Indice

UN MODELLO DINAMICO SU UN APPROCCIO BIG-DATA ALLA MOBILITÀ PER LO STUDIO DELLA DIFFUSIONE DEL COVID-19 NEL NORD ITALIA

Davide Micheli, Giuliano Muratore, Aldo Vannelli, Giuseppe Sola

La teoria dei Sistemi Dinamici ha proposto molti modelli per la diffusione delle epidemie e il recente sviluppo della Fisica dei Sistemi Complessi ha mostrato come le complesse reti di trasporto abbiamo un ruolo fondamentale nella diffusione a scala planetaria delle epidemie stesse. Tuttavia, le caratteristiche della pandemia da COVID-19 hanno messo in evidenza come la peculiarità della mobilità microscopica in aree con diverse caratteristiche di antropizzazione potrebbero avere profondamente influenzato la sua diffusione. I dati relativi al traffico telefonico radiomobile, essendo statisticamente correlabili con la mobilità della popolazione, sono attualmente un asset informativo importante un po' in tutto il mondo, e sono oggi utilizzati a supporto di molte decisioni, siano esse di tipo amministrativo o commerciale. La diffusione dei dispositivi mobili offre la possibilità di raccogliere grandi quantità di dati su un campione significativo della popolazione. In questo studio si è sviluppata la possibilità di utilizzare i dati relativi al traffico radiomobile TIM del mese di Febbraio 2020 in sinergia con un modello dinamico SEIR messo a punto sui dati disponibili dell'epidemia da COVID-19 nella Lombardia e nel Veneto.

Introduzione

In questo studio sono state utilizzate le cosiddette Matrici Origine-Destinazione, che permettono di ricostruire gli spostamenti in Italia tra le 9862 Aree Censuarie (ACE) con le quali l'ISTAT suddivide il territorio italiano.
Tale base informativa sfrutta in forma anonima e aggregata, come previsto dalle Normative vigenti in termini di Privacy, sia i dati di tassazione (o "cartellini" di traffico telefonico) sia i dati di segnalazione legati al traffico radiomobile sulla rete TIM.

Il mese di febbraio 2020 risulta particolarmente significativo per lo studio della diffusione del COVID-19 in Italia, in quanto evidenzia un periodo di possibile diffusione del virus ancor prima dell'individuazione di un focolaio d'epidemia.

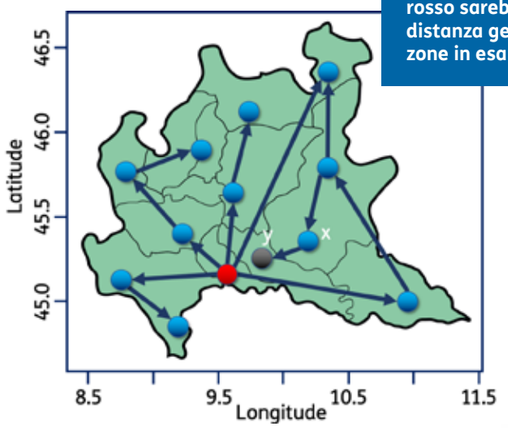
Lo studio affronta in primis un'analisi statistica (utilizzando il Tool Open Source R Studio [1]) della mobilità specifica che caratterizzava Codogno e Vo', prima dell'evidenza epidemiologica, al fine di isolarne le caratteristiche di mobilità utili nell'analisi epidemiologica.

Nell'articolo viene poi presentato uno studio che, incorporando dati di mobilità specifica in Lombardia e Veneto, adatta i modelli di diffusione del virus al caso specifico del nord Italia, disegnandone anche le possibili evoluzioni.

Mobilità e Grafi

La Lombardia e il Veneto rappresentano due regioni economicamente importanti in Italia, interessate da una grande mobilità interna (oltre che internazionale), sia di persone che di merci.

Favoriscono la mobilità regionale anche ragioni orografiche, come l'ampiezza della pianura Padana ed il fatto che tale pianura sia innervata da grandi strutture di comunicazione, stradali e ferroviarie.
Per valutare gli effetti della mobilità nell'ambito della diffusione del COVID-19 occorre però fissare criteri oggettivi di valutazione della mobilità di una regione, per procedere anche a confronti con zone differenti.



1
Semplice esempio di grafo basato sulla mobilità di una regione. Il nodo in esame (in rosso) presenta diversi cammini che lo congiungono a tutti gli altri nodi del grafo. I nodi direttamente collegati hanno "distanza" (numero di nodi attraversati) pari ad 1. Il numero di nodi attraversati tra il nodo rosso e quello nero è pari a 4. In questa situazione l'eccentricità del nodo rosso sarebbe appunto pari a 4. Ai fini dell'eccentricità non conta infatti la distanza geografica, ma l'esistenza o meno di spostamenti avvenuti tra le zone in esame.

Un aiuto in tal senso arriva dalla Teoria dei Grafi, sviluppata per l'analisi di strutture composte da nodi e connessioni tra gli stessi.

È noto innanzitutto che la trasmissione dei virus viaggi con le persone, e possa quindi insediarsi in ambiti differenti da quello d'origine per tramite di una successione di passaggi da persona a persona. Ciò che è significativo per la diffusione del virus è quindi stabilire se esistono connessioni (cioè mobilità di persone) tra una zona ed un'altra.

La mobilità del virus può infatti concretizzarsi in una nuova zona se portato da uno o più soggetti già infettati.

Successivamente il virus può essere nuovamente trasportato altrove, anche da altre persone della nuova zona, se nel frattempo sono state contagiate. Questa catena di passaggi finisce per essere assimilabile ad un grafo (vedi Figura 1), ed il maggiore o minore numero di connessioni (passaggi di persone) tra le diverse zone può concretamente contribuire al manifestarsi di crescite, anche esponenziali, dei contagi.

Considerando allora come ambito d'analisi la singola regione italiana, considerando come nodo del grafo su cui porre l'attenzione una specifica zona (Area Censuaria ISTAT [2]), ed infine considerando come connessione tra due nodi l'esistenza di spostamenti tra le due zone (Aree Censuarie) in questione, otteniamo una rappresentazione schematica delle possibili vie attraverso le quali potrebbe spostarsi un virus dentro quella regione.

Avendo trasformato le Aree Censuarie di una regione in nodi, ed avendo definito come connessione tra i nodi il fatto che ci siano stati spostamenti tra due Aree Censuarie, possiamo utilizzare i dati di mobilità tra le ACE di quella regione come il modo per verificare effettivamente quali Aree Censuarie sono state oggetto di mobilità, ed anche quando ciò è avvenuto.

Si può quindi, con le Matrici Origine-Destinazione, passare da un astratto concetto di Grafo di una regione italiana, al Grafo che corrisponde proprio alla situazione reale di mobilità in quella regione, in un dato periodo. Il cammino sul grafo non sarà altro che la sequenza di spostamenti [**nota 1**] che sono stati registrati nelle Matrici Origine-Destinazione, prima per passare da un nodo A ad uno nodo B, e poi per passare dal nodo B al nodo C, e così via. Più lungo sarà il cammino e più nodi saranno quindi stati attraversati [**nota 2**].

A questo punto dalla Teoria dei Grafi si può prendere in prestito il concetto di "eccentricità" di un nodo di un grafo. Il nodo che esamineremo sarà quello di Codogno (per il caso della Lombardia), e di Vo' Euganeo (per il caso del Veneto), misurando l'eccentricità di Codogno (e poi di Vo') rispetto al grafo della regione Lombardia (e poi Veneto). L'eccentricità di un nodo di un grafo è (dalla Teoria dei Grafi) definito come la misura del più lungo dei cammini più brevi che connettono quello specifico nodo a qualsiasi altro nodo del grafo.

L'Eccentricità misura così quanto un nodo risulti tendenzialmente isolato (alta eccentricità), piuttosto che fortemente centrale (bassa eccentricità), rispetto ai flussi di mobilità della sua regione.

Passando al caso concreto viene valutata l'eccentricità di Codogno (e poi di Vo'), misurando il più lungo dei cammini (più brevi) che connettono il nodo-ACE di Codogno (poi di Vo') a qualsiasi altro nodo-ACE della Lombardia (del Veneto). Per il calcolo dell'eccentricità di Codogno non ha quindi rilievo quale specifico percorso sia stato effettivamente seguito per lo spostamento di persone tra un'ACE all'altro, ma solo il fatto che tale mobilità sia avvenuta o non sia avvenuta nel periodo in esame.

Quanto maggiore risulterà l'eccentricità di un nodo di una regione, tanto più quel nodo risulterà isolato, perché i cammini risulteranno più lunghi, e quindi quel nodo sarà meno predisposto alla diffusione di un virus.
Viceversa, tanto minore risulterà l'eccentricità di un nodo, tanto più quel nodo sarà predisposto alla dif-

fusione del virus, a causa appunto di una elevata mobilità di quella zona, mobilità capace di creare molte connessioni con le altre zone della regione [**nota 3**].

I risultati delle analisi sui dati di mobilità territoriale di Lombardia e Veneto mostrano alcune somiglianze tra lo scenario di Codogno (in provincia di Lodi) e quello di Vo' (in provincia di Padova).

Per Codogno si misura un'eccentricità pari a 4, cioè risultano quindi necessari al più 4 passaggi, affinché dall'ACE di Codogno si tocchino tutte le altre zone (ACE) della Lombardia, composta da ben 1793 ACE. Le ACE a distanza unitaria da Codogno sono circa il 22% circa del totale delle ACE della Lombardia. Per queste ACE si registrano quindi spostamenti diretti di persone da Codogno.

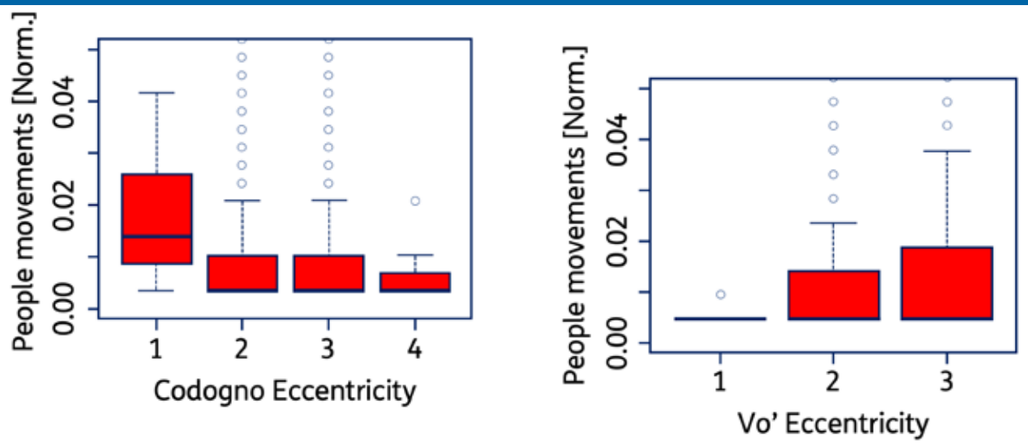
Nel caso di Vo' sono invece sufficienti al più 3 passaggi, affinché dall'ACE di Vo' si possano toccare tutte le altre 675 ACE del Veneto, ed inoltre le ACE oggetto di spostamenti diretti da Vo' risultano circa il 29% del totale delle ACE del Veneto.

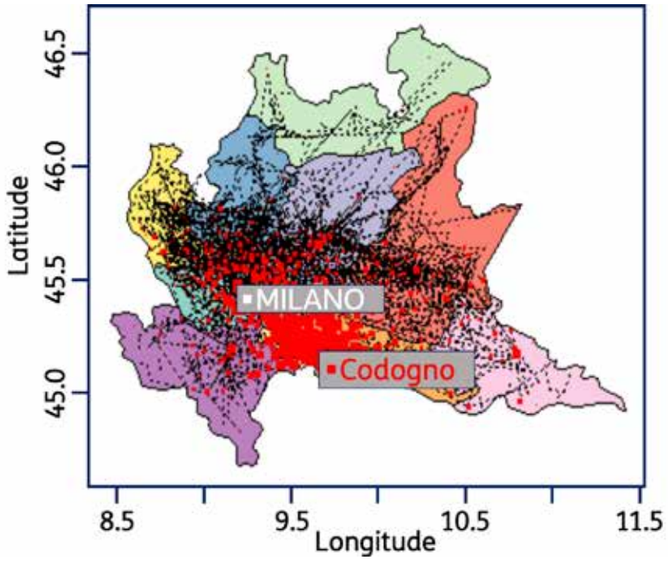
Siccome i passaggi da una ACE all'altra riguardano movimenti di persone che si sono appunto spostate tra quelle ACE, si potrebbe presumere che Vo' si trovasse in una situazione di maggiore predisposizione alla diffusione del virus, dovendo il virus seguire un minor numero di "salti" per raggiungere tutte le altre ACE venete, e contando Vo' più casi di Codogno (29% vs 22%) nella percentuale di ACE coinvolte in spostamenti diretti.

La situazione oggettiva è però quasi opposta, e ciò si può desumere (vedi Figura 2), osservando come siano distribuite le quantità di spostamenti di persone da Codogno e da Vo', in funzione delle diverse eccentricità dei relativi spostamenti tra ACE della regione. Nel caso di Codogno sono molto più popolati gli spostamenti diretti, che sono quelli a maggior probabilità di contagio (considerando Codogno l'epicentro), mentre risultano via via meno popolati gli spostamenti indiretti.

Nel caso di Vo' invece le connessioni dirette verso altre ACE risultano tante, ma si tratta di connessioni effettuate nel complesso da poche persone, mentre risultano più popolati i casi successivi, quelli per i quali le probabilità di contagio da Vo' (considerando Vo' l'epicentro) risultano via via più basse.

2
In figura è mostrata la notevole differenza che emerge tra la mobilità di Codogno e quella di Vo', normalizzando ad 1 il numero totale di spostamenti per una migliore confrontabilità. Anche se Codogno ha una eccentricità maggiore di Vo', è maggiore per Codogno l'incidenza degli spostamenti diretti dal focolaio del contagio rispetto a quelli indiretti (più lunga catena dei contagio). Di conseguenza una diffusione del contagio tra persone della stessa regione vede in partenza più critica la situazione di Codogno, rispetto a quella di Vo'.





3
Mobilità diretta (evidenziata in rosso) ed indiretta (evidenziata dalle linee nere tratteggiate) da Codogno. Si nota la differenza tra la minore mobilità periferica (linee nere più rade) e l'elevata concentrazione di mobilità regionale che attraversa la zona a Nord di Milano, in corrispondenza dell'asse viario padana da Torino a Venezia. Si nota altresì che la mobilità diretta da Codogno rimane in Lombardia prevalentemente sviluppata (quadratini rossi) lungo la congiungente con Milano. La dimensione dei quadratini, maggiore in prossimità di Codogno e minore in zone più lontane, descrive graficamente la maggiore numerosità di persone con mobilità diretta di corto raggio, rispetto alla minore numerosità di persone con mobilità diretta che raggiunge anche zone più lontane.

La diversa distribuzione della numerosità di spostamenti alle diverse eccentricità, rende quindi lo scenario di Codogno in partenza più critico rispetto a quello di Vo'.

Pur non affrontato in questo articolo, focalizzato sul tema epidemiologico di Codogno e Vo', l'applicazione del concetto di eccentricità appena espresso per queste due località si può considerare generalizzabile. Infatti, calcolando l'eccentricità per tutte le ACE di una regione si arriva a stimare il Coefficiente di Mobilità Regionale (CMR), semplicemente come valore medio dell'eccentricità delle ACE di cui è composta quella regione. La sinteticità di questo indicatore (eccentricità media in una regione) può infatti risultare utile anche per i confronti di mobilità su piccola scala tra differenti zone d'Italia (o in generale del mondo).

Mobilità in Lombardia

L'analisi dei dati telefonici della Lombardia a febbraio 2020, prima dell'istituzione della Zona Rossa (DPCM 23/02/2020 [3]), ci fornisce un quadro d'insieme della mobilità delle persone in una regione, elemento che influenza la diffusione di un virus.

Osservando (vedi Figura 3) gli spostamenti effettuati da Codogno nella prima ventina di giorni del mese di febbraio, in un periodo cioè durante il quale il virus poteva spostarsi con le persone, non essendo ancora entrate in vigore le misure di contenimento della mobilità, si può notare che la mobilità diretta da Codogno interessa in prevalenza la direttrice verso Milano, anche se la dispersione lungo il territorio lombardo della mobilità diretta da Codogno coinvolge comunque un'area abbastanza vasta. La mobilità indiretta (linee tratteggiate nere) ha maggiore concentrazione proprio lungo la direttrice che attraversa la Lombardia (asse Torino-Venezia).

La mobilità indiretta si estende a tutto il territorio lombardo, anche se ovviamente risulta inferiore in alcune zone periferiche (es. quelle alpine a Nord) per evidenti vincoli orografici.

UN MODELLO EPIDEMIOLOGICO INTEGRATO CON DATI DI MOBILITÀ TELEFONICA ORIGINE-DESTINAZIONE A PICCOLA SCALA SPAZIALE

Il modello proposto nell'ambito di una collaborazione tra TIM e l'Università di Bologna si basa sulla possibilità di raccogliere dati di mobilità attraverso l'attività di telefonia mobile alla scala delle aree censuarie (ACE) con frequenza oraria, per integrare tali informazioni entro modelli matematici di dinamica di popolazione e così simulare la diffusione di un'epidemia.

L'Università di Bologna ha quindi realizzato un sistema dinamico su grafo i cui nodi sono rappresentati dalle ACE georeferenziate nel territorio italiano con connettività pesata in proporzione alla mobilità Origine-Destinazione tra i nodi, mobilità inferita utilizzando dati, anonimi ed aggregati, delle Matrici Origine-Destinazione (MOD) di telefonia mobile.

La dinamica di popolazione all'interno di ciascuna ACE è simulata da un modello tipo SEIR (Susceptible, Exposed, Infectious, Removed) [4, 5], con la possibilità di specificare alcuni parametri in funzione delle caratteristiche sociali e del tessuto urbano dell'area. In particolare, le equazioni di evoluzione del modello epidemiologico per il nodo k-esimo al passo temporale Δt risultano essere:

$$
\begin{aligned}
S_k(t+\Delta t) &= S_k(t) + \mu^{\Delta t}_{S,k}(t) - \phi^{\Delta t}_{S\to E,k}(t)\\
E_k(t+\Delta t) &= E_k(t) + \mu^{\Delta t}_{E,k}(t) + \phi^{\Delta t}_{S\to E,k}(t) - \phi^{\Delta t}_{E\to I,k}(t) - \phi^{\Delta t}_{E\to A,k}(t)\\
I_k(t+\Delta t) &= I_k(t) + \mu^{\Delta t}_{I,k}(t) + \phi^{\Delta t}_{E\to I,k}(t) - \phi^{\Delta t}_{I\to R,k}(t)\\
A_k(t+\Delta t) &= A_k(t) + \mu^{\Delta t}_{A,k}(t) + \phi^{\Delta t}_{E\to A,k}(t) - \phi^{\Delta t}_{A\to G,k}(t)\\
R_{T,k}(t+\Delta t) &= R_{T,k}(t) + \phi^{\Delta t}_{I\to R_T,k}(t) - \phi^{\Delta t}_{R_T\to G,k}(t)\\
R_{H,k}(t+\Delta t) &= R_{H,k}(t) + \phi^{\Delta t}_{I\to R_H,k}(t) - \phi^{\Delta t}_{R_H\to G,k}(t)\\
G_k(t+\Delta t) &= G_k(t) + \mu^{\Delta t}_{G,k}(t) + \phi^{\Delta t}_{A\to G,k}(t) + \phi^{\Delta t}_{R\to G,k}(t)
\end{aligned}
$$

I modelli dividono la popolazione nelle diverse categorie: Suscettibili "S", Esposti "E", Infetti "I", Asintomatici "A", Ricoverati in ospedale (o in altre strutture) e in terapia intensiva "$R_{H,T}$", ed infine Guariti "G".

La dinamica delle popolazioni dipende dai flussi di scambio $\Phi_{X\to Y}$ introdotti secondo lo schema riportato in figura A.

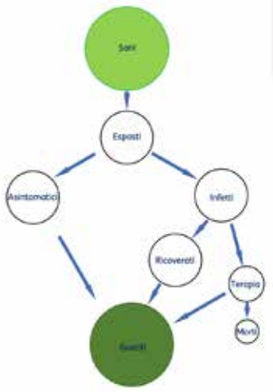
Tali flussi dipendono dai parametri e dalle diverse scale temporali legate ai meccanismi del contagio, dell'incubazione, dello sviluppo dell'infezione e della guarigione.

