui valori alti del livello del segnale rispetto a kT/q.

Dimensioni ridotte a pari livelli di tensione significano anche campi elettrici più intensi. Essi possono accelerare gli elettroni (elettroni caldi) che possono sfuggire al semiconduttore ed essere intrappolati nell'isolante che lo circonda (SiO_2) con effetti negativi sulla tensione di soglia dei FET. Essi inoltre, se acquistano sufficiente energia, possono innescare una scarica a valanga. In sostanza [22], assegnata una tensione massima, per i fenomeni suddetti, risultano fissati il minimo d e quindi la minima dimensione del componente.

Altro limite alla miniaturizzazione è la resistenza delle connessioni interne del circuito integrato. Per quanto detto al paragrafo 4, esse aumentano in lunghezza media relativa all'aumentare della complessità n e quindi in valore assoluto diminuiscono meno di *l*, la dimensione lineare del circuito logico elementare, mentre la loro sezione decresce

con l^2. La resistenza della connessione ($R=\rho L/s$) aumenta in modo più che proporzionale all'inverso di l. In generale la lunghezza della connessione L è assai inferiore alla lunghezza d'onda della massima frequenza del segnale, e quindi può essere vista come un circuito a parametri concentrati. All'aumentare della miniaturizzazione la connessione può trasformarsi da una linea LC a bassa attenuazione ad una ad alta attenuazione, non appena la dimensione minima dovesse ridursi sotto il micron [4]. L'attenuazione e le distorsioni potrebbero essere un'ulteriore difficoltà nella realizzazione di circuiti integrati su larghissima scala.

Un altro fenomeno contribuisce a rendere delicate le interconnessioni quando si va a piccolissime dimensioni: l'elettromigrazione prodotta dalle alte densità di corrente. La soluzione, come già detto, può trovarsi nelle connessioni su più strati e nel far funzionare i componenti a bassa temperatura (resistenza ed elettromigrazione più basse).

Il raggiungimento di bassi livelli di energia per operazione elementare ($\simeq 10^{-13}$J), rende sensibili gli apparati elettronici di memoria (RAM) agli effetti ionizzanti dei raggi cosmici e di particelle nucleari (α e protoni) con cammino medio nel silicio di qualche μ, ed energie paragonabili a quelle citate (10^{-13} J$\simeq$1 Mev). Le cariche liberate dalla ionizzazione prodotta dalle particelle nucleari possono cambiare il contenuto di carica nel condensatore elementare di una RAM, e falsare quindi il contenuto della memoria. Questi effetti hanno cominciato a diventare importanti con le memorie a 16 Kbit e a 64 Kbit, per le quali le energie in gioco per bit sono dell'ordine sopra indicato. L'inconveniente si ovvia con ridondanza nell'informazione immagazzinata (codici correttori d'errore) e con la scelta di materiali metallici e plastici a bassa radioattività naturale.

Quanto finora detto è stato riferito al Si come semiconduttore principe per la realizzazione dei circuiti integrati. Non sembra d'altra parte che altri materiali abbiano caratteristiche tali da poter sostituire il Si per le applicazioni digitali. L'AsGa ha alcuni vantaggi per una elettronica veloce, (registri a traslazione, contatori, moltiplicatori, convertitori A/D con frequenze di clock di alcuni GHz) che sembrano però non ancora sufficienti per la realizzazione di calcolatori general-purpose, anche se giustificano particolari applicazioni militari, nei sistemi per elaborazione segnali e nei sistemi per la trasmissione dell'informazione.

Più complesso diventa il confronto se si va alle basse temperature. Alla temperatura dell'azoto liquido (77K) i componenti al silicio sono ancora competitivi, con alcuni vantaggi: minori livelli di tensione, essendo $kT/q\simeq 7$ mV e quindi minori potenze dissipate e limiti meno stringenti posti dagli elettroni caldi e dalle scariche a valanga; maggiore mobilità dei portatori di carica e quindi minori resistenze nelle connessioni; elettromigrazione più contenuta; correnti inverse di giunzione più basse e quindi frequenze di rinfresco delle memorie dinamiche più basse e conseguente minor dispendio di energia. Ma altri limiti possono diventare rilevanti, quali le tolleranze di fabbricazione, le fluttuazioni della densità delle impurezze, i fenomeni di tunneling attraverso le barriere di potenziale, ecc.

Alla temperatura dell'elio liquido (4,2 K) domina il fenomeno della superconduttività e il confronto va fatto con i componenti ad effetto Josephson. La superconduttività presenta i seguenti vantaggi: campi elettrici molto deboli, livelli di tensione molto bassi ($KT/q=0.4$mV a 4.2 K), praticamente nessun limite per la potenza dissipata, utilizzo di leghe al posto di semiconduttori, senza problemi per fluttuazione di "piccoli" numeri di impurezze e per intrappolamento di cariche. Vi sono tuttavia alcuni inconvenienti legati ai bassi guadagni per dispositivo, in rapporto ai guadagni possibili con i transistori, e l'intrappolamento dei quanti di flusso magnetico, fenomeno da contrapporre all'intrappolamento delle cariche nei semiconduttori.

Per quanto già visto comunque al paragrafo 2, Eq. (3), vi sono meno vincoli tecnologici per raggiungere bassi livelli di energia per operazione elementare. Poco si può dire sulla tecnologia Josephson per produrre sistemi di elaborazione dell'informazione. L'esperienza IBM al proposito potrà essere importante per gli sviluppi nei prossimi anni [23]. A titolo d'informazione, M.J. Marcus [24] afferma che un cammino critico dell'unità centrale di un IBM 3033 che con l'attuale tecnologia genera un ritardo di circa 160 nsec, con la tecnologia Josephson richiederebbe circa 12 nsec, con un guadagno quindi di circa 13 ed una potenzialità di circa 50 MIPS. Sembra anche che il rapporto prestazioni/costo, tutto compreso, sia potenzialmente favorevole al calcolatore con tecnologia Josephson anche per macchine con potenza equivalente a quelle più potenti attualmente sul mercato, nonostante i problemi del criostato e della connessa tecnologia per mantenere i circuiti alla temperatura desiderata.

Questi fatti pongono un reale interrogativo su quale sarà l'hardware dell'unità centrale dei grandi calcolatori del futuro, dato che la tecnologia Josephson sembra promettere potenze dell'ordine di 100÷200 MIPS con architetture logiche tradizionali, consumi di potenza elettrica dell'ordine della decina di Kw, spazi occupati di una decina di m^2 e miniaturizzazione dell'ordine di quelle attualmente in uso per i circuiti integrati al silicio (2.5μ di larghezza di linea).

6. Unità di memoria

Un elaboratore ha prestazioni e costi che dipendono sensibilmente dalle unità di memoria di cui è dotato. Le innovazioni degli ultimi anni sono in gran parte conseguenza dell'evoluzione delle unità di memoria. Le memorie possono classificarsi in due grandi famiglie: statiche e non statiche a seconda che ci siano o meno parti meccaniche in moto. La prima categoria comprende le memorie a nuclei, a semiconduttore, a CCD e a bolle; la seconda le memorie a disco e a nastro.

La principale caratteristica che distingue la prima categoria dalla seconda è il tempo di accesso: nelle memorie non statiche attuali esso è superiore a 20÷30 msec, e quindi con modalità di trasferimento di dati a blocchi poco adatti per una buona compatibilità con i tempi e le modalità di operare dei circuiti logici delle unità centrali. Tempi di accesso, capacità e costi per bit hanno così creato una gerarchia delle memorie: memorie rapide (cache memories) a transistori ($t_s\simeq$ decine di nanosecondi) di capacità contenuta (qualche decina di migliaia di caratteri) sono in diretto contatto con i registri dell'unità centrale; memorie a semiconduttore con buoni tempi di accesso (qualche centinaio di nanosecondi) e discrete capacità (qualche milione di caratteri) comunicano direttamente con la "cache memory" e con le memorie di massa; per queste memorie il tempo medio di trasferimento per carattere o gruppi di caratteri (di solito non superiori alla decina) coincide con il tempo di accesso e in conseguenza si hanno frequenze di trasferimento per le cache memory dell'ordine del Gbit/sec, per le memorie centrali a semiconduttore dell'ordine delle centinaia di Mbit/sec. Le memorie di massa a disco con tempi di accesso dell'ordine dei 30 msec e velocità di trasferimento di $\simeq 10^7$ bit/sec (cioè velocità di trasferimento medie di circa 10^6bit/sec per blocchi di 30.000 bit) hanno capacità che superano facilmente le decine di miliardi di caratteri; recentemente il tempo di accesso è stato ridotto di un ordine di grandezza ricorrendo a memorie tampone a CCD o a bolle magnetiche. Vi sono poi le memorie su nastro magnetico con lunghi tempi di accesso e capacità pressochè illimitata, eventualmente organizzate in sistemi automatici di archiviazione.

L'informazione è trasferita avanti e indietro tra i vari componenti di questa gerarchia, con l'obiettivo di ottenere un adeguatamento delle prestazioni e una macchina, nel suo insieme, con il massimo rapporto prestazioni/costo. L'evoluzione della tecnologia, continua e rapida, causa una revisione frequente di questa architettura. Non dimentichiamo infatti che la gestione di una memoria gerarchica e la multiprogrammazione richiedono programmi di sistema complessi e pesanti (come tempo macchina ed occupazione di risorse), la cui struttura dipende dalle caratteristiche dei vari componenti la memoria gerarchica.

Allo stato attuale non è ancora chiaro se la tecnologia delle memorie CCD ed a bolle magnetiche e le politiche commerciali dei costruttori di calcolatori abbiano raggiunto la condizione atta a garantire un inserimento di queste memorie nella gerarchia prima illustrata, tra le memorie centrali a semiconduttore e quelle a disco rigido con testine mobili. Il loro inserimento sarà tanto più utile, quanto più servirà a semplificare i problemi di gestione della memoria gerarchica e a rendere meno drastica la differenza tra prestazioni delle unità a disco e quelle delle memorie a semiconduttore, differenza che l'evoluzione della tecnologia a semiconduttore ha recentemente esaltato.

Una recente tecnologia, quella dei dischi ottici a sola lettura con elevatissima densità di informazione ($6\cdot 10^5$ bit/mm^2 contro i 10^4 bit/mm^2 degli attuali dischi a testine mobili) sembra acquistare grande importanza in quelle applicazioni dove la lettura dell'informazione è molto più frequente della sua memorizzazione ed aggiornamento (archivi di dati di vita lunga, ecc.).

I limiti fisici prima visti per i componenti a semiconduttore, valgono anche le memorie a semiconduttore. Data però la loro struttura ripetitiva, esse si prestano ad una integrazione più spinta che non i circuiti logici generici.

Non sono impossibili integrati con 1Mbit di celle di memoria, a costi ragionevoli, quindi memorie centrali a semiconduttore di notevole capacità (decine o addirittura qualche centinaio di milioni di caratteri) che causeranno sensibili cambiamenti nella gestione della memoria centrale e quindi nel software di sistema e nelle caratteristiche del software applicativo (minore attenzione per i problemi di ottimizzazione dello spazio di memoria occupata, linguaggi elevati, ecc.). Più complessa è la valutazione dei limiti nelle memorie a disco e a nastro magnetico. La densità lineare di registrazione stà raggiungendo limiti molto spinti ($>$ 400 bit/mm), con conseguenti notevoli complicazioni per la tecnologia delle testine, dell'insieme supporto-testine di lettura e scrittura, dei film di materiale magnetico usato, dei circuiti di lettura e scrittura, ecc. Tuttavia i dati sulla recente e continua evoluzione, soprattutto per quanto riguarda la quantità di bit accumulabili in una unità a disco e quindi del costo per bit, sembrano indicare che si è ancora lontani dal raggiungimento di obiettivi limiti fisici e tecnologici e che quindi le unità a disco con testina mobile avranno ancora una lunga vita nel campo delle memorie di massa. I nastri, come è noto, sono sempre più destinati a fungere da unità per trascrivere l'informazione da e su un supporto facile da trasportare e da porre in magazzino. In linea essi vengono sempre più sostituiti dalle grandi unità di memoria di massa che, dopo le iniziali incertezze e la messa a punto del software di gestione, si diffondono sempre più negli impianti con grossi archivi e con problemi nella gestione del personale. Allo stato attuale, viste le densità di registrazione già raggiunte e le velocità di trasferimento dei dati (250 bit/mm e 10^7 bit/sec) non sembrano probabili, nè necessarie, grandi evoluzioni.

7. Architetture dei sistemi di calcolo

Da un lato l'evoluzione della tecnologia dei componenti con la corrispondente riduzione dei costi per funzione logica e dei tempi di propagazione del segnale e dall'altro lato la sempre più elevata richiesta di grandi potenze di calcolo, hanno spinto i costruttori di calcolatori per calcolo scientifico e per applicazioni in tempo reale a cercare soluzioni alternative alle architetture tradizionalmente usate [25]. Esistono applicazioni scientifiche, quali quelle connesse con la fisica nucleare, con la meteorologia, la sismica, la fluidodinamica, che richiedono potenze di calcolo che vanno dalle decine fino a qualche centinaio - migliaio di milioni di operazioni in virgola mobile (1000 MFLOPS): po-

tenze di calcolo cha vanno raffrontate a quelle disponibili per le macchine normalmente esistenti presso i Centri di calcolo nazionali (da qualche centinaio di migliaia a qualche milione di FLOPS); sono numeri spesso poco significativi se non vengono riferiti a precise condizioni operative; essi danno tuttavia un'idea dell'aumento della potenza di calcolo richiesto, pur mantenendo lo stesso livello della tecnologia dei componenti. Esistono pure applicazioni per elaborazioni in linea connesse con la guida ed il controllo di missili, sia per il lancio di satelliti che per altri scopi, e con l'elaborazione di segnali e immagini dove sono richieste analoghe potenze di calcolo, con in aggiunta una capacità di rispondere prontamente a sollecitazioni esterne che normalmente sono assenti nelle macchine usate per il calcolo scientifico.

Alla tradizionale soluzione di macchina di Von Neumann si sono così aggiunti i calcolatori microprogrammati e i sistemi parallelo comprendenti le architetture multicalcolatori, multiprocessori, vettoriali e pipeline. Soluzioni di questo tipo sono già ampiamente utilizzate anche nei calcolatori tradizionalmente presenti nei nostri Centri di calcolo, anche se solo come parziale modifica dell'architettura di Von Neumann per migliorare particolari prestazioni. Soluzioni più avveniristiche sono però in fase di sperimentazione per ottenere le prestazioni prima indicate, pur con una componentistica che in una macchina di Von Neumann non permetterebbe di raggiungere. A parte dovrebbe essere trattato il calcolatore basato sull'effetto Josephson perchè, date le prestazioni offerte dal componente, si potrebbero ottenere buoni risultati senza ricorrere a complicazioni eccessive dell'architettura della macchina, che si traducono sempre in complicazioni costruttive, hardware e software, e di gestione.

Si fa qui riferimento alle architetture variabili e a quelle basate sul flusso dei dati. La prima caratterizzata dal poter essere adattata alle caratteristiche dei programmi via software, la seconda dalla capacità di sfruttare al massimo il parallelismo intrinseco dei programmi senza incorrere nei problemi del controllo dell'esecuzione contemporanea di più programmi in un ambiente con risorse multiple. Ragioni di costo e di affidabilità sembrano condizionare scelte di questo tipo che, almeno da quanto dicono i risultati delle prime esperienze, offrono soluzioni con un numero di componenti più contenuto e con strutture più semplici delle soluzioni tradizionali.

I calcolatori con architettura adattiva adeguano la propria architettura all'algoritmo da eseguire attraverso l'uso di codici controllo che attivano determinate sequenze di istruzioni macchina. Questa proprietà si può ottenere con la microprogrammazione, la riconfigurabilità e l'adattamento dinamico. La microprogrammazione come strumento per adeguare l'hardware a varie funzioni attraverso modifiche software ha una lunga storia, risalendo al lavoro di M.V. Wilkes [26] del 1953 e con varie realizzazioni di interesse industriale. Con la microprogrammazione è possibile riconfigurare via software le interconnessioni tra le unità logiche elementari dell'unità centrale di un calcolatore, quali i registri, gli addizionatori, i contatori, ecc., ottenendo così un'ottimizzazione delle prestazioni per un'assegnata classe di algoritmi.

Un passo successivo è stato compiuto con le architetture riconfigurabili nelle quali è possibile modificare le interconnessioni tra varie unità funzionali della macchina, come unità centrale di elaborazione, memorie, unità di ingresso e uscita. Per questa via si aumenta la potenza di calcolo aumentando il parallelismo dei dati suddividendo l'elaboratore in una schiera di elaboratori più piccoli di potenza adeguata per gli algoritmi da eseguire [27]; minimizzando i tempi di trasferimento dell'informazione tra memorie ed elaboratori con connessioni dirette memorie-elaboratori [28-30]; calcolando diverse sequenze di operazioni matematiche con diverse strutture pipeline adatte a ciascuna sequenza [31-32]; adattando la topologia delle interconnessioni tra microcalcolatori alla struttura degli algoritmi da eseguire [33-34].

Con l'introduzione della tecnologia LSI e VLSI è diventato possibile riconfigurare anche le interconnessioni tra moduli logici, distribuendo in modo ottimale le risorse hardware disponibili tra i vari programmi in esecuzione. In altri termini le risorse hardware vengono dinamicamente organizzate in elaboratori indipendenti, con architetture adatte alle esigenze dei programmi in esecuzione (pipeline, array processor, multiprocessor, ecc.).

Queste tre manifestazioni di architettura adattiva possono trovare una sintesi nell'architettura modulare che può riconfigurare le interconnessioni ai tre livelli descritti sopra su comando del programmatore che per le varie sezioni del proprio programma richiede l'architettura più adatta. Vale a dire, una adattabilità al parallelismo delle istruzioni e dei dati, ridistribuendo le proprie risorse tra i vari programmi in modo da rendere massimo il numero di programmi elaborati dalla stessa risorsa; una adattabilità ai diversi tipi di elaborazione (pipeline, multielaboratore, elaboratore vettoriale, ecc.), riconfigurandosi dinamicamente in diversi tipi di architettura, adatti agli algoritmi in esecuzione; una adattabilità a differenti strutture di programma, con la capacità di attivare particolari insiemi di istruzioni macchina, ciascuno specializzato per particolari applicazioni.

In questo modo, a pari numero di risorse hardware [35], si può aumentare la produttività del sistema rispetto ad una architettura sia pure con elevato grado di parallelismo, ma non adattiva, estendendo i limiti raggiungibili come potenza di calcolo fino ai traguardi a cui si è fatto prima accenno senza ipotizzare costi impossibili o una componentistica attualmente non disponibile o non sufficientemente affidabile ed economica. I problemi hardware, ma soprattutto software a livello di sistema operativo, che l'architetture adattiva pone sono notevoli e non risolti e rappresentano un'interessante settore di indagine che richiama l'attenzione di molti.

Le architetture basate sulla struttura del flusso dei dati tendono a rendere possibile l'esecuzione di una istruzione quando sono disponibili i suoi operandi. Esse sono un'alternativa alle macchine che si basano sull'esecuzione sequenziale di una stringa di istruzioni. In esse non esiste un contatore che dica qual'è la prossima istruzione di sequenza, ma una coda di istruzioni pronte per l'esecuzione che attendono la disponibilità delle risorse necessarie. In tal modo si realizzerebbe in maniera naturale il parallelismo computazionale intrinseco di ogni programma istante per istante, eseguendo in parallelo tutte quelle istruzioni che sono pronte per l'esecuzione, cioè che dispongono dei propri operandi. Algoritmi da eseguirsi in tempo reale ricevono così le risorse necessarie per l'esecuzione in parallelo di tutte le istruzioni che non generano conflitti, con, sembra, un lavoro di supervisione e controllo assai limitato.

Una tale architettura, al di là dell'interesse scientifico che presenta, sembra suggerita dalla necessità di realizzare macchine con un largo numero di componenti (più del tipo LSI che VLSI) però di pochi tipi diversi, ciascuno con un elevato rapporto funzioni logiche contenute/numero di piedini dell'integrato e dall'aver a disposizione strumenti di programmazione [36-37] efficienti, che permettano un reale efficace utilizzo delle risorse. Sono noti infatti i problemi di programmazione dei supercomputer CRAY-1, CYBER 205, Burroughs BSP e dei precedenti STAR 100 e Texas Instruments ASC, la cui potenziale massima produttività viene raggiunta in particolari condizioni e piuttosto raramente [38-45].

Anche per questa architettura si pone il problema di realizzare strutture parallelo tra loro interagenti e quindi di ridurre il traffico delle comunicazioni tra le varie unità; la rete di interconnessione nelle strutture fisiche fin ora studiate, pur essendo meno critica di quella dei normali sistemi multiprocessori, ha una complessità che cresce più che linearmente con il numero di unità interconnesse e la lunghezza totale del cablaggio di interconnessione cresce con il quadrato. Il vantaggio dell'architettura controllata dal flusso dei dati è che la schedulazione e la sincronizzazione delle attività concorrenti sono realizzate a livello hardware, trattando ogni istruzione come attività indipendente, con un conseguente parallelismo molto fine, non ammissibile quando il controllo è realizzato in software o per mezzo di microprogramma. Un secondo vantaggio discende dal fatto che si conoscono regole ben definite per tradurre programmi scritti in opportuni linguaggi ad alto livello (ad es. ID [36] e VAL [37]) in codici per macchine controllate dal flusso dei dati.

I prototipi realizzati o in fase di sperimentazione presso l'Università di Manchester [40] e presso la Texas Instruments e gli studi svolti al MIT dal gruppo di J.B. Dennis mostrano che le data flow machines sono una strada percorribile, potenzialmente con notevoli risultati: una macchina con 512 elaboratori elementari e con circa 1000 integrati LSI con 100 piedini ciascuno per realizzare le necessarie interconnessioni, può raggiungere i 10^9 MIPS di picco, con circa 2 MIPS per elaboratore [42]. Risultati quindi potenzialmente notevoli con una tecnologia alla portata di questi anni.

Le innovazioni di architettura dei calcolatori possono espandere notevolmente la potenza di calcolo degli stessi, anche se non cambia sostanzialmente la tecnologia dei componenti. Le macchine per essere utilizzabili, devono essere affidabili e di costo adeguato alle prestazioni fornite e alle necessità dell'utente. La loro complessità deve quindi essere contenuta. Nelle architetture tradizionali la complessità rappresenta presto un limite difficile da superare, almeno con le attuali conoscenze.

8. Conclusioni

Questa analisi potrebbe continuare: dopo aver visto i problemi connessi con il componente e quelli con l'architettura, si potrebbero analizzare i problemi della trasmissione dell'informazione, del software di sistema, dei linguaggi, del software applicativo, della definizione dell'applicazione in modo tale che la sua automazione risulti in un concreto vantaggio per l'utente. Forse quest'ultimo punto è quello che più interessa l'utente e che assilla il venditore perchè quasi sempre non sono i MIPS a decidere il successo di un processo di automazione. Lasciando tuttavia ad altri il compito di trattare questi argomenti, molto vasti e complessi, si possono trarre ora alcune conclusioni da quanto prima esposto.

Esistono limiti fisici e tecnologici allo sviluppo dell'elaborazione dell'informazione. Alcuni di questi, a livello dei componenti, sono fondamentali, quali la necessità di dissipare energia, di utilizzare livelli di tensione relativamente elevati, di rimuovere il calore prodotto per mantenere la temperatura di funzionamento entro i limiti accettabili, essendo legati alle proprietà fisiche dei materiali e alle leggi fisiche che governano i fenomeni utilizzati. Altri limiti fondamentali discendono dalla fisica dei componenti, anche se spesso altre considerazioni, come il costo di produzione e l'affidabilità, diventano più importanti. I forti campi elettrici e i sottili spessori di dielettrico causano i fenomeni associati agli elettroni caldi, alle scariche a valanga, all'effetto tunnel; i limitati volumi in gioco rendono non trascurabili le fluttuazioni nel numero degli atomi droganti; la miniaturizzazione delle interconnessioni rende problematici gli effetti della elettromigrazione. Questi vincoli fondamentali si possono però rendere meno stringenti facendo lavorare i componenti a bassa temperatura.

Altri vincoli, sempre relativi ai componenti, nascono dalla complessità degli stessi ed è difficile classificarli come fondamentali in assenza di una adeguata teoria che li interpreti. Il problema delle interconnessioni tra circuiti inseriti sullo stesso integrato e delle connessioni tra circuiti integrati, sempre più grave quanto più elevata è la com-

plessità dei componenti e il grado di integrazione, è strettamente connesso con la capacità del progettista di trattare problemi complessi. Il limite potrà essere spostato sempre più verso nuove frontiere se le tecniche CAD (Computer Aided Design) si svilupperanno e si diffonderanno, nell'industria e nella scuola. Se tali metodi non si svilupperanno e non si diffonderanno, i componenti VLSI non saranno usati che per strutture molto regolari (memorie e certi microcalcolatori ad architettura fissa).

La produzione del software, in trent'anni di esperienza e con la realizzazione di sistemi estremamente complessi, si è convinta che l'unico modo per non essere sommersa dalla complessità dei sistemi stessi è di affrontare i problemi con metodo e ha richiesto lo sviluppo di criteri e tecniche quali il progetto gerarchico, i diversi livelli di astrazione, la programmazione strutturata, il progetto top-down, ecc. Questi criteri, anche se non formalizzati o formalizzabili in maniera precisa, sono stati introdotti quasi generalmente nella produzione con soddisfacenti risultati. Queste stesse idee sembrano fondamentali anche per il CAD per il progetto di circuiti VLSI e forse l'utente potrà così realizzare circuiti VLSI custom e rendere di pratica utilità l'idea della "fonderia di silicio" dove il progetto e la verifica del prodotto è lasciata al cliente.