Nello specifico, la definizione dei flussi tra le diverse categorie avviene secondo due schemi: la transizione tra suscettibili ed esposti è definita in base alla formula

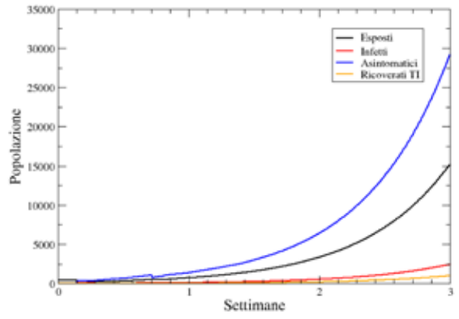
$$\Phi^{\Delta t}_{S\to E,k}(t) = [\beta_I \cdot I_k(t) + \beta_A \cdot A_k(t)] \cdot (m_k \cdot \Delta t) \cdot \frac{S_k(t)}{P_k(t)},$$

dove si evidenzia la dipendenza dai parametri β_I e β_A che definiscono la probabilità di avere un contagio quando un individuo sano incontra un infetto o un asintomatico, rispettivamente, e il numero m_k che definisce il numero medio di incontri con persone diverse per unità di tempo che un individuo effettua nell'ACE considerata (tale numero può tenere conto dell'attività sociale e della presenza di luoghi di incontro nell'area). La probabilità di un incontro è calcolata come il prodotto tra la popolazione suscettibile con la frazione di infetti o asintomatici presente nella popolazione. La presenza di una popolazione di esposti, per quanto piccola, può innescare la diffusione dell'epidemia, se il numero medio di contagiati per individuo infetto o asintomatico è superiore all'unità e la crescita è di tipo esponenziale. Solo la riduzione dell'attività sociale locale (ovvero del parametro m_k) e la riduzione del numero degli individui suscettibili (perché la popolazione dei guariti è cresciuta a sufficienza) può ridurre il flusso e permettere di raggiungere un massimo nella popolazione degli esposti dopo il quale l'epidemia comincia a regredire (cfr. Fig C). Tuttavia, affinchè la regressione sia efficace, la riduzione dell'attività sociale (ovvero la restrizione alla mobilità individuale attuata dal decreto '#io resto a casa') deve essere mantenuta fino a che la popolazione degli infettati si riduca a zero, altrimenti la diffusione ricomincia fino al raggiungimento dell'immunità di gregge, ovvero quando il numero dei guariti sarà così elevato da rendere improbabile che un infetto incontri persone da contagiare. Quest'ultima situazione è il solo equilibrio dinamico stabile che protegge la popolazione da un'eventuale nascita di futuri focolai di infezione. Resta anche possibile un cambiamento nel tempo dei parametri β_I e β_A per una diminuzione della virulenza dell'epidemia, che favorirebbe un'immunità di gregge con un numero totale di infettati inferiore durante l'epidemia stessa. Gli altri flussi definiti nel modello sono illustrati dalla transizione tra Esposti e Infetti:

$$\Phi^{\Delta t}_{E\to I,k}(t) = \begin{cases} 0 & \text{se } t < T_E \\ \alpha \cdot \Phi^{\Delta t}_{S\to E,k}(t - T_E) & \text{altrimenti} \end{cases},$$



A
Schema dei Flussi di scambio



B
Layout dell'interfaccia grafica del software per la simulazione del modello nell'area metropolitana a sud di Milano; l'evoluzione è in grado di assimilare i dati di una matrice OD per le aree ACE a scala oraria. A destra si riporta l'evoluzione delle caregorie Esposti, Asintomatici,Infetti e Ricoverati in terapia intensiva prevista dal modello durante le prime 3 settimane di febbraio.

Vediamo come tale formula dipenda dalla scala di tempo di incubazione T_E della malattia per cui gli individui che esposti al tempo t-T_E (che sono contati nel flusso S→E calcolato in quel momento), transitano nella categoria degli infetti in una percentuale α definita dal rapporto statistico tra infetti e asintomatici nella popolazione (per il COVID-19 tale rapporto sembra essere piccolo facendo sì che il contagio sia principalmente diffuso dagli asintomatici).

Avremo quindi una definizione analoga per il flusso tra Esposti e Asintomatici con una percentuale complementare 1-α. In modo simile vengono trattate le altre transizioni: ad esempio, gli individui Infetti vengono ricoverati o vanno in terapia intensiva dopo un tempo caratteristico T_i di sviluppo dei sintomi veri e propri dell'infezione. Il modello matematico si rappresenta mediante equazioni differenziali con ritardi che tengono conto delle varie scale di tempo necessarie per l'evoluzione del contagio e dell'infezione e la guarigione finale.

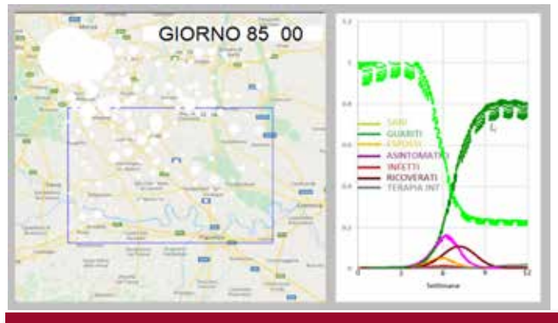
Da questo punto di vista stiamo parlando di un modello fisico-matematico adattabile a diverse tipologie di epidemia che potrebbe mettere in evidenza diverse caratteristiche dinamiche a seconda del valore attribuito ai vari parametri. Il modello si integra con dati della Matrice Origine-Destinazione (MOD) tra le varie ACE (Aree Censuarie ISTAT) mediante la definizione dei flussi di mobilità μ_X delle varie categorie considerate, nelle dovute proporzioni secondo le equazioni qui di seguito riportate:

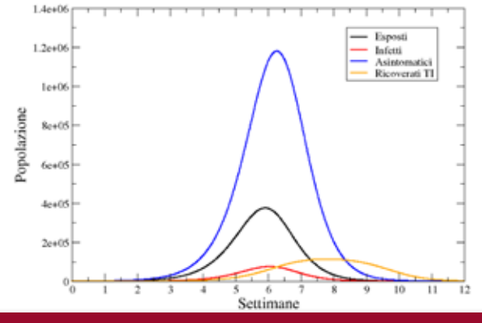
$$\mu^{\Delta t}_{X,k}(t) = \sum_{j \neq k} \left[X^{\Delta t}_{j \to k}(t) - X^{\Delta t}_{k \to j}(t) \right] \qquad X^{\Delta t}_{j \to k}(t) = P^{\Delta t}_{j \to k}(t) \frac{X_j(t)}{P_j(t)}$$

La prima equazione definisce μ_X come bilancio tra il flusso entrante ed uscente della categoria X nell'ACE k-esimo da e verso gli altri ACE connessi, essendo $P_{j \to k}$ l'elemento della MOD corrispondente. I valori dei flussi $\mu_{X,k}$ possono essere stimati dai dati di mobilità da telefonia mobile e determinano il meccanismo di diffusione dovuta alla mobilità nell'area considerata. La complessa struttura del territorio urbanizzato in aree come la Pianura Padana definisce un network di mobilità complesso che regola la diffusione dell'epidemia da scala locale (ACE) a quella globale. Il modello simulato risulta quindi una sovrapposizione di una dinamica locale, in cui il contagio si diffonde tra la popolazione per le attività sociali in ciascuna ACE (modello compartimentale SEIR) in funzioni di parametri caratteristici del virus e una dinamica di interazione tra ACE dovuta alla domanda di mobilità per le attività insediate nell'area considerata, che introduce uno scambio tra le popolazioni nei vari ACE secondo una struttura a network complesso. Politiche di isolamento di specifiche sotto-aree possono essere efficaci solo se adattate alla struttura del network di mobilità sottostante. Il risultato è un modello dinamico che simula la diffusione dell'epidemia tenendo conto della mobilità e dell'attività sociale degli individui contagiosi (asintomatici ed infetti). L'integrazione tra le dinamiche di diffusione dell'epidemia con dati di mobilità di telefonia mobile a piccola scala spaziale risulta un approccio innovativo nel campo della modellizzazione epidemiologica.

Il modello implementato in un pacchetto software per la simulazione ha consentito di realizzare uno strumento che consente un'analisi dello stato di fatto della diffusione dell'epidemia nel Nord Italia, basandosi sulle ipotesi che i focolai iniziali siano stati nella zona di Lodi-Codogno e Vo' Euganeo (PD) e considerando la prevedibilità dell'evoluzione delle varie popolazioni, in particolare della domanda di ospedalizzazione dovuta agli infetti con sintomi gravi. Inoltre, il modello implementato consente di studiare gli scenari possibili per comprendere gli effetti a lungo termine delle politiche di contenimento della mobilità individuale. Le caratteristiche di un tale strumento sono illustrate nelle figure B e C. I parametri utilizzati dal modello sono stati fissati in base agli andamenti osservati e ai dati riportati in letteratura [6].

Nel primo caso in Figura B si vede come un'area comprendente la provincia di Lodi e Milano possa essere stata soggetta ad una rapida diffusione del virus nelle prime due settimane di Febbraio 2020 per la stretta interconnessione dell'area dovuta all'intensa mobilità nella Pianura Padana il cui tessuto urbano non ha soluzione di continuità. Nella seconda figura mostriamo un'ipotesi di evoluzione delle popolazioni nella stessa area qualora non si fosse intervenuti con misure di contenimento della mobilità. In particolare, si evidenziano gli andamenti previsti per le classi degli infetti e ricoverati in terapia intensiva in un arco temporale di 12 settimane. Si nota un picco previsto di 10^5 individui che necessitano di una terapia intensiva dopo 9 settimane.

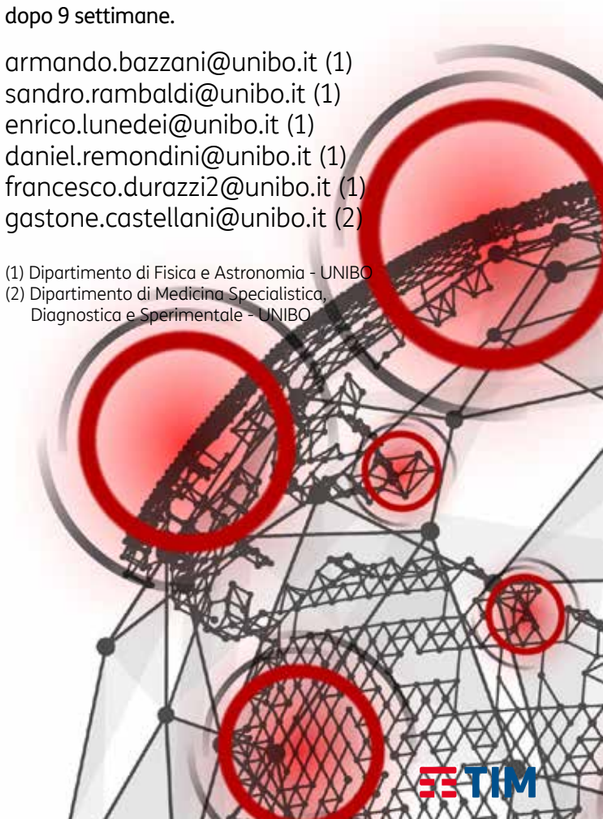


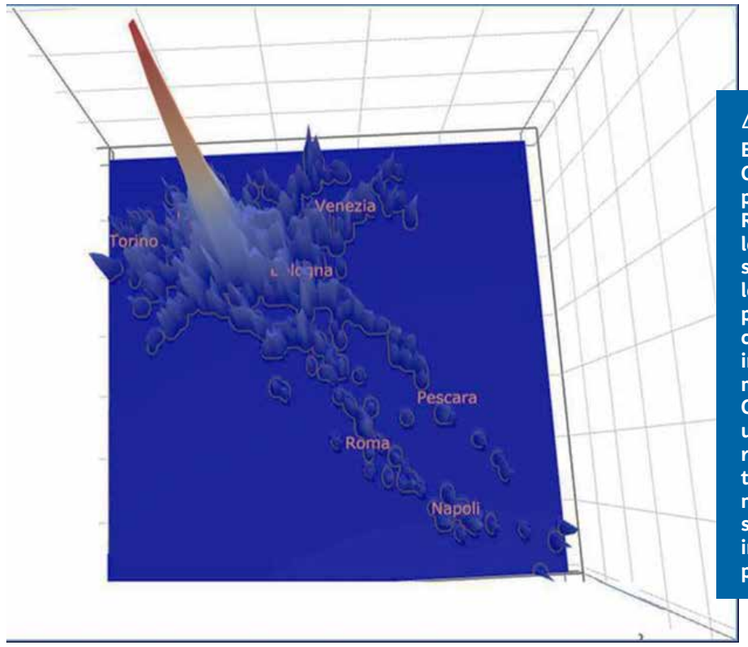


C
Lo stesso che nella Fig. B per un tempo d di evoluzione di 12 settimane supponendo che nessuna politica di restrizione della mobilità venisse introdotta. A destra si riporta l'evoluzione delle categorie Esposti, Asintomatici,Infetti e Ricoverati in terapia intensiva prevista dal modello evidenziando come si potesse prevedere un picco di 100000 persone ricoverate dopo 9 settimane.

armando.bazzani@unibo.it (1)
sandro.rambaldi@unibo.it (1)
enrico.lunedei@unibo.it (1)
daniel.remondini@unibo.it (1)
francesco.durazzi2@unibo.it (1)
gastone.castellani@unibo.it (2)

(1) Dipartimento di Fisica e Astronomia - UNIBO
(2) Dipartimento di Medicina Specialistica, Diagnostica e Sperimentale - UNIBO





4
Estensione nazionale della mobilità da Codogno durante il mese di Febbraio, prima dell'istituzione della Zona Rossa. L'asse verticale rappresenta la maggiore/minore intensità degli spostamenti (normalizzata in scala logaritmica per motivi grafici). Il picco maggiore rappresenta Codogno, cioè spostamenti interni al comune in esame, ma si notano anche una moltitudine di spostamenti da Codogno che arrivano ad abbracciare un po' tutta la pianura padana, e si ramificano poi lungo le due dorsali, tirrenica e adriatica, anche se per numerosità questi ultimi due rami di spostamenti da Codogno risultano inferiori a quelli registrati entro la pianura padana.

La mobilità da Codogno in Lombardia presenta quindi una chiara impronta regionale, con numerosità di spostamenti maggiore nelle zone intorno a Codogno.

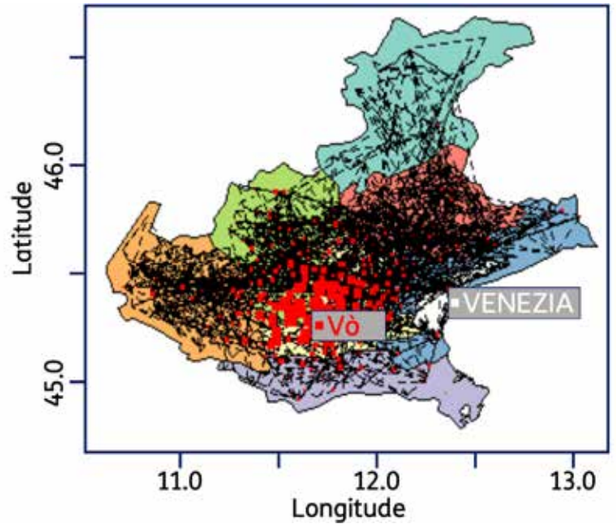
Considerata però l'estensione nazionale dell'epidemia, può risultare interessante dedicare un veloce sguardo anche all'osservazione della mobilità diretta da Codogno verso tutte le destinazioni italiane.
Questo fenomeno di mobilità diretta complessiva da Codogno può essere sinteticamente osservato nella Figura 4, la quale da un lato conferma quanto la mobilità diretta risulti prevalentemente regionale, ma dall'altro fa intuire anche l'esistenza di casi che si estendono verso la pianura padana, da Torino a Venezia, a cui si aggiungono due rivoli di mobilità che ricalcano le due dorsali italiane, quella adriatica e quella tirrenica.

La grande mobilità che interessa l'Italia, ed i relativi riflessi in termini di potenziale estensione del contagio, si vede quindi riflessa anche in questo fugace esempio, relativo alle tre settimane precedenti l'istituzione della zona rossa a Codogno.

Mobilità in Veneto

L'analisi dei dati telefonici in Veneto a febbraio 2020, prima dell'istituzione della Zona Rossa (DPCM 23/02/2020), ci fornisce un quadro d'insieme della mobilità delle persone in una regione, elemento che influenza la diffusione di un virus.

Osservando (vedi Figura 5) gli spostamenti effettuati da Vo' nella prima ventina di giorni del mese di febbraio, in un periodo cioè durante il quale il virus poteva spostarsi con le persone, non essendo ancora entrate in vigore le misure di contenimento della mobilità, si può notare che la mobilità diretta da Vo' intercetta solo marginalmente le principali direttrici (linee nere tratteggiate) del Veneto.
Queste ultime interessano prevalentemente il grande asse di comunicazione padano che va da Torino a Venezia. La mobilità in Veneto si estende comunque a tutto il territorio, anche se ovviamente risulta inferiore in alcune zone periferiche (es. quelle alpine a Nord) per evidenti vincoli orografici.



5
Mobilità diretta (evidenziata in rosso) ed indiretta (evidenziata dalle linee nere tratteggiate) da Vo'. Si nota la differenza tra la minore mobilità periferica (linee nere più rade) e la maggiore concentrazione di mobilità lungo le direttrici che convergono su Venezia (ad ovest da Torino-Milano, e a sud da Padova). Si nota altresì che la mobilità diretta da Vo' (quadratini rossi) rimane circoscritta alle zone limitrofe e solo marginalmente intercetta gli assi di mobilità del Veneto (linee nere tratteggiate, più dense dove la mobilità risulta maggiore). La dimensione dei quadratini rossi descrive graficamente la maggiore o minore numerosità di persone con mobilità diretta in quella specifica zona.

Valutazioni per i Modelli Epidemiologici

I dati di mobilità ricavati della Lombardia e del Veneto forniscono una misura degli spostamenti quotidiani durante il periodo di diffusione del COVID-19 nel nord Italia. Questi stessi dati quindi possono essere ulteriormente elaborati per diventare dati d'ingresso ad un modello in grado di descrivere il fenomeno di diffusione del contagio nel suo complesso, affiancandosi alla molteplicità di dati epidemiologici e di contesto che sono necessari per sviluppare una tale applicazione.
Il modello di diffusione, sviluppato dall'Università di Bologna, è accuratamente descritto nel box che ne approfondisce anche gli aspetti matematici.

Ulteriori sviluppi

Nello studio presentato si è mostrato come le Matrici Origine-Destinazione consentano di stimare i flussi di mobilità tra le diverse celle della rete di accesso di telefonia mobile, calando gli studi sulla diffusione planetaria dell'epidemia COVID-19 nell'ambito nazionale di specifico interesse, in primis i focolai di Codogno e Vo' con l'associata mobilità della Pianura Padana.

La profondità dell'analisi, arrivata fino al livello delle singole aree censuarie ISTAT in Italia, può traguardare una vista ancora più fine.
Per l'affinamento occorre entrare nelle aree censuarie e stimarne i flussi di mobilità interna, in modo da valutare anche singole situazioni di maggiore o minore aggregazione delle persone (e relativo rischio di contagio).

Tale vista ulteriormente raffinata sarà attuabile sia ricorrendo all'integrazione nei modelli epidemiologici di informazioni posizionali (GPS) raccolte da varie App installabili sui terminali, sia facendo leva sulle misure radio (anche georeferenziate GPS) che tutti i terminali mobili in-

viano come measurement report, continuativamente, per consentire il funzionamento ottimale della rete di accesso radiomobile.
La funzionalità specifica che abbina le misure radio prodotte e periodicamente riportate dai terminali alla relativa posizione GPS è stata introdotta negli standard radiomobili 3GPP internazionali a partire dal 3G UMTS, poi evoluta nel 4G LTE ed ora nel 5G. Tale prestazione, nota con la sigla MDT (Minimization of Drive Test) [7], rende la raccolta di posizioni GPS indipendenti dalla presenza di App nei terminali mobili, ereditando i vantaggi tipici delle prestazioni standard. MDT può operare infatti anche per i milioni di terminali mobili che ogni anno dall'estero arrivano in Italia (roaming internazionale) e che potrebbero non avere disponibile l'App giusta nella propria lingua. Inoltre, le misure radio nascono negli standard internazionali e sono progettate in modo da ridurre al minimo possibile l'impatto sui vari modelli di terminale (anche in termini di consumo di batteria), rendendo possibile la generazione di amplissime quantità di misure radio senza mai intaccare i bundle di traffico a pagamento dei clienti (come invece succede con le App).
Lo studio della mobilità integrando le Matrici Origine Destinazione con le misure georeferenziate, tra cui i dati MDT, costituisce quindi la naturale evoluzione per i modelli epidemiologici presentati in questo articolo, grazie anche al livello di diffusione della prestazione MDT sui terminali che rende robusta la relativa base statistica [8-14].