Il successo di una tecnologia è legato all'economia dei grandi numeri. Nel campo dei circuiti integrati al silicio sembra che due tipi di prodotto soddisfino a questa legge: quei circuiti di controllo di vasta applicazione quali quelli per il controllo dei motori delle automobili, e quei circuiti non specializzati che vengono adattati all'applicazione dall'utente, quali i microprocessori e le memorie. I vincoli sopraricordati, uniti a quelli non fondamentali, perchè in continua evoluzione, del costo e dell'affidabilità, rendono probabile lo sviluppo della componentistica al silicio nelle seguenti direzioni [21]:

- produzione dei blocchi funzionali, già oggi realizzati, con una più elevata miniaturizzazione e quindi con prestazioni più spinte, rese migliori, costi minori e, forse, una maggiore affidabilità;
- produzione di blocchi non specializzati, con maggiori funzioni incorporate, ad esempio fino a contenere l'intera CPU di un midi o maxi elaboratore attuale;
- blocchi funzionali speciali, ad esempio suggeriti da particolari architetture, che debbano esser prodotti su larga scala.

Per la prima direzione esistono sostanzialmente solo i limiti fisici e tecnologici prima citati, per la seconda e la terza invece possono esistere anche vincoli economici di difficile valutazione legati all'evoluzione delle tecniche CAD (silicon compilers [43], librerie di celle funzionali elementari, tecniche di rappresentazione di celle elementari non legate alla tecnologia, progettazione strutturata top-down con servizi per la gestione del progetto, la simulazione, la prova e la verifica del prodotto, ecc.) e al fatto che quanto più complesso e potente è un integrato non specializzato, tanto maggiore e difficile è il lavoro di personalizzazione lasciato alla produzione del software, lavoro costoso che oltre un certo limite può far preferire integrati "custom".

Non sembra che altri semiconduttori possano competere con il silicio, troppo limitati e particolari essendo i vantaggi che alcuni di essi presentano. In particolare non pare che il GaAs possa sostituire il silicio, anche se, come già detto, ha trovato applicazioni di laboratorio e militari ove erano necessarie prestazioni particolarmente avanzate e alcune case costruttrici finanziano ricerche nel settore [44].

Un'incognita è invece la tecnologia dei componenti ad effetto Josephson, che finora è stata ampiamente coltivata soprattutto presso i laboratori IBM e impiegata in apparecchi prototipali. La sua potenzialità è notevole, ma tali sono le innovazioni tecnologiche, che si riflettono su problemi di affidabilità, manutenzione ed esercizio, che è difficile prevedere quando la tecnologia Josephson sarà matura per interessare il mercato, sia pure solo dei supercalcolatori.

Le innovazioni dell'architettura degli elaboratori indicano che molto può attendersi da esse nel senso che i limiti imposti dalla tecnologia dei componenti da soli non fissano i limiti della potenza di calcolo di un elaboratore. Dette innovazioni peró sono il frutto di molti anni di studio e di sperimentazione e non avvengono con la rapidità con cui evolve la tecnologia dei componenti. Il limite con cui devono confrontarsi è comunque quello della complessità della rete di interconnessione [41], per la quale non esistono soluzioni ottimali e molta ricerca è ancora necessaria per definire fino a qual punto deve intendersi un limite fondamentale.

9 - Bibliografia

[1] R.W. KEYES: *"Fundamental limits in digital information processing"* Proc. IEEE 69 *(1981), 267-278*

[2] R.W. KEYES: *"Physical limits in digital electronics"* Proc. IEEE 63 *(1975), 740-767*

[3] B. HOENUSIN, C.A. MEAD: *"Fundamental limitations in microelectronics" Part. 1: MOS technology,* Solid State Electronics 15 *(1972), 819-829; Part. 2: Bipolar technology,* Solid State Electronics 15 *(1972), 981-987*

[4] R.W. KEYES: *"The evolution of digital electronics towards VLSI"* IEEE J. Solid State Circuits, SC - 14 *(1979), 193-201*

[5] C.S. SHANNON: *"The mathematical theory of communication"* Bell Syst. Tech. J. 27 *(1948), 379-423*

[6] A.D. WYNER: *"Fundamental limits in information theory"* Proc. IEEE, 69 *(1981), 239-251*

[7] R.W. LANDAUER: *"Irreversibility and heat generation in the computing process"* IBM Res. Dev. J.5 *(1961), 183-191*

[8] R.E. EDELSON: *"Voyager telecommunications: the broadcast from Jupiter"* Science 204 *(1979), 913-921*

[9] J.L. GOULD: *"The case for magnetic sensitivity in birds and bees"* Ame. Scientist 68 *(1980), 256-267*

[10] C.H. BENNET: *"Logical reversibility of computation"* IBM Res. Dev. J. *17 (1973), 525-532*

[11] H. MATISOO: *"Overview of technology logic and memory"* IBM Res. Dev. J 24 *(1980), 113-129*

[12] W.R. HELLER ET AL: *"Prediction for wiring space requirements for LSI"* Proc. 14 Design Automation Conf., *New Orleans (LA) (1977), 32-42*

[13] B.S. LANDMAN, R.L. RUSSO: *"On a pin versus block relationship for partition of logic graphs"* IEEE Trans. Comp. C-20 *(1971), 1469-1479*

[14] R.L. RUSSO: *"On the tradeoff between logic performance and circuit-to-pin ratio for LSI"* IEEE Trans. Comp., C-21 *(1972), 147-153*

[15] C.F. STEVENS: *"The neuron"* Scientific Am. 241 *(1979), nr. 3 pp. 54-65*

[16] R.J. BLUMBERG, S. BREMER: *"A 1500 gate random logic LSI master slice"* IEEE Solid State Circuit J. SC-14 *(1979), 818-822*

[17] L.C. WU: *"VLSI and mainframe computers"* Spring COMPCON IEEE Dig. *(1978), 26-29*

[18] T. CHIBA: *"Impact of the LSI on high speed computer packaging"* IEEE Trans. Comp. C - 27 *(1978), 319-325*

[19] C. CLIFFORD: *"Tendenze nella tecnologia VLSI"* Elettronica Oggi, *aprile '81 33-40*

[20] V. CANTONI, I. DE LOTTO: *"Evoluzione dei sistemi per l'elaborazione dell'informazione: aspetti hardware"* Informatica nelle piccole e medie aziende, *Pavia (1979), 376-439*

[21] J.R. TOBIAS: *"LSI/VLSI building blocks"* Computer, 14, *nr. 8 (1981), 83-101*

[22] F.M. KLAASEN: *"Design and performance of micro-size devices"* Solid State Electronics 21 *(1978), 565-571*

[23] IBM RES. DEVEL. J. 24 (1980) nr. 2: *"Josephson computer technology"*

[24] M.J. MARCUS: *"Some system implications of a Josephson computer techonolgy"* Proc. IEEE 69, *(1981), 404-409*

[25] COMPUTER, IEEE Computer Society, 13, nr. 11 (1980): *"Supersystems for the 80's" con ricca bibliografia*

[26] M.V. WILKES, J.B. STRINGER: *"Microprogramming and the design of the control circuits in an electronic digital computer"* Proc. Cambridge Phil. Soc., part 2, vol. 49 *(1953), 230-238*

[27] Y. OKAGA ET AL.: *"A novel multiprocessor array"* 2nd Euromicro Symp., *Venezia (1976), 83-90*

[28] W.A. WULF ET AL.: *"C. mmp - A multi-mini-processor"* AFIPS 41 *(1972), 765-777*

[29] R.J. SWAN ET AL.: *"Cm - A modular multi-microprocessor"* AFIPS 46 *(1977), 637-643*

[30] J.L. BAER: *"Multiprocessing systems"* IEEE Trans. Comp. C-25, *(1976), 1271-1277*

[31] R.M. RUSSEL: *"The CRAY-1 computer system"* Comm. ACM 21 *(1978), 63-72*

[32] W.J. WATSON: *"The TI-ASC: a highly modular and flexible supercomputer architecture"* AFIPS 41 *(1972), 221-228*

[33] G.A. ANDERSON, E.D. JENSEN: *"Computer interconnection: structures, taxonomy characteristics and examples"* Computing Surveys 7 *(1975), 197-213*

[34] Y. PAXER, M. BOZYIGIT: *"Variable topology multicomputer"* 2nd Euromicro Symp. *Venezia (1976), 141-149*

[35] C.R. VICK ET AL.: *"Adaptable architectures for supersystems"* Computer 13, *nr. 11, (1980), 17-35*

[36] K.P. ARVIND ET AL.: *"An asynchronous programming language and computing machine"* Inf. and Comp. Science Dept., University of California, Irvine, TR 114 a *(1978)*

[37] W.B. ACKERMAN, J.B. DENNIS: *"VAL: a value oriented algorithmic language"* Comp. Science Dept., MIT,TR 218 *(1979)*

[38] E.W. WOZDROWICKI ET AL.: *"Second generation of vector supercomputers"* Computer 13, *nr. 11 (1980), 71-83*

[39] J.E. RUMBAUGH: *"A data flow multiprocessor"* IEEE Trans. Comp. C-26 *(1977), 138-146*

[40] I. WATSON ET AL.: *"A prototype data flow computer with token labelling"* AFIPS 48 *(1979), 623-628*

[41] W.W. CHU ET AL.: *"Task allocation in distributed data processing"* Computer 13 *nr. 11 (1980), 57-69*

[42] JB. DENNIS: *"Data flow supercomputers"* Computer 13 *nr. 11 (1980), 48-56*

[43] J.P. GRAY: *"Introduction to silicon compilation"* Proc. 16° Design Automation Conference, *giugno 1979, 305-306*

[44] B.C. BOSCH: *"Gigabit Electronics"* Proc. IEEE 63 *(1979), 340-379*

[45] G. RONDRIGUE ET AL.: *"Perspective on large scale scientific computation"* Computer 13 *nr. 10 (1980), 65-80*

Il calcolatore controlla il motore dello "Space Shuttle"

ROD COPA
JERRY SAMBERG

Honeywell
Avionics Division
St. Petersburg (USA)

1. Introduzione

Il modulo di controllo del motore principale dello Space Shuttle (*SSMEC = Space Shuttle Main Engine Controller*) è stato progettato e costruito dalla Avionics Division della Honeywell nell'ambito del programma di sviluppo di tale motore. Dall'istante del lancio fino al collocamento in orbita del veicolo, lo SSMEC controlla la velocità di combustione del motore principale, in modo da ottenere i livelli di potenza via via prescritti dal calcolatore generale dello Shuttle. Esso inoltre esegue il monitoraggio continuo delle prestazioni del motore, raccogliendo dati su temperature, pressioni, velocità delle turbopompe e flusso del carburante mediante sensori montati nel motore, e mette in atto contromisure in caso di anomalie di funzionamento.

Sullo Space Shuttle ci sono in totale tre SSMEC, uno per ciascuno dei motori principali. Dalla progettazione alla fabbricazione, un accento particolare è stato messo sull'aspetto affidabilità. Poichè il modulo di controllo è montato direttamente sul motore, esso è stato progettato in modo da resistere a condizioni ambientali estremamente severe, in termini di vibrazioni e temperatura. Inoltre, per assicurare il funzionamento in ogni possibile evenienza, è stato introdotto un elaborato sistema di ridondanze. Ciascun SSMEC possiede due canali ridondanti (canale A e canale B), ognuno dei quali costituito da cinque sottosistemi. I sottosistemi di ciascun canale sono collegati con quelli corrispondenti dell'altro canale, in modo da potersi scambievolmente sostituire in caso di guasto. Il modulo di controllo può quindi continuare a funzionare anche in presenza di più guasti. Nel caso di guasto contemporaneo dello stesso sottosistema in entrambi i canali, entra in funzione una procedura di sicurezza che porta allo spegnimento del motore.

2. Il motore principale dello Space Shuttle

Il motore principale dello Shuttle è il primo motore a razzo effettivamente regolabile ed è il primo dotato di un sistema di controllo digitale ad anello chiuso. Una funzione specifica del modulo di controllo è di far sì che il motore possa rispondere a comandi di riduzione della spinta durante l'ascensione. Se, infatti, durante tale fase l'accelerazione supera i 3 g, la struttura del veicolo può essere danneggiata. D'altra parte, man mano che il veicolo diventa più leggero a causa del consumo di propellente, questo limite di 3 g verrebbe superato se i motori non fossero regolabili.

Questi motori, concepiti tra l'altro per essere riutilizzati, funzionano con propellente liquido. Essi bruciano idrogeno come combustibile, e usano ossigeno come comburente. (Idrogeno liquido è inoltre fatto circolare attraverso molteplici condotti e cavità del motore, per raffreddare le superfici esposte ai gas bollenti di scarico). Questi propellenti forniscono la maggior energia per unità di peso e inoltre non sono inquinanti poichè generano, come prodotto della combustione, soltanto calore e vapore acqueo. Anche se altri motori a razzo usano questi propellenti, il motore dello Shuttle presenta caratteristiche mai prima raggiunte. In particolare, la pressione nella camera di combustione è di circa 3.000 libbre per pollice quadro, di gran lunga maggiore di quella di tutti i precedenti motori a razzo utilizzanti idrogeno e ossigeno liquidi.

Questo motore, pur generando una spinta enorme, presenta dimensioni relativamente modeste. Ciò è il risultato di una concezione di progetto del tutto nuova. Esso infatti è il primo esempio di motore a idrogeno e ossigeno liquidi avente un ciclo di combustione a due stadi. Buona parte del combustibile viene infatti precombusta in due apposite camere: una di esse circonda la turbina della turbopompa di alta pressione del combustibile, mentre l'altra circonda la turbina della turbopompa analoga del comburente. Il gas che esce da queste camere è vapore saturo di idrogeno surriscaldato. Questo gas aziona le turbine delle pompe di alta pressione e poi entra nella camera principale di combustione dove, insieme con ulteriore idrogeno vaporizzato, viene combinato con l'ossigeno per dare luogo alla reazione.

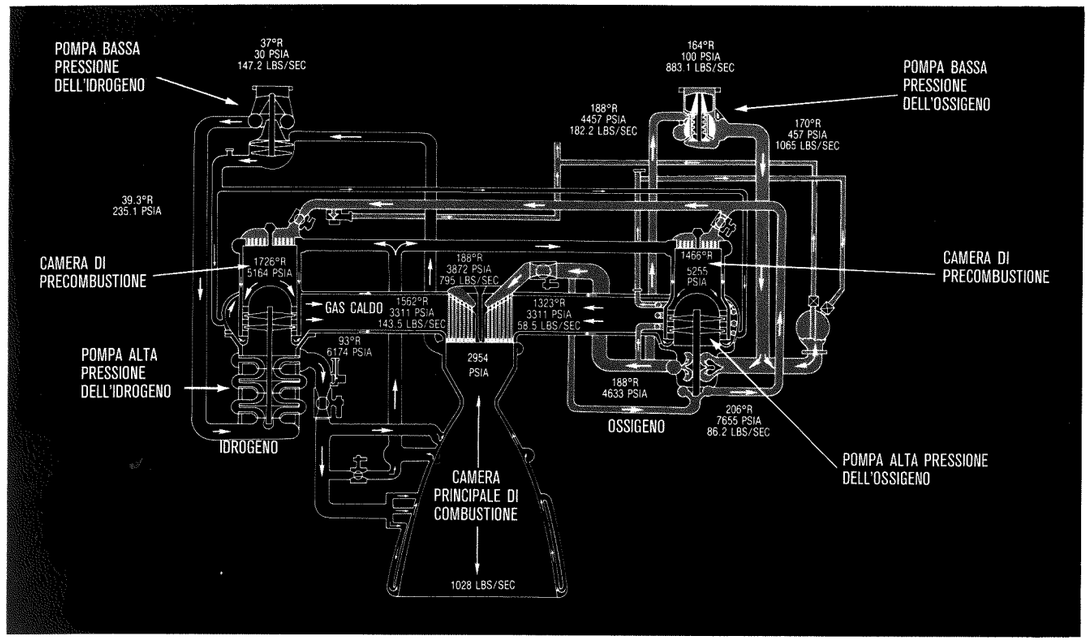


Fig. 1 - L'enorme potenza sviluppata dal motore principale dello Space Shuttle è principalmente il risultato del suo ciclo di combustione a due stadi. Buona parte del combustibile passa infatti attraverso due camere di precombustione. I gas in uscita da queste azionano, prima di entrare nella camera principale di combustione, delle turbopompe che portano a valori elevatissimi la pressione dei propellenti (idrogeno e ossigeno). Questi ultimi bruciano nella camera principale di combustione ad una pressione di circa 3.000 PSIA (Pounds per Square Inch Absolute), un valore molto maggiore di quello raggiunto in ogni precedente motore a razzo.

Il segreto di tanta potenza sviluppata da un motore di queste dimensioni deriva dalla capacità delle turbopompe di immettere in tempi brevissimi enormi quantità di propellente nella camera di combustione. Per dare una idea della capacità di tali pompe, esse potrebbero vuotare una piscina di medie dimensioni nel giro di qualche secondo. Al livello prescritto di potenza, la combustione del propellente in ciascuno dei tre motori produce una spinta di 375.000 libbre al livello del mare e di 470.000 nello spazio.

3. Le turbopompe

Ciscuno dei motori possiede quattro pompe azionate da turbine: una coppia di pompe in cascata (bassa e alta pressione) serve per l'alimentazione del combustibile, mentre l'altra coppia serve per l'alimentazione del comburente. Il primo stadio di ciascuna coppia, cioè la turbopompa a bassa pressione, è collegata al condotto del propellente. Le turbine di queste pompe sono fatte girare dal propellente (all'inizio per caduta e successivamente dal propellente che ricicla nel tragitto verso la camera di combustione).

Il secondo stadio di ciascuna coppia è una turbopompa ad alta pressione azionata dal vapore surriscaldato proveniente dalle camere di precombustione. Le due pompe ad alta pressione, una per l'idrogeno liquido e l'altra per l'ossigeno liquido, operano sullo stesso principio ma sono dimensionate in modo diverso perchè la turbopompa dell'idrogeno deve fornire un volume di propellente maggiore di quanto non debba fare quella dell'ossigeno. La proporzione ideale di idrogeno e ossigeno in questo motore è di uno a sei in peso. Tuttavia, poichè l'ossigeno liquido è assai più denso dell'idrogeno liquido,il volume di idrogeno liquido è circa 2,5 volte quello dell'ossigeno liquido.

Il progetto e la fabbricazione delle turbopompe ad alta pressione hanno richiesto la soluzione di diversi difficili problemi. Per esempio, l'albero rotante della turbopompa, che

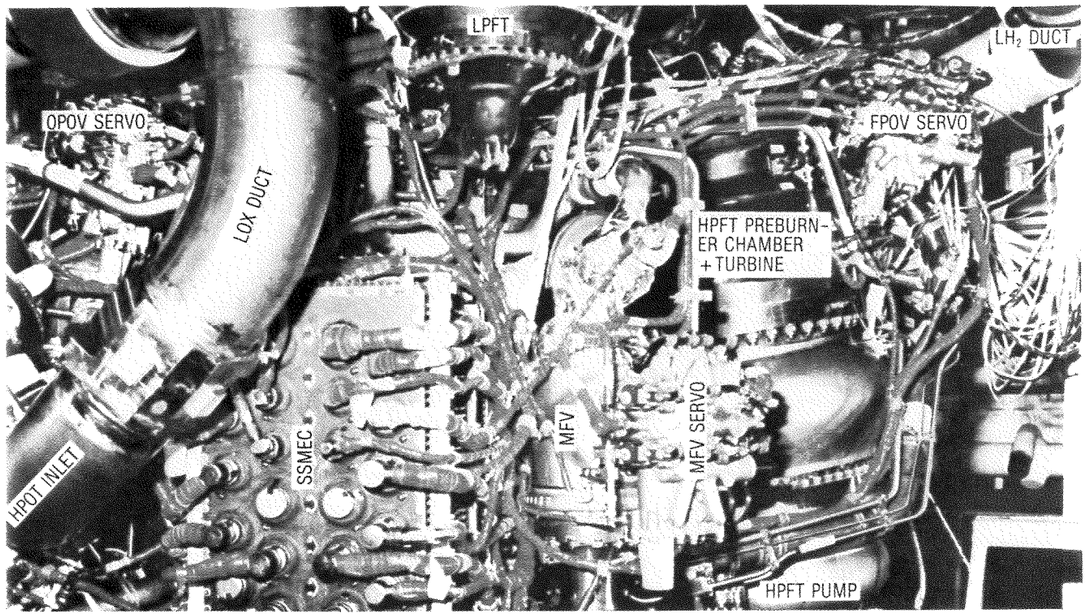


Fig. 2 - Il modulo di controllo del motore principale dello Shuttle (SSMEC = Space Shuttle Main Engine Controller) è montato direttamente sul motore stesso. Nella fotografia è visibile la parte superiore dello SSMEC, coi bocchettoni per le connessioni elettriche.

è relativamente corto, è superfreddo ad un estremo (lato pompa) e caldissimo dall'altro (lato turbina). Oltre a ciò, è necessario fare in modo che i gas caldi che escono dalla turbina non si mescolino coi propellenti criogenici che si trovano ad una estremità della pompa. Ciò è particolarmente importante per la pompa ad alta pressione dell'ossigeno, poichè anche l'acciaio speciale di cui sono fatte alcune parti del motore brucerebbe in un ambiente ricco di ossigeno. Probabilmente la cosa peggiore che possa capitare a questo motore è che parti metalliche calde siano esposte a ossigeno puro. Se ciò avvenisse, il metallo brucerebbe o volatizzerebbe anche se non fosse caldo abbastanza per fondere.

4. I servocomandi

Lo SSMEC, cioè il modulo di controllo del motore principale dello Shuttle, riceve le informazioni relative sia alla fase operativa (preparazione, accensione, regime, arresto) che al livello di potenza da erogare, direttamente dall'elaboratore generale dello Shuttle. Lo SSMEC risponde a questi comandi essenzialmente regolando la portata di cinque valvole ad azionamento idraulico servo-assistite che controllano il flusso dei propellenti nel motore. Le cinque valvole regolano rispettivamente il flusso principale del combustibile, il flusso principale del comburente, il flusso del fluido di raffreddamento, e infine il flusso nei due prebruciatori. Queste ultime due valvole giocano un ruolo essenziale, in quanto controllano la velocità delle turbopompe ad alta pressione.

Consideriamo infatti la valvola che controlla il flusso dell'ossigeno liquido nella camera di precombustione della pompa ad alta pressione dell'ossigeno. Quanto maggiore è la quantità di ossigeno nel prebruciatore, tanto maggiore risulta la temperatura e la pressione dei gas che pilotano la turbina della pompa in questione. Quanto più velocemente gira tale turbina, tanto più ossigeno liquido viene spinto attraverso la valvola principale nella camera di combustione e quindi tanto più intensa risulta la combustione. In sostanza, la valvola che controlla il flusso dell'ossigeno nel prebruciatore regola *il livello di potenza del motore,* essendo questa più una funzione della quantità di comburente (ossigeno) che di combustibile (idrogeno).

L'altra valvola controlla invece, in modo analogo, la velocità della turbopompa del combustibile, e quindi il flusso del combustibile nella camera di combustione. Questa valvola viene usata per ottenere o ripristinare *il valore ottimale della miscela* idrogeno-ossigeno nella camera di combustione.

Ogni valvola è in effetti costituita da un complicato insieme di solenoidi, di elementi idraulici e pneumatici e di valvole a scatto. Una parte essenziale dell'insieme è costituita dai servo-attuatori (due per ciascuna valvola, i canali A e B già

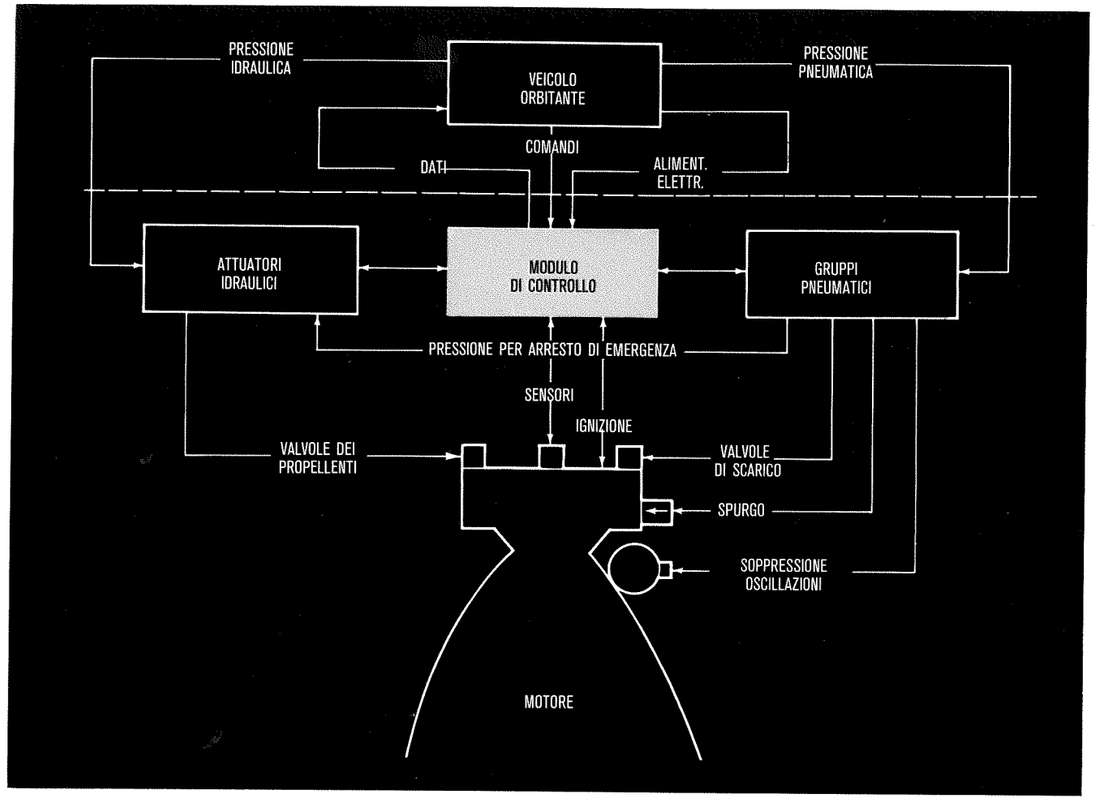


Fig. 3 - Ogni modulo di controllo ha due interfacce, una verso il sistema di controllo del veicolo, l'altra verso il motore controllato. Per garantire il funzionamento e la sicurezza anche in caso di guasti, è stato introdotto un elevato grado di ridondanze.

citati). Tutti questi servomeccanismi sono controllati dallo SSMEC. Di norma è il canale A che controlla il posizionamento idraulico delle valvole. Se però lo SSMEC, sotto il controllo del suo "software di missione", scopre un guasto nel canale A, allora energizza solenoidi e valvole a scatto in modo che sia il canale B a prendere il controllo del posizionamento idraulico della valvola.