Conclusioni

La teoria dei Sistemi Dinamici ha proposto molti modelli per la diffusione delle epidemie e il recente sviluppo della Fisica dei Sistemi Complessi ha mostrato come le complesse reti di trasporto abbiamo un ruolo fondamentale nella diffusione a scala planetaria delle epidemie stesse. Tuttavia, le caratteristiche della attuale epidemia da COVID-19 hanno messo in evidenza come la peculiarità della mobilità microscopica in aree con diverse caratteristiche di antropizzazione potrebbe avere profondamente influenzato la sua diffusione.
Lo studio presentato mostra l'applicabilità pratica di adattare i modelli per la diffusione delle epidemie a contesti di mobilità su piccola scala (regionale), avvalendosi di dati telefonici (Matrici Origine-Destinazione) e traguardando benefici anche delle relative evoluzioni.
I modelli di diffusione epidemiologica, integrati con la conoscenza della effettiva mobilità territoriale, possono risultare molto utili non solo nella gestione delle problematiche durante la diffusione di un'epidemia per supportare le decisioni degli stakeholders, avvalendosi di tutte le conoscenze disponibili in un dato momento, ma anche aiutare lo sviluppo di politiche di prevenzione e gestione delle epidemie che facciano tesoro dell'esperienza vissuta. I benefici che possono derivare da una modellizzazione sempre più accurata dei fenomeni legati alla mobilità delle persone, saranno d'ausilio anche in futuro non solo nel contrasto sempre più efficace di scenari drammatici come quello indotto dal COVID-19, ma anche nelle auspicate fasi di ripresa della vita nelle città e, in generale, per lo sviluppo economico della nazione ■

Bibliografia

1. RStudio: Integrated Development Environment for R, www.rstudio.com
2. ISTAT – Istituto Nazionale di Statistica – Le Aree Censuarie sono costruite per somma delle unità minime di rilevazione su cui è organizzata la rilevazione censuaria. L'unità minima, la Sezione Censuaria, è costituita da un solo corpo delimitato da una linea spezzata chiusa. A partire dalle Sezioni di Censimento sono ricostruibili, per somma, le entità geografiche ed amministrative di livello superiore come le Aree Censuarie (località abitate, aree sub-comunali, collegi elettorali ed altre). Ciascuna Sezione di censimento è completamente contenuta all'interno di una ed una sola località. Il territorio comunale viene da ISTAT esaustivamente suddiviso in sezioni di censimento, in modo che la somma di tutte le sezioni di censimento ricostruisce l'intero territorio nazionale. www.istat.it
3. DPCM 23/02/2020 - Disposizioni attuative del decreto-legge 23 febbraio 2020, n. 6, recante misure urgenti in materia di contenimento e gestione dell'emergenza epidemiologica da COVID-19. https://www.gazzettaufficiale.it/eli/id/2020/02/23/20A01228/sg
4. M. Chinazzi et al., Science The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak10.1126/science.aba9757 (2020)
5. F. Brauer and C. Castillo-Chavez, Mathematical Models in Population Biology and Epidemiology, Springer-Verlag. New York, (2001)
6. R. Li et al., Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2) Science 10.1126/science.abb3221 (2020).
7. Reference Specification 37.320. Universal Terrestrial Radio Access (UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRA); Radio measurement collection for Minimization of Drive Tests (MDT); Overall description; Stage 2. https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2602
8. D Micheli, G Muratore, "Smartphones Reference Signal Received Power MDT Radio Measurement Statistical Analysis Reveals People Feelings during Music Events," Published in: 2019 PhotonIcs & Electromagnetics Research Symposium - Spring (PIERS-Spring). Date of Conference: 17-20 June 2019, Publisher: IEEE, DOI: 10.1109/PIERS-Spring46901.2019.9017606. https://ieeexplore.ieee.org/document/9017606
9. Davide Micheli, Riccardo Diamanti, "Statistical Analysis of Interference in a Real LTE Access Network by Massive Collection of MDT Radio Measurement Data from Smartphones," Published in: 2019 PhotonIcs & Electromagnetics Research Symposium - Spring (PIERS-Spring). Date of Conference: 17-20 June 2019. Publisher: IEEE. DOI: 10.1109/PIERS-Spring46901.2019.9017353. https://ieeexplore.ieee.org/document/9017353.
10. Andrea Scaloni, Pasquale Cirella, Mauro Sgheiz, Riccardo Diamanti, Davide Micheli. "Multipath and Doppler characterization of an electromagnetic environment by massive MDT measurements from 3g and 4g mobile terminals, " Published in: IEEE Access (Volume: 7), 21 January 2019, Page(s): 13024 – 13034. DOI: 10.1109/ACCESS.2019.2892864. https://ieeexplore.ieee.org/document/8620498
11. Chiara Mizzi, Alessandro Fabbri, Sandro Rambaldi, Flavio Bertini, Nico Curti, Stefano Sinigardi, Rachele Luzi, Giulia Venturi, Micheli Davide, Giuliano Muratore, Aldo Vannelli, Armando Bazzani, "Unraveling pedestrian mobility on a road network using ICTs data during great tourist events," EPJ Data Sci. (2018) 7: 44, Regular article, https://doi.org/10.1140/epjds/s13688-018-0168-2.
12. Davide Micheli, Giuliano Muratore, Aldo Vannelli, "Clima, MDT e Machine Learning per osservare il comportamento delle città" Telecom Italia, Notiziario Tecnico n.2-2019. https://www.telecomitalia.com/content/tiportal/it/notiziariotecnico/edizioni-2019/n-2-2019/N6-Clima-MDT-Machine-Learning-per-osservare-comportamento-citta.html
13. Davide Micheli, Giuliano Muratore, Aldo Vannelli, "La mobilità di breve e lungo raggio con le innovative misure radiomobili e l'Intelligenza Artificiale" Telecom Italia, Notiziaro Tecnico n.3-2018. https://www.telecomitalia.com/content/tiportal/it/notiziariotecnico/edizioni-2018/n-3-2018/N8-La-mobilita-breve-lungo-raggio-innovative-misure-radiomobili-e-Intelligenza-Artificiale.html
14. Davide Micheli, Giuliano Muratore, Aldo Vannelli, "Big Data georeferenziati MDT per servizi digitali nelle Smart Cities" Telecom Italia, Notiziaro Tecnico n.1-2018. https://www.telecomitalia.com/content/tiportal/it/notiziariotecnico/edizioni-2018/n-1-2018/N9-Big-Data-georeferenziati-MDT-per-servizi-digitali-Smart-Cities.html

Note

1. Per la Teoria dei Grafi la sequenza di spostamenti tra un nodo è l'altro è detta cammino lungo i "rami" di un grafo, di conseguenza la lunghezza di un cammino è la somma dei rami attraversati.
2. Per la Teoria dei Grafi il risultato della trasformazione delle Matrici O/D è rappresentato da Grafi orientati (il verso della freccia indica se la mobilità è in uscita o in ingresso nel nodo specifico) e pesati (dove il peso che si attribuisce a ciascun ramo che interconnette due nodi è proporzionale agli eventi di mobilità osservati in un dato periodo).
3. In Teoria dei Grafi si utilizza il concetto di "grado di interconnessione" di un nodo rispetto ad un determinato network/grafo. Questo parametro indica quanto è innervata l'area (cluster o network) di cui il nodo di riferimento fa parte. Il grado di interconnessione è un parametro che si usa anche nella Social Media Analysis per capire quanto un cluster di utenti è interconnesso.

Davide Micheli *davide.micheli@telecomitalia.it*

Laureato in Ingegneria Elettronica e delle Telecomunicazioni e in Ingegneria Aerospaziale e Astronautica, è entrato in azienda nel 1989 dove si occupato fino al 2001 di Progettazione, Realizzazione impianti, Esercizio e Qualità nell'Area Territoriale di Ancona. Dal 2002 si è traferito a Roma dove lavora tuttora nel settore di Ingegneria della Rete di Accesso Radio occupandosi di varie tematiche connesse con l'ingegnerizzazione della rete tra cui quelle legate allo studio della propagazione elettromagnetica. Negli ultimi anni, dopo aver conseguito un Dottorato di Ricerca in Ingegneria Aerospaziale, ha iniziato ad approfondire nell'ambito del suo lavoro le tecniche di machine Learning, in particolare, sui Big Data di tipo elettromagnetico statistico disponibili nella rete di accesso radio. E' inoltre autore di numerosi articoli scientifici su riviste internazionali ■

Giuliano Muratore *giuliano.muratore@telecomitalia.it*

Laureato in Ingegneria Elettronica, è entrato in azienda nel 1987, ricoprendo responsabilità prima nel nascente mercato liberalizzato dei servizi di messaggistica interpersonale (1990) ed in seguito nello sviluppo della Rete e dei Servizi Radiomobili di TIM (1995), con incarichi nell'evoluzione del Piano di Numerazione Nazionale (1997) e nell'introduzione in Italia della Mobile Number Portability (2001), per poi seguire il Mobile Roaming business (2010) e successivamente progetti internazionali TIM in GSMA. Negli ultimi anni ha messo la sua esperienza a disposizione della formazione Big Data e dello sviluppo delle tecniche di Machine Learning applicate a dati radiomobili ■

Giuseppe Sola *giuseppe.sola@olivetti.com*

Laureato in Scienze dell'Informazione, nel 1996 entra TIM dove nel corso degli anni ha ricoperto diversi ruoli sia ambito tecnologico che commerciale sviluppando competenze specifiche su sistemi, piattaforme e processi nell'ambito delle telecomunicazioni e dei servizi applicativi. Ha sempre seguito con interesse l'innovazione contribuendo alla costruzione prima e al lancio poi dei primi servizi a valore aggiunto di TIM e avviando la trasformazione digitale della Customer Interaction del Gruppo Telecom Italia. Nel 2013 passa in Telecom Italia Digital Solutions a capo della Linea di Business Web & Applications Services dedicata alle soluzioni di Customer Digital Management e Advanced Analytics. In Olivetti da Gennaio 2016 assume la responsabilità prima del Marketing&Sales della divisione Smart Retail e successivamente della Business Unit Data Monetization Solutions. Oggi a capo della direzione vendita Digital Services continua a seguire con interesse le evoluzioni tecnologiche e applicative in ambito Big Data Analytics & IoT ■

Aldo Vannelli *aldo.vannelli@telecomitalia.it*

Laureato in Fisica e in Ingegneria dell'Informazione, in azienda dal 1988. Dopo un'ampia esperienza nell'ambito dell'Ingegneria e dell'Innovazione delle Reti Dati (Frame Relay, ATM e IP), nel 2001 passa in TIM per occuparsi dello sviluppo e dell'innovazione di applicazioni e servizi multimediali su tecnologie 2.5G/3G. In questo ambito ha coordinato numerosi progetti riguardanti l'integrazione multiservizio di voce/video/dati su mobile e lo sviluppo di soluzioni per il Mobile Content Distribution. Dal 2012 lavora nella Direzione Business & TOP Clients dove si occupa dello sviluppo di iniziative e progetti innovativi per Aziende di rilevanza Nazionale. Da alcuni anni si interessa dello sviluppo di iniziative finalizzate alla realizzazione di Proof of Concept basati su tecniche Big Data Analytics & Machine Learning ■

QUANTUM COMPUTING PER L'OTTIMIZZAZIONE DELLE RETI MOBILI (4.5G E 5G)

Andrea Boella, Michele Ludovico,
Giuseppe Minerva, Mauro Alberto Rossotto

Il quantum computing è tra le tecnologie innovative di maggior interesse, accompagnato da un "hype" sui mezzi di comunicazione per le potenzialità che sembrano promettere un impatto disruptive in diversi campi.

La competizione per la leadership tecnologica nel quantum computing è partita da alcuni anni e l'interesse è cresciuto negli ultimi tempi grazie ai progressi della ricerca e al recente annuncio sul raggiungimento della "quantum supremacy" da parte di Google, cioè l'individuazione di un primo esempio reale di algoritmo implementato sul Quantum Computer, che non può essere elaborato in tempi di calcolo accettabili sui computer classici.

Questi risultati hanno creato grandi aspettative, ma, allo stato attuale di sviluppo della tecnologia, si prevede che la disponibilità "commerciale" di quantum computers con potenze di calcolo elevate e la loro applicazione su larga scala, si verificherà su un orizzonte temporale stimabile tra i 5 e i 10 anni.

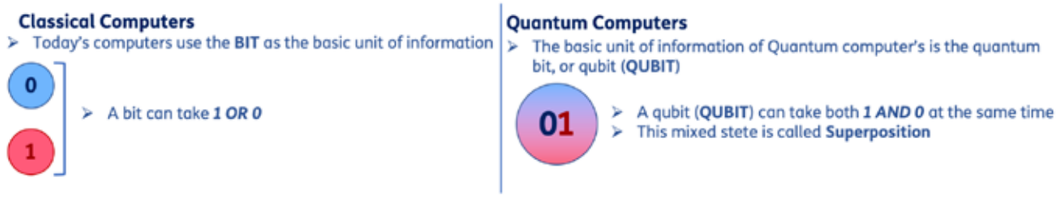
Ciò non toglie che il quantum computing possa già essere applicato ad alcuni casi d'uso, sfruttandone opportunamente le capacità computazionali attualmente disponibili. È quanto è avvenuto nel contesto TIM, in cui è stato applicato ad un problema di pianificazione delle reti mobili 4.5G e 5G. I risultati sono incoraggianti avendo ottenuto una maggiore rapidità di esecuzione (con un fattore 10x) e una qualità media superiore rispetto ai metodi tradizionali di ottimizzazione.

TIM TIM

Introduzione al quantum

L'idea che ha dato origine allo sviluppo del quantum computing nasce dall'osservazione che i fenomeni della natura e le aree scientifiche che dal loro studio sono state derivate (fisica, chimica, biologia...), sono governati dai fenomeni della meccanica quantistica. Pertanto, per progredire ulteriormente nella conoscenza di questi campi, è necessario disporre di un computer che abbia una logica computazionale che riproduca gli stessi principi, cioè – in altri termini – un'analoga modalità di funzionamento. In questo modo, tutte le attività di ricerca e sviluppo (elaborazione di dati, definizione di modelli, simulazioni, analisi predittive, ecc.) potranno essere effettuate più facilmente e in maniera più accurata, perché più precisa potrà essere la rappresentazione dei fenomeni da analizzare.
La caratteristica di questi problemi è quella di avere una complessità computazionale che cresce esponenzialmente all'aumentare delle dimensioni dei dati di input.
La complessità di questi problemi, dovuta alla numerosità delle variabili, impedisce ai computer classici di eseguire simulazioni o ricerche di ottimizzazione efficaci a causa del loro approccio "lineare e sequenziale". La crescita della capacità computazionale degli elaboratori classici non è una soluzione praticabile in questo ambito, per cui si è guardato ai fenomeni della meccanica quantistica per realizzare degli elaboratori che cambiassero l'approccio e la metodologia di lavoro. Sono stati così teorizzati e poi progettati i primi computer quantistici che, grazie ad alcune caratteristiche peculiari della meccanica quantistica, permettono un'elaborazione parallela e probabilistica in grado di affrontare e risolvere questi casi complessi.[1]
Questo nuovo approccio "quantistico" è generalizzabile, perché questa tipologia di problemi si incontra in vari contesti come la chimica dei materiali, la meteorologia, la finanza.
Pur disponendo di una potenza di calcolo elevata, i quantum computers non sostituiranno i computer classici, ma lavoreranno in sinergia con questi secondo un modello di computazione ibrida classica – quantistica. Le due tecnologie, infatti, sono adatte a classi distinte di problemi: alle famiglie di problemi elaborabili con computer classici, si aggiungono quindi le famiglie adatte ai quantum computers.
Nei computer classici, l'unità di informazione attraverso cui sono codificati i dati da processare è rappresentata dal bit che può assumere due valori in maniera fra loro esclusiva (0 oppure 1). Nei computer quantistici l'unità di informazione è, invece, il qubit (quantum bit) che, grazie ai principi della meccanica quantistica che ne sottendono l'implementazione, può assumere non solamente i valori 0 o 1, ma anche entrambi contemporaneamente (0 e 1), in uno stato di sovrapposizione intermedio tra i due stati fisici corrispondenti ai valori 0 e 1. Per usare un'analogia, si può pensare ad una moneta che a seconda di come è posizionata può essere rivolta verso il lato "testa" o lato "croce". Se la moneta viene fatta roteare, finché non si ferma entrambe le facce "testa" e "croce" sono contemporaneamente possibili; in questa condizione la moneta è come se si trovasse in uno stato di sovrapposizione. Questo stato, che nel mondo quantum si chiama superposition, conferisce ai quantum computer quelle caratteristiche probabilistiche alla base delle loro potenze di calcolo esponenziali. Le capacità computazionali dei quantum computers si misurano sulla base del numero di qubit che sono in grado di gestire.
Nella famiglia dei quantum computer, si possono distinguere alcune tipologie di processori - anche denominati QPU (quantum processing unit) - in base alla tecnologia e alla finalità:

- Quantum gate array: anche denominato universal quantum computer per intendere la macchina in grado di elaborare qualunque processo computazionale complesso (qualunque algoritmo quantistico) in forma massiva (senza limitazioni sulle dimensioni dei dati trattati) in tempi rapidi. Si tratta del quantum computer così come lo immaginiamo per analogia con i computer classici, versatile con prestazioni massime, ma anche molto complesso da realizzare,



1
Confronto tra Classical and Quantum Computing
Il bit, unità di informazione classica, può assumere solo uno dei possibili valori alla volta (0 OR 1) mentre i qubit, unità di informazione quantistica, possono assumere entrambi i valori (0 AND 1) contemporaneamente (superposition)

a causa dell'elevato numero di qubit necessari. Per avere un'idea del grado di interesse attorno al quantum gate array, i vendor che stanno investendo in questa tecnologia sono IBM, Google, Microsoft, Intel, Alibaba. I quantum computer attualmente più potenti supportano circa 50 qubit e per arrivare a disporre di macchine che ne equipaggino alcune centinaia, quantità necessaria per le dimensioni dei problemi che dovrebbero trattare, occorrerà, in base alle attuali previsioni, aspettare alcuni anni (5 – 10). Per coprire questo gap temporale e poter già applicare il quantum computing a casi d'uso reali, la tecnologia si è orientata anche su soluzioni alternative, descritte nei punti seguenti. [2]

- Quantum annealer: è un computer che utilizza una modalità di elaborazione che riproduce un fenomeno fisico specifico: la tendenza naturale che porta un sistema a raggiungere la condizione di equilibrio a minima energia. Modellando quindi un fenomeno ben definito, è un computer adatto a trattare una classe specifica di problemi, quelli di ottimizzazione di natura combinatoria. Il fornitore più rappresentativo della categoria quantum annealer è canadese, D-Wave che fornisce un processore di 2000 qubit e ha in roadmap il rilascio commerciale della sua evoluzione a 5000 qubit nel corso del 2020. La numerosità dei qubit così elevata rispetto al quantum gate array, è dovuta al fatto che i quantum annealer non sono macchine universali, ma ottimali per una specifica famiglia di problemi e nella loro modalità di processing la gestione dei qubit è più semplice, semplificando di conseguenza anche la tecnologia di base. Occorre anche considerare che la potenza di calcolo del quantum annealer non è in relazione diretta col volume di qubit supportati, ma ridotta perché i qubit non sono tra loro connessi a maglia completa.

Attualmente entrambe le tipologie funzionano come dei quantum computer solo se protette in un ambiente pressoché isolato e prossimo alla temperatura dello zero assoluto. Qualunque minimo disturbo (vibrazione, variazione di temperatura, rumore, ecc.) provoca la perdita delle loro proprietà quantistiche e di conseguenza le loro capacità computazionali. Queste tipologie di quantum computer, quindi, devono essere installate in ambienti che garantiscono quelle specifiche condizioni di isolamento tipiche dei laboratori. Per questa ragione l'accesso ai quantum computers attualmente è fornito in modalità cloud attraverso un'interfaccia, tramite la quale si forniscono i dati di input e il codice dell'algoritmo da eseguire scritto in un linguaggio di alto livello e si ottengono i risultati dell'elaborazione.

D-Wave's vision:
the future of quantum computing and its applications

Quantum computing has the potential to fundamentally change society and how we interact with technology.
Quantum is poised to disrupt a number of major industries, including transportation, mobility, materials science, medicine, and more, enabling powerful new applications that aren't possible with classical supercomputers.

While it's still early in the lifespan of quantum computing, promising applications with practical business value are emerging.

Cloud access and hybrid quantum/classical computing approaches have made it much easier for virtually any organization or developer to get started building quantum applications.