Il software di missione prescrive che se un anello di servocomando del canale A si guasta, vengano attivati tutti gli anelli del canale B. Se successivamente si guasta anche uno dei servocomandi del canale B mentre il motore sta funzionando a pieno regime, lo SSMEC blocca idraulicamente tutte le cinque valvole del motore nella posizione raggiunta e si limita ad effettuare il monitoraggio del motore per controllare che non si arrivi ad una situazione di emergenza che richieda lo spegimento del motore stesso. In tale evenienza, poichè lo spegnimento non può più essere fatto chiudendo le valvole per via idraulica, lo SSMEC fa intervenire un apposito sistema di soccorso di tipo pneumatico.

Questo sistema di riserva chiude tutte le valvole usando elio pressurizzato ogni qual volta lo SSMEC non è in grado di fare ciò per via idraulica. In definitiva, il sistema di controllo pneumatico è sempre pronto a intervenire per spegnere il motore in caso ciò non possa essere effettuato nel modo normale.

5. Operazione a prova di guasto

Lo SSMEC è collegato al sistema centrale di controllo del veicolo mediante un canale triplo, via appositi convertitori di comandi e di dati, attraverso cui riceve gli ordini relativi alle varie fasi operative ed ai livelli di potenza da erogare. I progettisti dello SSMEC hanno preso una serie di misure per assicurare il riconoscimento dei comandi, che possono risultare disturbati durante la trasmissione, e per mettere in atto appropriate azioni correttive in caso di guasto.

I convertitori di comandi e di dati presentano, a tale fine, ridondanza tripla, anzichè doppia come il resto dello SSMEC. Ciascun convertitore riceve separatamente una parola di comando di 31 bit (16 bit per i dati

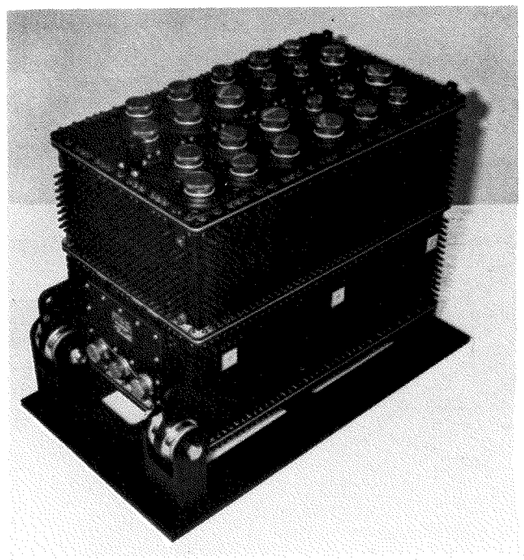


Fig. 4 - La fotografia mostra un modulo di controllo. Le dimensioni sono di 14.5 x 18.24 x 24.5 pollici. È visibile tutto attorno il dissipatore per smaltire il calore generato all'interno del modulo durante le lunghe sequenze di pre-lancio. Una volta nello spazio, occorre invece mettere in funzione un riscaldatore interno per impedire che la temperatura scenda a valori troppo bassi, che possono danneggiare i componenti elettronici.

e i rimanenti 15 per il codice di errore) dall'elaboratore generale dello Shuttle e memorizza la parte dati del comando. Le due unità di elaborazione dello SSMEC prendono, a questo punto, decisioni indipendenti sul triplice insieme di comandi. Il software operativo fa in modo che un comando venga accettato se almeno due comandi su tre sono tra loro in accordo. Inoltre, il comando accettato per essere "validato" deve risultare parte del repertorio di comandi dello SSMEC ed essere compatibile con la fase operativa in atto in quel momento.

Se, per qualche ragione, un comando non viene validato, l'elaboratore generale dello Shuttle viene messo al corrente che si è verificato un guasto nella trasmissione, attraverso il doppio canale che collega lo SSMEC a tale elaboratore (ed anche al sistema di telemetria del veicolo, che trasmette agli elaboratori di terra). Ogni 40 millisecondi l'unità di elaborazione abilitata trasmette su questi canali un blocco di dati riassuntivi della situazione. In questo modo, il modulo di controllo può segnalare all'elaboratore generale che un comando non è stato validato.

I cinque tipi di sottosistema che compongono ciascun SSMEC sono: le unità di elaborazione digitale, l'elettronica di interfaccia con l'elaboratore generale, l'elettronica di ingresso e quella di uscita verso il motore e, infine, l'alimentatore.

Ognuno di tali sottosistemi è duplicato in ciascun SSMEC; per esempio, esiste una unità di elaborazione digitale (DCU = Digital Computer Unit) come parte del canale A (DCUA) ed una identica nel canale B (DCUB). Sia DCUA che DCUB possono assumere il ruolo di DCU nel controllo di entrambi i canali della elettronica di ingresso e di uscita. Di norma DCUA è inserita mentre DCUB è di riserva. Però, entrambe le DCU ricevono i dati relativi al motore tramite le doppie unità di ingresso (che convertono i segnali analogici forniti dai sensori del motore in dati digitali). Pertanto, DCUB segue tutte le operazioni dello SSMEC ed effettua i calcoli parallelamente a DCUA, in modo da essere pronta ad assumere il controllo se DCUA non risulta più abilitata a tale compito.

La DCU che possiede il controllo è collegata ad entrambi i blocchi elettronici di uscita (quelli cioè che convertono i comandi digitali in tensioni analogiche di controllo) e può comandare le valvole del motore attraverso l'uno o l'altro di tali blocchi. Essi sono infatti entrambi attivabili, ma è la DCU che possiede il controllo a decidere se il posizionamento idraulico della servovalvola deve essere effettuato tramite il canale A o il canale B.

6. Un progetto per un ambiente difficile

Lo SSMEC è montato direttamente sul motore ma non è esposto al calore da esso generato poichè tale parte del motore è raffreddata con idrogeno liquido. Tuttavia lo SSMEC genera esso stesso calore, che deve essere asportato per prevenire il sovrariscaldamento dei suoi componenti. In funzione delle variazioni del carico, lo SSMEC dissipa da 575 a 675 watt di potenza trifase a 400 Hertz. Il calore generato viene dissipato mediante un radiatore che costituisce buona parte della superficie esterna del modulo. Dopo la fase di salita del veicolo, il problema è rappresentato dal freddo piuttosto che dal caldo. Due riscaldatori interni al modulo, che dissipano 250 watt a partire da una tensione di 28 volt cc, intervengono allorchè la temperatura interna allo SSMEC scende sotto −10 °C, onde prevenire il danneggiamento dei componenti.

Nel progetto dello SSMEC sono stati impiegati tutti gli accorgimenti necessari per conseguire la massima densità di impaccamento dei componenti, onde minimizzare il peso e le dimensioni dell'unità. Il peso complessivo risulta di 213 lib-

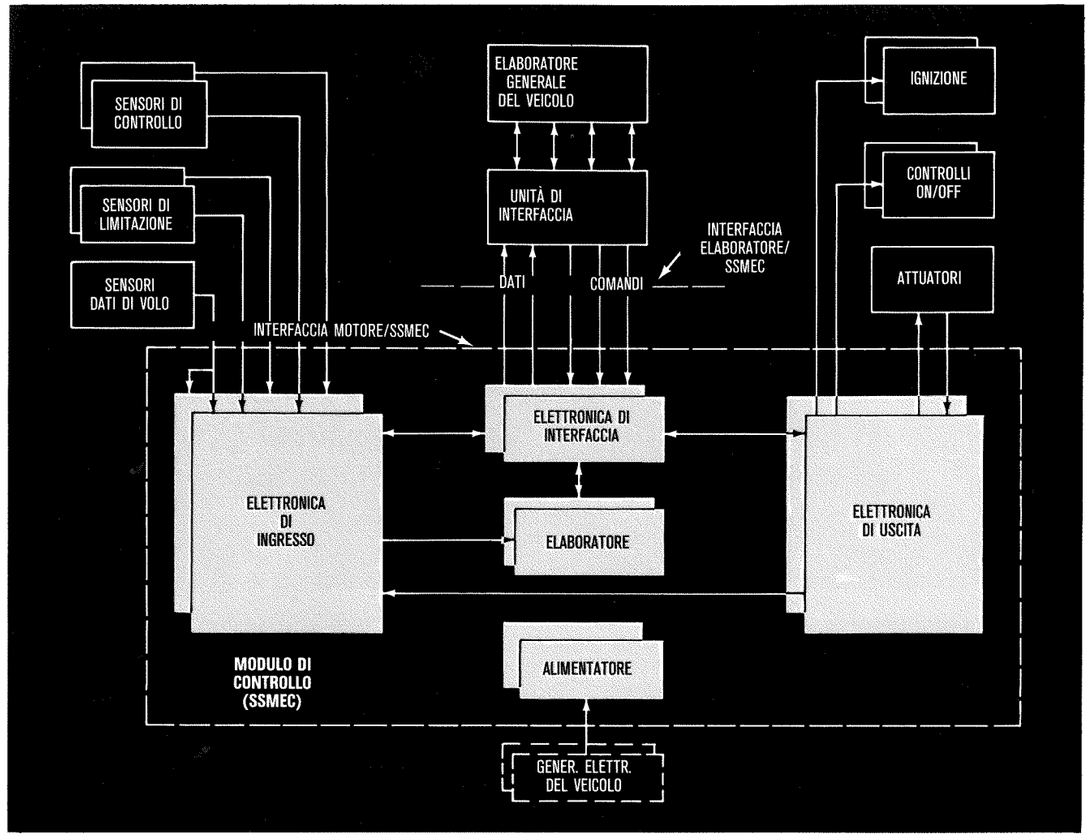


Fig. 5 - Lo schema mostra i cinque sottosistemi di cui è composto il modulo di controllo e cioè: l'elaboratore digitale, l'interfaccia verso il calcolatore generale del veicolo, l'elettronica di ingresso e di uscita verso il motore, l'alimentatore. Come è indicato in figura, ognuno dei sottosistemi è duplicato per ragioni di affidabilità.

bre, mentre le dimensioni sono di 14,5 x 18,24 x 24,5 pollici.

Costruttivamente, lo SSMEC consiste di due chassis, uno interno ed uno esterno, congegnati in modo da costituire un'unità pressurizzata. Ogni chassis contiene un certo numero di cavità, ed in ogni cavità è inserita una coppia di piastre a circuito stampato. Le pareti della cavità conducono all'esterno il calore dissipato dalle piastre. Queste piastre (per un totale di 71), oltre a due alimentatori e a quattro moduli di memoria, costituiscono i cinque sottosistemi dello SSMEC. I connettori per l'interfaccia col mondo esterno sono posti sul coperchio dell'unità.

7. Le memorie del modulo di controllo

Un criterio generale seguito nella progettazione dello Shuttle è stato quello di usare parti altamente affidabili, già collaudate dall'esperienza. Tuttavia un paio delle tecnologie usate nello SSMEC sono relativamente nuove. La memoria, per esempio, è costituita da sottili fili di rame-berillio ricoperti con una pellicola magnetica di ferro-nichel. Questo tipo di memoria era stato sviluppato originariamente dalla Honeywell per l'elaboratore installato sulla sonda spaziale Viking.

Dei cinque sottosistemi dello SSMEC, quello che ha richiesto forse la maggior inventiva è stato l'alimentatore di potenza. La maggior funzione di questo dispositivo è di convertire la corrente trifase a 115 volt, 400 Hertz, fornita dal sistema elettrico dello Shuttle, nelle tensioni continue stabilizzate richieste dall'elettronica del modulo nonchè dalle valvole e dai dispositivi di accensione del motore. L'alimentato-

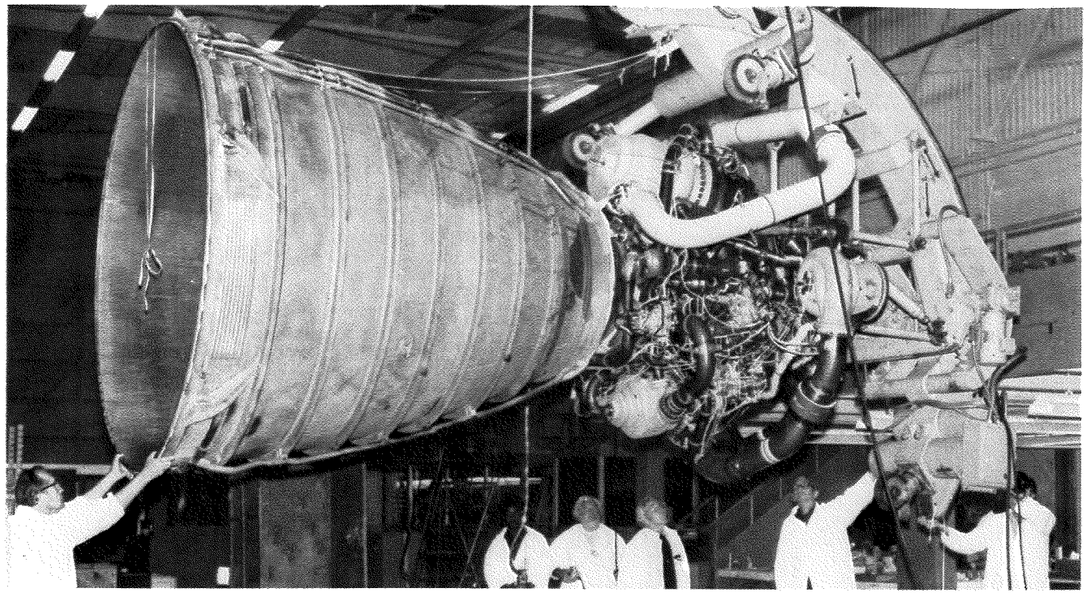


Fig. 6 - Il software operativo impiegato durante il volo dello Shuttle viene usato anche a terra per il collaudo del motore. Questa fotografia mostra uno dei motori principali dello Shuttle durante uno di questi test.

re in questione deve fornire nove tensioni stabilizzate, sei per le operazioni interne allo SSMEC e tre per le funzioni esterne. Esso deve inoltre rispondere appropriatamente a variazioni della potenza di ingresso e proteggere il carico da tensioni abnormi.

I progettisti dell'alimentatore dovettero risolvere il non facile problema di comprimere le funzioni richieste in uno spazio molto limitato o, come essi ebbero a dire, di mettere venti libbre di hardware in una scatola per sole dieci libbre. Sono stati pertanto adottati sistemi di conversione elettronica di potenza invece dei soliti trasformatori. Questa soluzione non solo ha ridotto lo spazio ed il peso richiesto, ma ha anche diminuito la dissipazione di potenza.

8. Il software del modulo di controllo

Due differenti programmi software sono stati sviluppati dalla Honeywell in connessione con lo SSMEC, e cioè il "Controller Acceptance Test Program" (CATP) e l'"Operational Program" (OP).

Il primo viene usato in produzione, durante la fase di collaudo del sistema. Insieme col software dell'apparecchiatura automatica di collaudo, questo programma verifica l'hardware dello SSMEC, incluso il meccanismo di scambio delle parti ridondanti. Durante alcune fasi del collaudo, vengono simulati da consolle gli input/output relativi al motore e gli input dall'elaboratore generale dello Shuttle. Per esempio, durante una prova, la consolle invia gli input simulati dei sensori del motore, i quali vengono confrontati coi segnali in uscita dall'elettronica di input dello SSMEC.

Il programma operativo OP viene usato oltre che durante il volo effettivo del veicolo, anche durante il collaudo a terra. Questo programma ha un ciclo fondamentale di 20 millisecondi. Durante tale periodo esso riceve dall'elaboratore generale dello Shuttle i comandi che definiscono la fase ed il modo operativo del motore. Il compito maggiore del software operativo è di controllare questi due aspetti, ma nel contempo esso esegue il monitoraggio continuo dei sensori del motore e dello stesso SSMEC, inviando i dati all'elaboratore generale dello Shuttle.

Si distinguono sei differenti fasi del motore: controllo, preparazione alla accensione, accensione, regime, spegnimento e post-spegnimento.

La fase di controllo consiste di cinque differenti sequenze di verifica. I sensori del motore vengono controllati inviando un segnale di calibrazione a ciascuno di essi e confrontando la risposta con dei valori registrati in memoria. Ciascuno degli attuatori delle cinque valvole del motore viene pure controllato, sottoponendolo ad una specifica sequenza di aperture, bloccaggi e chiusure, durante cui si verificano le prestazioni e la capacità di rivelare guasti. I sei solenoidi pneumatici sono controllati energizzandoli e verificandone il funzionamento. Durante il controllo del sistema di ridondanze, vengono verificate tutte le possibili combinazioni. Per esempio, si passa il controllo a DCUB e si verificano tutti i relativi percorsi di controllo. Durante questa fase viene usata una serie di valori preregistrati dei vari parametri del motore per simulare l'accensione del medesimo. Questi valori sostituiscono quelli che si avranno effettivamente durante la combustione del propellente.

La fase di preparazione all'accensione predispone il motore per la combustione. Consiste di quattro sequenze di spurgo del motore, che asportano l'aria presente nei condotti del propellente.

Durante la fase di accensione, lo SSMEC apre le varie valvole nella sequenza appropriata e attiva i dispositivi di accensione nella camera principale di combustione e nelle camere di precombustione, dando inizio al processo e portando rapidamente il motore nella fase di funzionamento di regime.

Durante lo spegnimento, che può essere provocato o da un comando oppure da un guasto del motore o dello SSMEC, quest'ultimo chiude le valvole secondo una sequenza prefissata.

Nella fase successiva allo spegnimento, allorchè il motore è completamente arrestato, si spegne il modulo di controllo.

Come si è detto, il software dello SSMEC è in grado di eseguire il monitoraggio di quest'ultimo oltre che del motore. Qualche esempio di questa capacità di autoverifica può servire a dare un'idea della cura posta durante tutto il progetto dello SSMEC per assicurare che esso fosse capace di rispondere a qualsiasi evenienza. Uno dei controlli che lo SSMEC esegue su stesso consiste in un programma di prova che fa eseguire all'elaboratore tutte le sue istruzioni in una determinata sequenza. I risultati di questa elaborazione vengono successivamente confrontati con dei valori preregistrati. Allorchè si trova un errore, il test viene ripetuto; se non viene superata neanche la seconda prova, l'elaboratore viene disabilitato. Un altro esempio è costituito dal test con cui l'elettronica di ingresso viene provata, applicando cioè tensioni note e confrontando l'uscita digitale con valori registrati. Se si verifica un errore, il sottosistema di ingresso in questione viene disabilitato ed i dati da esso forniti non sono più usati nel controllo del motore. Un alteriore esempio è dato da un test progettato per controllare la risposta dell'unità di elaborazione ad un errore di parità. Questo test simula tale tipo di errore in una delle parole registrate in memoria e verifica che si realizzi il processo di riconoscimento. Se ciò non accade, l'unità di elaborazione in questione viene fermata.

Questi test, insieme con tutti gli altri che vengono effettuati in continuazione, sono alla base della caratteristica fondamentale dello SSMEC, di poter cioè funzionare correttamente anche in presenza di guasti e di garantire, in ogni caso, la sicurezza del sistema.

NOTA.

Il materiale qui presentato è stato tratto dall'articolo "The Space-Shuttle Main Engine Controller" degli stessi Autori, pubblicato su "SCIENTIFIC HONEYWELLER, vol. 2, n. 4, dic. 1981.

Agricoltura e computer

TITO GAUDIO

E.E.D.
Volpiano (Torino)

1 - Introduzione

L'agricoltura, questa fondamentale tra le attività umane, è da tempo al centro di molte considerazioni in chiave critica.

L'esodo dalle campagne, l'emigrazione verso la città con le note conseguenze di inurbanizzazione massiccia e caotica, non sono che i fenomeni più appariscenti nel mondo rurale.

L'impiego non sufficientemente controllato della chimica per la concimazione e la difesa dei raccolti, ha determinato un pericoloso aumento dell'inquinamento ambientale, l'impoverimento dei suoli e uno scadimento biologico del prodotto agricolo.

La tipologia delle colture, la particolare configurazione del territorio e la rapida obsolescenza delle macchine agricole, hanno portato a un parco vario e in continuo rinnovamento contribuendo al fenomeno tipico dell'indebitamento medio.

Stretta in una morsa, l'agricoltura italiana e gran parte di quella europea, è per così dire "dominata". Da un lato l'industria controlla gli input con la fornitura di energia e prodotti industriali, dall'altro il commercio decide gli acquisti, condiziona i prezzi e pilota la distribuzione.

L'obiettivo di una agricoltura "protagonista", produttrice di cibo ed energia, passa necessariamente attraverso l'evoluzione dei sistemi agricoli. Uno sviluppo basato sulla massima considerazione delle qualità della vita, dell'ecologia, della biologia, delle nuove fonti energetiche; sull'uso delle "tecnologie povere" e sull'impiego ottimale delle risorse locali.

E' necessario attuare, sostengono gli esperti, nuove forme di intervento rivolte a rinnovare ed accrescere le potenzialità specifiche dell'agricoltura. Occorre riconsiderare il ruolo che essa ha nel soddisfacimento dei bisogni primari dell'uomo, non solo come fonte di cibo per le nostre tavole ma soprattutto come grande *collettore solare,* produttore potenziale di risorse rinnovabili (cibo, materiali, energie), strumento di conservazione, recupero e riciclaggio delle risorse.

2 - Nasce l'Agronica

Assai più che altre tecnologie, l'elettronica ha la capacità di stimolare tematiche sociali di notevole importanza. Queste possibilità derivano soprattutto dagli attributi che meglio la caratterizzano, ovvero "capacità di controllo" e "miniaturizzazione".

Il controllo può concorrere alla soluzione di molte problematiche sociali come il risparmio energetico, l'inquinamento ambientale, ecc.

La miniaturizzazione, consente, in particolare, un risparmio di materiali, spazi ed energie; comporta basse potenze in gioco e quindi consumi contenuti.

Ora però c'è da chiedersi se entrambe le proprietà siano sempre usate con finalità positive e se talvolta non accada che vengano sfruttate come mezzi di condizionamento e di conquista. Qualche esempio: si sfrutta la miniaturizzazione per invadere il mercato di alcuni prodotti alienanti o superflui, si coglie il momento ecologico per intervenire con soluzioni costose e complesse.

Interessi economici e politici, vedute compartimentali, limitatezza di orizzonti rischiano nel futuro di creare una divisione tra elettronica e società con grossi problemi di interazione verso l'uomo e l'ambiente.

Gli interventi, quindi, in settori come la difesa ambientale, la produzione e il controllo energetico, il riciclaggio delle risorse costituiscono delle opportunità in linea con il fluire dei tempi.

Seguendo, dunque, linee e tendenze di questa nuova cultura che si va rapidamente affermando, opportuno appare l'intervento dell'elettronica nell'agricoltura.

Applicazioni avanzate e prospettive avvincenti già emergono nelle realtà agricole più evolute. Il computer arriva in campagna, entra nelle serre e nelle stalle, offre nuovi servizi alle aziende e alle comunità.

Nasce così l'*Agronica*, tecnologie e prodotti elettronici orientati all'agricoltura per ottimizzare il lavoro dei campi.

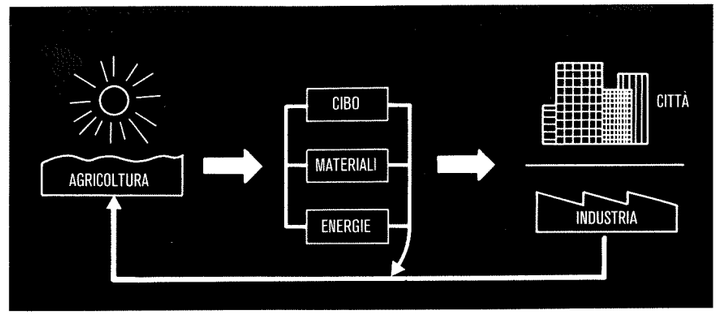


Fig. 1 - *L'Agricoltura.*
L'agricoltura, un grande collettore solare, produce cibo, materiale ed energia che cede alle città e alle industrie.
L'industria rende prodotti industriali, energia e anche inquinamento.