To date, D-Wave users and customers have built over 200 quantum applications in diverse fields.

Major companies, including Menten AI, Volkswagen, the German Aerospace Center, and the Italian State Railway, have used D-Wave quantum computers to optimize various parts of their R&D, resource allocation, and supply chain.

As quantum computers increase qubit counts and processing power – such as D-Wave's upcoming 5000 qubit Advantage system – applications will become more robust and lead to breakthroughs in critical fields, including drug discovery, transportation, and machine learning, among others.

D-Wave's Quantum Annealing Computers

D-Wave quantum computers use a process called quantum annealing to find solutions to problems expressed in terms of optimizing a certain function. A problem is represented by a graph with real-valued weights on nodes and edges. The graph is mapped onto an arrangement of qubits (nodes) and couplers (edges) inside the quantum processing unit (QPU). During the computation (called an anneal), gradually-evolving forces are applied to the qubit system, to drive it into an energy state that corresponds to an optimal solution to the original problem. Typical anneal times range from 2 to 200 microseconds per input.

A current-generation D-Wave 2000QTM system contains 2000+ qubits with up to six couplers per qubit (fewer on circuit boundaries). This allows the QPU to solve problem graphs with between roughly 64 nodes (dense) and 2000 nodes (sparse), depending on qubit yields.

The QPU operates within a highly shielded environment --- at temperatures below 15 millikelven, and experiencing less than 50,000x of Earth's magnetic field --- to protect the computation from external noise and improve the probability of a successful outcome. Typical anneal times range from 2 to 200 microseconds per input.

The next-generation AdvantageTM system, to be launched later this year, will contain 5000+ qubits and approximately 15 couplers per qubit, corresponding to problem graphs with between roughly 182 and 5000 nodes.

Dr. Catherine McGeoch
cmcgeoch@dwavesys.com

Per la modalità on-premises, largamente diffusa con i computer classici, occorrerà attendere l'evoluzione tecnologica.

In questa fase transitoria, in attesa della piena maturità del quantum computing, si stanno affacciando sul mercato soluzioni che possiamo definire "QUANTUM INSPIRED". Si tratta di emulatori Hardware e/o Software che permettono di sviluppare codice Quantum Ready, eseguendolo inizialmente su processori lineari come CPU e GPU. Le potenzialità equivalenti di queste soluzioni sono nell'ordine di alcune decine di qubit e quindi comparabili con le potenze di calcolo degli attuali quantum computer di tipo gate array. Rispetto a questi hanno il vantaggio che i loro qubit sono più stabili, visto che utilizzano hardware classico. Ricordando che la potenza di calcolo dei quantum annealer è inferiore rispetto al dato di targa, i quantum inspired sono comparabili - in termini di prestazioni – anche con questa tipologia di quantum computer.

Oltre ai fattori prestazionali, le soluzioni quantum inspired sono interessanti per una serie di ulteriori motivi:

- il software sviluppato per queste piattaforme sarà riutilizzabile sui computer quantistici di potenza elevata, quando si renderanno disponibili
- offrono un ambiente di sviluppo sul quale acquisire esperienza nella modellizzazione e programmazione quantistica
- la programmazione quantistica degli algoritmi porta già dei primi vantaggi negli scenari di ottimizzazione, poiché la corrispondente modellizzazione è diversa e abilita la ricerca non lineare di soluzioni al problema
- utilizzando processori classici (CPU e GPU), ammettono anche la modalità di installazione on-premises o in Edge (sempre che le loro potenze di calcolo siano sufficienti per le applicazioni su cui si intende impiegarli)

La disponibilità di computer quantici di capacità di calcolo elevate, potrà sicuramente ridurre i tempi di elaborazione di queste stesse procedure sviluppate su quantum inspired e contemporaneamente aiutare a trovare le soluzioni ottime, cioè di maggiore qualità.

Applicazioni del quantum computing

I computer quantistici sono adatti per trattare problemi complessi, che sono rappresentabili con modelli multidimensionali nei quali intervengono un numero elevato di variabili e che richiedono, di conseguenza, tempi di elaborazione molto lunghi, talvolta non compatibili con le tecnologie tradizionali.
I casi d'uso che presentano queste caratteristiche sono tipicamente:

- problemi di ottimizzazione
  - ottimizzazione nell'assegnazione/riuso di risorse soddisfacendo certi vincoli, come per esempio nei casi assimilabili al problema di colorazione delle mappe
  - ottimizzazione di percorsi (esempio tipico è il travel salesman problem) per gestione di veicoli, robot, droni, aerei, flusso di traffico ecc.
- ricerche esaustive su spazi di ricerca di grosse dimensioni
  - applicazioni legate alla crittografia, per sviluppare nuovi protocolli di cifratura resistenti ai quantum computer
- classificazione di dati di grosse dimensioni o per identificazione di pattern ricorrenti
  - applicazioni per il Machine Learning di image processing, analisi predittive, diagnosi
  - set ottimale dei pesi di una rete neurale.

Le applicazioni descritte sono non solo complesse a livello computazionale, ma anche – nella maggior parte dei casi – time-critical.
La gestione real-time di questi casi d'uso, che tipicamente viene soddisfatta collocando la logica di controllo in prossimità del punto di attuazione, al momento non è ancora possibile con i quantum computers ma lo diventerà col miglioramento della tecnologia.

Nella prospettiva che il quantum computing diventi una tecnologia adatta per un'installazione on-premises, si può pensare ad un suo utilizzo a livello di edge, proprio per la gestione delle applicazioni time-critical.

Un'alternativa, per disporre di capacità computazionali quantistiche all'edge, è rappresentata dalle soluzioni quantum inspired, presumibilmente realizzabili in tempi più brevi e compatibilmente con le prestazioni che sono in grado di garantire [3].

Un esempio di ottimizzazione della rete mobile: la pianificazione degli identificativi di cella (PCI)

Nell'ambito delle telecomunicazioni, problemi di ottimizzazione sono presenti in differenti contesti in cui le risorse sono limitate e condivise. Un primo esempio che è stato affrontato e sperimentato con le tecniche di quantum computing è la pianificazione degli identificativi di cella (PCI: Physical Cell Identifier) nelle reti mobili 4.5G e 5G. I

n ambito radiomobile, si definisce "cella" l'unità elementare di territorio corrispondente ad una antenna trasmittente e ad una specifica banda frequenziale. L'identificazione della cella, proprio attraverso il PCI, consente ai terminali mobili di gestire le procedure di mobilità nel passaggio tra celle geograficamente adiacenti.

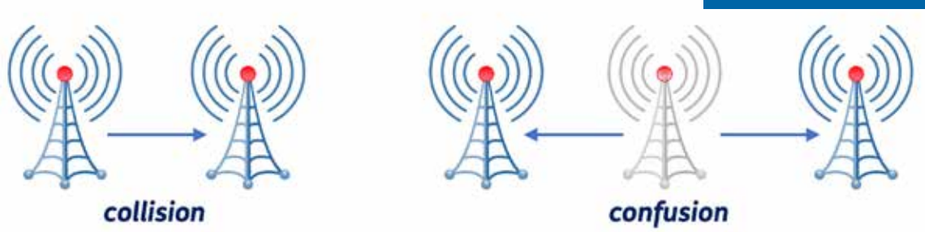
L'identificativo PCI è assegnato a ciascuna cella della rete mobile all'interno di un set predefinito di valori, variabile da sistema a sistema in base alle specifiche caratteristiche della rete. In termini generali, il PCI è composto da due differenti campi, denominati – rispettivamente – Group_ID e Cell_ID; il valore risultante dell'identificativo è espresso, in funzione delle due componenti, mediante la relazione:

[Eq1] $PCI = 3 \cdot Group\_ID + Cell\_ID$

La coppia Group_ID e Cell_ID è usata sia per la generazione dei canali di sincronizzazione utilizzati dal terminale mobile per agganciarsi correttamente alla cella sulla quale intende operare, sia per generare il segnale di riferimento della cella medesima (un vero e proprio "faro") la cui individuazione è necessaria affinché il terminale possa decodificare le informazioni di servizio trasmesse dal sistema.

L'obiettivo finale della pianificazione dei PCI è la definizione di un "piano" che minimizzi i problemi correlati ai cosiddetti casi di "collision" (PCI uguali assegnati a celle adiacenti) e "confusion" (PCI uguali assegnati a celle aventi un'adiacenza in comune) illustrati sinteticamente in Figura 2.
In presenza di tali situazioni, infatti, il terminale mobile può fallire le procedure di mobilità nel passaggio da una cella all'altra e – di conseguenza – subire una caduta di chiamata. Un ulteriore obiettivo, meno prioritario rispetto al precedente, è evitare l'assegnazione di PCI con lo stesso Group_ID a nodi geograficamente vicini ("adiacenti") e operanti sulla stessa banda frequenziale, condizione utile – insieme all'uso obbligatorio di un medesimo Group_ID all'interno di ciascun nodo – al fine di facilitare le funzionalità di decodifica.

2
Condizioni di "collision" e "confusion" fra celle delle reti 4.5G o 5G



I problemi di "collision" e "confusion" corrispondono, a tutti gli effetti, a veri e propri vincoli di assegnazione (comunemente detti di "riuso"), che discendono dall'esistenza delle adiacenze di rete e/o elettromagnetiche specifiche del sistema 5G o LTE e delle adiacenze di rete dovute alla presenza, nel medesimo territorio, di celle delle reti 2G e/o 3G.

In tale contesto è importante evidenziare come la gestione del passaggio di una connessione in essere dalle reti di precedente generazione alla rete LTE, sia spesso garantita semplicemente definendo "in adiacenza" la cella 2G o 3G di interesse e la banda frequenziale LTE verso la quale si intende far migrare la generica chiamata, senza indicazione di specifiche celle della rete 4.5G.

In termini generali, la definizione di un piano dei PCI che soddisfa il maggior numero possibile di vincoli di "riuso" permette di migliorare le prestazioni della rete di accesso e, quindi, la qualità del servizio sperimentata dagli utenti.
In particolare, una buona pianificazione consente di ridurre significativamente i tassi di caduta delle connessioni in corso, il tutto grazie alla diminuzione dei fallimenti nell'esecuzione delle procedure di Hand-Over (cioè il passaggio della connessione da una cella ad un'altra). Ciò implica notevoli benefici – in particolare – per le connessioni "voce" (chiamate VoLTE, nel contesto LTE) per le quali la velocità di esecuzione e il corretto completamento della procedura sono di fondamentale importanza per ottenere prestazioni adeguate.

Il numero di differenti PCI disponibili ai fini della pianificazione degli identificativi di cella dipende dalla rete considerata ed è pari a 1008 (con numerazione 0÷1007) nel caso 5G e 504 (con numerazione 0÷503) nel caso LTE. Tali valori sono generati mediante la relazione [Eq 1] in base al Group_ID e al Cell_ID, i cui campi di esistenza sono i seguenti:

per 5G
$Group\_ID \in [0; 355]$ $Cell\_ID \in [0; 2]$

per LTE
$Group\_ID \in [0; 167]$ $Cell\_ID \in [0; 2]$

Al fine di meglio quantificare il potenziale peggioramento delle prestazioni della rete, è associato – al mancato rispetto dei vincoli di "riuso" – un "costo" dipendente sia dalla tipologia del vincolo coinvolto, sia dalla numerosità delle violazioni introdotte (una generica coppia di celle, infatti, può violare contemporaneamente più vincoli di "riuso", anche della stessa tipologia).

Il costo relativo a ciascun tipo di violazione dipende, quindi, sia dalla conformazione geografica del territorio – attraverso la copertura del segnale radio garantita da ciascuna cella che incide direttamente sulla definizione delle adiacenze – sia dall'importanza che si attribuisce a ciascun vincolo non rispettato, che è propria di ogni Telco Operator poiché derivante dalla lunga esperienza accumulata nel corso dello sviluppo delle reti.

In conseguenza di quanto descritto e data una generica coppia di celle (i,j), i due fattori di costo sono i seguenti:

$C_{i,j}$ costo ("degrado") dovuto all'assegnazione dello stesso PCI alle celle i e j

$S_{i,j}$ costo ("degrado") dovuto all'assegnazione dello stesso gruppo alle celle i e j.

Il costo complessivo (cioè il "degrado" totale) di un piano di assegnazione, è dato da una funzione di costo contenente i contributi (alcuni di valore nullo) dovuti alla potenziale violazione dei vincoli di "riuso" relativi ai PCI e ai Group_ID di ciascuna possibile coppia di celle oggetto della pianificazione.

La funzione assume, quindi, la seguente forma:

$Tot_{cost}=\Sigma_i\Sigma_j C_{i,j} \cdot v_{i,j}+\Sigma_i\Sigma_j S_{i,j} \cdot w_{i,j}$ [Eq 2]

dove:

$v_{i,j}$ = 1 se lo stesso PCI è assegnato alle celle i e j
$v_{i,j}$ = 0 altrimenti
$w_{i,j}$ = 1 se lo stesso Group_ID è assegnato alle celle i e j
$w_{i,j}$ = 0 altrimenti.

La pianificazione degli identificativi PCI effettuata in coordinamento con l'algoritmo automatico di individuazione delle adiacenze di rete (operante in closed looop sulla base delle misure raccolte dai terminali) è gestita in TIM attraverso la piattaforma Software Open SON (Self Organizing Network) descritta in [4]. In [5] sono illustrati, inoltre, i risultati ottenuti – in termini di miglioramento delle cadute VoLTE – attraverso la pianificazione dei PCI in "closed loop" integrando, all'interno del framework, gli algoritmi di ricerca operativa sviluppati da TIM e impiegati anche nelle fasi di progettazione della rete.
Nel caso della pianificazione dei PCI, in particolare, è utilizzato un algoritmo di tipo "Fast Greedy" [6].

L'approccio alla pianificazione dei PCI basato sul Quantum Computing si inserisce nel contesto Open SON, come descritto in Figura 3.

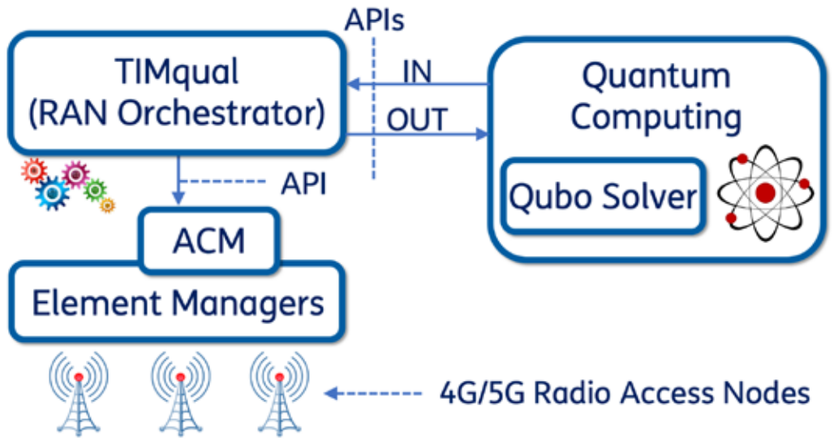
In particolare, il modulo RAN Orchestrator (corrispondente all'applicativo TIMqual sviluppato "in house" da TIM) si occupa di gestire lo scambio di informazioni con gli elementi di rete, attraverso le API (Application Programming Interfaces) rese disponibili dai sistemi di ACM (Automatic Configuration Management) forniti dai RAN vendors nell'ambito dei sistemi di gestione della rete (Element Managers).

L'applicazione del Quantum Computing è stata pensata con un ulteriore scambio di informazioni verso sistemi esterni (nel caso specifico "hosted" presso D-Wave). Un maggiore livello di integrazione sarà valutato nell'ipotesi di disponibilità di quantum computer gestiti "on premises" all'interno della rete TIM.

Modellizzazione dell'algoritmo di pianificazione PCI in ottica quantum Computing

Modello QUBO

L'attività di pianificazione degli identificativi di cella (PCI), ha l'obiettivo, come descritto, di ottimizzare



3
Framework Open SON di TIM e interazione con la piattaforma di Quantum Computing

REPLY'S VISION: ALGORITMI E CASI D'USO PER IL QUANTUM COMPUTING

Il Quantum Computing è una disciplina al confine tra la fisica, la matematica, l'informatica e l'ingegneria in continuo sviluppo, il cui scopo è quello di svolgere operazioni computazionalmente costose in tempi molto ridotti e, allo stesso tempo, raggiungendo risultati di alta qualità.
La gamma di problemi che può essere affrontata attraverso il formalismo quantistico è molto ampia e spazia dal Machine Learning e Intelligenza Artificiale all'Ottimizzazione Combinatoria.
Nel primo caso si vedono come maggiori protagonisti i cosiddetti Universal Quantum Computer, ovvero dei calcolatori quantistici in cui è possibile accedere direttamente allo stato dei qubit e, pertanto, modificarlo in modo arbitrario. Questo tipo di approccio ha permesso lo sviluppo di algoritmi come il Variational Quantum Eigensolver (VQE) o il Quantum Approximate Optimization Algorithm (QAOA) particolarmente adatte al Machine Learning e alla simulazione di sistemi quantistici. L'idea alla base di queste due strategie consiste nella cooperazione tra macchine classiche e quantistiche: la velocità della Quantum Processing Unit (QPU) viene sfruttata per misurare l'energia del sistema, mentre la parte classica viene sfruttata attraverso un sistema che di fatto impara la strategia migliore per codificare i dati in una macchina quantistica. Di conseguenza diventa possibile trovare, ad esempio, pattern nascosti nei dati per risolvere problemi di Supervised Learning, oppure ottenere il minimo di una funzione arbitraria in un problema di ottimizzazione. È su questo principio che si fondano algoritmi quali Quantum Support Vector Machines (QSVM) e Quantum Neural Networks (QNN).
Per quanto riguarda i problemi di ottimizzazione combinatoria, risultano adatti i cosiddetti Quantum Annealers. Questi incapsulano un diverso tipo di QPU in cui non è possibile manipolare direttamente lo stato dei qubit, bensì occorre programmare la funzione di energia cui il sistema quantistico è sottoposto. Queste macchine lavorano lasciando che i qubit evolvano naturalmente verso il minimo della funzione d'energia data dal modello di Ising.
L'approccio che fa uso dei Quantum Annealers si presta particolarmente alla soluzione di problemi che difficilmente possono essere riscritti mediante modelli lineari, come, ad esempio, la pianificazione dei Physical Cells Identifier (PCI) per l'ottimizzazione delle reti mobili 4.5G e 5G. Inoltre, dato che il modello riguarda un caso reale e quindi complesso, risulta fondamentale utilizzare il Quantum Annealer applicando tutte le possibili strategie che ne consentono il miglior utilizzo. Questo si rende necessario in quanto modalità diverse impattano significativamente sulle caratteristiche della soluzione trovata, come approfondito in questo lavoro in pubblicazione su Springer Quantum Machine Intelligence [8].
Nel caso della pianificazione dei PCI, la velocità di esecuzione dell'algoritmo unito alla possibilità di trovare soluzioni di qualità molto alta a problemi non-lineari ha permesso di verificare come gli algoritmi Quantistici possano essere applicati con successo a problemi reali.

L. Asproni l.asproni@reply.it
D. Caputo da.caputo@reply.it
M. Magagnini m.magagnini@reply.it

l'assegnazione rispettando i vincoli di "riuso".
È quindi un problema di natura combinatoria analogo, in linea di principio, al problema di colorazione delle mappe geografiche, nelle quali due paesi confinanti devono assumere colori diversi.
Nella classe dei problemi di ottimizzazione combinatoria, un framework che trova ampia applicazione è il QUBO (Quadratic Unconstrained Binary Optimization)[7].

In questi problemi, generalmente, intervengono una serie di variabili binarie, che possono quindi assumere due soli valori corrispondenti a due stati "netti" (0/1, si/no, on/off...). Ogni combinazione del set di variabili determina un risultato che, a seconda del problema, rappresenta un costo o un guadagno.
In termini generali, il QUBO è un modello matematico per ottimizzare un problema che può essere rappresentato nella seguente forma:

minimizzare $y = x^T Qx$ [Eq 3]

Analizzando la formulazione matematica si può notare che:

- il problema viene descritto matematicamente in una forma quadratica, nella quale il vettore x rappresenta il set delle variabili di cui occorre trovare la combinazione di 0 e 1 ottimale
- la matrice Q è composta da elementi di valore costante che, oltre a modellare matematicamente il problema, tengono conto anche dei vincoli a cui le variabili devono sottostare. L'artificio matematico per includere questi vincoli porta ad avere ad una formulazione "compatta", autoconsistente e quindi "unconstrained"
- l'ottimizzazione consiste nell'individuare il minimo della funzione di costo quadratica.