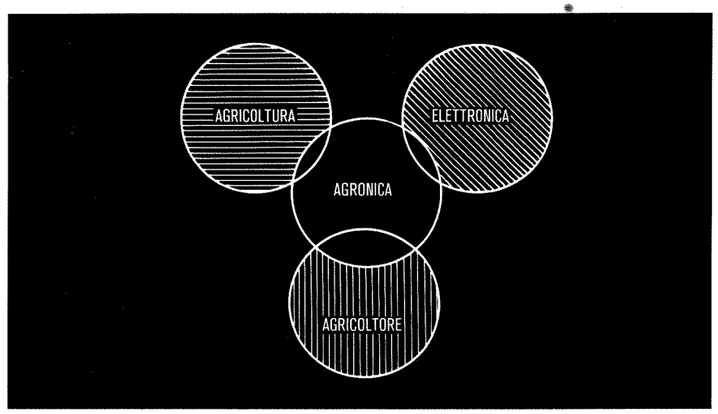


Fig. 2 - *L'agronica.*
L'agronica è il nuovo anello che tende a congiungere tecnologie elettroniche (informatica, telecomunicaz., ecc.) e agricoltura.
Molto più rilevante che in altre integrazioni di tipo industriale (robotica, avionica, ecc.) assume il fattore utente (l'agricoltore).

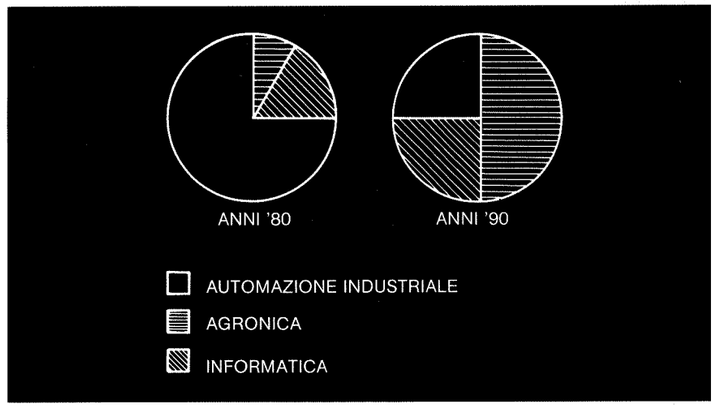


Fig. 3 - *L'evoluzione dell'agronica.*
L'elettronica nell'agricoltura è attualmente caratterizzata soprattutto da automazione di tipo industriale adattata a problemi agricoli. Significativa è già la presenza dell'informatica specie nelle aziende agro-zootecniche di grandi dimensioni.
L'introduzione di prodotti agronici e sistemi informativi per piccole e medie aziende apporterà una modifica sostanziale nei prossimi 20 anni.

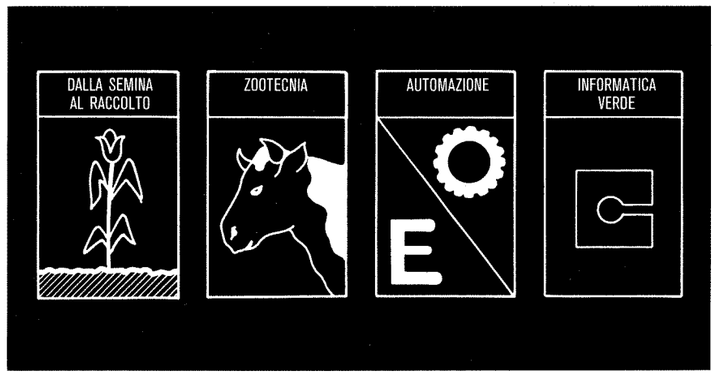


Fig. 4 - *Possibilità di intervento.*
Quattro sono, in sintesi, le aree di intervento dell'elettronica orientata all'agricoltura.
Nel regno vegetale (dalla semina al raccolto), nella zootecnia, nell'automazione dei processi paragricoli e la produzione di energie alternative ed integrative, nel management aziendale attraverso la diffusione dell'informatica.

Un corretto approccio all'agricoltura e alle applicazioni elettroniche che ne potranno conseguire nei prossimi anni deve tenere conto, attraverso una rapida riflessione, di alcune problematiche che caratterizzano l'elettronica applicata.

Ci sono atteggiamenti culturali non corretti nei confronti dell'elettronica. Averne paura, ad esempio, o il ritenere che l'elettronica risolva tutto (o al contrario che sia superflua), concepire il prodotto elettronico complicato, delicato, difficile da manutenere e riparare.

Ci sono anche aspetti più concreti, ancora non completamente risolti, come la difficoltà di accesso economico e tecnologico, tecniche e apparecchiature non tutte facilmente comprensibili e dominabili, rapida obsolescenza dei prodotti, reperimento del personale specializzato per eseguire manutenzioni e riparazioni.

Ma una serie di problematiche che l'operatore elettronico incontra nell'affrontare lo sviluppo di un prodotto agronico nasce dalla diversità di linguaggio, un aspetto assai più articolato e complesso di quello ricorrente nel rapporto con altri settori applicativi.

L'eccessivo specialismo che sempre più caratterizza la quasi totalità dei progettisti elettronici, ostacola di fatto la corretta conoscenza di altre tematiche (per non parlare poi di quelle agricole) e rende di conseguenza l'approccio lungo e difficile.

Qui è importante, oltre a recepire correttamente e nella sua completezza il problema da risolvere, saper conciliare metodi di valutazione e obiettivi finali da raggiungere. L'intervento dell'elettronica potrà contenere (ma non annullare) il margine di errore per molte operazioni oggi svolte in modo soggettivo dall'utente. Si tratta, in altri termini, di passare da interventi eseguiti sulla base di valutazioni personali (ma non per questo sempre grossolani) a interventi controllati e pilotati con un margine di errore contenuto.

Nell'agricoltura sono poco significativi termini come "microsecondi" o "milligrammi", mentre si parla spesso di "un certo tempo" o di "mezzo sacco". Hanno più significato i risultati espressi in termini qualitativi che quantitativi; sono più efficaci le indicazioni grafiche che quelle numeriche. Per un allevatore, ad esempio, è più importante conoscere lo stato di salute del singolo capo (se è regolare, se va verso l'infezione o il collasso ecc.) che il valore della temperatura corporea espressa in decimi di grado.

Di fronte alla costante evoluzione tecnologica, ci sono poi le difficoltà di realizzare un prodotto con un grado di obsolescenza contenuto. Questo aspetto deve essere oggetto

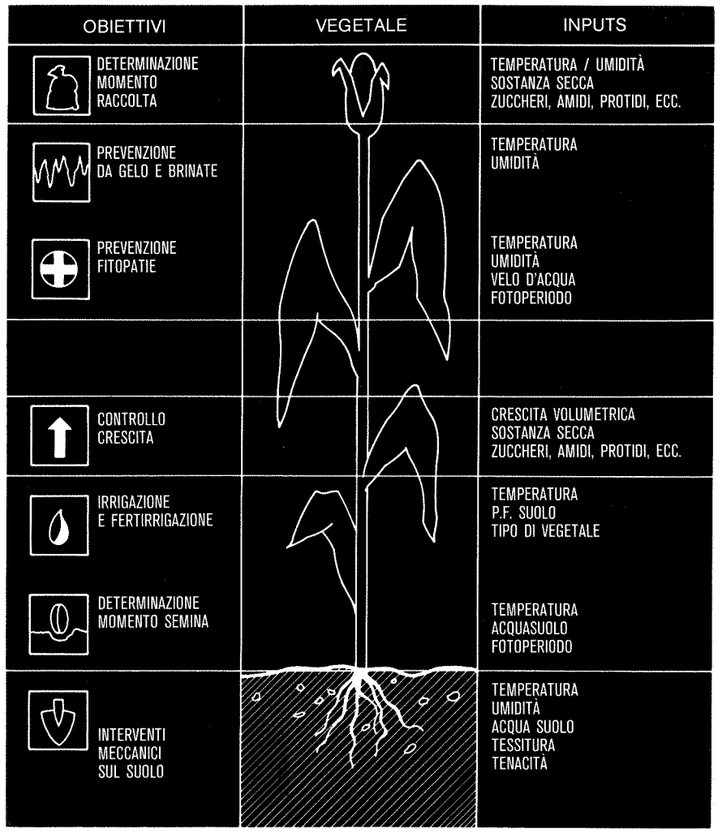


Fig. 5 - *Dalla semina al raccolto* Molteplici sono le possibilità di intervento nel regno vegetale sia attraverso l'uso di centraline microclimatiche installate nei campi che mediante computer portatili orientati alla soluzione di problemi agricoli.

Fig. 6 - *Controllo climatico delle serre.*

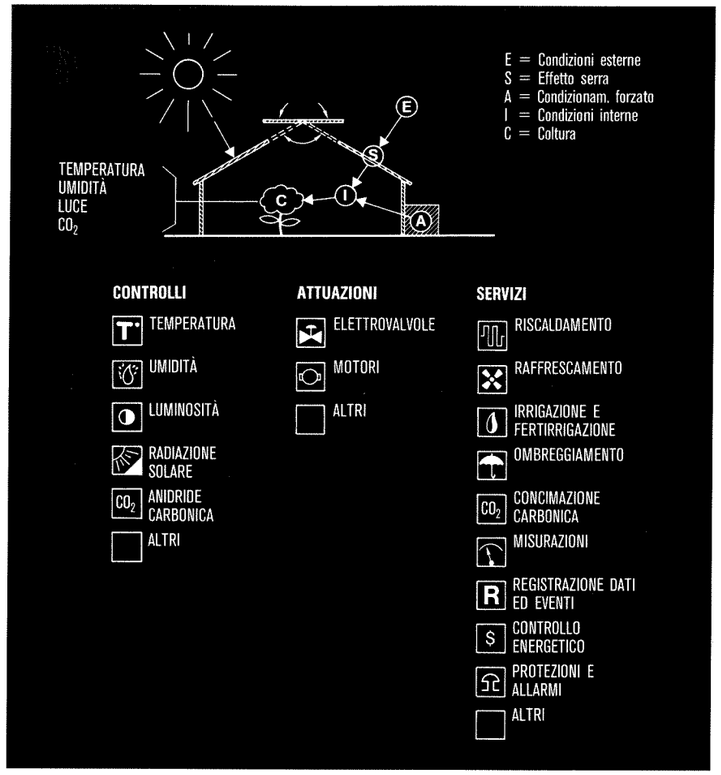
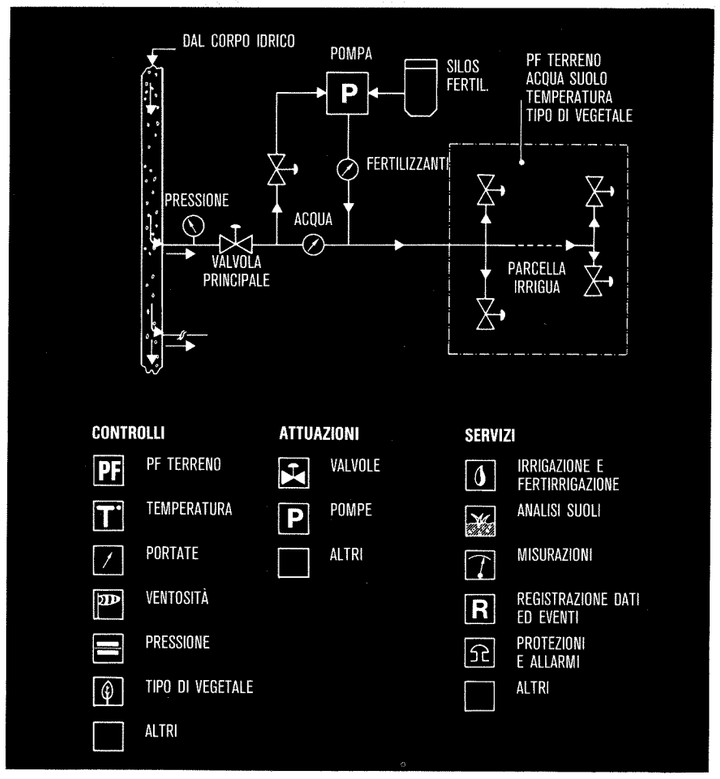


Fig. 7 - *Controllo irrigazione e fertirrigazione.*



di attenta valutazione. Se da un lato una possibile riduzione dei costi, minori ingombri e migliori prestazioni, possono costituire elementi motivanti di vendita, dall'altro occorre considerare che l'utente richiede un prodotto duraturo oltrechè rustico e affidabile. L'agricoltore è attaccato al bene.

Restio a cambiare (anche di fronte a sensibili vantaggi) preferisce ripararlo (se necessario) ma mantenere il prodotto nel quale ha riposto piena fiducia.

I prodotti elettronici per l'agricoltura devono dunque essere soprattutto semplici, economici, facili da usare, coerenti con l'evoluzione dei sistemi agricoli, opportuni se permettono l'ottimizzazione di precise funzioni, dannosi se esasperano le problematiche cioè se spingono verso automazioni complesse e costose che non trovano giustificazioni nella nostra agricoltura.

I vantaggi che ne possono derivare da una applicazione coerente e funzionale alle diverse problematiche dell'agricoltura sono indubbiamente tanti.

La disponibilità di *intelligenza a basso costo,* caratteristica peculiare dell'elettronica, può dare un contributo significativo apportando un deciso arricchimento di know-how e inducendo una positiva retroazione culturale.

L'adozione di un adeguato linguaggio "uomo-macchina" può facilitare la diffusione della logica e migliorare le capacità razionalizzatrici dell'uomo dei campi senza per questo apportare sostanziali limitazioni alla sua fantasia.

Così come è avvenuto in altri settori, l'elettronica può migliorare il grado di affidabilità, (intesa come capacità di ripetizione), offrire maggiore precisione (cioè minori tolleranze), rapidità e congruenza di risposta.

La capacità di esplicare funzioni di controllo a livelli differenziati di intervento, passivo (segnalazione o registrazione istantanea dell'evento), logico, analogico, a soglia (con o senza attuazioni), di allarme (in caso di emergenza), consente di avvertire, quidare e a volte sostituire l'operatore agricolo.

La conseguente possibilità di attuare comandi e azionamenti, siano essi derivanti da verifiche automatiche o da interventi manuali semplificati, limitano le operazioni noiose, ripetitive e pericolose, offrono affidabilità di esecuzione e precisione nei tempi di attuazione, ottimizzano l'energia dissipata.

Attraverso l'ottimizzazione si potrà conseguire, per le singole operazioni, un più alto grado di affidabilità, precisione e sensibilità. Proprio nell'agricoltura, dove spesso l'uomo non ha in modo spiccato alcune attitudini, dove è necessario ridurre il margine di errore in tutte quelle operazioni svolte in modo soggettivo, dove è imperativo realizzare risparmi di risorse (materiali, energie, tempo e lavoro).

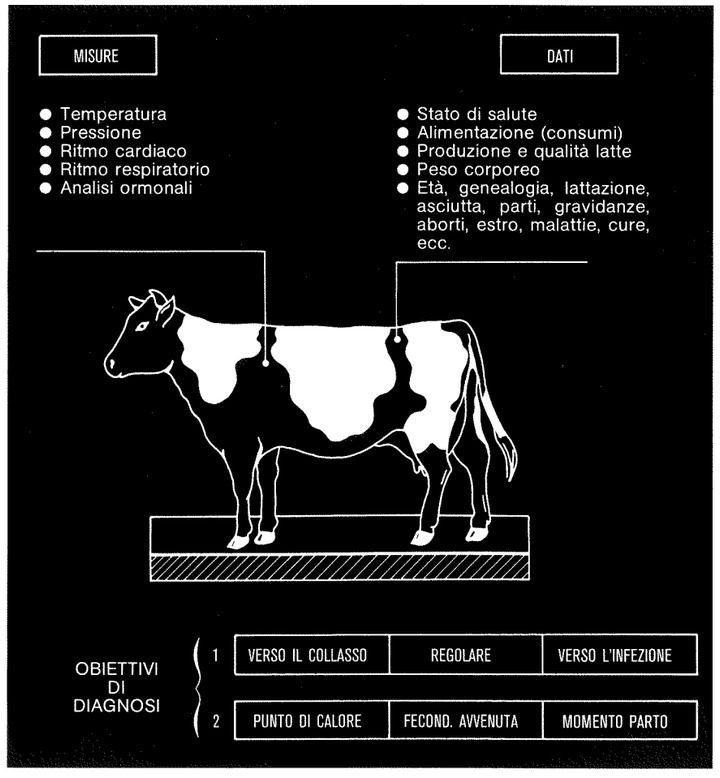


Fig. 8 - *Diagnostica animale.* Lo stato di salute degli animali (bovini, suini, ecc.) e alcune manifestazioni particolari (gravidanza, parto, ecc.) potranno essere diagnosticate direttamente dall'allevatore mediante computer portatili o da apparecchiature installate nei locali. Il rilevamento di alcuni parametri fisiologici potrà essere effettuato in modo tradizionale (misurazione a contatto, analisi, ecc.) o automaticamente attraverso sonde sottocutanee di tipo intelligente.

Non ultima, la capacità di elaborare un notevole numero di informazioni in tempo reale trova largo spazio anche nel settore agricolo, dove sono anche richiesti calcoli particolarmente complessi e con notevole variabilità dei dati, la multifunzionalità delle operazioni, la trasmissione veloce a distanza.

3. Aree di intervento

3.1 Dalla semina al raccolto

Una prima classificazione dei tipi di intervento è caratterizzata da una serie di obiettivi finalizzati a ridurre il margine di errore in molteplici attività agricole oggi svolte in modo soggettivo.

Sia che si parli di colture in pieno campo che di ambienti confinati come le serre, si tratterà di pilotare direttamente alcune lavorazioni e di fornire indicazioni attendibili circa il momento ottimale per effettuare interventi sul suolo e sulle vegetazioni.

Da un lato le stazioni meteorologiche disseminate sul territorio e dall'altro la climatizzazione delle serre, rappresentano in questa ottica le realizzazioni più convenzionali e conosciute. Le attuali potenzialità dell'elettronica e il notevole sviluppo in atto, consentono di delineare in prospettiva nuovi e più efficaci interventi.

Attraverso l'impiego di centraline microclimatiche si potrà conoscere il momento ottimale per effettuare interventi sul suolo (aratura, semina, ecc.) e sulle colture (crescita, raccolta, ecc.). Si potrà attuare una lotta guidata per combattere diversi tipi di fitopatie (come la ticchiolatura delle pomacee e la peronospora della vite) evitando trattamenti anticrittogamici inutili o superflui; sarà possibile prevenire danni alle colture derivanti da agenti atmosferici come il gelo e la brina. Nell'irrigazione e fertirrigazione si potrà ottimizzare l'uso delle risorse idriche, migliorare la resa delle aree coltivate e ridurre i costi degli impianti.

L'azione, in sintesi, consiste nell'attuare una serie di interventi che concorrono ad aumentare complessivamente la resa dell'agricoltura, attraverso l'applicazione di tecniche appropriate sul fronte del recupero dei suoli, sull'uso corretto dell'acqua e dell'humus, sull'estensione delle serre: tecniche che l'elettronica può rendere convenientemente praticabili e più efficienti.

3.2 Nella stalla

Nella zootecnia, l'applicazione più ricorrente è il controllo dell'alimentazione delle bovine da latte negli allevamenti a stabulazione libera. Ogni mucca è dotata di un collare che identifica l'animale nell'ambito del gruppo. Quando accede alla stazione di alimentazione (la mangiatoia), il collare emette un segnale di identificazione in codice che viene trasmesso all'elaboratore. Controllato che il codice sia compatibile con il programma di alimentazione precedentemente inserito, viene attivata la distribuzione del mangime. Il computer può controllare, quindi, il consumo giornaliero di ogni bovina, il numero di soggetti che assumono una percentuale minima programmata, il consumo globale dell'impianto, la distribuzione del concentrato per ogni singola stazione di alimentazione.

Nelle sale di mungitura, macchine, impianti e sistemi automatici sono operanti da anni e oggetto di continuo perfezionamento. Dispositivi elettronici controllano i tempi di pulsazione, lo svuotamento delle mammelle, la rimozione automatica del gruppo di mungitura, la quantità di latte appena munto e la percentuale di grassi contenuti.

Guardando al futuro, l'obiettivo più ambizioso è quello di consentire direttamente all'allevatore, attraverso l'introduzione di nuove metodologie diagnostiche, il monitoraggio dello stato di salute degli animali. Oltre alla prevenzione e al controllo dei singoli soggetti (attività regolare, infezione, collasso, ecc.), indicazioni di notevole utilità riguardano il momento ottimale per l'inseminazione artificiale (punto di calore), l'avvenuta fecondazione e l'approssimarsi del parto. In funzione dello stato di salute, del peso e di altre caratteristiche tassonomiche e genealogiche, si potrà migliorare il controllo computerizzato dell'alimentazione in linea, con un giusto rapporto costi/benefici.

3.3 L'automazione

Automazione agronica significa sinergismo di controlli e azionamenti, frutto di un insieme coordinato di tecniche e tecnologie diverse, dove l'elettronica svolge un ruolo determinante di controllo e ottimizzazione.

Si tratta, in genere, di aumentare il margine di sicurezza degli impianti, ridurre i costi di gestione, ottenere la massima produttività aziendale, migliorare la qualità dei prodotti.

Una tipica applicazione è il controllo dei mangimifici dove l'elettronica può rendere automatico il dosaggio dei componenti riducendo tempi di preparazione e costi di produzione, migliorando la precisione e il grado di omogeneità del mangime.

Ci sono in prospettiva ipotesi di sviluppo interessanti: conoscere alcune indicizzazioni tipiche dei prodotti insilati come il grado di ammuffimento, di conservabilità, di salubrità ecc.; controllare essiccatoi di derrate agricole e celle frigorifere ottimizzando il problema in termini di energia/spazio/tempo. Così come notevoli possibilità di intervento vanno ricercate nei sistemi di produzione di energie da biomasse: pilotare direttamente il compostaggio attraverso misurazioni automatiche convalidate da analisi chimico-biologiche integrative; aumentare la resa dei digestori anaerobici condizionando il funzionamento alla variabilità dei dati in input (volume, composizione, temperatura del compost, ecc.); automatizzare il processo di compressione e distribuzione del biogas.

Tutta una serie di attività, insomma, (dalla trasformazione e conservazione dei prodotti agricoli, alla produzione di energie rinnovabili e integrative), potranno ricorrere all'impiego dell'elettronica per automatizzare in tutto o in parte i loro cicli produttivi.

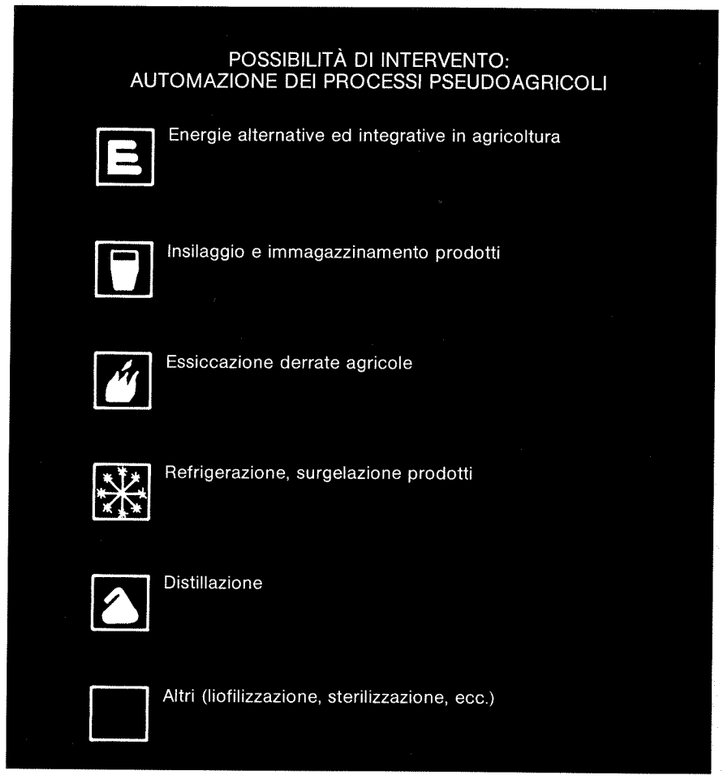


Fig. 9 - *Automazione dei processi paragricoli.*
Questo tipo di intervento, destinato alle aziende di trasformazione e conservazione dei prodotti agricoli, è il maggiormente diffuso poichè automazioni e strumentazioni sono affini al settore industriale. L'intervento è tuttavia ancora modesto, orientato soprattutto al controllo del ciclo produttivo mentre sono ancora poche le applicazioni sul controllo qualità e risparmio energetico.