Mettendo assieme gli elementi descritti, si ottiene il "Quadratic Unconstrained Binary Optimization" model, cioè il QUBO.
La generalizzazione del modello ai casi in cui anziché il minimo occorra individuare il massimo, si deriva riformulando la funzione obiettivo del problema nel suo opposto negativo; in questo modo il metodo di risoluzione continua ad essere lo stesso, perché il valore minimo che si ottiene corrisponde al valore massimo del problema originario.

L'attività di sviluppo di un algoritmo di ottimizzazione si può così ricondurre alla riformulazione del problema secondo il modello QUBO e all'utilizzo delle librerie che risolvono questi casi in maniera efficiente.
Questa metodologia di lavoro consente di standardizzare e rendere efficiente l'intero processo.

La grande varietà di problemi a cui il modello è applicabile, unitamente all'ampia gamma di piattaforme (CPU e GPU) che supportano le librerie software di risoluzione, danno al framework QUBO un ruolo distintivo nell'ambito dei problemi di ottimizzazione combinatoria.

A rafforzare ulteriormente l'importanza del modello, contribuisce il fatto che anche nelle librerie software delle piattaforme di quantum computing sono stati integrati solver di risoluzione per il QUBO.
Per quanto visto nell'introduzione, i quantum annealer sono particolarmente adatti per i problemi di ottimizzazione di natura combinatoria; per questo motivo l'algoritmo di pianificazione PCI, sviluppato secondo il modello QUBO, è stato eseguito sul quantum computer di D-Wave 2000QTM.
Lo stesso algoritmo QUBO è stato elaborato anche su CPU e GPU.

Modellizzazione pianificazione identificativi di celle

Riprendendo la formulazione matematica degli identificativi di cella [Eq 1] unitamente ai vincoli di "riuso", la modellizzazione del problema – secondo il framework QUBO – si traduce nella seguente forma:

$$QUBO=\lambda_1 H_{PCI_{HC}}+\lambda_2 H_{GROUP_{HC}}+\lambda_3 H_{PCI_{SC}}+\lambda_4 H_{GROUP_{SC}}+\lambda_5 H_{GROUP\_yHC}$$ [Eq 4]

Nella quale intervengono 5 contributi:

$H_{PCI_{HC}}$ esprime il vincolo di utilizzare PCI diversi per celle adiacenti

$H_{GROUP_{HC}}$ esprime il vincolo di utilizzare Group_ID differenti per nodi diversi

$H_{PCI_{SC}}$ esprime il costo che deriva dall'utilizzo dello stesso PCI su celle adiacenti

$H_{GROUP_{SC}}$ esprime il costo che deriva dall'utilizzo dello stesso Group_ID su nodi adiacenti

H_{GROUP_yHC} esprime il vincolo di utilizzare Cell_ID diversi per celle dello stesso nodo.

I pesi di ognuno (le λ) devono essere calibrati per individuare l'area ottimale di lavoro.

L'elaborazione di questo algoritmo "completo", comprensivo di tutte le componenti, ha però presentato alcuni problemi:

- la calibrazione dei pesi (le λ) è risultata complessa per la quantità di parametri (5)
- il modello è poco scalabile
- la qualità della soluzione non è migliore rispetto al metodo tradizionale.

Queste criticità hanno portato così allo sviluppo di un nuovo modello seguendo una strategia alternativa:

- il modello è stato scomposto in fasi computazionali, definite in modo che la complessità computazionale di ognuna fosse gestibile con i computer classici e quantistici attuali, utilizzandoli anche entrambi con flussi di elaborazione ibrida
- le λ sono state suddivise nelle varie fasi, semplificandone il tuning.

Così facendo l'algoritmo QUBO è stato strutturato in due fasi:

1. primo layer: assegnazione degli identificativi di gruppo (Group_ID) con la definizione di un modello QUBO che coinvolge i soli due termini: $H_{GROUP_{HC}}$, $H_{GROUP_{SC}}$
2. secondo layer: assegnazione dei Cell_ID di ogni cella all'interno dei gruppi, utilizzando l'assegnazione dei gruppi del primo layer, ottenendo così la distribuzione dei PCI di ogni cella. Il corrispondente modello QUBO contiene quindi i tre termini: $H_{PCI_{HC}}$, H_{GROUP_yHC}, $H_{PCI_{SC}}$.

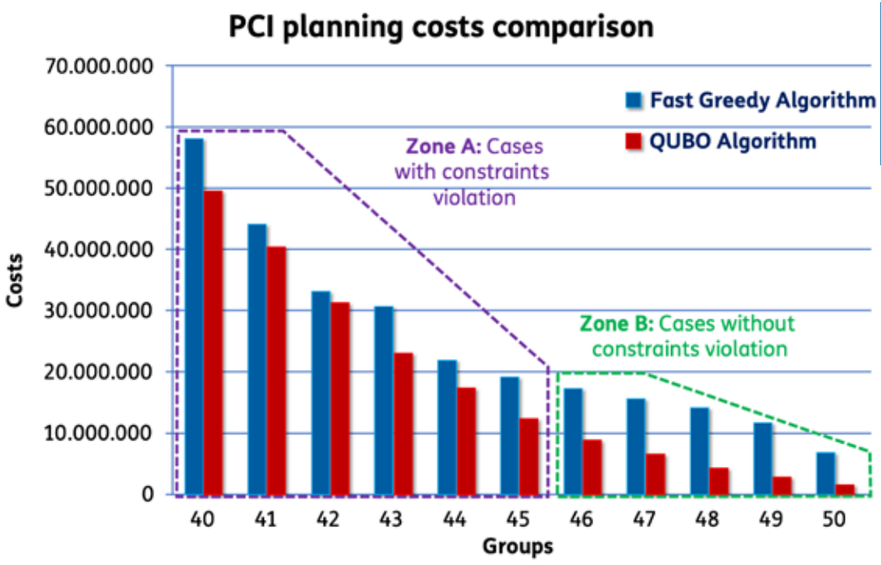
I due layer così costituiti, hanno singolarmente una complessità computazione inferiore rispetto all'algoritmo "completo".
Con questo nuovo modello è stato possibile elaborare set di celle di dimensioni significative, ottenendo una qualità migliore rispetto alla procedura di pianificazione dei PCI basata su tecniche combinatorie tradizionali.
Una sintesi dei risultati ottenuti è riportata nella sezione sui risultati dell'elaborazione.

Segmentazione dei dati di input

Lo sviluppo dell'algoritmo non si è limitato alla sola modellizzazione QUBO, ma è stato necessario tarare i set di celle affinché le loro dimensioni fossero gestibili con le potenze di calcolo dei computer a disposizione (CPU, GPU, QPU).
Questo perché la complessità computazionale del problema cresce esponenzialmente all'aumentare del numero delle celle.
In funzione dei vincoli sulla potenza di calcolo, sono stati creati dei set di dimensioni massime di alcune centinaia di celle, sulla base delle loro adiacenze.

Risultati dell'elaborazione dell'algoritmo QUBO di pianificazione PCI

L'applicazione dell'algoritmo QUBO al contesto della pianificazione dei PCI delle reti 5G e LTE ha fornito risultati incoraggianti dal confronto con quelli garantiti dal Fast Greedy Algorithm già utilizzato nell'ambito del framework Open SON di TIM.
Dal punto di vista dei tempi di calcolo necessari per definire un piano dei PCI, non è facile effettuare un confronto "omogeneo" poiché i criteri di stop dei due algoritmi sono profondamente diversi. Ciononostante, da misure indicative e tenendo conto del processo complessivo emerge la chiara indicazione di una riduzione del tempo di un fattore 10x, con la prospettiva di ulteriori margini di miglioramento.
La riduzione dei tempi di calcolo è di particolare interesse nell'ottica dell'evoluzione della pianificazione



4
Grafico di confronto su insieme campione

dei PCI verso modalità "real-time" (ad esempio attraverso l'integrazione in sistemi di tipo RIC – Radio Intelligent Controller nell'ambito delle architetture definite dalla O-RAN Alliance).
Oltre ai tempi di calcolo, al fine di analizzare le potenzialità dell'algoritmo su piani di complessità variabile, sono state effettuate una serie di prove diminuendo il numero di gruppi disponibili (rispetto al massimo consentito dallo standard).

Al diminuire del numero di gruppi, infatti, aumenta la probabilità di violazione dei vincoli. In tal modo, è stato individuato il minimo numero di PCI necessario per definire un piano ottimale, cioè tale da soddisfare tutti i vincoli di importanza "prioritaria" (relativi al PCI completo) e sono stati ottenuti i costi complessivi del piano (sia "primari", sui PCI, sia "secondari", sul GROUP_ID, valutati mediante la [Eq 2]) quando questo viola uno o più vincoli di "riuso" sull'identificativo di cella completo.

A titolo di esempio, si presentano i risultati ottenuti nel caso di un insieme campione di 450 celle della rete TIM, riportati in Figura 4:

Come si può osservare, in tutti i casi si ha un miglior comportamento dell'algoritmo QUBO.

Nella zona A, quest'ultimo ottiene soluzioni con un numero di vincoli inferiore rispetto al precedente algoritmo Fast Greedy (che conferma comunque la propria efficacia), mentre nella zona B – cioè in assenza di vincoli violati – esso garantisce funzioni costo inferiori grazie al minor numero di Group_ID utilizzati (46 contro 50) e ad una più efficace gestione del vincolo di "riuso" relativo al GROUP_ID.

Verso un applicazione sistematica del Quantum Computing

Allo stato attuale, sono in corso sviluppi dell'algoritmo di pianificazione dei PCI mirati a far evolvere l'algoritmo QUBO in modo da gestire insiemi sempre maggiori di celle.

Con questo obiettivo è stata definita una pipeline di lavoro, articolata su tre fasi:

- una di clustering da realizzare attraverso una procedura di decomposizione dei set di celle in base alle loro adiacenze
- due fase successive che corrispondono ai due livelli di elaborazione presentati nella sezione descrittiva del modello

prevedendo anche il tipo di hardware (CPU, GPU, QPU) più appropriato per ogni fase, rispetto alla complessità computazione che dovranno gestire.

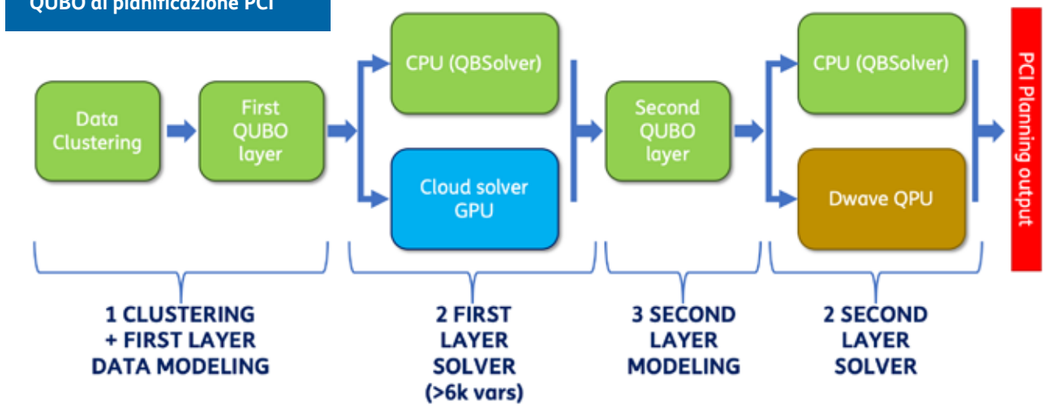
Nel secondo layer di elaborazione è prevista, in particolare, l'invocazione al quantum computer di D-Wave.

L'architettura della pipeline è schematizzata nella figura 5.

La pianificazione dei PCI della rete mobile è stato il primo caso d'uso affrontato in TIM con un approccio quantistico, utile per verificare l'applicabilità, le implicazioni e i vantaggi di questa nuova metodologia di programmazione.
È importante evidenziare come in ambito "Telco" sia possibile individuare ulteriori problemi adatti a tale tipologia di approccio e come altri use case siano stati identificati come potenziali candidati per la modellizzazione secondo la metodologia quantum.

Se gli studi di fattibilità daranno esito positivo – un possibile esempio sono gli algoritmi di Coverage and Capacity Optimization, anche ove basati sull'impiego dell'AI – questi casi d'uso verranno trasportati nel mondo della computazione quantistica.
Tra gli use case candidati si stanno considerando alcune applicazioni non real-time e near real-time proprie dei RAN Intelligent Controller (RIC) che, secondo il paradigma della open RAN, potrebbero integrarsi con le funzionalità real-time proprietarie dei vendor, con le quali opererebbero in sinergia a fini di ottimizzazione della gestione delle risorse radio.
Un fattore che favorirà l'utilizzo del quantum è legato al possibile impiego on-premises, per gestire la complessità delle applicazioni time-critical e tra queste, come già accennato, il suo utilizzo a livello di edge.
L'obiettivo di un'azienda come TIM, è – in conclusione – quello di estendere ulteriormente l'utilizzo del quantum computing alle applicazioni in ambito Telco, per gestire e risolvere problemi reali, trarne beneficio e migliorare la qualità dei servizi offerti alla clientela grazie a questa tecnologia innovativa.

5
Pipeline obiettivo di evoluzione dell'algoritmo QUBO di pianificazione PCI



Bibliografia

1. State of Quantum Computing in 2020 for Business Leaders. (s.d.). Tratto da Blog Aimultiple: https://blog.aimultiple.com/quantum-computing/
2. [2] KATWALA, A. (s.d.). Quantum computers will change the world (if they work). Tratto da Wired: https://www.wired.co.uk/article/quantum-computing-explained
3. [3] Quantum Computing Applications in 2020. (s.d.). Tratto da Blog AIMultiple: https://blog.aimultiple.com/quantum-computing-applications/
4. [4] Bertotto, P., Epifani, F., Ludovico, M., & Zarba, G. (s.d.). Open Self-Organizing Network: a continuous development for Radio Access Network performance optimization. Tratto da Telecom Italia: https://www.telecomitalia.com/tit/it/notiziariotecnico/edizioni-2019/n-3-2019/N3-Open-Self-Organizing-Network-continuous-development-for-Radio-Access-Network-performance-optimization.html
5. [5] Bertotto, P., & Zarba, G. (s.d.). Applicazioni "Open SON" nella rete TIM. Tratto da https://www.telecomitalia.com/content/tiportal/it/notiziariotecnico/edizioni-2018/n-1-2018/N6-DigiRAN-valore-automazione-accesso-radio/approfondimenti-2.html
6. [6] B., C. (s.d.). The Greedy Algorithms Class: Formalization, Synthesis and Generalization. Tratto da Université catholique de Louvain: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.34.5676&rep=rep1&type=pdf
7. [7] Glover, F. K. (2019, May). A Tutorial on Formulatingand Using QUBO Models. Tratto da Leeds Faculty of Colorado: : http://leeds-faculty.colorado.edu/glover/511%20-%20QUBO%20Tutorial%20-%20updated%20version%20-%20May%204,%202019.pdf
8. [8] Asproni, et al., Accuracy and minor embedding in subqubo decomposition with fully connected large problems: a case study about the number partitioning problem, arXiv:1907.01892v2

Andrea Boella *andrea.boella@telecomitalia.it*

Ingegnere elettronico, è in Telecom Italia dal 1997, dove è entrato nella ex direzione territoriale del Piemonte e Valle d'Aosta iniziando con la pianificazione della rete di trasporto regionale, per passare in qualità di responsabile prima al supporto specialistico trasmissioni e successivamente al secondo livello di assistenza tecnica per la clientela enterprise. Dal 2003 al 2005, come espatriato, è stato responsabile del settore esercizio in Telecom Italia France e al ritorno in Italia entra a far parte del gruppo ex Global Network coordinando, come project manager, attività di supporto verso le partecipate estere inizialmente su tematiche di rete fissa e a partire dal 2008 sulla core network di rete mobile. Da fine 2018 lavora nel gruppo di Service Innovation con attività di ricerca su servizi per l'IOT, blockchain, quantum computing e quantum communication ■

Michele Ludovico *michele.ludovico@telecomitalia.it*

Ingegnere elettronico, ha iniziato ad operare nel gruppo TIM progettando sistemi a micro-onde per comunicazioni via satellite. Dal 2001 si occupa di strumenti e metodologie di progettazione ed ottimizzazione dell'accesso radio, che TIM sviluppa "in house" a supporto dell'evoluzione della rete mobile. Dal 2014 è responsabile della funzione di TIM che assicura l'ingegneria e lo sviluppo delle soluzioni di automazione per la rete di accesso radio, secondo il paradigma "Self Organizing Network". Ha svolto, inoltre, attività di formazione e consulenza in Italia ed all'estero ed è co-inventore di diversi brevetti nel campo della progettazione wireless e della gestione delle risorse radio ■

Giuseppe Minerva *giuseppe.minerva@telecomitalia.it*

Laurea in Ingegneria Elettronica nel 1991 e assunzione nel gruppo TIM nel marzo del 1992. Servizio prestato nei settori di Innovazione e di Ingegneria di Rete. Attività nel campo della definizione di algoritmi e metodologie di pianificazione e ottimizzazione dell'accesso radio fin dai tempi della nascita della rete GSM e dell'ideazione dell'algoritmo di assegnazione frequenziale usato da TIM. Attività di formazione interna e di brevettazione nel campo della progettazione radio. Attualmente coinvolto nella transizione delle metodologie di pianificazione e ottimizzazione della rete TIM verso i paradigmi SON e di Full-Automation ■

Mauro Alberto Rossotto *mauro.rossotto@telecomitalia.it*

Laureato all'Università di Torino in Informatica con specializzazione Intelligenza Artificiale. Entrato in Telecom Italia nel 1995 ha partecipato a diversi progetti realizzativi legati a Data Mining Lab, analisi dati a scopo Antifrode e Marketing, Push-to-talk, Smart Inclusion. Dal 2012, in Innovation, ha seguito in qualità di responsabile di struttura le attività legate allo sviluppo di servizi innovativi su device connessi, a partire da TIM Vision e TIM Cloud su Smart TV e Game Console.
Nel 2014 in Strategy & Innovation sono iniziate le prime attività sul mondo IoT, ed in particolare sui verticali Smart Home, Wellness, Smart City, Smart Retail, Energy e Industry affrontando tutti gli aspetti tecnologici del servizio. Oggi in Service Innovation è responsabile del Program "Internet of Everything" dove vengono seguite tematiche relative a Droni, Robotics, IoT, Blockchain e Quantum Technologies con attività di scouting, prototipazione, sperimentazione e pipeline verso ingegneria ■

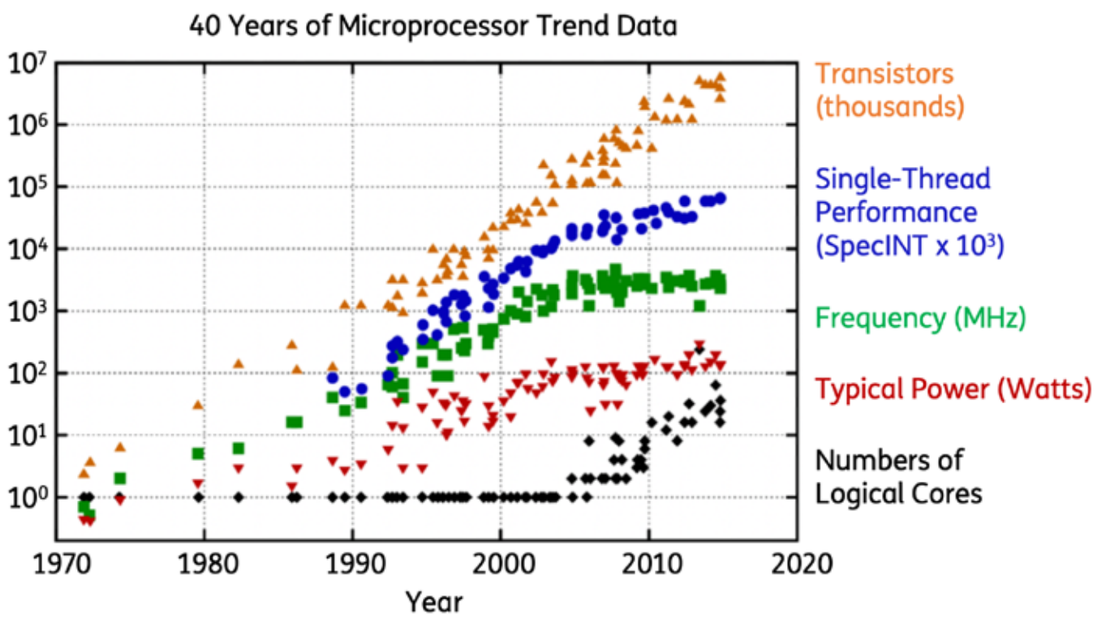
IL RUOLO DELL'ACCELERAZIONE HARDWARE NELLE RETI DI TELECOMUNICAZIONI DI NUOVA GENERAZIONE

Luciano Lavagno, Roberto Quasso , Salvatore Scarpina

Per raggiungere gli sfidanti obiettivi posti, il 5G ha bisogno di appoggiarsi ad una piattaforma di rete agile ed intelligente, in grado di offrire le caratteristiche di flessibilità, autonomia e performance richieste. Le tecnologie cloud diventano di grande interesse per gli operatori Telco in quanto possono fornire queste capability essenziali: in particolare, l'utilizzo degli acceleratori hardware rappresenta un punto di svolta abilitante per nuovi scenari ad elevato carico computazionale.