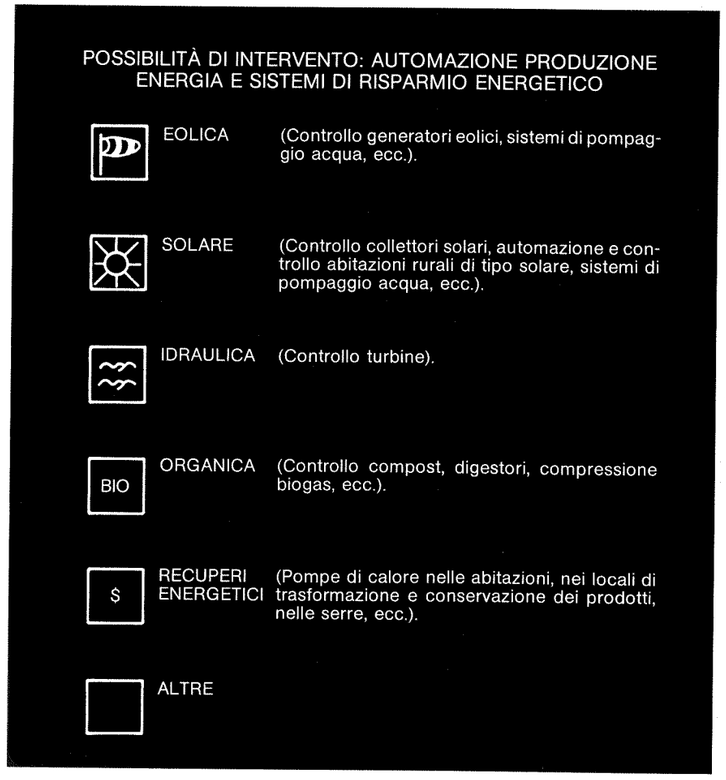


Fig. 10 - *Automazione sistemi di produzione energia.*
L'agricoltura non produce solo cibi e materiali ma anche alcuni tipi di energie da considerarsi alternative e integrative. L'intervento dell'elettronica è destinato sia all'automazione dei processi produttivi sia al recupero degli specchi energetici attraverso l'uso delle pompe di calore.

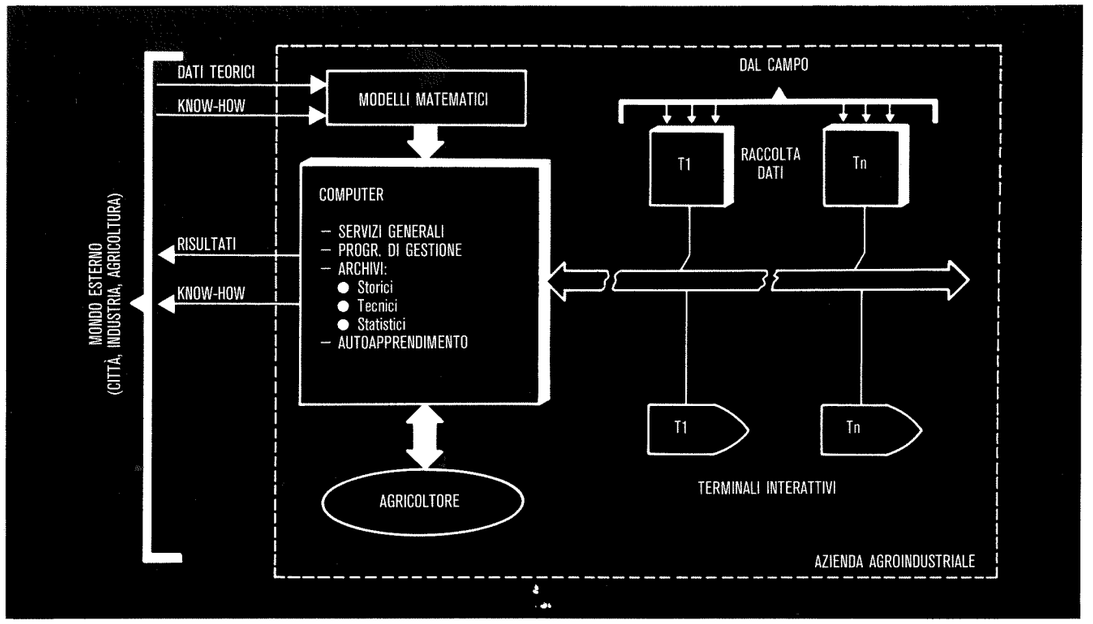


Fig. 11 - *Gestione computerizzata delle aziende agricole.*
Una unità di servizio centralizzata gestisce i servizi generali (contabilità, fatturazione, ecc.), pianifica il ciclo produttivó agricolo, gestisce il parco macchine operatrici, raccoglie i dati dal campo. L'unità costituisce anche un centro di esperienze significativo capace anche di interagire con il mondo esterno.

3.4 L'informatica verde

Dopo i successi conseguiti nel mondo industriale, l'elettronica introduce il trattamento automatico delle informazioni anche nel settore agro-zootecnico. E' nata così l'informatica verde: uomini e computers per un nuovo modello di management dell'impresa agricola.

Dalle classiche attività gestionali (contabilità, fatturazione, ecc.) ai problemi di vera strategia imprenditoriale, il computer può dare un contributo efficace per contenere il progressivo aumento dei costi, migliorare i margini di profitto e svolgere ogni pratica amministrativa.

Oltre alla zootecnia (gestione di allevamenti bovini, suini, ecc.), l'impiego dell'informatica trova larghi spazi in quelle situazioni dove elaborazioni tecnico-scientifiche e problematiche contabili-amministrative sono fortemente caratterizzate dalla continua variabilità dei dati produttivi.

Prestazioni tipicamente agricole, come la formulazione di un corretto piano di alimentazione animale, gestione e pianificazione del ciclo produttivo, proposte alternative nella rotazione delle colture, statistiche e proiezioni sulla resa delle aree coltivate, archivi storici e geografici, possono operare da elementi razionalizzatori e costituire un valido supporto alla conduzione aziendale.

La quasi totalità delle applicazioni è realizzata con grandi calcolatori, orientati quindi a centri di elaborazione o comunque a una agricoltura ricca e industrializzata. Un servizio oggi destinato a pochi che può essere esteso solo se prospettato in chiave collettiva cioè presso Associazioni di categoria e Cooperative agricole.

Maggior successo potrà avere invece il "personal computer". E' la strada già intrapresa da alcune software-house impegnate a sviluppare programmi applicativi per l'agricoltura. Economico, compatto, affidabile, sembra il tipo di calcolatore più indicato per attività manageriali in questo campo.

Certo gli obiettivi sono ben altri anche se lontani. L'introduzione di computer portatili ("hand computer"), per esempio, specificatamente sviluppati e orientati alla soluzione di problemi tipicamente agricoli.

Si tratta cioè di prodotti ad alto grado di integrazione tecnologica, estremamente compatti, capaci di elaborare in base a dei programmi di utilità, dati raccolti a mezzo di sonde esterne e dati introdotti dall'agricoltore, fornendo una soluzione razionale a ciascun problema proposto.

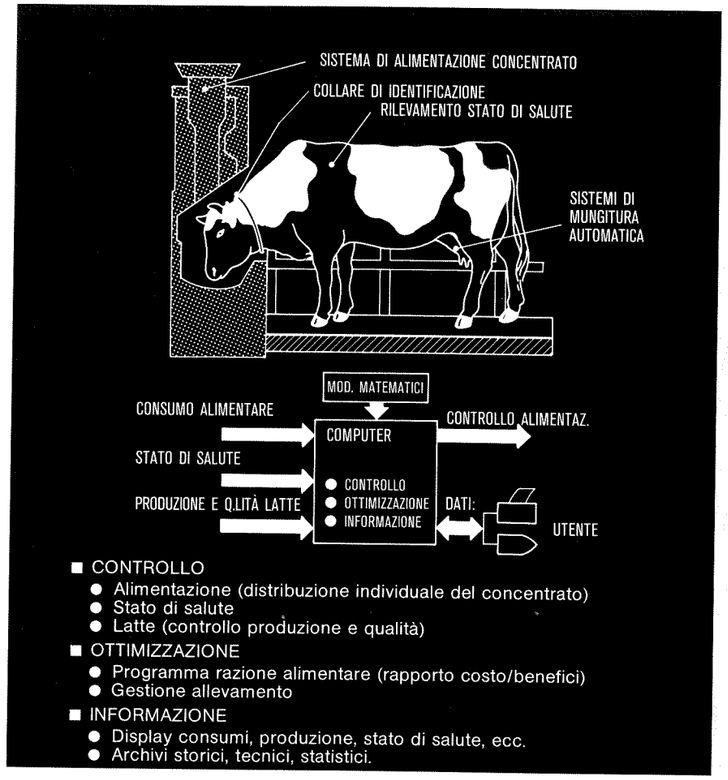


Fig. 12 - *Gestione computerizzata di un allevamento di bovine da latte a stabulazione libera.*

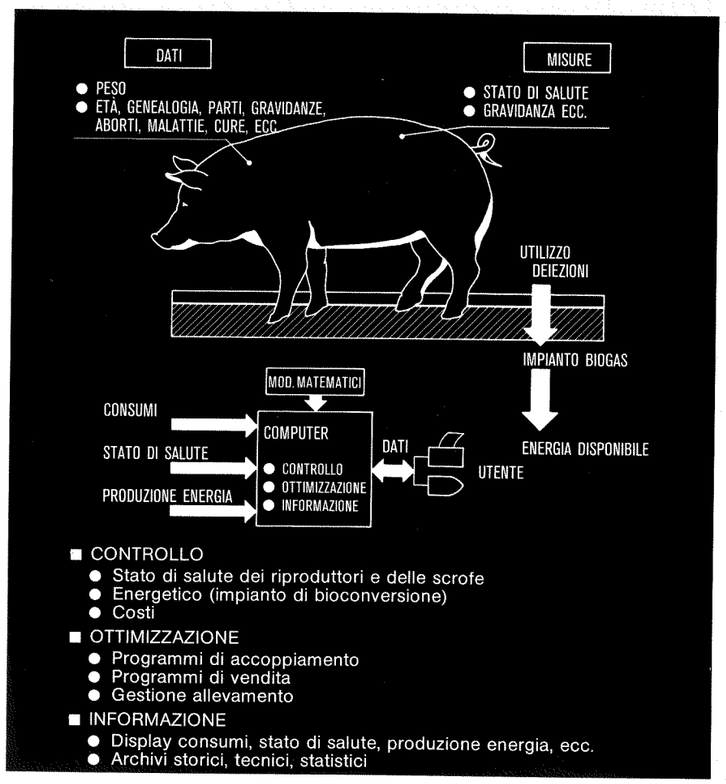


Fig. 13 - *Gestione computerizzata di un allevamento suinicolo.*

4. Problemi per lo sviluppo

Molti problemi, alcuni noti altri nuovi, condizionano lo sviluppo dell'elettronica per l'agricoltura.

Sono difficoltà tipiche dell'agronica che derivano dalla natura specifica dei problemi, dal rapporto con il nuovo tipo di utente e dalle particolari condizioni ambientali.

4.1 I trasduttori

In primo luogo la necessità di "trasduttori", la cui disponibilità contenuta ne condiziona fortemente lo sviluppo.

Poichè i sensori presenti sul mercato (temperatura, pressione, ecc.) consentono un numero limitato di applicazioni, nasce la necessità di sviluppare dispositivi specifici per il settore agricolo. Se questo può sembrare (e in parte lo è) un fattore limitante di sviluppo, occorre però non arrivare ad affrettate conclusioni. L'impiego di uno o più sensori oggi disponibili, consente già di effettuare un discreto numero di interventi, specie nel controllo microclimatico. Si tratta semmai di realizzare, a costi contenuti, raggruppamenti di sensori come ad esempio piccole stazioni meteo. Nè bisogna dimenticare come spesso basta rilevare un solo parametro (quello giusto) al posto di più elementi difficilmente rilevabili e come nuovi trasduttori possono essere realizzati abbinando uno o più sensori disponibili a dispositivi fisico-chimici.

Limitazioni e condizionamenti tecnologici non consentiranno per molto tempo ancora il controllo della sensorialità sul territorio e nell'ambiente, già oggi ritenuto necessario, anche se nuovi rilevatori (specie nel settore optoelettronico e microonde) sono in fase di avanzata applicazione in molti laboratori di ricerca.

Nella tabella sono elencati alcuni degli elementi la cui rilevazione automatica potrà consentire sviluppi estremamente interessanti.

Idrogeno naturale	H_2
Ossigeno naturale	O_2
Acqua	H_2O
Anidride Carbonica	CO_2
Acido Solfidrico	H_2S
Ammoniaca	CH_3
Metano	CH_4
Alcool Etilico	C_2H_5OH
Alcool Metilico	CH_3COOH
Glucosio	$C_2H_{16}O_6$
Amido e Cellulosa	$(C_6H_{10}O_5)_n$

4.2 L'utente

Altra tematica importante concerne l'interazione *uomo-macchina* ovvero gli aspetti che riguardano l'interfaccia fisica e le metodologie operative di input-output con l'agricoltore.

Le attività interattive non differiscono sostanzialmente da quelle normalmente in uso: dati e informazioni vengono introdotti a mezzo tasti specifici o tastiere, i risultati visualizzati con spie luminose o display. Raramente sono pensabili supporti per l'introduzione automatica dei dati (floppy-disk, badge, ecc.) mentre talvolta può essere necessaria la stampa (ricevuta, grafici, ecc.).

Le caratteristiche dell'interfaccia fisica dipendono dal tipo e complessità del problema da risolvere: potranno bastare dei semplici "led" per indicare l'andamento di un impianto irriguo; sarà magari necessario un display grafico per visualizzare lo stato di crescita di una coltura vegetale.

Queste tematiche risultano assai complesse perchè proposte in una situazione ambientale e culturale molto delicata. Il colloquio tra l'agricoltore e la macchina deve svolgersi con modalità semplici, intuitive, quasi confidenziali, attraverso elementi le cui caratteristiche ergonomiche e funzionali devono essere curate nei minimi particolari.

Manovre effettuate con procedure formalmente corrette ma secondo modalità non ortodosse (ad esempio azionare un tasto con forza), manovre accidentali, oppure eseguite con procedure errate o in contemporaneità non compatibili, manovre abusive (o dolose) sono tipicamente più ricorrenti che in altre situazioni.

Accorgimenti ergonomici a parte, nuove soluzioni vanno anche ricercate verso forme avanzate di dialogo. La dinamicità nella presentazione dei messaggi, la composizione sintetica ma estremamente chiara dei testi, la grandezza e la luminosità dei caratteri, la disposizione logistica dei comandi, l'uso di una simbologia appropriata, sono alcuni elementi che possono concorrere a semplificare e mitigare il rapporto agricoltore-computer. L'introduzione guidata delle informazioni, la presentazione "in chiaro" dei messaggi, indicazioni visive semplici e inequivocabili sono in questo contesto attributi indispensabili.

Particolarmente interessante in alcune applicazioni è la trasmissione a distanza delle informazioni (dal campo all'abitazione agricola, ai centri di assistenza tecnica, ecc.) dei dati più significativi e dello stato di funzionamento dell'unità. In questo caso, problematiche inerenti il tipo di collegamento (via radio, via filo, ecc), le potenze in gioco, possibili interferenze esterne, costituiscono fattori da non sottovalutare.

4.3 L'ambiente

In tema di resistenza alle sollecitazioni ambientali, è evidente che le apparecchiature destinate all'agricoltura debbono presentare particolari caratteristiche. Ciò vale sia nelle condizioni di funzionamento che durante lo stazionamento dei prodotti (a volte anche di intere stagioni), quando cioè l'unità pur non operando rimane installata per lunghi periodi. Le caratteristiche strutturali del prodotto, oltre che a robustezza e semplicità di design, devono quindi tener conto delle particolari sollecitazioni esterne di natura termica, chimica, elettrica e meccanica.

Il caldo d'estate (oltre i 40°C) e il freddo d'inverno (−20°C) impongono campi di temperatura e umidità piuttosto ampi, mentre la possibile dislocazione geografica degli apparecchi può dare luogo a eventuali criticità derivanti dalla pressione atmosferica.

La presenza di sostanze chimiche in concentrazioni più o meno elevate nell'area circostante il prodotto (si pensi ai trattamenti anticrittogamici) può danneggiare la carrozzeria così come particolari condizioni ambientali possono dar luogo a problemi di polverosità e illuminamento. Analoga importanza riveste la resistenza agli urti e alle vibrazioni, specie durante il trasporto delle apparecchiature nei campi. Scariche atmosferiche, infine, soprattutto nei periodi estivi, possono se non distruggere, compromettere seriamente il funzionamento dell'unità.

Accorgimenti per un rapida installazione e manutenibilità sono elementi essenziali così come la rigorosa applicazione di tutta la normativa di sicurezza a livello di protezione antinfortunistica (impianto elettrico, carrozzeria, sovratemperatura delle parti accessibili, perturbazioni verso l'esterno, ecc.).

4.4. Il modello matematico

Anche il modello matematico, cioè l'algoritmo che regola il funzionamento dell'apparecchiatura consentendo la "soluzione" di determinati problemi di natura agricola, necessita di alcune considerazioni.

Come per i trasduttori, la questione si pone innanzitutto in termini di "disponibilità" (oggi abbastanza limitata) e in termini di affidabilità. Il programma di utilità, inoltre, deve essere di tipo dinamico, flessibile alle varie situazioni microclimatiche e caratterizzato da un notevole grado di sicurezza in termini di "copertura".

Alcuni modelli, particolarmente semplici, possono essere sviluppati dallo stesso progettista elettronico, cablati in hardware o implementati a livello firmware senza particolari difficoltà. La maggior parte, invece, per la natura e complessità del problema, sono affidati a specialisti del settore ed Enti di ricerca che provvedono alla formulazione e alla successiva verifica sperimentale in campo. Evidente, in questo caso, come lo sviluppo di un prodotto agronico induca una pluralità di rapporti tra elettronica e mondo esterno coinvolgendo esperti di altre discipline (agronomia, geologia, chimica, biologia, ecc.). Nella diagnosi animale, per esempio, si collaborerà con esperti in veterinaria con know-how sulla patologia e fisiologia dei ruminanti, nello sviluppo di centraline atte a prevenire infezioni crittogamiche si opererà con fitopatologi e così via.

4.5 L'alimentazione delle apparecchiature

L'alimentazione delle apparecchiature è normalmente realizzata con batterie di tipo commerciale. Sono rare e riguardano per lo più automazioni di impianti pseudo-agricoli (insilaggio, essicazione, ecc.) le applicazioni con rete alternata.

L'alimentazione a batteria, mentre evita la presenza di disturbi in ingresso, ha in genere altri tipi di inconvenienti come l'autonomia limitata, maggiori ingombri, difficoltà di connessione, resistenza alle sollecitazioni ambientali, ecc.

La disponibilità di una sola tensione (di solito 12V), specie in presenza di circuiti analogici, comporta la generazione di altre alimentazioni con valore e polarità diversa; però, per ragioni pratiche, è sempre preferibile a più batterie tra loro interconnesse.

Nelle apparecchiature agroniche, limitare il consumo è ovviamente imperativo. Ciò è ottenibile ricorrendo soprattutto a tecnologie low-power (CMOS, ecc.), ad una accurata ottimizzazione delle funzioni circuitali e a originali soluzioni di progetto. Nel caso di centraline microclimatiche, per esempio, si può ipotizzare una logica di controllo che, oltre a gestire le funzioni principali, abbia il compito di isolare le alimentazioni dei circuiti ausiliari attivandole solo quando è necessario.

4.6 In campo

Superati i problemi in sede progettuale, altri devono essere attentamente valutati in relazione alle attività in field.

Le difficoltà di installazione sono facilmente intuibili così come particolarmente difficoltosi si presentano gli interventi di manutenzione e riparabilità del prodotto in campo. L'intervento sulle macchine da parte dell'utilizzatore, infatti, non può che essere limitato alle normali operazioni d'uso, alla sostituzione globale in caso di guasto e alla pulizia della carrozzeria. Difficilmente può essere richiesto all'agricoltore di garantire la continuità di funzionamento attraverso interventi sul prodotto anche se per questi non sono previste particolari cognizioni tecniche.

E' difficile, inoltre, disporre di un feed-back corretto dal campo per elaborazioni statistiche, specie se mancano sistemi diagnostici implementati nel prodotto. E' importante (ma purtroppo solo augurabile) che in caso di non funzionamento, il tipo di anomalia non incida sensibilmente sul livello di confidenza con la macchina. Un guasto permanente, anche se di grave entità, è preferibile a guasti saltuari e intermittenti.

5. Conclusione

L'agricoltura, in tutti i suoi molteplici aspetti, costituisce un enorme potenziale campo di applicazione delle tecniche dell'elettronica e dell'informatica. E' lecito affermare che l'uso di queste tecniche, oggi ancora praticamente all'inizio, consentirà di modificare in modo sostanziale approcci e soluzioni tradizionali e di dare un nuovo impulso a questo fondamentale settore delle attività dell'uomo.

Tecnologie informatiche e mass-media

ENRICO CARITÀ

Editrice La Stampa
Torino

1 - Introduzione

Quale sia stato il ruolo svolto dalle tecnologie elettroniche nel trasformare dall'interno il mondo dell'editoria, nel corso dell'ultimo decennio, è fatto che ormai appartiene alla storia, più che alla cronaca, delle applicazioni industriali nei paesi più evoluti.

Vi sono stati, è vero, degli sfasamenti notevoli, anche di parecchi anni, fra i vari paesi e, nell'ambito di una stessa nazione, fra le aziende più avanzate e le più inerti a cogliere l'apporto delle tecnologie all'ottimizzazione della gestione.

Ne discende che, ad esempio, scelte attuate negli USA da case editrici di avanguardia agli inizi degli anni '60 non hanno ancora sfiorato le più statiche aziende italiane alla fine degli anni '70.

Per non parlare dei paesi dell'Europa orientale che, da questo punto di vista, sono oggi al livello raggiunto negli anni '30 dai paesi più progrediti.

Nonostante le notevoli distanze che, da un paese all'altro caratterizzano il livello di modernità delle aziende editrici, la situazione oggi è tale da consentire, a fronte dei necessari investimenti e di un contesto sociale sufficientemente ricettivo, di recuperare i ritardi anche per i paesi e per le imprese che stanno in zona di retroguardia.

Questa opportunità, veramente senza precedenti, che si presenta in questo singolare momento storico, di bruciare le tappe del tempo perduto, è da cogliere senza esitazioni, affrontando con coraggio e determinazione anche le inevitabili ripercussioni sociali. E ciò non solo in ordine alle necessità, di giorno in giorno più pressanti, di conferire anche alle aziende editrici, come ogni altro tipo di impresa, gli adeguati livelli di efficienza e di produttività che ne garantiscano l'indipendenza economica, soprattutto a fronte di una trasformazione di ben più vasta portata, e dalle conseguenze difficili da definire nell'arco di un ragionevole orizzonte temporale.

In una recente intervista apparsa su La Stampa [1] Giovanni Giovannini, presidente della FIEG (Federazione Italiana Editori Giornali) afferma infatti "non ci troviamo solamente di fronte all'evoluzione tecnologica di un settore, per quanto importante: siamo all'inizio di un nuovo modo di comunicare, e quindi di vivere insieme, con implicazioni culturali, politiche e storiche che si cominciano appena ad avvertire.... Noi pensiamo troppo agli aspetti tecnologici di ogni nuovo mezzo per poter veramente dire che cosa ci potrà dare in futuro questa nuova tecnologia. E' certo che essa creerà dei nuovi bisogni, dei nuovi servizi di informazione e penso sia nostra responsabilità di imprenditori dell'informazione di prevedere il più possibile il futuro per evitare quegli errori che abbiamo fatto in passato, perdendo prima il treno della radio e poi quello della televisione".

E, ne "La sfida mondiale" Jean Jacques Servan Schreiber riprende quanto sostenuto da André Danzin, secondo il quale l'analisi del mondo contemporaneo porta ad evidenziare un'evoluzione da un modello agricolo a uno industriale, per arrivare poi ad una società di servizi e, in futuro, a una "società dell'informazione".

Partendo da queste premesse, il "Club di Parigi" arriva ad anticipare che, a fronte dei radicali mutamenti tecnici che sono ormai in mezzo a noi, non possiamo che attenderci come naturale conseguenza una vera e propria rivoluzione sociale. L'informatizzazione della società verrà a determinare nuove necessità di scambi, di conoscenze, di interdipendenze, che apriranno nuove opportunità di lavoro e stimoleranno nuove attività professionali.

Se, in prospettiva, questo quadro di riferimento è corretto, ne consegue, come ricordavo nel libro "Una sfida per la stampa", che l'editoria, e quella quotidiana in particolare, è chiamata a confrontarsi con una realtà estremamente mutevole che l'avvolge dall'esterno e la modifica dall'interno. E' opportuno che essa si concentri meno sul prodotto e più sulla funzione che viene assolta dal prodotto; in altri termini è necessario che essa reinterpreti le proprie finalità alla luce delle necessità informative del tempo e della società che si propone di servire.

Per convincersene basterà considerare l'effetto di contrazione delle dimensioni spazio-temporali prodotte anche solo dai meno recenti fra i media di natura elettrica, quali il telefono e la radio, e come attraverso essi si siano modificati i concetti di tempestività e di esigenze informative che ci si aspetta da un medium stampato.

Una delle cause determinanti di questo mutamento di "ambiente" si ritrova senza alcun dubbio nella evoluzione delle tecnologie di base.

Fra queste l'informatica, fino alle sue più recenti evoluzioni di tele-informatica, sta giocando un ruolo che non è ancora completamente definito nei suoi contorni. Secondo alcuni l'incontro fra l'editoria e l'informatica è destinato ad avere conseguenze maggiori della stessa invenzione della stampa.

Cercheremo nel seguito di valutare il fondamento di una tale convinzione, che ai più potrà sembrare eccessiva, sotto tre diversi punti di vista.