Con “accelerazione hardware” si intende l’utilizzo di dispositivi hardware dedicati per eseguire alcune funzioni in maniera più veloce e/o più efficiente (ad esempio consumando meno energia) rispetto all’utilizzo di un processore (CPU – Central Processing Unit) tradizionale “general purpose”.
L’accelerazione hardware ha una lunga storia. Nei primi anni ’80 i sistemi x86 venivano equipaggiati con FPU (Floating Point Unit), coprocessori matematici che incrementavano le capacità computazionali in virgola mobile delle CPU; successivamente questi componenti sono stati inclusi direttamente nelle CPU. Negli stessi anni erano già in circolazione chip dedicati alle funzioni video, per pilotare schermi sempre più performanti (CGA, EGA, MDA fino al VGA), o alle funzioni audio, come i DSP (Digital Signal Processor) per l’audio stereo. Spesso si trattava di schede voluminose da inserire all’interno dei computer dell’epoca e in generale erano dedicate ad applicazioni professionali. Oggi l’audio multicanale, elaborato mediante DSP dedicati, è dato per scontato e non è raro che un computer sia equipaggiato con una scheda GPU (Graphics Processing Unit) per il rendering delle immagini 3D su schermi full-HD.
Gli acceleratori hardware esistono perché la CPU, per sua natura, non può svolgere qualsiasi funzione in maniera efficiente; infatti la sua architettura è così flessibile da permettere l’implementazione di qualsiasi funzione software, ma questa flessibilità necessita di un livello di complessità cresciuto al punto da aver raggiunto il limite tecnologico di alcune soluzioni. La miniaturizzazione è arrivata a livelli estremi (si parla ormai di transistor grandi 2 nm), ma il problema principale dei processori è rappresentato dalla dissipazione di calore del packaging, che impedisce di fatto l’aumento della frequenza (“clock”) di funzionamento: l’energia dissipata dipende dal quadrato della frequenza, oltre certi valori non è più possibile estrarre il calore dal chip e si arriva alla sua fusione (problema noto a chi pratica l’overclocking). La legge di Moore (“il numero di transistor in un circuito integrato raddoppia ogni 2 anni”) non è più valida da tempo (fig. 1), in particolare perché il costo di fabbricazione per transistor aumenta, invece di diminuire come in passato.

1
40 years of microprocessor trend data (original data up to the year 2010 collected by M. Horowitz, F. Labonte, O. Shacham, L. Hammond and C. Batten. New plot and data collected for 2010-2015 by K. Rupp)



L’aumento delle prestazioni, che prima era garantito dall’aumento della velocità e della densità dei transistor, adesso viene perseguito implementando nuove soluzioni, come la tecnologia multi-core, e architetture sempre più complesse (instruction pipelining, vectorization, caching, branch prediction, speculative execution, etc.)

Tipi di acceleratori: GPU, ASIC e FPGA

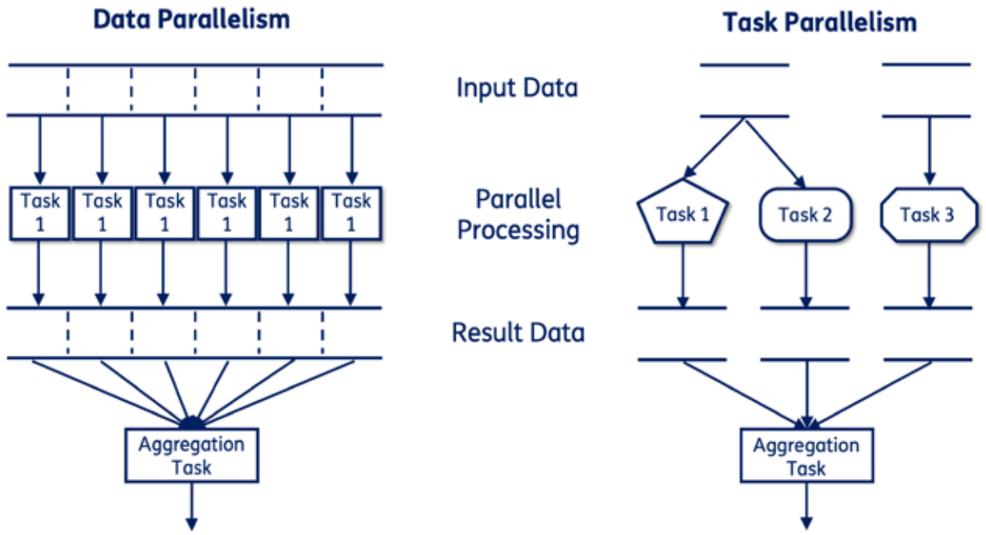
L’acceleratore oggi più diffuso è sicuramente la GPU, nata a metà degli anni ’90 per elaborare le informazioni video, cioè le caratteristiche di ogni singolo pixel che compone un’immagine, con risoluzione, numero di colori e frequenza dei fotogrammi sempre maggiori: il rendering realtime di immagini 3D ad alta risoluzione ha trovato nei videogiochi e nel CAD (Computer Aided Design) due settori trainanti. Il punto di forza delle GPU è rappresentato dall’enorme quantità di elementi computazionali che possono eseguire le stesse operazioni “elementari” (il concetto di elementare va contestualizzato al tipo di applicazione) contemporaneamente su un numero enorme di dati (“SIMD” - single instruction, multiple data), seguendo l’approccio “data parallelism”: ad esempio il calcolo delle caratteristiche di tutti i poligoni elementari in cui viene suddivisa una immagine.

Al contrario, le CPU lavorano utilizzando più unità di elaborazione in parallelo (i “core”) per applicare istruzioni diverse su dati diversi (“MIMD” - multiple instructions, multiple data): un moderno processore può eseguire programmi come i “socket server” che gestiscono contemporaneamente sessioni con più client.

Verso la metà degli anni 2000 nascono le GPGPU (General Purpose GPU) quando le GPU vengono impiegate per accelerare operazioni che non hanno nulla a che fare con il rendering video, sfruttando l’estremo parallelismo dell’architettura hardware applicato ad operazioni computazionalmente onerose, tipiche ad esempio dell’algebra delle matrici.
Oggi non è possibile pensare ad applicazioni di intelligenza artificiale senza l’impiego delle GPU per l’addestramento dei modelli di machine learning (Box approfondimento: Acceleratori hardware e AI) basato su enormi quantità di dati. Ma ci sono anche altri settori che si basano sull’utilizzo delle GPU, ad esempio i database, la biologia (studio dei P systems), la medicina (ricostruzione delle immagini delle scansioni).

Allo stesso tempo, questi algoritmi facilmente parallelizzabili sono spesso gli unici che possono sfruttare in modo efficiente l’elevatissimo parallelismo delle GPU, e quindi gli unici che oggi possono “scalare” traendo ancora vantaggio come un tempo dall’aumento del numero di transistor per chip.

Tutto questo è stato reso possibile dallo sviluppo di linguaggi di programmazione (ad esempio OpenCL) e ambienti di sviluppo come CUDA (Compute Unified Device Architecture) che consentono di sfruttare il parallelismo insito della GPU: OpenCL consente al programmatore di continuare a scrivere codice in maniera sequenziale ma andando ad individuare alcune porzioni (i kernel), che poi vengono automaticamente replicate su molte, anche migliaia, unità di elaborazione contemporaneamente.
Ad esempio, le singole iterazioni di un ciclo (loop), se indipendenti l’una dall’altra (come nel caso ad esempio del calcolo matriciale), possono essere gestite come kernel ed essere eseguite in parallelo (unrolling).



2
Data Parallelism vs Task Parallelism [rif. Parallel Programming Concepts GPU Computing with OpenCL – Frank Feinbube]

Questo tipo di approccio è fondamentalmente diverso, ad esempio, dalle tecniche di multithreading implementate nelle CPU (approccio "task parallelism", fig. 2) dove si cerca l'esecuzione concorrente di più set di istruzioni sequenziali all'interno della stessa unità di elaborazione: essendo il thread composto da più fasi (fetch-decode-execute), un singolo processore pur eseguendo effettivamente una sola operazione per volta può ottimizzare i tempi di latenza di operazioni esterne (ad esempio l'accesso alla memoria) per dare l'impressione di eseguire operazioni diverse in parallelo.
Un'ulteriore tecnica introdotta nel corso degli anni dai produttori di CPU è la vectorization che sfrutta l'applicazione di operazioni su vettori (array) invece che su singoli elementi: i moderni compilatori per CPU provano automaticamente ad effettuare unrolling dei cicli in tal senso.
Questa tecnica, come il multithreading, impatta l'organizzazione ed il funzionamento interno delle CPU e punta ad ottenere migliori performance senza richiedere modifiche alla programmazione, come invece occorre fare per sfruttare le GPGPU con OpenCL e CUDA.

Gli ASIC (Application Specific Integrated Circuit) sono un altro tipo di acceleratore, specializzati per una specifica funzionalità fin dalla progettazione del circuito integrato su silicio e sono utilizzati per accelerare specifiche funzioni, scaricando quindi la CPU dal relativo carico computazionale: ad esempio i DSP utilizzati per l'elaborazione del segnale audio, i NIC (Network Interface Card) utilizzati per l'accelerazione dei protocolli di rete, i TPU (Tensor Processor Unit) utilizzati per l'accelerazione della libreria TensorFlow dedicata al machine learning.
Sono acceleratori molti efficienti e danno le massime prestazioni, ma il loro costo di sviluppo ed i tempi realizzazione sono talmente elevati da renderli convenienti solo per la produzione di milioni di esemplari.

ACCELERAZIONE HARDWARE E INTELLIGENZA ARTIFICIALE

gianluca.francini@telecomitalia.it

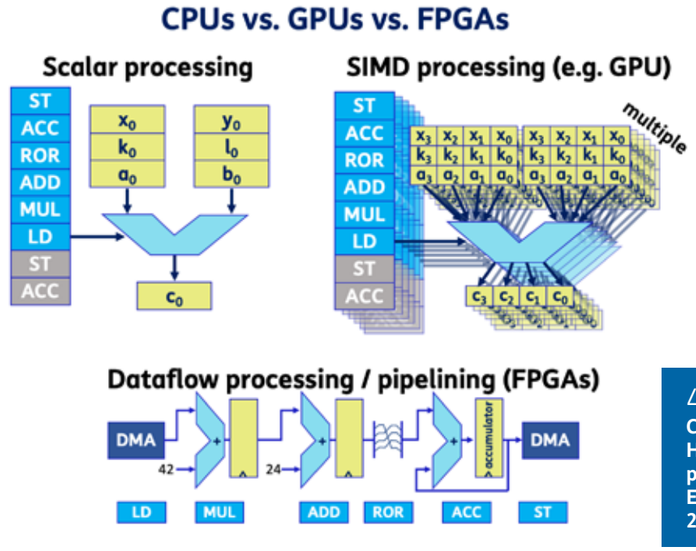
Siamo in un periodo storico in cui si parla molto di Intelligenza Artificiale e questo è dovuto al fatto che i computer hanno iniziato a svolgere dei compiti che fino a poco tempo fa sembravano essere affrontabili solo da esseri umani, come guidare un veicolo, tradurre da una lingua ad un'altra, interagire con le persone mediante il linguaggio naturale. Questo incremento nelle abilità delle macchine è stato piuttosto repentino ed ha alla base una specifica tecnologia: le reti neurali profonde, conosciute con il nome di Deep Learning.
Le reti neurali sono dei modelli matematici che nascono per mimare il comportamento del cervello umano e sono state introdotte a metà del secolo scorso. Sono passati quindi numerosi anni dalla loro ideazione, ma solo di recente il loro utilizzo ha consentito di realizzare le applicazioni sofisticate che abbiamo oggi a disposizione. Ci sono diversi motivi che giustificano questo lungo tempo di attesa. C'è stata un'evoluzione teorica delle reti neurali e un incremento dei dati a disposizione per addestrarle, ma c'è stato anche un fattore che si è rivelato cruciale: l'adozione della Graphics Processing Unit (GPU) come acceleratore hardware per l'addestramento delle reti neurali.
L'esplosione del Deep Learning si può infatti ricondurre alla vittoria della rete neurale AlexNet, progettata da Alex Krizhevsky, studente dell'Università di Toronto del gruppo del prof. Geoffrey Hinton, che nel 2012 vinse la competizione internazionale ImageNet Large Scale Visual Recognition Challenge, diminuendo drasticamente l'errore commesso dal sistema automatico nel classificare gli oggetti presenti in un esteso dataset di immagini. Krizhevsky fu in grado di sviluppare una rete neurale sofisticata grazie all'adozione di due GPU per addestrare il modello e il risultato ottenuto, che aveva un errore di più del 10% inferiore rispetto al secondo classificato, è considerato da molti come il punto di svolta del Deep Learning, in cui la tecnologia ha catturato non solo l'attenzione della comunità scientifica ma anche quella dell'industria.

Le reti neurali profonde sono composte da un numero molto elevato di neuroni, unità che eseguono somme pesate degli input di altri neuroni a cui sono collegate. Si tratta quindi di eseguire numerose operazioni elementari come somme e moltiplicazioni, oltre a passare i risultati attraverso semplici funzioni non lineari. Le reti neurali più complesse hanno un numero di pesi nell'ordine delle centinaia di miliardi di parametri: questo fa capire come sia fondamentale utilizzare delle unità elaborative in grado di accelerare i calcoli. Le elaborazioni eseguite dalle reti neurali sono per fortuna altamente parallelizzabili: l'esecuzione dei calcoli di un neurone può essere eseguita in parallelo con quella di moltissimi altri neuroni e per questo motivo le GPU sono molto adatte ad accelerare le applicazioni di Intelligenza Artificiale.

Un tipo di acceleratore che recentemente sta conquistando un posto di rilievo all'interno dei datacenter è la FPGA (Field Programmable Gate Array).
Questi dispositivi sono caratterizzati dalla presenza di un elevato numero elevato di blocchi di logica programmabile (fino a quasi 4 milioni), una fittissima rete di interconnessioni a loro volta programmabili ed una crescente varietà di "accessori" quali memorie RAM multiporta, ALU, PLL e porte seriali multigigabit. Di fatto sono dispositivi su cui è possibile istanziare e collegare funzionalità con livello di granularità variabile, da quello macro (per esempio una Fast Fourier Transform) fino a scendere al livello di singole porte logiche elementari, esattamente come si fa nel progetto di un dispositivo ASIC.
Ma l'enorme vantaggio rispetto agli ASIC è che le funzionalità possono essere modificate nel corso del tempo "sovrascrivendo" tutte od in parte quelle precedenti.
Trattasi, quindi, di dispositivi molto efficienti dal punto di vista energetico (rispetto alle CPU e GPU, ma meno degli ASIC dedicati, fig. 3), in quanto sostituiscono il ciclo fetch/decode/execute tipico dei processori con una struttura di controllo direttamente realizzata in hardware, ed adattano sia le dimensioni degli operatori aritmetici (somma, moltiplicazione, etc.) sia l'architettura all'applicazione.
Finora sono stati utilizzati soprattutto nell'ambito della prototipazione hardware, ma anche nella realizzazione di dispositivi in serie quando i numeri non giustificano la realizzazione di un ASIC dedicato (questo accade per esempio negli apparati TLC dedicati). Come per le GPU, anche per le FPGA il loro avvicinamento al mondo della programmazione software è stato possibile nel momento in cui gli sviluppatori hanno avuto a disposizione un ambiente di sviluppo e un linguaggio di programmazione (ancora OpenCL, che di fatto è un'estensione di C/C++) che permettesse di tradurre i blocchi di codice in "programmazione" hardware. Per la sua granularità, la FPGA è adatta a lavorare con "pipeline parallelism" (fig. 4), nel quale i task sono organizzati secondo una logica "consumer-producer" (i dati elaborati da un blocco diventano input per il blocco successivo). Gli ambiti di applicazione delle FPGA sono in costante evoluzione, ed annoverano già settori quali quello finanziario, ricerca sulla genomica, accelerazione dei motori di ricerca (ad esempio Bing), accelerazione di stack radio (ad es. livello fisico 5G).

Type	Device	GFLOPS (SP)	Cost (€)	Power (W)	GFLOPS/€	GFLOPS/W
Multi-core	Intel E5-2630v3 8x2.4 GHz	600	700	85	0.85	7.05
	Intel E5-2630v3 10x2.3 GHz	740	1250	105	0.59	7.04
Many-core	Xeon Phi, Knights corner, 16 GB	2416	3500	270	0.69	8.94
	Xeon Phi, Knights landing, 16 GB	7000	3500	300	2.00	23.3
GPU	Nvidia GeForce Titan X	7000	1000	250	7.00	28
	Nvidia Tesla K80	8740	7000	300	1.24	29.13
	Nvidia Tegra X1	512	450	7	19.42	73
	Radeon firepro S9150	5070	3500	235	1.44	21.5
	ARM Mali T880 MP16	374	?	5?	?	74
FPGA	Altera Arria 10	1500	3000	30	1.00	50
	Altera Stratix 10	10000	2000?	125?	5.00?	80
	Xilinx Ultrascale+	4600	2000	40?	2.30	115

3
Confronto performance e consumo di energia tra diversi acceleratori hardware [rif. Bringing Hardware Acceleration closer to the programmer - Iakovos Mavroidis (EcoScale-ExaNest joint workshop Rome, 17 February 2017)]



4
Confronto modelli operativi [rif. Bringing Hardware Acceleration closer to the programmer - Iakovos Mavroidis (EcoScale-ExaNest joint workshop Rome, 17 February 2017)]

L'accelerazione nei datacenter TLC

L'introduzione degli acceleratori hardware rappresenta la naturale evoluzione dei data center che, per rispondere alla domanda crescente di prestazioni, non possono più fare affidamento sull'aumento delle prestazioni delle singole CPU e, d'altro canto, non possono ricorrere alla sola scalabilità orizzontale per fronteggiare i volumi sempre crescenti di dati da trattare.

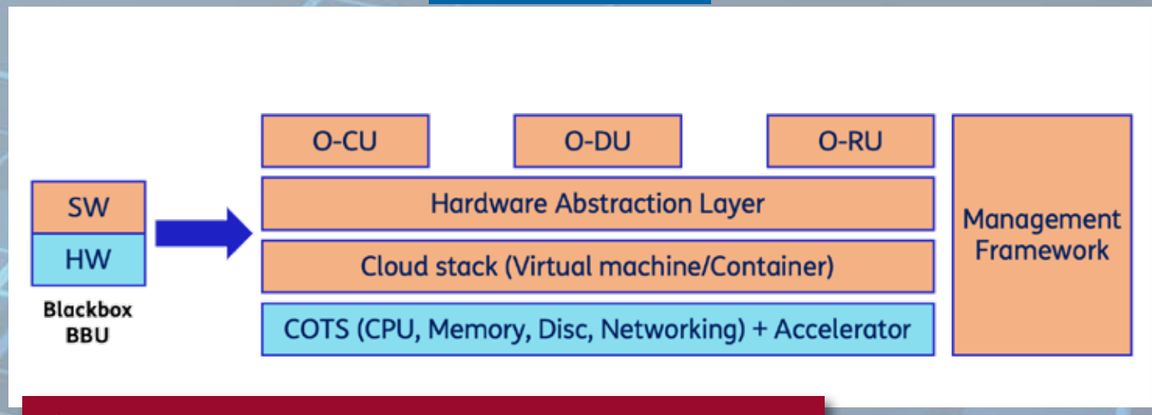
Gli acceleratori hardware offrono non solo maggiori velocità ma anche minori consumi di energia e di spazio per unità di informazione trattata e, quindi, consentono di contenere i costi infrastrutturali di spazio, alimentazione e raffreddamento.
L'evoluzione attuale dei data center ha spinto la definizione di architetture comuni fino alla standardizzazione degli apparati hardware (COTS – Commercial Off the Shelf), con il risultato di una compressione dei costi da un punto di vista sia dell'infrastruttura IT fisica che della sua gestione.
Le reti di telecomunicazioni hanno già iniziato a prendere in "prestito" alcuni di questi concetti di efficientamento, ad esempio definendo le SDN (Software Defined Network), in particolare con la separazione fisica tra control-plane e data-plane.