Dapprima si esaminerà lo stadio di sviluppo dei "sistemi editoriali", cioè dei sistemi di tele-elaborazione adattati alle esigenze della redazione e della composizione di prodotti editoriali, per comprendere come viene a modificarsi *dall'interno* l'azienda editrice, ed il quotidiano in particolare.

In seguito si cercherà di valutare le conseguenze indotte dall'ambiente tecnologico esterno per effetto della diffusione di "reti di trasmissione" e di "banche di dati" a prevalente contenuto giornalistico e di documentazione.

Infine, si daranno alcuni elementi per valutare l'insorgere di *nuove esigenze del mercato*, a fronte del crescente volume di informazioni che tende a saturare le capacità umane di ricezione dei messaggi.

2 - I sistemi editoriali

Con questo termine ci si riferisce a quei sistemi di teleelaborazione che svolgono tutte le fondamentali funzioni, dall'arrivo delle notizie all'elaborazione redazionale, alla composizione dei testi e dei titoli fino, se possibile, alla impaginazione con generazione del prototipo di pagina chiusa, con eventuale incisione diretta della lastra di stampa.

Va detto che all'inizio degli anni '80 un sistema editoriale che assolva tutte queste funzioni in modo completo non esiste ancora. Alcuni sistemi editoriali sono più completi o più efficienti di altri: tutti tendono comunque ad assolvere una varietà di funzioni e a includere sia le funzioni redazionali che quelle produttive.

Se si eccettua la parte finale del processo, cioè l'impaginazione e l'incisione della lastra, si osserva che già alla fine degli anni '70 le caratteristiche dei sistemi editoriali hanno raggiunto un livello di maturazione notevole, al punto che non si prevede cambieranno nei prossimi anni in modo significativo le loro funzioni fondamentali.

Si assisterà certamente ad un appiattimento delle caratteristiche dei diversi sistemi, nel senso che le prestazioni dei meno completi tenderanno ad allinearsi a quelle dei più efficienti.

Terminali video intelligenti, piccoli sistemi editoriali, facilità di collegarli fra loro e con sistemi editoriali di più grandi dimensioni, tutto ciò renderà agevole la soluzione di collegamenti fra redazioni periferiche e redazioni centrali, così come sarà possibile realizzare un immediato e permanente scambio di informazioni con gli abituali collaboratori che svolgono la loro attività a domicilio.

Per i sistemi editoriali ci si attende nei prossimi anni anche una maggiore facilità d'uso e una maggior completezza nell'esecuzione delle funzioni redazionali.

I sistemi editoriali potranno integrare una quantità crescente di altre funzioni, oltre a quelle tradizionali, mediante la progressiva standardizzazione sia di componenti fisici (hardware) che di pacchi di programmi (software); attraverso il processo di standardizzazione sarà anche possibile avere degli sviluppi comuni fra editori in diversi settori, quale la raccolta della pubblicità, lo sviluppo di banche di informazioni, l'utilizzo di servizi redazionali comuni.

Nell'ambito dei sistemi editoriali non ha ancora trovato un'adeguata soluzione tutta l'area dell'impaginazione. A tutt'oggi esistono i singoli moduli di un sistema in grado di consentire una completa impaginazione elettronica, fino all'uscita diretta su lastra di stampa. Tuttavia, alla luce delle passate esperienze, sarà necessario almeno un quinquennio per ottenere una completa trasferibilità alla produzione di quelli che oggi sono dei prototipi di apparecchiature.

E' opinione diffusa che per la metà degli anni '80 tutte le fasi di lavorazione riguardanti il trattamento di testi, titoli, immagini, pubblicità, fino all'incisione delle lastre di stampa, saranno effettuabili senza interventi manuali. Ovviamenti, non tutti i sistemi saranno così completi da includere il trattamento delle immagini e l'incisione diretta delle lastre di stampa; ma anche nel caso dei sistemi più semplici si potrà arrivare alla generazione di pagine intere di giornale su film.

Per concludere, è prevedibile che l'evoluzione dei sistemi editoriali si svilupperà secondo le seguenti linee di tendenza:

- possibilità di crescita modulare a partire dalle funzioni fondamentali già note e consolidate;
- capacità di assolvere ad una globalità di funzioni gradualmente integrabili fra loro: dalla notizia alla lastra;
- maggiore interattività (cioè dialogo più facile ed esauriente fra uomo e macchina), anche per effetto del miglioramento delle caratteristiche dei terminali video;
- più facile personalizzazione, tarata sulle esigenze di ciascun utente;
- possibile integrazione con le elaborazioni tipicamente gestionali e con quelle di "office automation";
- collegabilità a banche di dati interne o esterne e, più in generale a reti informatiche complesse;

Da quanto finora esposto, emerge chiaramente come il ricorso a sistemi editoriali trasformi profondamente gli strumenti e le modalità di lavoro in quello che è il nucleo centrale della funzione editoriale: ricezione-redazione-composizione delle notizie, fino alla produzione del prototipo per la stampa.

Essi non solo consentono di realizzare drastiche riduzioni nei costi di produzione, anche rispetto alla fotocomposizione tradizionale, ma sono indispensabili come mezzi di accesso a strutture di banche di dati e a reti di trasmissione dati.

Per questo è vitale la loro adozione, anche se non sono di lieve entità i problemi sociali che ne conseguono. Esistono, inoltre, dei motivi che travalicano di molto per importanza i pur essenziali obiettivi di economicità della gestione aziendale.

Infatti, in quanto sistemi di teleinformatica, i sistemi editoriali diventeranno veicolo di innesto delle tecniche di produzione editoriali in un più ampio contesto di evoluzione tecnologica.

Il sistema editoriale tenderà a configurarsi all'interno di una casa editrice come la terminazione nervosa di un sistema globale dell'informazione, costituito da computer, banche di dati, reti di trasmissione, di cui tutto l'attuale "mondo dell'informazione" si troverà a far parte.

Sui sistemi editoriali va fatta una notazione sulla presenza, o meglio, sulla totale assenza di fornitori nazionali di una qualche dimensione.

Mentre l'industria italiana sembra avere acquisito un notevole grado di consapevolezza sull'importanza dell'OI (Office Automation) e sugli strumenti del CAD (Computer Aided Design) e del CAM (Computer Aided Manufacturing), sembra invece essere sfuggito ai più che i sistemi editoriali, riunendo simultaneamente le tre tecniche, possono costituire un formidabile terreno di sperimentazione.

La conseguenza di questa disattenzione porta il nostro paese ad una totale dipendenza tecnologica dall'estero, non senza conseguenze sotto il duplice aspetto culturale e industriale.

3 - Reti di trasmissione e Banche Dati

L'utilizzo nelle aziende editrici di sistemi editoriali consente di integrare le risorse interne con quelle di altri sistemi esterni, tramite reti di trasmissione a livello nazionale e sovranazionale.

La disponibilità di reti di trasmissione dati a diversi livelli di prestazione e di compatibilità con le strutture di comunicazione preesistenti, è ormai universalmente riconosciuta come una condizione essenziale per la transizione da una società industriale ad una società dell'informazione.

Quando negli anni scorsi nel nostro paese si sono di fatto create le premesse per un sostanziale rallentamento degli investimenti nelle reti di trasmissione, si è data una brusca frenata ad una evoluzione positiva della società.

Secondo i componenti del "Club di Parigi", le infrastrutture essenziali della società informatizzata non rappresentano quelle rigide strutture la cui acquisizione costò alle società industrializzate numerosi decenni. Alle reti ferroviarie, fluviali, stradali corrisponderanno le reti di telecomunicazione, di informazione, di istruzione e di formazione secondo le più moderne tecnologie.

Le reti di trasmissione sono strumenti insostituibili per la condivisione delle risorse fra diversi sistemi e, sopprattutto, per lo scambio reciproco di informazioni fra diverse banche di dati.

L'abbinamento dei "sistemi editoriali" con "banche di dati" esterne, alle quali sia possibile interconnettersi mediante reti di comunicazione di elevate prestazioni e di basso costo, consentirà, in una prospettiva temporale che ci auguriamo ragionevole anche per il nostro paese, di avere dei "nodi" con proprie capacità elaborative e di memorizzazione, dai quali si potranno ottenere, a seconda delle necessità o dei desideri, le notizie per il quotidiano o il servizio per il periodico o il notiziario per la stazione radio televisiva o la risposta sul video di casa a domande fatte dal privato tramite il suo televisore.

Questo tipo di considerazioni ci porta a spostare la nostra attenzione sulla disponibilità di archivi di redazione su computer. In effetti il continuo sviluppo di impianti di elaborazione a costi decrescenti, e lo sviluppo di nuovi tipi di memorie di massa, porteranno ad una crescente diffusione di banche di informazioni giornalistiche.

Banche di dati di questa natura, affiancandosi a quelle già esistenti o in corso di sviluppo, di tipo giuridico, economico, scientifico, ecc., andranno ad arricchire il crescente patrimonio di conoscenze reso accessibile in tempo reale.

In attesa che si sviluppino adeguatamente gli strumenti software di trattamento automatico del linguaggio naturale, rimarrà scoperto uno spazio di mediazione rispetto all'utente generico, in cui anche l'editoria qualificandosi fra i "fornitori di informazioni" potrà trovare un proprio ruolo.

Gli attuali archivi di redazione sono costituiti, nella generalità dei casi, da ritagli di quotidiani o di riviste raccolti in buste, identificate per argomento. Le fotografie sono trattate in modo analogo. Questo metodo di immagazzinamento delle informazioni ha certamente il pregio di fare ricorso a soluzioni semplici e a bassa intensità di capitale investito. Ma d'altro canto, chi può dichiararsi sinceramente soddisfatto del servizio che viene offerto da un archivio di redazione tradizionale?

Pur essendo largamente soggettiva la valutazione sulla "bontà" di un archivio ad uso redazionale, va detto che esistono delle difficoltà strutturali nella gestione di un archivio tradizionale, che vengono ai nostri giorni progressivamente aggravate dai crescenti volumi di informazioni da trattare.

L'archivio di redazione può oggi trasformarsi in una moderna banca di dati dell'azienda editrice avvalendosi delle tecniche dell'elaborazione elettronica dei dati e della microfilmatura degli originali.

Una volta che tutto il contenuto del quotidiano, o di altra pubblicazione, sia già in forma elaborabile dal computer per la composizione, è immediato immaginare la realizzazione di un sistema computerizzato d'archivio che abbia i testi immagazzinati nelle sue memorie di mas-

sa. In pratica è piuttosto frequente nelle banche dati giornalistiche non ricorrere alla memorizzazione integrale dei testi, ma utilizzare una soluzione di tipo misto, per cui il computer viene utilizzato per le operazioni di ricerca, mentre la memorizzazione degli originali viene effettuata su microfilm o su microschede (generalmente chiamate microfiches). In questo modo nel computer non viene fatta la memorizzazione dei testi completi, che comporterebbe una quantità di memorie di massa dai costi impraticabili; è sufficiente memorizzare le parole chiave ed un estratto (abstract) di ciascun testo, oltre alle informazioni necessarie per rintracciare l'originale su microfilm.

La soluzione mista computer-microfilm, pur essendo in teoria meno sofisticata di un archivio completamente elettronico, in pratica presenta una serie di vantaggi in termini di costo, flessibilità di applicazione, gradualità di approccio per l'automazione dell'archivio di redazione.

Il livello di automazione di un archivio è certamente frutto di un compromesso fra la quantità e la qualità di servizi che si vogliono realizzare ed i corrispondenti costi di realizzazione e di esercizio.

Attualmente, le banche dati giornalistiche accessibili in tempo reale all'interno delle redazioni non vanno oltre la ventina in tutto il mondo, mentre quelle accessibili anche ad utenti esterni non superano la mezza dozzina.

Fra le banche dati che non hanno la memorizzazione integrale dei testi, ma utilizzano estratti e parole-chiave prodotti da specialisti, le più significative sono la "New York Times Information Bank", che risale agli inizi degli anni '70, e la Banca dati di "Gruner & Jahr" di Amburgo, in funzione dalla metà degli anni '70. In Italia opera sulla base di analoghe metodologie l'unica banca dati di contenuto giornalistico creata dall'ANSA, la nostra maggior agenzia di stampa che appartiene agli editori di giornali.

Fra le banche dati a memorizzazione integrale dei testi, abbiamo quelle del "Toronto Globe and Mail" e del "Boston Globe". In queste la estrazione delle parole chiave viene effettuata automaticamente. Le tecniche di elaborazione del linguaggio per l'estrazione delle parole-chiave sono ancora all'inizio del loro sviluppo, per cui le capacità di risposta univoche di questi sistemi sono in genere limitate ad interrogazioni molto semplici.

In ogni caso è necessaria sia l'aggiunta alle parole-chiave di una serie di informazioni complementari sugli articoli memorizzati, sia l'intervento di specialisti di documentazione nel caso di domande complesse.

La costruzione di banche di informazioni giornalistiche di grandi dimensioni comporta degli investimenti e dei costi di esercizio che trovano difficilmente una giustificazione economica all'interno di una singola azienda editrice.

E' necessario quindi considerare le informazioni memorizzate su computer come la base per nuovi tipi di prodotti/servizi, in modo da poter trasferire sulla vendita delle informazioni a clienti esterni una buona parte dei costi di sviluppo e di esercizio.

Attorno alla costruzione di banche di informazioni sembra naturale che si sviluppino delle forme di collaborazione fra aziende editrici non solo per ripartire i costi, ma anche per evitare assurde duplicazioni di lavoro e di archivi.

Questo di per sè potrà modificare l'habitat in cui ciascuna impresa editrice viene ad operare. Anche per questo si è ritenuto di dover fare qualche accenno alle caratteristiche di banche di dati di tipo giornalistico.

La costruzione di banche di dati giornalistiche non ha valide alternative per chi domani non voglia restare escluso dall'accesso a "nuovi media", quali il videotex, e condizionato nella gestione dei vecchi.

Esse non costituiranno che una parte numericamente limitata, anche se qualitativamente importante, nell'ambito di uno scenario in cui, negli anni futuri, si assisterà ad una proliferazione di banche dati, in termini di quantità e diversificazione di contenuti.

Le possibilità di accesso multiplo e in tempo reale a questi "serbatoi" informativi, attraverso le maglie di reti di trasmissione ad elevate prestazioni, alterando in modo rilevante le potenzialità e le caratteristiche dei flussi informativi, lasciano intravedere delle profonde modificazioni nelle modalità di memorizzazione e di diffusione delle conoscenze.

Soprattutto la selezione rapida dell'esatta informazione, nel momento desiderato, fra una quantità crescente di dati disponibili corrisponderà ad una esigenza fondamentale per l'evoluzione delle società industrializzate: la quantità di informazioni cresce infatti di giorno in giorno, fino al punto di spingere alcuni a parlare di inquinamento da informazioni.

Il problema riguarda gli individui, ma ancora più le organizzazioni complesse, che rischiano di essere schiacciate dalla mole di materiale informativo da esse stesse prodotto.

4 - Nuove esigenze del mercato

Prima di chiudere questa panoramica sulle conseguenze, già in atto o prevedibili, delle tecnologie elettroniche sui mass-media e, in particolare, sui prodotti editoriali, è opportuno accennare all'emergenza di nuove esigenze nel mercato dell'informazione.

I mezzi di comunicazione disponibili appartengono ad aree specifiche: interpersonali, per piccoli gruppi, per il mercato di massa locale, regionale, nazionale. Tecnologie diverse sono associate con specifici tipi di contenuto: il telefono non si è sviluppato (come qualcuno avrebbe potuto pensare) come un mezzo di intrattenimento, ma è stato usato per affari e per l'invio di messaggi personali; né è stato sviluppato come uno strumento di propaganda di massa ma per la conversazione. Altre tecnologie coprono l'area dell'intrattenimento e della propaganda. Oggi, tuttavia, una ridistribuzione di ruoli è possibile. Il cavo telefonico diventerà quasi certamente il supporto di trasmissione per certi tipi di informazione pubblica, forse anche per forme speciali di intrattenimento. La radio e la televisione potrebbero essere usate per la distribuzione di materiale stampato, e nuove specie di fibre ottiche possono essere alla base del video-telefono o, forse, di qualche nuovo tipo di comunicazione televisiva locale.

Antony Smith, in un recente libro, ci ha fatto conoscere i risultati di una ricerca condotta dal Ministero giapponese delle Poste e Telecomunicazioni.

Questa ricerca ha portato a tracciare una mappa dei sistemi di informazione esistenti mettendo in evidenza la loro estensione che va dai mezzi di comunicazione interpersonali ai mezzi di comunicazione di massa (fig. 1).

Emerge chiaramente un'area scoperta per l'informazione specializzata (cioè di piccola "audience"); è in questo terreno vuoto che i nuovi sistemi elettronici di informazione faranno la loro comparsa durante il periodo iniziale della loro introduzione. Ovviamente, ci sono dei vuoti nell'estensione dei mezzi di informazione nella società, vuoti che infatti sono stati parzialmente riempiti, indipendentemente dal computer.

L'enorme crescita dei periodici specializzati, in Nord America in particolare, e la crescita delle comunicazioni via etere in Europa sono entrambe manifestazioni dello stesso fenomeno. I mezzi di informazione e di intrattenimento non stanno soddisfacendo adeguatamente gruppi socialmente importanti che richiedono qualche cosa di intermedio fra la conversazione di salotto ed il materiale annacquato e di basso livello prodotto per il mercato di massa.

I periodici specializzati, che hanno tratto profitto dal fatto di soddisfare qualcuna delle esigenze di questo pubblico crescente, non riforniscono una élite intellettuale, ma danno informazioni dettagliate che servono a chi segue un particolare stile di vita.

I cittadini delle società più evolute della fine del XX° secolo sono pronti a spendere una gran parte dei loro ricavi e del loro tempo in "hobby" e in occupazioni secondarie.

I mezzi di comunicazione tradizionali non sono necessariamente sull'orlo di qualche sorta di caduta: piuttosto, essi si trovano nella posizione di aver aiutato a stimolare una gamma di gusti molto più ampia di quanto essi possano ora soddisfare.

L'ipotesi sottesa nel diagramma giapponese è che esista una "audience" di grandi quantità di informazioni di natura specialistica o semispecialistica che può essere distribuita sulla base di scelte individuali.

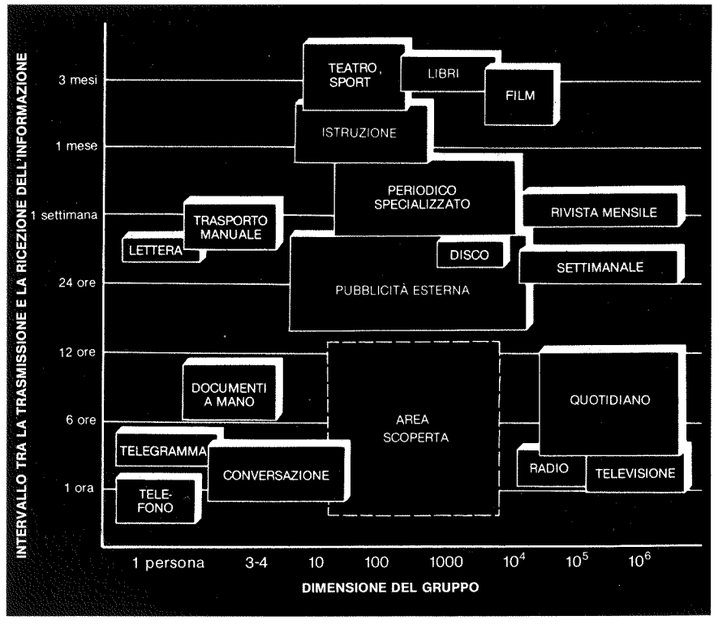
Simultaneamente, sono stati scoperti modi del tutto nuovi di distribuire queste informazioni.

Si tratta dei cosidetti "nuovi media", per i quali nei vari paesi si pone oggi l'interrogativo se e quando portarli a livello di servizio pubblico.

Per concludere, possiamo osservare come il mondo dell'editoria, ma soprattutto quello dei quotidiani, è stato sottoposto negli ultimi anni ad una quantità di pressioni e forze contrastanti, in parte provenienti dal mondo esterno, e in parte generate all'interno delle aziende. Fra queste, anche le spinte tecnologiche hanno svolto un ruolo determinante, e ancora più importante sarà il loro peso nei prossimi anni.

Informazione e informatica - o meglio, telematica - sono destinate a produrre una sorta di sinergismo ri-

Fig. 1 - *Mappa dei sistemi di informazione (da "Scienza '81").*



sonante, che conferisce alla prima i ritmi di espansione che sono propri della dinamica della seconda, per cui l'assetto atteso per il "nuovo" mondo dell'informazione tenderà a riconfigurare - in un arco di tempo che sta comunque al di fuori di ogni ragionevole capacità di previsione - con strutture assai più ricche, e quindi anche assai diverse da quelle attuali.

Questa accelerazione che è stata iniettata dall'esterno al preesistente mondo dell'informazione postula la capacità di svolgere un ruolo attivo nel definire modalità e strumenti di lavoro e nel ridisegno di responsabilità e di strutture organizzative.

Soprattutto il ricorso a sistemi editoriali presuppone una analisi delle caratteristiche dell'aziende editrice che faccia riconoscere da un lato gli elementi di somiglianza con gli altri media, dall'altro l'esigenza di rivedere l'organizzazione della struttura redazionale. E sotto questo aspetto sarebbe utile avvalersi della "teoria dei sistemi".

Una cosa è certa: che le tradizionali divisioni in cui viene ripartito il mondo delle comunicazioni di massa si andranno gradualmente dissolvendo. Si ha una crescente interdipendenza fra i diversi media ed un compenetrarsi dei contorni di ciascun settore, soprattutto nella prospettiva ormai imminente dell'avvento di nuovi media quali il videotex, e della possibilità di accesso a strutture sempre più imponenti di banche di dati.

Tentando di valutare le conseguenze dell'innesto dell'informatica nel mondo editoriale, il rischio maggiore che si corre è di sottovalutarne la portata. E ciò per il fatto che il settore in esame ha delle peculiarità proprie che lo differenziano in modo essenziale da altri settori industriali.

La differenza di fondo rispetto ad altre imprese consiste nel fatto che l'informazione nel nostro caso non è solo veicolo di conoscenza e di controllo dell'ambiente esterno (mercato) e di quello interno (prodotto), ma diventa anche "materia prima" del processo di trasformazione che porta al prodotto finito.

5 - Bibliografia

[1] GIOVANNI GIOVANNINI, *La stampa del 29/1/82* - Inserto Telematica - pag. 15

[2] JEAN JACQUES SERVAN SCHREIBER, *La Sfida Mondiale* - Mondadori - 1980

[3] ENRICO CARITÀ, *Una sfida per la stampa* - Etas Libri - 1981

[4] ANTONY SMITH, *Goodbye Gutenberg* - Oxford University Press - 1980

Il computer assiste il guidatore nella scelta dei percorsi urbani

PASQUALE FIORE

Borsista A.T.A.
Centro Ricerche FIAT
Orbassano (Torino)

1. Introduzione

La crescita sempre più rapida e caotica degli odierni agglomerati urbani, ed il conseguente aumento del parco dei veicoli circolanti sulle reti viarie cittadine, sta facendo assumere dimensioni enormi al problema del traffico urbano.

Ingorghi e code ai semafori sono molto frequenti, ed a farne maggiormente le spese sono gli utenti abituali della rete urbana: taxisti, guidatori di mezzi di soccorso, etc.

Può essere allora molto utile studiare un "sistema informatico" particolare, che automaticamente ricerchi quale è il percorso che è preferibile seguire per muoversi da un punto di partenza ad un punto d'arrivo, fissati dall'utente, evitando il più possibile le zone di traffico pesante o addirittura saturo.

Usualmente, fra tutti i possibili percorsi che congiungono i due punti suddetti, alcuni consentono una maggiore facilità e rapidità di spostamento di altri.

Il modo più immediato di quantificare il concetto di "facilità e rapidità di spostamento", è quello di definire un "costo" o, più correttamente, un "funzionale di costo".

(*) Il materiale di questo articolo è tratto dalla tesi di laurea dell'Autore presso il Politecnico di Torino. A tale tesi è stato assegnato un premio della Honeywell Information Systems Italia.

Quest'ultimo associa un particolare valore numerico ("costo") a ciascuno dei possibili percorsi tramite cui si può raggiungere il punto d'arrivo desiderato, dal punto di partenza fissato.

È evidente allora che il tragitto seguito per spostarsi tra i due punti scelti è tanto migliore quanto più basso è il suo costo, e di conseguenza il percorso ottimo è quello che ha il "costo minimo".

Molteplici sono i fattori che influenzano il costo di un particolare percorso.

I più importanti sono:
- numero di semafori presenti
- numero e tipo di svolte da effettuare (una svolta a sinistra richiede un tempo notoriamente maggiore di una a destra)
- tipo di fondo stradale
- distanza totale da percorrere
- condizioni di traffico lungo il percorso

I primi quattro fattori sono costanti per il percorso fissato, mentre l'ultimo, come è facilmente intuibile, è variabile nel corso della giornata.

Dunque il costo totale di un qualunque percorso nella rete stradale urbana varia nel tempo, per cui la scelta del tragitto effettuata "a naso" dal guidatore può facilmente rivelarsi la peggiore, anzichè la migliore.

Un sistema informatico "di assistenza" che ricerchi automaticamente il percorso migliore per ciascun utente, elimina evidentemente questo problema.

Si noti che questo sistema deve poter misurare in continuazione i flussi di traffico nella rete (per mezzo di opportuni sensori), al fine di garantire l'affidabilità del servizio fornito.

2. Descrizione del sistema di assistenza

Come già accennato, il *servizio* che il sistema automatico deve attuare, è quello di fornire all'utente, su un terminale a bordo del veicolo, il percorso in quel momento ottimo fra i due punti preventivamente indicati dall'utente stesso.

Le parti fondamentali componenti il sistema devono quindi essere:
- insieme di "canali di comunicazione" fisici
- sistema di calcolo centrale
- sistema di "acquisizione dati".

Un primo schema funzionale a blocchi può perciò essere quello mostrato in fig. 1. In esso sono messe in evidenza le tre parti fondamentali menzionate, e le loro interazioni con il "mondo esterno", costituito dal "processo traffico" e dall'insieme degli utenti del servizio.

Il sistema di calcolo centrale è l'insieme degli elaboratori elettronici ("hardware") e dei programmi ("software") che elaborano la risposta alla richiesta di ciascun utente.

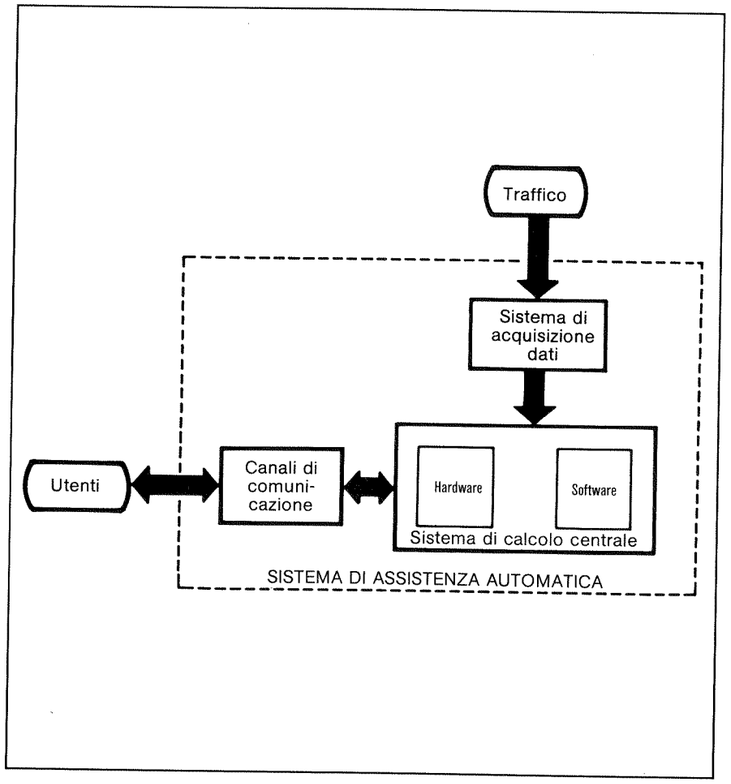


Fig. 1 - Schema funzionale del sistema di assistenza automatica.

Esso, per le sue funzioni di coordinatore dell'intero sistema, deve essere dislocato in posizione opportuna nella città.
Il suo hardware, per il carico di lavoro a cui dev'essere presumibilmente soggetto, è previsto che si basi su due "mini-computer", cosicchè ciascuno si accolli una parte dell'intero carico di elaborazioni da svolgere.

Il software "applicativo" che è eseguito dal sistema di calcolo, implementa gli algoritmi per la ricerca del percorso di costo minimo tra due punti della rete viaria (simulata), e aggiorna i dati dei costi di percorrenza dei singoli tratti di percorso.

Per "canali di comunicazione" si intendono tutte quelle catene di dispositivi atte a trasmettere le informazioni necessarie, dall'utente alla centrale di calcolo e viceversa.

Ciascun canale può essere visto come una sequenza di dispositivi posti uno in cascata all'altro.

Partendo dall'utente, il primo dispositivo deve essere un terminale posto in vettura, come già accennato. Questo deve interpretare i comandi dell'utente, e deve trasmetterli ad appositi ricevitori "locali" dislocati lungo le strade della rete viaria. Peraltro esso deve ricevere i segnali di risposta, e li deve presentare in modo opportuno al guidatore del veicolo.

Per le comunicazioni fra vettura e rice-trasmettitori locali ci si deve servire, come evidente, di un "canale hertziano", per cui sono necessarie ancora una antenna a bordo del veicolo, ed alcune antenne fisse a terra.

Naturalmente sono necessarie più stazioni ("centraline") rice-trasmittenti locali, dislocate sulla rete viaria reale, ed esse devono essere tante più, quanto più fitto e complesso si vuole sia l'insieme di strade assistito.

In questo sistema di assistenza è molto pressante la necessità di sgravare il più possibile la unità di calcolo centrale di tutti quegli oneri computazionali che potremmo chiamare secondari. La soluzione migliore in tal senso è quella di inserire più "micro-computer" nel sistema, proprio al livello delle "centraline locali" menzionate sopra. I compiti principali di questi microelaboratori sono quelli di codificare, decodificare, e compattare nel formato più idoneo sia i messaggi di richiesta provenienti dall'utente, sia i messaggi di risposta che giungono dal sistema di calcolo.
La architettura dell'intero sistema risulta allora di tipo gerarchico-distribuito, in quanto si ha una compenetrazione fra una struttura informatica gerarchica, in cui il sistema di calcolo centrale è ad un livello superiore mentre i micro-computer locali sono ad un livello inferiore, ed una struttura distribuita, in cui l'intero onere delle elaborazioni viene suddiviso fra numerosi dispositivi intelligenti. Fanno ovviamente parte del blocco "canali di comunicazione" anche i collegamenti tra centraline locali e sistema di calcolo centrale. Questi è preferibile realizzarli via cavo, come sarà

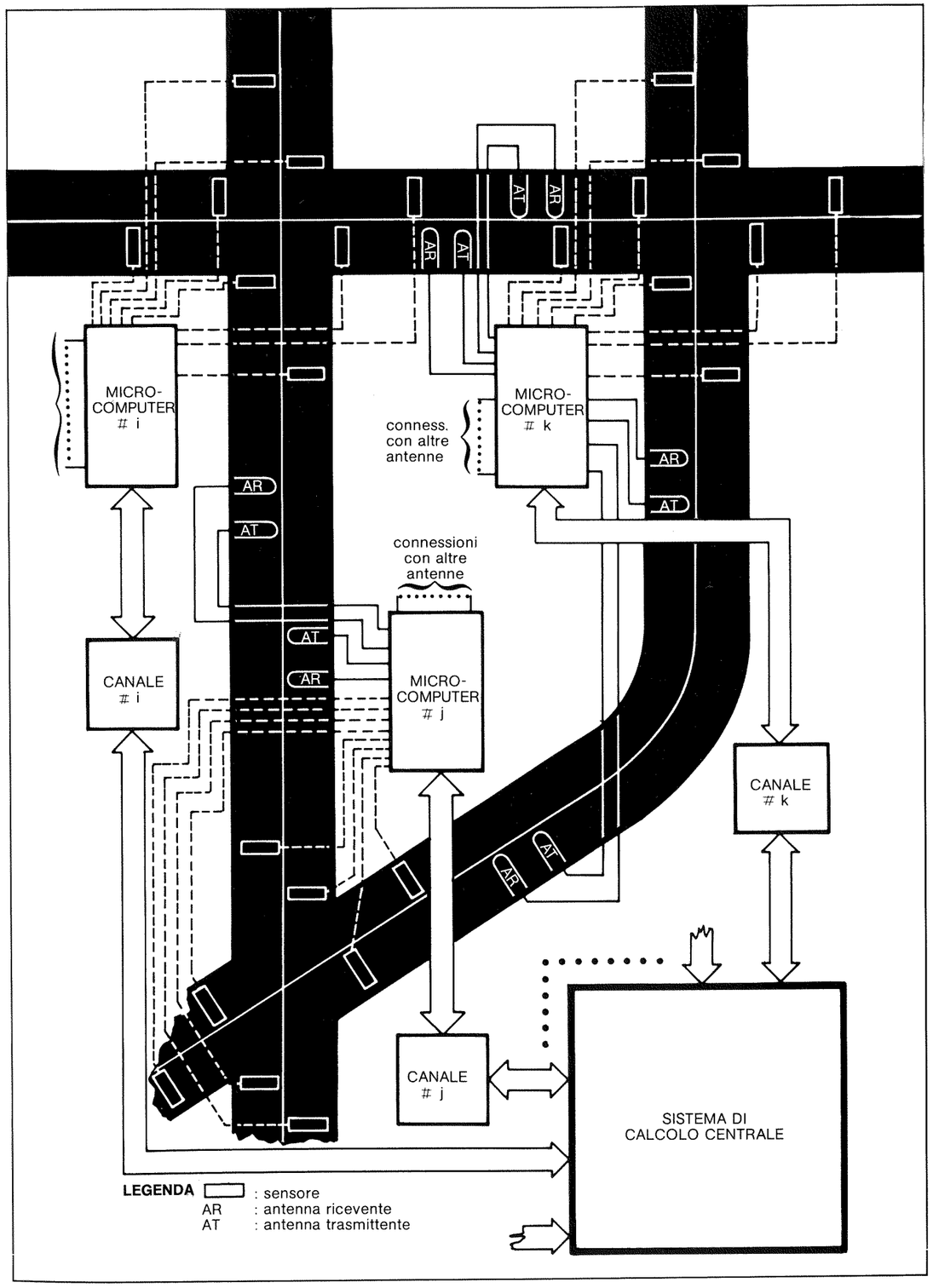


Fig. 2 - Schema reale del sistema di assistenza.

spiegato nel paragrafo 4.

Il sistema di "acquisizione dati" è poi l'insieme di sensori e trasduttori che rilevano e misurano "on line" i flussi di traffico nei vari punti della rete. Tramite le informazioni così ottenute il sistema di calcolo centrale può facilmente e costantemente tenersi aggiornato sulle condizioni di traffico su tutta la rete, e quindi può consigliare agli utenti dei percorsi che sono sempre i migliori.

Per trasmettere al sistema di calcolo centrale i segnali uscenti dai sensori, si può molto semplicemente utilizzare i canali già predisposti, collegando tali segnali ai microcomputer locali.

Lo schema reale di tutto il sistema di assistenza viene dunque a configurarsi come nella fig. 2. Si è supposto, nella figura, di mandare ad uno stesso processore tutti i segnali provenienti dai sensori di uno stesso incrocio. Questa è una cosa ragionevole per il fatto che la dislocazione del generico microelaboratore è proprio in prossimità del crocevia.

3. Software applicativo del sistema di assistenza

Passiamo ad esaminare ora quella parte che è stata poco prima chiamata "software applicativo", cioè quell'insieme di programmi che permettono al sistema di calcolo centrale di espletare il suo particolare servizio.

Come già accennato sono due le funzioni necessarie:

- aggiornamento dei dati dei costi di percorrenza dei singoli tratti di strada
- ricerca del percorso di costo minimo da un punto di partenza ad un punto di arrivo comunicati dall'utente, operando su una opportuna schematizzazione della rete viaria.

L'interazione del software applicativo con il "modo esterno" può dunque essere esemplificata dallo schema di fig. 3. Il primo gruppo di programmi manipola i dati provenienti dai sensori, aggiorna i costi di percorrenza, ed immagazzina i risultati in un'area di memoria comune. Il secondo gruppo di programmi attinge i dati di base da questa area comune, e ricerca il percorso di costo minimo relativo alla particolare richiesta dell'utente.

Questi ultimi programmi operano su una particolare schematizzazione della rete viaria urbana, in cui i tratti "elementari" di percorso non sono i tronchi di strada ("link") compresi tra due incroci, bensì i tragitti di attraversamento di un incrocio, in uno dei modi possibili. Più correttamente, essi sono i tragitti necessari per portarsi da un certo senso di marcia di un "link" ad un altro senso di marcia, appartenente ad un "link" vicino.

La fig. 4 mostra uno di questi percorsi elementari, p_k, che per convenzione parte ed arriva nei punti di mezzo dei due "link" interessati.

Fig. 3 - Interazioni del software applicativo con il mondo esterno.

Fig. 4 - Percorso elementare nella rete viaria urbana.

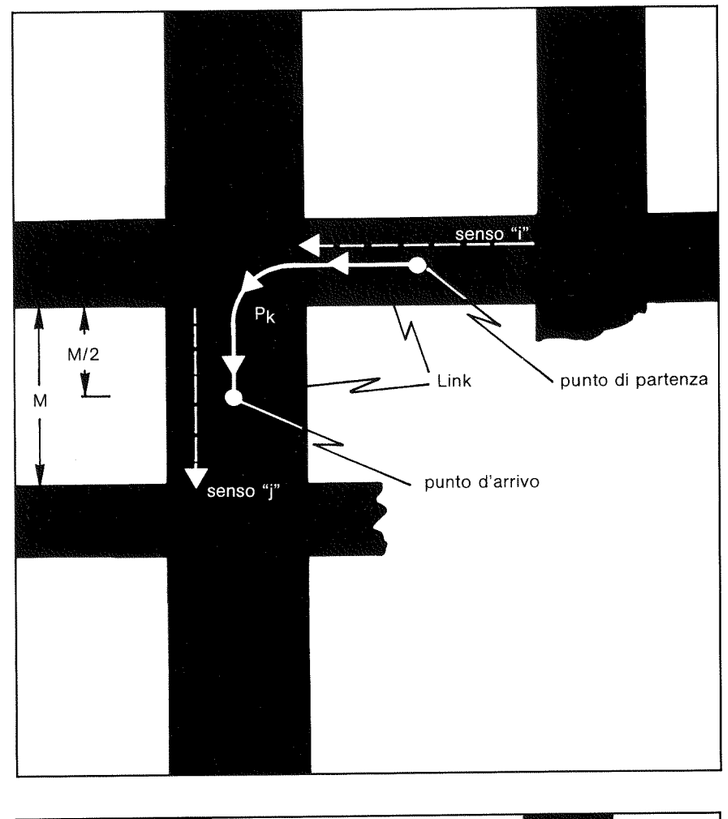
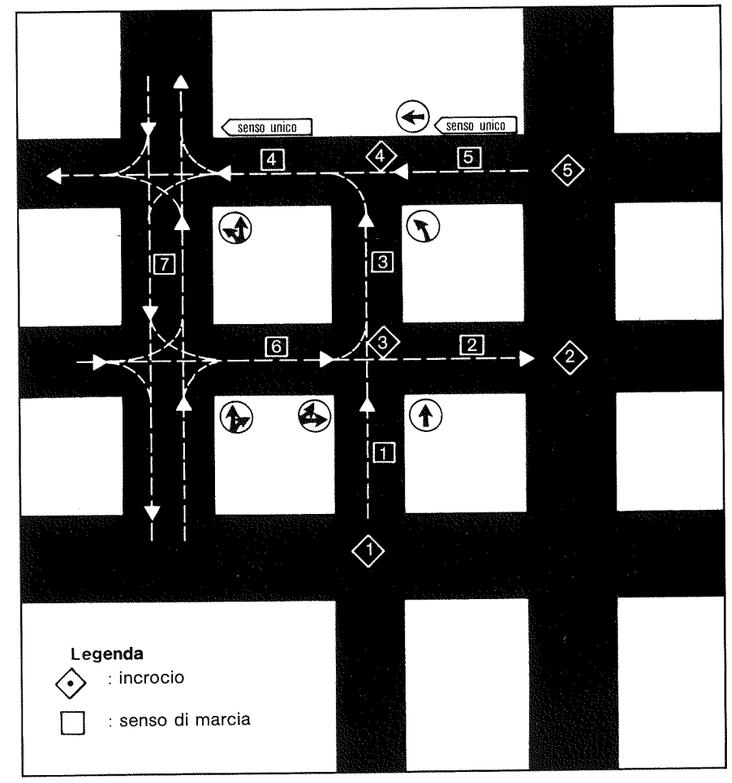


Fig. 5 - Esempio di rete viaria regolamentata da sensi unici e divieti di svolta.



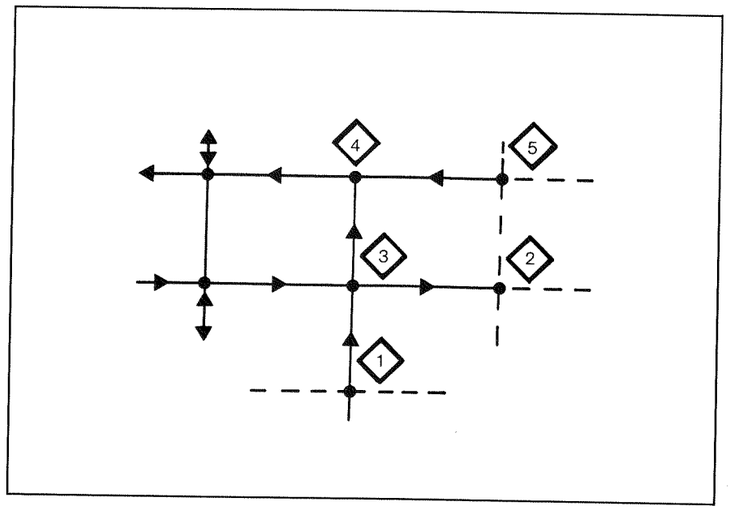


Fig. 6 - Grafo orientato errato.

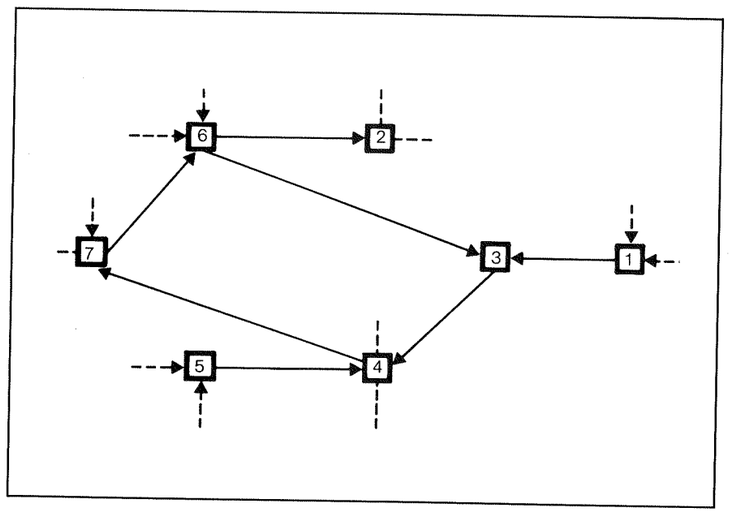


Fig. 7 - Grafo orientato corretto.

L'insieme di tutti i percorsi elementari di questo tipo esistenti nella rete viaria forma un cosiddetto "grafo orientato", che costituisce la schematizzazione sopra menzionata.

Il motivo della scelta dei percorsi elementari come in fig. 4, è che in una rete viaria urbana sono solitamente numerosi gli interventi di regolamentazione della circolazione veicolare, quali istituzioni di sensi unici e divieti di svolta.

Ne consegue che numerosi "link" della rete sono percorribili solo in un senso, oppure li si può imboccare solo se si proviene da particolari direzioni.

Si esamini come esempio la piccola rete mostrata in fig. 5, dove sono indicati i sensi di marcia e le svolte possibili, in ossequio alla segnaletica presente.

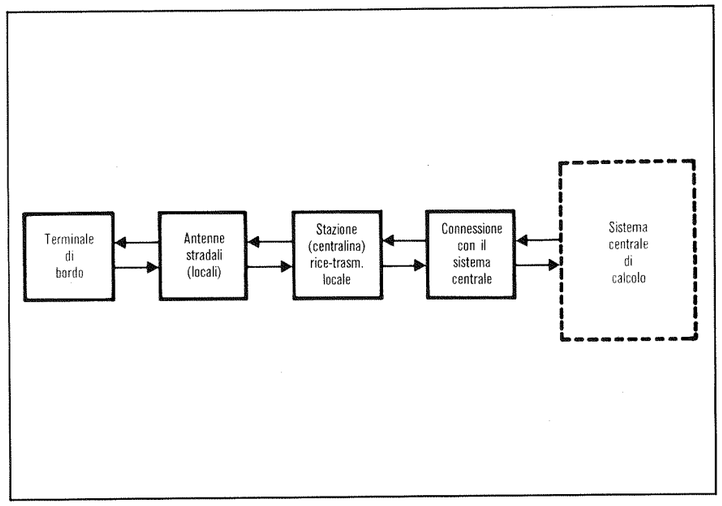
Se si prendessero come percorsi elementari i singoli "link", si otterrebbe il grafo orientato di fig. 6, in cui i "nodi" sono gli incroci e gli "archi" orientati rappresentano i "link" esistenti, con i sensi di marcia permessi al loro interno.

Secondo tale grafo parrebbe dunque possibile passare direttamente dall'incrocio 1 al 2 , attraverso l'incrocio 3 . Ciò è evidentemente errato, in quanto esiste divieto di svolta dal senso di marcia 1 nel senso di marcia 2 (fig. 5).

Dunque il grafo orientato da ricavare è quello che ha come nodi tutti i sensi di marcia esistenti nella rete, e come archi orientati le svolte che è possibile effettuare da un senso di marcia in un altro. Se da un senso di marcia "i" non si può in alcun modo svoltare (o passare) direttamente in un altro qualsivoglia senso "j" della rete, nel grafo corrispondente non esiste alcun arco orientato che dal nodo "i" porti direttamente al nodo "j". Nel caso dell'esempio, questo grafo risulta come in fig. 7.

Risulta evidente dunque che, per

Fig. 8 - Schema del canale di comunicazione bidirezionale.



definire il costo di un qualsivoglia tragitto nella rete reale, basta assegnare un "costo elementare" a ciascun percorso elementare, cioè a ciascun arco orientato del grafo.

Allora per rappresentare sinteticamente tutto il "grafo orientato ad archi quotati" così ottenuto, basta definire una "matrice dei costi elementari" C = [c(i,j)] i cui elementi siano convenzionalmente:

c(i,j)
- = 0 se i=j
- = ∞ se non esiste l'arco orientato dal nodo "i" al nodo "j"
- > 0 se esiste l'arco orientato da "i" a "j".

Risulta chiara adesso la funzione dei programmi di "aggiornamento dati" di cui s'è parlato prima. Essi infatti ricalcolano di volta in volta gli elementi della matrice C, usando sia i dati fissi sui costi dei singoli percorsi elementari, memorizzati una volta per tutte, sia i dati variabili dipendenti dal traffico, provenienti dai sensori.

Si noti a tal proposito che è ragionevole considerare le condizioni del traffico nella rete "stazionarie nella media" per un intervallo di tempo almeno superiore al più lungo tempo di percorrenza totale della rete.

Infatti i microelaboratori locali effettuano un prefiltraggio (media nel tempo) dei segnali uscenti dai sensori, per cui il sistema di calcolo centrale non è costretto a "riaggiornare" continuamente la matrice C.

I programmi di ricerca del percorso di costo minimo operano sulla matrice C memorizzata nella area comune, e implementano l'algoritmo proposto nel 1959 da Dijkstra ([1], [2]). Esso, che è il più efficiente per questo tipo di problemi, riceve in ingresso la matrice C e i nodi x_p (partenza) e x_d (destinazione) comunicati dall'utente, e fornisce in uscita il percorso cercato.

L'intero algoritmo si compone di due parti distinte. Un primo algoritmo, operando sulla matrice C del grafo e sul nodo x_d fissato, ricerca i costi minimi complessivi per giungere a x_d, partendo da tutti i restanti nodi. Successivamente un secondo algoritmo utilizza questi risultati ed il particolare nodo x_p, fissato, per fornire la successione ordinata di tutti i nodi toccati dal percorso minimo tra x_p e x_d, che è quel percorso a cui corrisponde il costo minimo complessivo fra x_p e x_d (già calcolato dal primo algoritmo).

4. Hardware del sistema di assistenza

Come s'è già detto, tutta la parte hardware del sistema completo può pensarsi suddivisa in tre blocchi fondamentali:

- canali di communicazione
- parte hardware del sistema di calcolo centrale
- sistema di acquisizione dati.

Ricordiamo che il primo blocco è quello che permette agli utenti di "colloquiare" col sistema di calcolo centrale, mentre il secondo blocco è quello che si interessa della gestione di tutto il software applicativo. Il terzo blocco, infine, serve a raccogliere i dati sulla domanda di traffico nella rete viaria urbana, e ad inviarli al sistema di calcolo centrale.

A) Canali di comunicazione

Per quanto si è detto nel paragrafo 2, il canale di comunicazione (bidirezionale) fra utente e sistema di calcolo centrale, si può rappresentare schematicamente come in fig. 8.

Come si è già mostrato (fig. 2), è opportuno disporre una centralina locale a "microcomputer" in prossimità di ciascun incrocio della rete assistita.

Ogni centralina gestisce sia i sensori di quell'incrocio, sia le antenne (riceventi e trasmittenti) poste su alcuni dei "link" vicini.