Da un punto di vista computazionale, nel dominio delle reti di telecomunicazioni le CPU sono lo strumento adatto per la gestione del control-plane e dei livelli più alti dello stack protocollare di trasporto; al contrario soffrono nella elaborazione dei carichi di lavoro bit-intensive e packet-based che sono tipici dei livelli OSI più bassi (dal livello 4 in giù), per i quali si fatica a raggiun-

L'ACCELERAZIONE HARDWARE NELLA CLOUD RAN

marco.caretti@telecomitalia.it

Il concetto di "Open RAN" si sviluppa intorno all'idea di ricercare nuove soluzioni ed approcci che permettano di realizzare un'architettura più flessibile, più rapidamente dispiegabile e potenzialmente più sostenibile economicamente, grazie anche alla definizione di sistemi multivendor; con le richieste di nuovi investimenti associati ai nuovi sistemi 5G, questa ricerca ha assunto ancora più importanza ed ha portato a diverse attività in ambito di consorzi internazionali, uno tra i quali è l'O-RAN Alliance (https://www.o-ran.org/).
L'approccio cloud, esteso anche alle funzionalità radio mediante l'impiego di opportune soluzioni di Iaas/PaaS, è visto in questo contesto come un importante abilitatore per il raggiungimento dei diversi obiettivi. Per far ciò, uno strumento fondamentale è la separazione tra hardware e software per le varie componenti che compongono l'architettura O-RAN (figura A), che consenta il dispiegamento delle componenti su "commodity server" (o hardware COTS) integrati con acceleratori hardware programmabili quanto necessario.



A
Separazione tra hardware e software nella definizione architetturale di orchestrazione e cloudificazione della rete 5G [rif. Cloud Architecture and Deployment Scenarios for O-RAN Virtualized RAN - O-RAN-WG6.CAD-V01.00.00]

La scelta di utilizzare acceleratori hardware (e.g. GPU, FPGA, ASIC) dipende sostanzialmente dalla necessità di rispettare i requisiti di performance minimi richiesti per l'esercizio ottimale delle funzionalità, come ad esempio la latenza massima accettabile tra O-DU (Distributed Unit) e O-RU (Radio Unit), ma anche da considerazioni di natura tecnica: ad esempio l'uso di acceleratori hardware può portare ad una riduzione significativa del consumo di energia e dell'emissione di calore, quindi ad una riduzione dei costi di esercizio del datacenter o, addirittura, abilitare il dispiegamento della funzionalità in ambienti con restrizioni infrastrutturali (ad esempio nodi edge o siti radio).
Low-Density Parity-Check (LDPC) Forward Error Correction (FEC) nella O-DU, Wireless Ciphering nella O-CU (Central Unit) e algoritmi di Artificial Intelligence nel RIC (RAN Intelligent Controller) sono alcune delle funzioni "computing-intensive" che potrebbero beneficiare dell'utilizzo degli acceleratori hardware.

Tuttavia l'uso degli acceleratori hardware deve rispettare comunque il principio di separazione del software dall'hardware: per questo in O-RAN si sta procedendo alla definizione di un "Acceleration Abstraction Layer" (AAL) che consenta di definire un framework di gestione degli acceleratori stessi, indipendente dalla tipologia e dalle caratteristiche specifiche del singolo acceleratore, separando l'hardware dalla funzionalità che viene accelerata. Questo framework dovrà evitare l'utilizzo di acceleratori specializzati per singole funzionalità e piuttosto favorire la "neutralità" degli acceleratori (ovvero fornitori diversi potranno fornire il loro dispositivo per accelerare le stesse funzionalità) e la loro "condivisione" (uso di uno stesso acceleratore per accelerare funzionalità diverse). Inoltre, esso dovrà garantire che questi acceleratori possano essere integrati all'interno dei processi e delle piattaforme di gestione e orchestrazione della piattaforma cloud utilizzata, in maniera del tutto simile alle altre risorse hardware a disposizione nel cloud (concetto di commodity).

gere livelli di throughput adeguati per i moderni servizi; per questo motivo fino ad oggi si è ricorso all'utilizzo di apparati dedicati all'implementazione delle funzioni di rete (PNF – Physical Network Function). L'introduzione degli acceleratori hardware nei moderni data center cambia radicalmente questo scenario, perché rende questi ambienti adeguati alla gestione dei carichi di lavoro tipici del data-plane: ecco, quindi, che l'idea di "virtualizzare" le reti di telecomunicazioni, compresi quei segmenti come la RAN (Radio Access Network) finora toccati marginalmente dalla centralizzazione all'interno di data center, assume una nuova prospettiva di fattibilità.

Tuttavia, poiché gli acceleratori hardware rappresentano una novità all'interno dei data center, è necessario pensare a nuove interfacce e nuove modalità per la gestione di questi dispositivi; infatti, se si vuole trarre vantaggio dalla loro riprogrammabilità e capacità di evolvere, occorre adeguare gli strumenti di gestione esistenti. Da qui il lavoro di specifica svolto da alcuni enti e associazioni quali ad esempio O-RAN nel dominio dell'accesso radio (Approfondimento: L'accelerazione hardware nella Cloud RAN).

In definitiva, nella sfida per la realizzazione di un'infrastruttura in grado di sostenere i tre pillar del 5G – enhanced mobile broadband, massive m2m communications, ultra-reliable and low latency communications - e che al contempo sia economicamente sostenibile, scalabile nelle dimensioni e flessibile in ottica evolutiva, la virtualizzazione delle componenti infrastrutturali potrà giocare un ruolo importante. In questo processo gli acceleratori hardware (GPU e FPGA in particolare) saranno fondamentali per il raggiungimento delle performance richieste, con conseguenze ad oggi ancora difficilmente identificabili a partire dalle eventuali nuove architetture hardware fino agli strumenti di gestione. Inoltre è evidente che gli sviluppatori delle funzioni di rete virtualizzate dovranno evolvere verso una sempre maggiore consapevolezza dell'architettura hardware e software delle macchine che le devono eseguire, se vorranno sfruttare al meglio le loro potenzialità ■

Luciano Lavagno *luciano.lavagno@polito.it*

Luciano Lavagno si è laureato in Ingegneria Elettronica presso il Politecnico di Torino nel 1983 e ha ricevuto il dottorato dall'University of California a Berkeley nel 1992. Dal 1984 al 1988 ha lavorato presso lo CSELT di Torino, dove ha partecipato allo sviluppo di un sistema completo per la sintesi architetturale di circuiti digitali. Dal 1993 ad ora è stato professore al Politecnico di Torino. Ha collaborato con Cadence Design Systems, partecipando alla creazione dello strumento di sintesi ad alto livello CtoSilicon, e attualmente collabora con Xilinx alla realizzazione dello strumento di compilazione per FPGA Vitis. I suoi interessi di ricerca includono la sintesi ad alto livello da modelli in C++ ed OpenCL verso RTL ed i sensori capacitivi a lungo raggio per localizzazione in ambienti chiusi ■

Roberto Quasso *roberto.quasso@telecomitalia.it*

Laureato in Ingegneria Elettronica presso il Politecnico di Torino, è in TIM dal 1992 in ambito innovazione e ricerca. Ha lavorato alla progettazione HW e SW di strumenti, prototipi e dimostratori per l'innovazione della rete fissa e mobile, utilizzando la tecnologia FPGA e seguendo la sua evoluzione in termini dimensionali e funzionali. Dal 2017 si occupa della progettazione di funzioni di livello fisico radio 5G su sistemi SDR basati su FPGA, utilizzate per l'accelerazione hardware delle misure che TIM effettua sulle antenne attive 5G sia in camera anecoica sia in campo. Si interessa inoltre dell'impiego delle tecniche di accelerazione hardware alla virtualizzazione delle funzioni di rete di accesso mobile ■

Salvatore Scarpina *salvatore.scarpina@telecomitalia.it*

Laureato in ingegneria elettronica presso Polito, lavora in TIM dal 2005 nell'ambito dell'innovazione e della ricerca. Inizialmente focalizzato sullo sviluppo di tecnologie sui terminali mobili, successivamente si è orientato allo sviluppo di servizi web ottimizzati per diverse tipologie di terminali (mobili e non); in questo contesto ha sviluppato competenze in ambito di project management e di ambienti cloud riapplicate successivamente in ambito Accesso Mobile (nel quale lavora d 2017), insieme a nuove skill in ambito Management di reti mobili, Machine Learning applicato al SON, accelerazione hardware, Edge Computing. Fin dal 2007 partecipa ad attività di standardizzazione presso diversi enti, tra cui OMA DM (chair), ETSI MEC e O-RAN ■

DENTRO LO SMARTPHONE: BANDA BASE E PROTOCOLLI RADIO

Bruno Melis, Damiano Rapone, Giovanni Romano

Questo articolo descrive i blocchi funzionali di uno smartphone: la catena di ricetrasmissione, dove sono esplicitate le principali funzionalità introdotte da 5G NR, ed i protocolli radio necessari per consentire la comunicazione tra "telefonino" e rete.

Introduzione

Il nostro smartphone è ormai parte integrante del nostro modo di vivere tanto che si dà ormai per scontata tutta la tecnologia che si cela dietro di esso: ma cosa permette alle nostre app di funzionare in Italia e nel resto del globo?
Tutto questo è possibile grazie allo sforzo congiunto dell'industria che ha riconosciuto l'importanza di creare un unico standard per la telefonia cellulare e di investire fortemente nella ricerca di nuove soluzioni per ottimizzare l'uso delle risorse radio.Uno smartphone è un insieme di molte tecnologie: lo schermo, le fotocamere, il sistema operativo e le applicazioni che forniscono i servizi sono solo alcuni esempi. Tutto questo però non funzionerebbe senza la componente di comunicazione, ovvero la capacità di operare in una rete cellulare. A tale scopo, questo articolo intende spiegare il funzionamento della banda base di uno smartphone ed i protocolli di comunicazione radio. Nell'articolo "Dentro lo smartphone - SoC e testing" [1] si affrontano invece gli aspetti più implementativi e propedeutici alla commercializzazione.

La standardizzazione tecnica

Il 3GPP è l'ente di riferimento per le tecnologie radiomobili.

Tutti i telefoni cellulari si basano su quanto definito da questo ente. In particolare, il gruppo RAN1 specifica il livello fisico, ovvero le tecniche di modulazione, i codici per la correzione degli errori introdotti dal canale di trasmissione, il numero di risorse utilizzabili (nel tempo e nella frequenza) per la trasmissione dei dati e dei segnali di controllo.

La gestione delle risorse radio (Radio Resource Management, RRM), le misure e le procedure per la mobilità nonché i protocolli che regolano il funzionamento del terminale sono invece definiti dal RAN2.

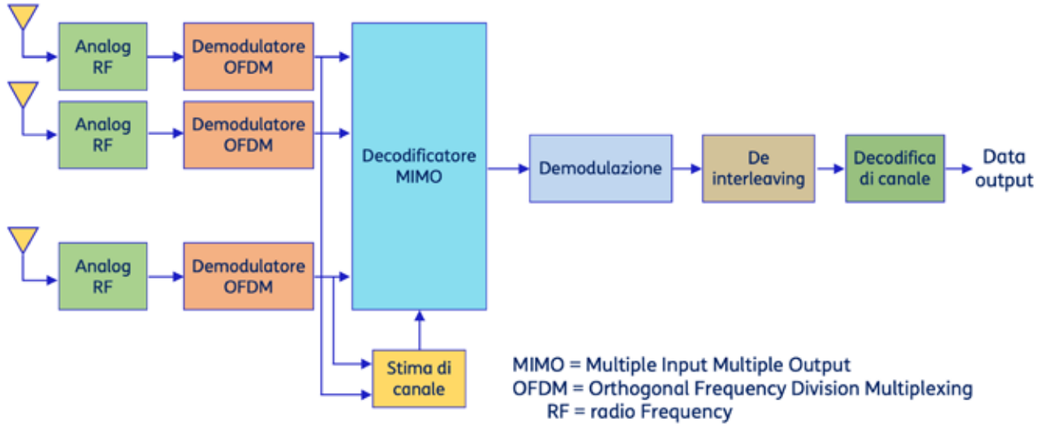
Nel seguito sono descritte le principali caratteristiche di quanto specificato dal 3GPP RAN1 e RAN2.

La banda base: la catena ricetrasmissiva

Il livello fisico di NR è descritto nelle specifiche del 3GPP RAN1, in particolare in [2][3]. La tecnica OFDM è stata scelta dal 3GPP come forma d'onda per la trasmissione del segnale 5G NR.
La modulazione OFDM suddivide il flusso di dati ad elevato bit rate in N flussi ciascuno dei quali modula una delle N sottoportanti a banda stretta del segnale OFDM.

Il risultato è che ciascuna sottoportante OFDM vede un canale non selettivo in frequenza, ovvero di semplice equalizzazione. NR introduce alcune migliorie rispetto alla tecnica OFDM adottata per LTE:

- Numerologia flessibile con spaziatura in frequenza Δf delle sottoportanti OFDM configurabile da 15 kHz (valore fisso utilizzato da LTE) fino a Δf=120 kHz per i canali di traffico. L'utilizzo di una spaziatura in frequenza maggiore è utile per frequenze portanti elevate, ad esempio ad onde millimetriche sopra i 20 GHz, per rendere il sistema più robusto rispetto alle caratteristiche del canale di propagazione e per ridurre la durata del simbolo OFDM (TS) e quindi la latenza a livello fisico (vale infatti la relazione fondamentale TS=1/Δf). Inoltre, una spaziatura elevata permette di utilizzare canali con banda maggiore (fino a 400 MHz per frequenze sopra 24.25 GHz) utilizzando un numero N di sottoportanti gestibile in termini di complessità (il numero N di sottoportanti corrisponde alla dimensione dell'operazione di FFT che è alla base della generazione del segnale OFDM).
- Tecniche di filtraggio per contenere lo spettro del segnale OFDM. Il contenimento spettrale del segnale NR permette di raggiungere un utilizzo del canale radio allocato fino al 98%, rispetto a LTE che si ferma invece al 90%.



1
Schema a blocchi funzionale di un ricevitore NR per la parte di livello fisico

La tecnica MIMO svolge un ruolo fondamentale per NR.

Essa consiste nell'utilizzo di antenne multiple al trasmettitore ed al ricevitore al fine di aumentare sia la velocità di trasmissione sia la copertura radio del sistema. Il MIMO in NR è stato progettato nativamente per supportare il beamforming in trasmissione e ricezione.

Il beamforming consente di indirizzare il segnale radio verso direzioni preferenziali in cui sono localizzati gli utenti con il risultato di aumentare il raggio di copertura e l'efficienza spettrale.

Inoltre, il beamforming è fondamentale per l'applicazione della tecnica MU-MIMO, dove due o più utenti condividono le stesse risorse trasmissive ma i relativi segnali sono separati trasmettendoli in direzioni diverse mediante antenne attive.

Il 3GPP ha inoltre definito alcuni vincoli per garantire un livello di prestazioni elevate nella tratta downlink come l'utilizzo di 4 antenne riceventi nel terminale.

Al contrario non è previsto un vincolo sul numero di antenne trasmittenti nel terminale.

Un'ulteriore innovazione di NR è l'utilizzo dei codici di canale LDPC che grazie alla loro particolare struttura di decodifica parallela permettono di raggiungere velocità dell'ordine di molti Gbit/s.

Lo schema a blocchi funzionale del ricevitore NR per la parte di livello fisico è illustrato in figura 1. I segnali ricevuti dalle varie antenne sono prima convertiti in banda base e poi digitalizzati.

I segnali sono quindi soggetti alla demodulazione OFDM e poi sono inviati al MIMO decoder che elabora il segnale ricevuto e ricostruisce il flusso dati trasmesso.

L'utilizzo della tecnica OFDM congiuntamente alla tecnica MIMO è particolarmente interessante perché permette di separare nel ricevitore l'operazione di equalizzazione

del canale da quella di decodifica della trasmissione MIMO.
La separazione di queste due operazioni permette di semplificare la struttura del ricevitore con una conseguente riduzione della complessità. Le successive operazioni di demodulazione, de-interleaving e decodifica di canale servono a migliorare l'affidabilità dei segnali ricevuti correggendo gli errori introdotti dal canale di propagazione.

La pila protocollare radio 5G

Nel sistema 5G coesistono due tecnologie di accesso: l'evoluzione di LTE e la nuova interfaccia radio NR.

Entrambe le tecnologie costituiscono la Next Generation Radio Access Network (NG-RAN) attraverso la quale il terminale (UE) si connette alla rete tramite sia nodi di rete LTE (ng-eNB) che NR (gNB) [5] [6].

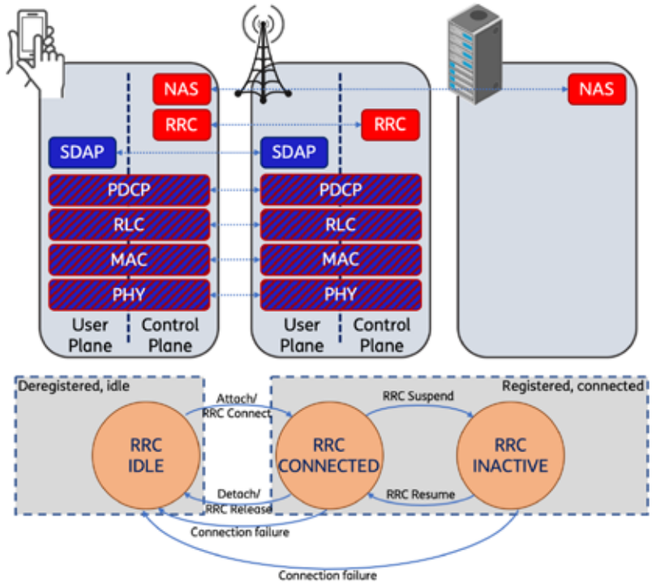
La comunicazione tra la rete di accesso NG-RAN e lo UE è gestita dalla pila protocollare di Access Stratum (AS), implementata sia lato nodo di rete sia lato terminale.

La pila protocollare del 5G è stata specificata dal 3GPP in Release 15 ed è illustrata in Figura 2.

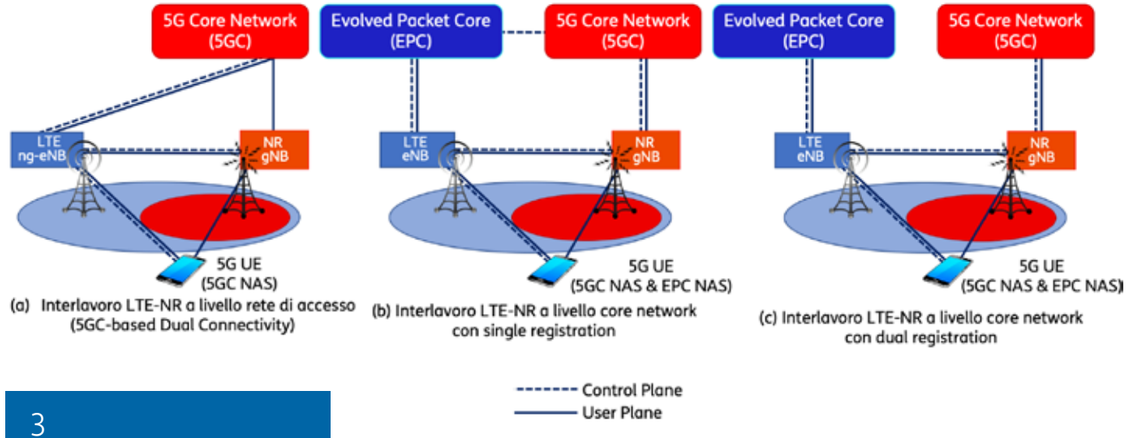
La comunicazione di AS interessa sia il Control Plane (CP, ossia la segnalazione di rete) che lo User Plane (UP, ossia i dati d'utente).

Le procedure di CP prevedono il protocollo di Radio Resource Control (RRC) che consente ad es. l'invio a tutti i terminali presenti nell'area di copertura di un sito delle informazioni (System Information) necessarie allo UE per stabilire una comunicazione.

Altre funzionalità tipiche del protocollo RRC sono la gestione della mobilità del terminale (procedure di cell reselection e di handover) e la configurazione delle misure ad essa associate, il controllo di connessione (procedure di creazione/modifica/rilascio di una connessione RRC, attivazione delle procedure di security, configurazione dei radio bearers, recupero della connessione conseguente la Radio Link Failure (RLF), ecc.), il trasferimento della segnalazione che si richiude esclusivamente tra lo UE e la core network (segnalazione di Non Access Stratum, NAS) ed infine la gestione delle feature supportate dallo UE (capability), dato che non tutti gli smartphone implementano tutte le funzionalità standardizzate dal 3GPP.