La affidabilità del servizio di assistenza fornito dipende, come ovvio, dal corretto scambio di messaggi fra il singolo utente ed il sistema. Ecco perciò la necessità di progettare un canale di comunicazione che

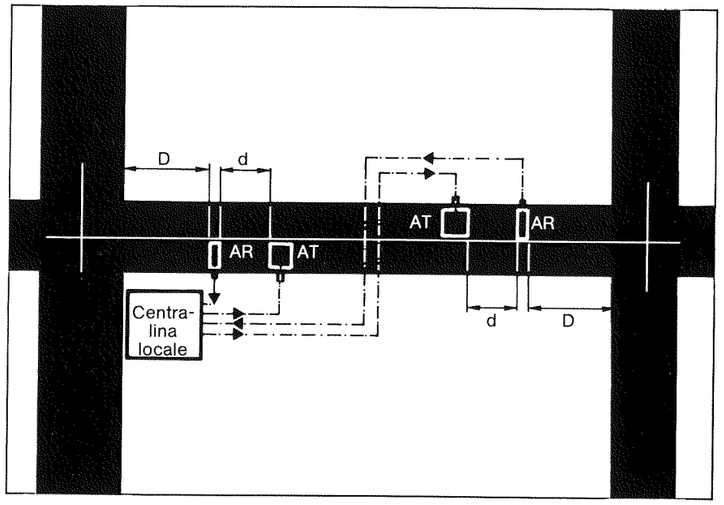


Fig. 9 - Dislocazione delle antenne stradali su un "link".

sincronizzi perfettamente la trasmissione con la ricezione, in ambo i sensi.

Le antenne stradali giocano un ruolo fondamentale in questo.

Molti sono i motivi che inducono a scartare le antenne aree quali:

- eventualità che qualche richiesta (degli utenti) non venga ricevuta dal sistema, se molti utenti richiedono contemporaneamente il servizio sullo stesso "link" (congestione del ricevitore di quel "link")
- possibilità che una stessa richiesta sia captata da antenne su "link" diversi
- presenza di forte "inquinamento" elettromagnetico e di interferenze radio, dovute a trasmissioni di diverse emittenti.

Decisamente preferibili per questa applicazione sono invece le antenne a "loop" interrate nell'asfalto, che sfruttano la induzione elettromagnetica a bassa frequenza per la trasmissione con il veicolo [3].

Infatti, allorquando il mezzo dell'utente passa sopra queste antenne, disposte orizzontalmente rispetto al piano stradale, si viene a creare un "accoppiamento" fra la spira a terra e l'antenna sul veicolo, anch'essa a "loop", disposta sul pianale della vettura, stabilendo così il contatto radio. Ciascuna antenna a terra è collegata alla centralina locale tramite cavo coassiale.

I vantaggi offerti da questa soluzione sono molteplici:

- ricezione in ordine strettamente sequenziale delle richieste degli utenti su uno stesso "link"
- possibilità di trasmettere correttamente la risposta al singolo utente destinatario, indipendentemente dagli altri utenti e dall'inquinamento elettromagnetico
- semplificazione nelle operazioni dell'utente (egli deve indicare solo la *destinazione,* perchè il senso di marcia di partenza è automaticamente individuato dall'antenna a terra che capta la richiesta).

La dislocazione delle antenne stradali su un "link" a doppio senso di marcia, può essere come in fig. 9.

L'uso della antenna ricevente (AR) distinta da quella trasmittente (AT) è vantaggioso soprattutto perché rende meno stringenti le specifiche sul massimo intervallo di tempo intercorrente fra l'avvenuta trasmissione della richiesta da parte dell'utente, e la successiva comunicazione della risposta, tramite AT.

Ciò vuol dire principalmente tempi di esecuzione meno stringenti per il software applicativo del sistema.

Le due distanze D e d in fig. 9 si possono quindi calcolare in funzione delle velocità presunte di arrivo dei veicoli e dei tempi di risposta del sistema. Le dimensioni longitudinali di AR e AT devono essere diverse perchè c'è da comunicare sequenze di segnali numerici codificati di lunghezza diversa (la risposta è molto più lunga della richiesta dell'utente).

La antenna posta sul veicolo dell'utente è anch'essa a "loop". È montata nella parte inferiore della vettura, orizzontalmente rispetto al pianale e quindi alla superficie stradale. Essa viene commutata alternativamente sul trasmettitore o sul ricevitore di bordo ed inoltre le sue caratteristiche elettriche e meccaniche sono simili a quelle delle antenne a terra [4].

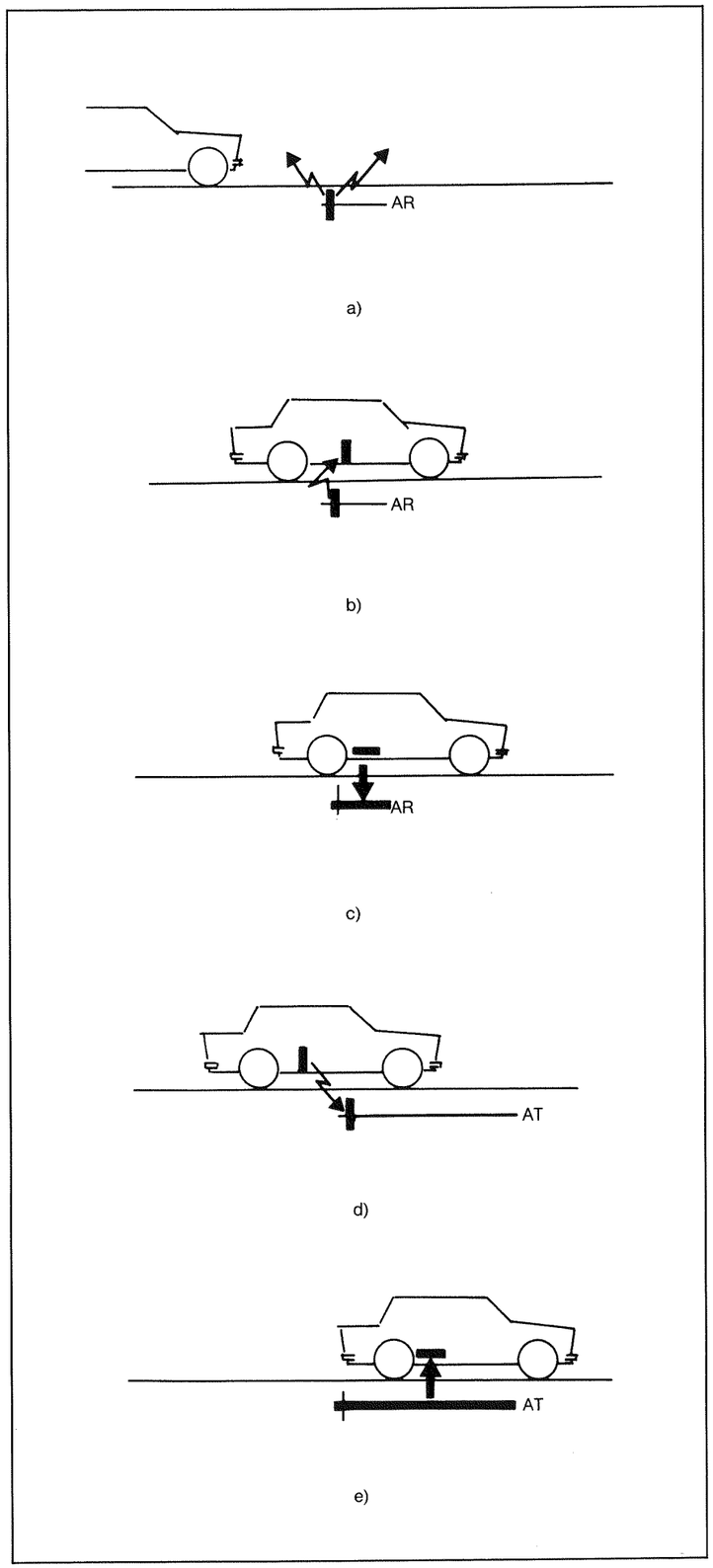
È interessante ora esaminare più in dettaglio l'organizzazione ed il funzionamento del colloquio fra utente e centralina locale, che è basato sul metodo del cosiddetto "handshaking". In altri termini, trasmissione e ricezione sono sincronizzate per mezzo di opportuni segnali di "sgancio" ("triggering"), come sarà chiaro tra breve.

A terra, all'interno di ciascuna delle antenne AR e AT, disposte orizzontalmente alla superficie stradale, si deve inserire un'altra antenna a "loop" più piccola posta verticalmente.

Una antennina del tutto simile dev'essere montata sul veicolo dell'utente in aggiunta alla antenna rice-trasmittente di cui si è parlato prima.

Queste antenne più piccole a terra e sul veicolo, dette di "triggering",

Fig. 10 - "Colloquio" fra terminale di bordo e centralina locale. Sono indicate le diverse antenne a terra e a bordo e la successione degli interventi.



trasmettono e ricevono un segnale non modulato a bassa frequenza, necessario per realizzare l'"handshaking". L'antennina di "triggering" contenuta in AR è sempre in trasmissione, quella contenuta in AT è sempre in ricezione, mentre quella dell'utente viene opportunamente commutata.

Quando un utente del servizio di assistenza parte per recarsi in un dato punto della città, ricerca prima di tutto in una apposita tabella (letta sul "display" del terminale di bordo) il numero che identifica il senso di marcia in cui desidera giungere e quindi lo imposta sulla tastiera del suo terminale.

Questo numero viene codificato in una sequenza di "bit" dalla logica di controllo (a microprocessore) del terminale di bordo, e quindi posto "in attesa" che il veicolo giunga sopra l'antennina di "triggering" contenuta in AR (fig. 10a).

Appena ciò avviene, il segnale di "sgancio" è captato dall'apposita antennina di bordo (fig. 10b), e questo dà il via alla trasmissione del segnale numerico (opportunamente modulato) tra antenna di bordo e antenna AR (fig. 10c).

A fine trasmissione un piccolo segnale acustico avverte l'utente che la sua richiesta è stata trasmessa.

Il "micro-computer" che gestisce la centralina locale aggiunge al segnale ricevuto (destinazione dell'utente) il codice del senso di marcia di partenza (sapendo da quale "porta di Input/Output" giunge il dato, e quindi da quale AR). Quindi esegue l'opportuno "formatting" dei bit, mandandoli poi sulla interfaccia col trasmettitore collegato al sistema di calcolo centrale.

Eseguite le debite elaborazioni, il segnale numerico di risposta viene mandato indietro alla stessa centralina locale, che può mantenerlo pronto per la trasmissione verso l'utente.

Nel contempo, appena l'utente supera AR, la antennina di "triggering" di bordo viene commutata su un oscillatore a frequenza fissa che emette una portante non modulata.

Quando il veicolo giunge su AT, la antenna di "triggering" a terra capta questo segnale (fig. 10d), e dà la abilitazione alla trasmissione di tutto il segnale numerico di risposta.

La antenna "di commutazione" di bordo, che è stata preventivamente commutata sul ricevitore, capta così la risposta (fig. 10e).

I circuiti logici di bordo quindi la decodificano, e la presentano sul terminale come sequenza di tutti i sensi di marcia del percorso di costo minimo richiesto, scritti in forma esplicita ed immediatamente comprensibile al guidatore.

In definitiva dunque, la trasmissione dei segnali numerici verso terra e verso il veicolo viene abilitata soltanto quando arriva l'opportuno segnale di "triggering": in tal modo ricevitori e trasmettitori sono sempre sincronizzati e le antenne di comunicazione vengono a trovarsi alla distanza ottimale per l'accoppiamento induttivo.

Come già accennato, le trasmissioni fra veicolo e terra sfruttano la induzione elettromagnetica a BF fra le antenne a "loop". I segnali da trasmettere sono di tipo numerico ed è comunque necessaria una modulazione per renderli adatti a transitare sul canale hertziano disponibile.

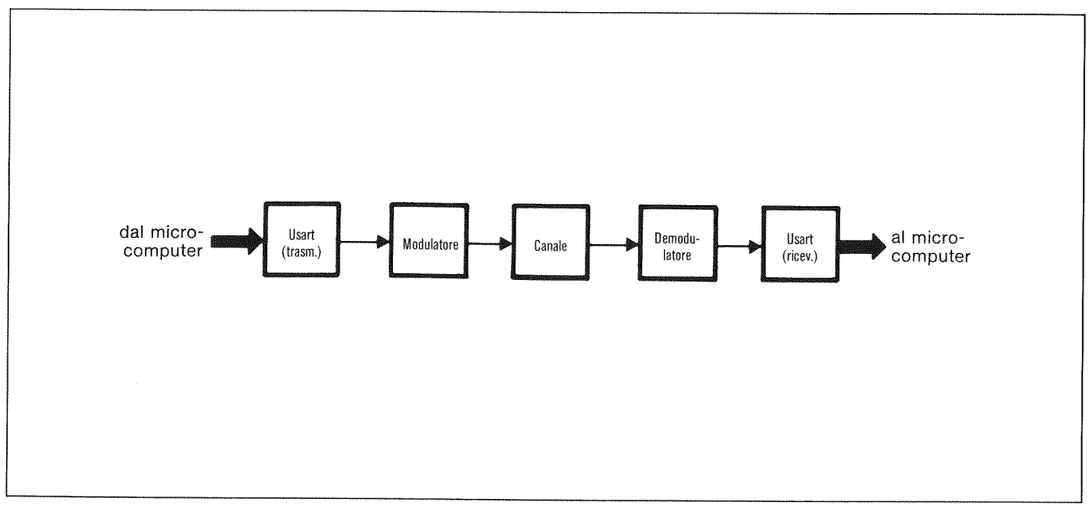
I metodi senz'altro preferibili per trasmissioni di questo tipo, sono quelli che usano le cosiddette "modulazioni numeriche" [3].

Lo schema di principio di tutto questo è mostrato in fig. 11.

Per quanto si è detto, gli elementi fondamentali necessari sui veicoli degli utenti sono dunque:

- una antenna di "triggering"
- una antenna di comunicazione
- un micro-computer
- dei dispositivi di interfaccia con l'utente guidatore.

Fig. 11 - Schema di trasmissione dei segnali numerici.



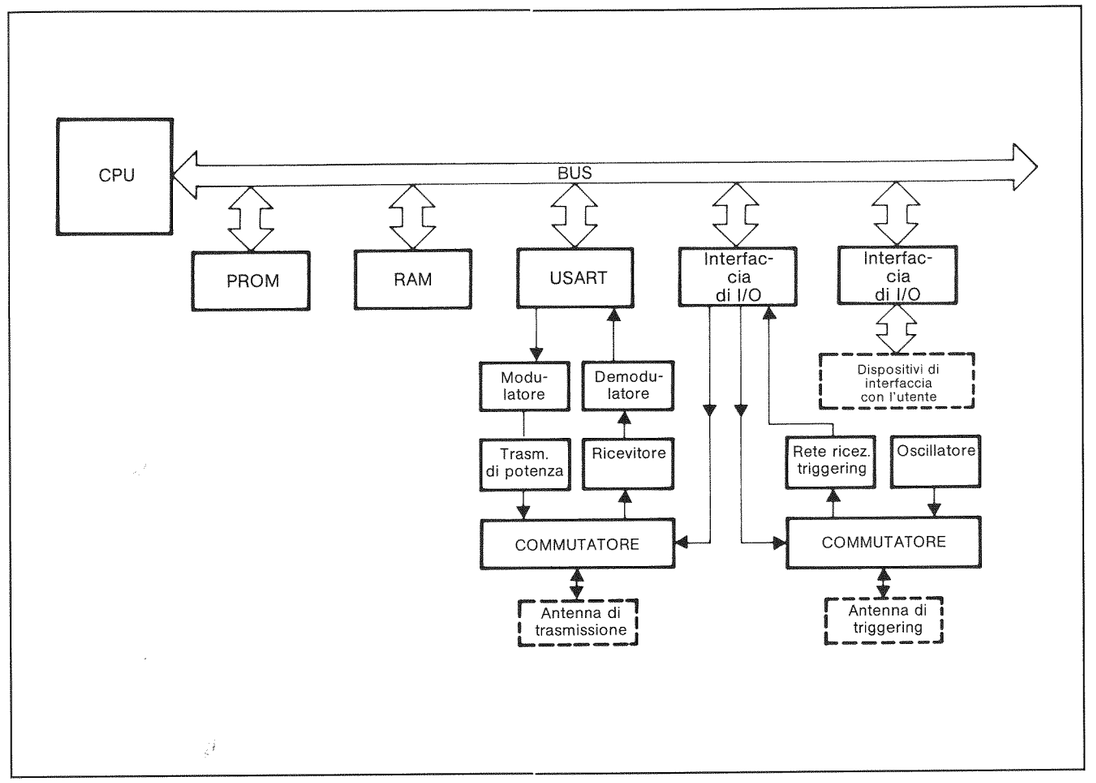


Fig. 12 - "Micro-computer" del terminale di bordo.

Delle antenne e dei dispositivi di interfaccia s'è già parlato prima. Si ricordi però che i dispositivi di I/O indispensabili verso l'utente sono una tastierina, un "display" alfanumerico, un piccolo avvisatore acustico.

Il micro-computer di bordo è quello che, opportunamente programmato, gestisce le informazioni da utente a sistema e viceversa. Il suo schema hardware completo è mostrato in fig. 12.

La CPU gestisce il funzionamento di tutto il micro-computer, colloquiando con i vari "chip" tramite il BUS.
Oltre alla memoria (PROM e RAM) è necessaria una USART (Universal Synchronous/Asyncronous Receiver/Transmitter) e delle interfaccie di Input/Output.

La USART serve per trasmettere e ricevere i segnali numerici di richiesta e risposta, in quanto essa opera le "conversioni" parallelo-serie e serie-parallelo sui dati che transitano nei due sensi fra micro-computer e antenna di trasmissione. In particolare, il segnale numerico di richiesta dell'utente viene serializzato e mandato al modulatore, mentre il segnale di risposta viene dapprima demodulato, e quindi parallelizzato dalla USART.

Le porte di I/O servono invece a gestire il segnale di "triggering" ricevuto o emesso dalla apposita antennina di bordo, a comandare i commutatori delle due antenne, ed a controllare i dispositivi di comunicazione con l'utente.

Prendiamo ora in esame la "centralina rice-trasmittente locale" realizzata con micro-computer. Questo è l'elemento più importante dei canali di comunicazione, in quanto deve gestire contemporaneamente N coppie di antenne stradali AR, AT con le rispettive antenne di "triggering", una interfaccia di comunicazione verso il sistema di calcolo centrale, M sensori di traffico, ed eventualmente un certo numero di interfacce con altre centraline vicine.

Come noto in generale, un elaboratore può gestire le operazioni di I/O (in questo caso verso le antenne stradali, verso il sistema centrale ed eventualmente verso le altre centraline) con il metodo del "polling" oppure con quello dell' "interrupt".

Secondo il primo metodo, la CPU interroga ciclicamente i vari dispositivi collegati, per sapere se hanno bisogno di trasferire dati. Quando viene individuata una "richiesta di servizio", questa viene subito esaudita, e quindi viene ripreso il ciclo di "polling".

Il metodo dell'interrupt invece, che è di gran lunga preferibile in questa applicazione (anche se richiede una maggiore complessità sia nell'hardware che nel software), comporta che il microprocessore venga interrotto solo quando effettivamente

giunga una richiesta di servizio da uno dei dispositivi collegati. All'arrivo della suddetta richiesta, la CPU sospende il programma in esecuzione, e passa ad eseguire l'opportuna "routine di servizio dell'interrupt".

Utilizzando la filosofia dell'interrupt, l'hardware della centralina locale può essere realizzato come descritto in [5], [6].

Oltre alla CPU e alla memoria, ci sono N "chip" di USART (programmati per rice-trasmissioni asincrone) collegate ad altrettante coppie AT, AR, in modo simile a quanto visto in fig. 11. Un'altra USART (programmata per rice-trasmissioni sincrone) deve essere usata per l'interfacciamento col sistema di calcolo centrale, ed eventuali altre sono necessarie per collegamenti con le centraline vicine. I segnali provenienti dagli M sensori di traffico sono gestiti tramite porte di I/O e contatori. Esiste inoltre un "chip" ("Interrupt Priority Handler") che effettua lo "scheduling", cioè l'ordinamento secondo priorità, di tutte le richieste di interrupt dei periferici, e colloquia con la CPU per indirizzarla alle opportune routines di servizio dello interrupt.

Per quanto riguarda la connessione della centralina locale con il sistema di calcolo centrale, è preferibile utilizzare un canale fisico (cavo coassiale) per il fatto che quello hertziano ha, come noto, grossi problemi di disturbi e di attenuazione. Per realizzare il collegamento è allora necessaria una USART, quindi un MODEM ed una interfaccia con linea telefonica. Il canale vero e proprio è dunque una linea telefonica "punto a punto" dedicata, che giunge fino alla centrale di calcolo.

B) Parte "Hardware" del sistema di calcolo centrale

Per come è strutturato l'intero sistema, è necessario che la unità centrale di calcolo possegga i due requisiti di alta velocità d'esecuzione (per ridurre i tempi di risposta), e grande capacità di memoria interna (per consentire rappresentazioni più dettagliate della rete viaria, o permettere espansioni notevoli della rete assistita).

Una possibile soluzione è costituita da due minicalcolatori operanti in configurazione "master-slave" [7].

I due "mini" colloquiano tra di loro e accedono alle memorie di massa. Lo "slave" si interessa delle operazioni di I/O con le centraline locali e con le "console" degli operatori del centro di calcolo.

C) Sistema di acquisizione dati: i sensori di traffico

Come si è già ampiamente descritto, l'acquisizione dei dati reali di traffico è effettuata da sensori localizzati opportunamente nella rete viaria stradale e collegati alle centraline locali.

I dati di traffico acquisiti consistono nell'individuazione, vale a dire nel rilevamento del passaggio, nel campo di rivelazione, del singolo veicolo.

I sensori di traffico da prendere in considerazione non devono fornire alcuna discriminazione sulla classe, velocità, etc. dei veicoli, dati di traffico questi non necessari alle strategie di controllo e di calcolo "on line" dei costi elementari c(i,j).

Numerosi sono i tipi di sensori esistenti: a "loop" induttivo, magnetici, magnetometrici, radar, ultrasonici, a pressione.

Tutti sono stati ampiamente studiati ([3], [8], [9]), ma sulla base di numerosi parametri di funzionamento, di installazione, di manutenzione e di costo, è stata verificata una netta superiorità dei sensori a "loop" induttivo.

Il loro principio di funzionamento è basato sulla variazione dell'induttanza, dovuta al passaggio della massa metallica del veicolo, all'interno del campo di rivelazione di alcune spire poste sotto il manto stradale. L'elemento di rilevamento è costituito appunto da queste spire nell'asfalto; tale trasduttore è collegato mediante cavo al dispositivo elettronico di rilevamento.

La localizzazione ed il numero dei sensori di traffico su un incrocio è funzione della particolare strategia che si vuole adottare, e della affidabilità che si richiede al sistema di acquisizione dati.

4. Conclusione

L'esposizione fin qui svolta ha dovuto necessariamente essere molto sintetica: parecchi argomenti sono stati soltanto accennati ed alcuni completamente trascurati.

In ogni caso si può facilmente notare che il sistema di assistenza qui descritto presenta alcuni attributi molto interessanti:

- flessibilità: piccole variazioni sul software e sull'hardware consentono di adattare il sistema a variazioni imposte sulla regolamentazione della rete viaria;

- struttura "orientata alla destinazione": da qualunque punto della rete assistita si può ottenere il percorso di costo minimo per giungere ad una stessa destinazione;

- protocollo di colloquio tra veicolo e centralina locale basato sul reciproco scambio di segnali di abilitazione ("handshaking"): ciò riduce moltissimo la probabilità di errore nelle trasmissioni, in quanto queste avvengono nelle condizioni ottimali (minima distanza tra le antenne), e risultano sincronizzate;

- struttura hardware-software di tipo gerarchico-distribuito: questo eleva notevolmente il "throughput" dei calcolatori centrali e la affidabilità dell'intero sistema.

Il sistema di assistenza qui proposto risulta allora tecnicamente fattibile e realizzabile.

Chiaramente esso è ragionevolmente applicabile solo in grandi città dove la grossa mole di traffico ed il grande numero di potenziali utenti possano giustificare il costo iniziale dell'impianto, oppure in quelle città dove la rete di centraline e sensori possa servire anche per un controllo in "loop" chiuso del traffico cittadino, tramite attuazione automatica coordinata degli impianti semaforici (per esempio a Torino si sta ultimando il "Progetto Torino" per la semaforizzazione automatizzata della città).

5. Bibliografia

[1] N. CHRISTOFIDES - *Graph theory: an algorithmic approach,* Academic Press - 1975 - London

[2] T.C. HU - *Integer programming and network flows,* Addison - Wesley - 1969 - Reading, Mass. (USA)

[3] A.S. PALATNICK, H.R. INHELDER - *Automatic vehicle identification systems: methods of approach,* IEEE Transactions on Vehicular Technology - 1970 - Vol. VT-19

[4] D.A. ROSEN et al. - *An electronic route guidance system for highway vehicles,* IEEE Transactions on Vehicular Technology - 1970 - vol. VT-19

[5] M. RONDO, L. GILLI - *Un microelaboratore per il controllo del traffico urbano,* Memoria presentata al Convegno "Evoluzione del controllo Semaforico" - Torino - 27/28 marzo 1980

[6] C. LONARDO - *Automazione del traffico stradale: centralini a microcalcolatore per la regolazione semaforica,* Rivista "Raggruppamento Ansaldo" n. 4 - 1978

[7] F. FERRARIS - *Sistema centrale ad alta affidabilità per il controllo del traffico urbano,* Memoria presentata al Convegno di cui [5].

[8] M.K. MILLS - *Future vehicle detection concepts,* IEEE Transactions on Vehicular Technology - 1970 - vol. VT-19

[9] J.L. BARKER - *Radar, acoustic, and magnetic vehicle detectors,* IEEE Transactions on Vehicular Technology - 1970 - vol. VT-19

[10] M.RONDO - *I sensori di traffico: stato dell'arte,* Memoria presentata al Convegno di cui [5].