2
Pila protocollare del 5G (AS e NAS) e diagramma di transizione tra stati RRC del terminale 5G



3
Opzioni di connettività con la core network

I protocolli di Layer-2 – Packet Data Convergence Protocol (PDCP), Radio Link Control (RLC), Medium Access Control (MAC) – interagiscono quindi con la catena rice-trasmissiva descritta precedentemente.
Lo UP è gestito dalla stessa pila protocollare radio, con le seguenti differenze:

- Il protocollo RRC non è coinvolto, agendo esclusivamente sul CP
- Ai protocolli di Layer-2 prima menzionati si aggiunge, al di sopra del PDCP, il nuovo protocollo Service Data Adaptation Protocol (SDAP) che gestisce il nuovo modello di qualità del servizio della core network 5G (5GC).

La comunicazione NAS, invece, è esclusivamente di CP ed è gestita dal relativo protocollo implementato nello UE e nell'entità Access and mobility Management Function (AMF) della 5GC. Il protocollo NAS gestisce la segnalazione di procedure come network registration, authentication, security, paging, location update e session management.

Gli stati RRC e la gestione delle risorse radio

L'attività del terminale e la gestione delle risorse (RRM) è distinta sulla base degli stati RRC (Figura 2).
Uno smartphone 5G, infatti, può trovarsi in uno dei seguenti stati,

a cui sono associate funzionalità e procedure distinte [6]:

- RRC_IDLE – il terminale non riceve né trasmette dati. Lo UE è in uno stato 'dormiente' per ridurre il consumo di batteria, tranne per ricevere le System Information o per monitorare le richieste di connessione (paging). La mobilità si basa sulle procedure di cell (re)selection per identificare la cella migliore sulla quale accamparsi. La posizione del terminale è nota alla reta a livello di core network su base Tracking Area e lo UE trasmette solo per indicare il cambio di Tracking Area e quando esegue la procedura di Random Access per il passaggio allo stato RRC_CONNECTED.
- RRC_CONNECTED – il terminale riceve e trasmette dati. La rete può configurare funzionalità di Discontinuous Reception (DRX) per ridurre il consumo di batteria nei casi di traffico dati particolarmente variabile. La posizione del terminale è nota a livello cella e la mobilità è controllata dalla rete (handover) sulla base di misure fornite dal terminale.
- RRC_INACTIVE – nuovo stato introdotto per gestire la trasmissione frequente di piccole quantità di dati. Se per un certo tempo il terminale non scambia dati con la rete esso può sospendere la sua sessione senza però perdere la connessione verso la rete; il terminale passerà allo stato RRC_CONNECTED semplicemente riattivando la sua connessione. Il comportamento e la mobilità sono simili a RRC_IDLE ma la posizione del terminale è nota a livello di rete di accesso NG-RAN su base RAN Notification Area.

Opzioni di connettività verso la core network

Il 3GPP ha definito diverse configurazioni di rete 5G [7], che prevedono un interlavoro in rete di accesso – Dual Connectivity tra LTE e NR – e in core network (Figura 3).

Quest'ultimo si verifica nei casi di architetture 5G di tipo standalone in cui LTE ed NR operano in maniera indipendente l'una dall'altra.

Il terminale deve poter supportare sia il NAS della Evolved Packet Core (EPC) sia il NAS della 5GC e sono possibili i casi di single e dual registration [8].

Uno UE in single registration si registra ad una sola delle due core network (EPC o 5GC) e la gestione della mobilità intersistema richiede un'interfaccia dedicata tra le due core network.

Nel caso di dual registration, lo UE si registra contemporaneamente sia alla EPC che alla 5GC per cui non è necessaria l'interfaccia tra le due core network.

Tuttavia, per migliorare le prestazioni di mobilità, il terminale potrebbe dover implementare una doppia catena rice-trasmissiva.

Conclusioni

Il nostro "telefonino" funziona grazie alle funzionalità ed ai protocolli radio standardizzati dal 3GPP.

Grazie allo standard, il cliente può scegliere tra molteplici brand di smartphone ed è sicuro di fruire del servizio in qualsiasi parte del mondo.

Questo articolo ha descritto le principali funzionalità definite dal 3GPP per quanto riguarda il livello fisico ed i protocolli radio.

Sono state evidenziate le principali innovazioni introdotte da 5G NR in termini di modulazione, codici per la correzione degli errori introdotti dal canale radio e utilizzo della tecnica MIMO.

Sono stati descritti i protocolli radio e gli stati logici in cui si può trovare un terminale, illustrando anche come è specificata la connettività con la Core Network.

L'impatto di queste funzionalità è visibile a tutti: il throughput è definito dalle funzionalità di livello fisico (es. numero di antenne nella tecnica MIMO), la latenza è influenzata dalla durata di simbolo OFDM e dalla "velocità" con cui lo smartphone riesce a passare da uno stato logico all'altro.

I diversi stati logici impattano la durata della batteria in quanto agiscono sulla segnalazione di Control Plane ed ottimizzano il tempo in cui lo smartphone è attivo ■

Bibliografia

[1] D. Arena, C. Carlini, M. Ubicini, “Dentro lo smartphone: SoC e testing”, Notiziario Tecnico n.1 - 2020

[2] TS 38.211, “NR - Physical channels and modulation”. Release 15, V15.8.0 (2019-12)

[3] TS 38.212, “NR - Multiplexing and channel coding”. Release 15, V15.8.0 (2019-12)

[4] TS 38.306, “NR - User Equipment (UE) radio access capabilities”. Release 15, V15.8.0 (2019-12)

[5] TS 38.300, “NR - Overall description (Stage-2)”. Release 15, V15.8.0 (2019-12)

[6] TS 38.401, “NG-RAN - Architecture description”. Release 15, V15.7.0 (2019-12)

[7] TR 38.801, “Study on new radio access technology: Radio access architecture and interfaces”. Release 14, V14.0.0 (2017-03)

[8] TS 23.501, “System architecture for the 5G System (5GS)”. Release 16, V16.3.0 (2019-12)

Acronimi

3GPP	Third Generation Partnership Project
5GC	5G Core network
AMF	Access and mobility Management Function
AS	Access Stratum
CP	Control Plane
DRX	Discontinuous Reception
EPC	Evolved Packet Core
FFT	Fast Fourier Transform
LDPC	Low Density Parity Check
LTE	Long Term Evolution
MAC	Medium Access Control
MIMO	Multiple Input Multiple Output
MU-MIMO	Multi User MIMO
NAS	Non Access Stratum
NG-RAN	Next Generation Radio Access Network
NR	New Radio
OFDM	Orthogonal Frequency Division Multiplexing
PDCP	Packet Data Convergence Protocol
SoC	System on Chip
RAN	Radio Access Network
RAN1	RAN Working Group 1
RAN2	RAN Working Group 2
RLC	Radio Link Control
RLF	Radio Link Failure
RRC	Radio Resource Control
RRM	Radio Resource Management
SDAP	Service Data Adaptation Protocol
UE	User Equipment
UP	User Plane

Bruno Melis *bruno1.melis@telecomitalia.it*

Ingegnere Elettronico, è entrato in azienda nel 1995 occupandosi delle tecniche di trasmissione a livello fisico per i sistemi radiomobili. Attualmente è nella funzione Technology Innovation. Si è occupato dell’analisi delle prestazioni e del dimensionamento dei sistemi radio tramite tecniche di simulazione numerica applicate a diversi sistemi fra i quali il GSM, UMTS/HSDPA, LTE e recentemente per il sistema 5G NR. È coautore di diversi brevetti relativi ad algoritmi di elaborazione del segnale e a tecniche di trasmissione/ricezione basate su antenne multiple ■

Damiano Rapone *damiano.rapone@telecomitalia.it*

Laureato con lode all’Università degli Studi di Roma “Tor Vergata” in Ingegneria Elettronica nel 2011, specializzazione Elettronica a Radio Frequenza. Nel 2013 ottiene il Master di II° livello “Innovazione di reti e servizi nel settore dell’ICT”, con lode, presso il Politecnico di Torino. Nello stesso anno entra in Telecom Italia nel gruppo Wireless Access Innovation con sede a Torino (TILAB) per occuparsi di analisi di prestazioni dei sistemi LTE/LTE-A. Dal 2016 è delegato Telecom Italia nel 3GPP RAN2 che standardizza i protocolli Layer 2 e Layer 3 delle tecnologie radio 5G (NR ed evoluzione di LTE). Dal 2019 è delegato in 5G Automotive Association (5GAA), nel WG3, dove segue tematiche di interoperabilità e di conformance testing dei dispositivi C-V2X. Dal 2020 è delegato nel WG5 di O-RAN Alliance che definisce i profili di interoperabilità delle interfacce di rete (X2, Xn, F1, W1, E1) in ottica multi-vendor. È coinvolto attivamente in vari progetti finanziati dall’Unione Europea ■

Giovanni Romano *giovanni.romano@telecomitalia.it*

Ingegnere elettronico, si occupa del coordinamento delle attività di standardizzazione tecnica su accesso radio, terminali mobili e frequenze. Rappresenta Telecom Italia in 3GPP RAN, con incarico della gestione dei rapporti tra 3GPP ed ITU-R. Dal 2013 al 2017 ha inoltre rivestito la carica di vice presidente del 3GPP RAN; dal 2016 rappresenta TIM come Alternate Board Director in NGMN. Ha iniziato a lavorare nel mondo degli standard nel 1996, partecipato ai lavori di ETSI, 3GPP, ITU-R e NGMN.

Fino al 2004 è stato project manager per le attività radio su UMTS e nel 1999-2001 è stato responsabile tecnico del trial UMTS a Torino ■

DENTRO LO SMARTPHONE: SOC E TESTING

Domenico Arena, Camillo Carlini, Massimiliano Ubicini

Questo articolo descrive le componenti HW di uno smartphone, necessarie per consentire la comunicazione tra il "telefonino" e la rete cellulare. Le prestazioni di queste componenti impattano la qualità del servizio percepita dal cliente (come raggiungibilità e throughput). L'articolo descrive quindi il processo di verifica di conformità alle specifiche 3GPP necessario per assicurare il corretto funzionamento nella rete TIM di uno smartphone.

Introduzione

Uno smartphone è un insieme di molte tecnologie: lo schermo, le fotocamere, il sistema operativo e le applicazioni che forniscono i servizi sono solo alcuni esempi.

Tutto questo però non funzionerebbe senza la componente di comunicazione, ovvero la capacità di operare in una rete cellulare. In particolare, la capacità di calcolo e di memoria di un "telefonino" rivaleggiano con i supercomputer degli anni '90 del secolo scorso, con un livello di integrazione che permette di concentrare in un unico chip (System on Chip, SoC) le funzionalità fondamentali.

Questo articolo analizza come i protocolli radio descritti nell'articolo "Dentro lo smartphone - Banda base e protocolli radio" [1] sono realizzati nel cuore HW del terminale e illustra i passi necessari per verificarne il corretto comportamento in rete e quindi la commercializzazione.

In particolare, per poter arrivare sugli scaffali di un negozio TIM, uno smartphone deve seguire alcune procedure standardizzate. Come descritto in [1], i protocolli e le funzionalità radio sono specificate dai gruppi RAN1 e RAN2 del 3GPP.

Sulla base di queste specifiche, il 3GPP RAN4 definisce le prestazioni radio, in termini di sensibilità del ricevitore, potenza trasmessa ed emissioni spurie fuori banda. In altre parole, il RAN4 permette di definire le prestazioni minime oltre le quali il device non è più in grado di funzionare correttamente.

Le specifiche del RAN4 sono poi la base del lavoro di ETSI, CEN e CENELEC per definire le regole di commercializzazione di uno smartphone in Europa, cioè per poter apporre il marchio CE.

Questo risultato è raggiunto se l'apparato non disturba i sistemi operanti nelle bande radio adiacenti (emissioni fuori banda), se è in linea con le regole di compatibilità elettromagnetica. Infine, il 3GPP RAN5 specifica le metodologie di test del device, in modo da assicurarne la rispondenza alle specifiche tecniche.

Quanto fa il 3GPP, seppur necessario, non è però sufficiente a garantire il corretto funzionamento di uno smartphone in rete. Per assicurare questo, occorre certificare che il device sia effettivamente rispondente alle specifiche e non introduca comportamenti anomali.

Il Global Certification Forum (GCF) definisce il set di test minimo che permette ai costruttori di autocertificare la rispondenza alle specifiche 3GPP.

In modo analogo il gruppo della GSM Association che si occupa di terminali mette a disposizioni le reti dei diversi operatori per fare verifiche in campo del funzionamento.

L'aver passato questi test e certificazioni, dimostra che lo smartphone è rispondente alle specifiche 3GPP ed è in grado di interoperare correttamente con gli apparati di rete.

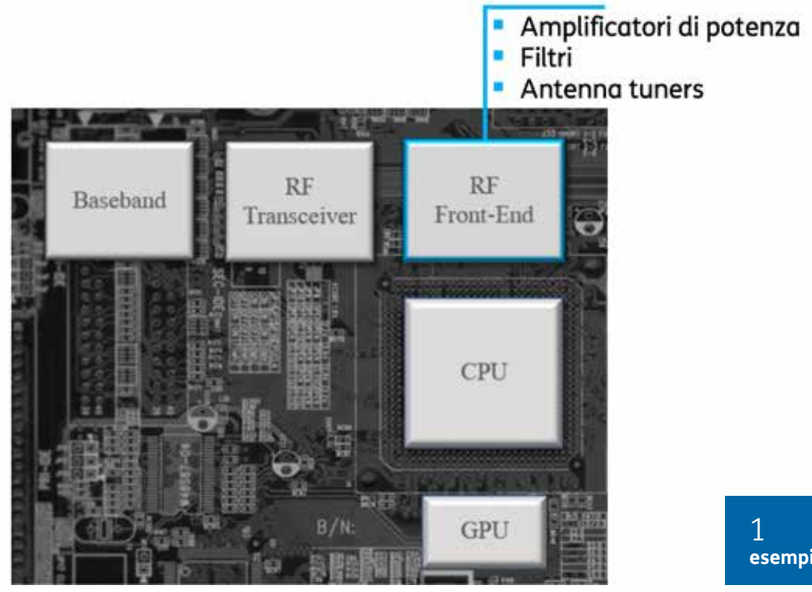
SoC - I principali componenti

L'elemento centrale dei più recenti smartphone 4G e 5G è costituito dal "System on Chip" (SoC), un circuito integrato che racchiude i componenti fondamentali rappresentati in Fig. 1.

SoC commerciali molto diffusi nell'attuale mercato degli smartphone 4G e 5G sono per esempio: Qualcomm® Snapdragon™, Samsung Exynos, Huawei Kirin. Questi moderni SoC sono ormai realizzati con tecnologia produttiva inferiore a 10 nm, integrando quindi miliardi di transistor.

Vediamo i componenti fondamentali di un SoC:

- il chip "Baseband" è il vero e proprio Modem per la gestione della connettività 5G e 4G, responsabile per l'esecuzione delle tecniche OFDM e MIMO. Negli ultimi anni la capacità computazionale dei Baseband è notevolmente incrementata,



1
esempio di modem SoC

al fine di poter gestire canali dati di comunicazione con la rete mobile aventi ampiezze di banda di frequenza (canalizzazione) sempre maggiori, nonché tecniche di modulazione e MIMO di ordine superiore. In 4G e per le frequenze FDD, la canalizzazione massima per portante è di 20 MHz per downlink e 20 MHz per uplink (era 5 MHz in 3G). Grazie alla funzionalità di Carrier Aggregation, introdotta nelle reti commerciali fin dal 2014, è possibile aggregare anche 4 o 5 portanti di differenti frequenze, ottenendo canalizzazioni dell'ordine dei 50 - 100 MHz (tipicamente per downlink). L'adozione della modulazione 256QAM e MIMO4x4 permettono di innalzare il throughput di picco nella trasmissione in downlink tra rete mobile e smartphone (tipicamente nel range 500 Mbps – 1 Gbps di picco), ma anche comportano un incremento della capacità computazionale richiesta al Baseband. L'aumento delle ampiezze di banda da gestire è particolarmente rilevante nel caso 5G NR, in cui la singola portante presenta una canalizzazione fino a 100 MHz (Frequency Range 1, fino a 6 GHz), mentre nel range di frequenze delle onde millimetriche (e.g. 26 GHz) è già oggi possibile aggregare fino a 800 MHz di ampiezza di banda totale

- l'RF Transceiver è il componente che determina le prestazioni misurate prima che intervengano le antenne dello smartphone. Parametri misurati sono la sensibilità in ricezione (per downlink) e la potenza in trasmissione (per uplink). Queste caratteristiche radio sono cruciali per la qualità della comunicazione: se per esempio la sensibilità in ricezione fosse inferiore di 3 dB rispetto ai requisiti target definiti dagli operatori, a parità di condizioni di rete mobile, l'utente avrebbe per i propri servizi un livello di segnale 4G / 5G del 50% inferiore a quello garantito da uno smartphone conforme ai requisiti target.

DEFINIZIONE DEI REQUISITI OTA NELL'AMBITO DELLA NORMATIVA EUROPEA

Nel 2017 l'Amministrazione Olandese [3] ha richiesto ad ETSI la definizione di requisiti minimi di prestazione dei sistemi d'antenna dei terminali misurati over the air in "modalità chiamata vocale (BHH mode)". ETSI ha creato un gruppo di lavoro all'interno di MSG-TFES per finalizzare i relativi Harmonized Standards.

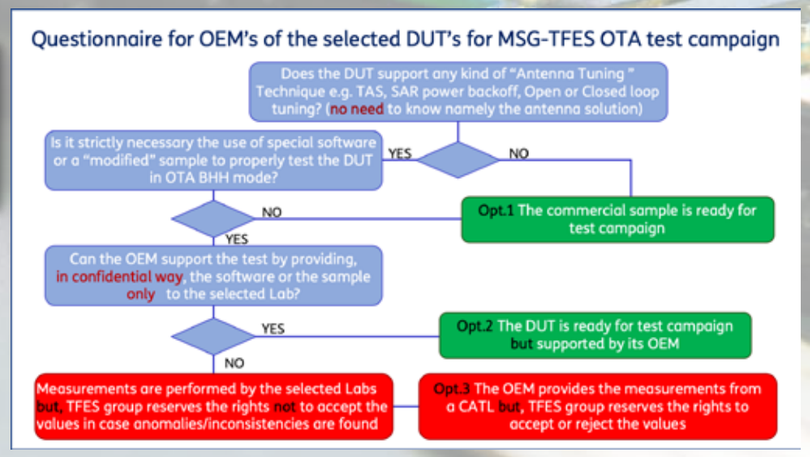
MSG-TFES vede la partecipazione attiva di TIM, dei maggiori costruttori di terminali ed apparati TLC e dei principali Operatori Mobili e Fissi europei, oltre all'Amministrazione Olandese.

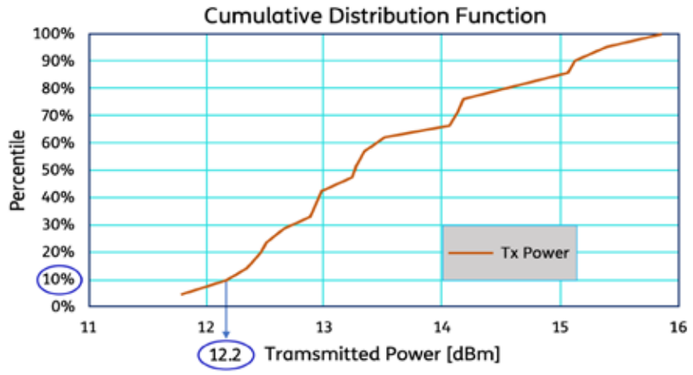
La misura delle prestazioni in irradiata dei terminali è tecnicamente molto complessa ed i risultati delle misurazioni possono dipendere fortemente dalle procedure di misura, dalla qualità delle apparecchiature di testing e dalla qualità del laboratorio stessa.

Il gruppo di lavoro ha deciso di derivare i requisiti minimi delle prestazioni OTA dalle misure effettuate su di un campione di terminali concordato [4] (data driven approach): 21 terminali (20 smartphone + 1 industrial device), acquistati open market (no terminali MNO branded), sono stati distribuiti a 5 laboratori mondiali accreditati ISO 17025, che li hanno misurati secondo la metodologia prevista per il BHH mode [5].

Un altro aspetto importante è stato la possibilità di distinguere tra i terminali prescelti quelli che supportano particolari tecniche di ottimizzazione del diagramma di irradiazione in presenza dell'utente umano (antenna tuning) e che richiedevamo quindi, particolare attenzione e/o il supporto del Costruttore per la corretta misura in camera anecoica.

A
Questionario per classificazione dei terminali prescelti per la campagna di misure ETSI MSG-TFES





B
Esempio di derivazione del "minimum requirement" dalla distribuzione cumulativa della potenza trasmessa (data driven approach)

Il processo di accettazione e classificazione dei terminali da parte di ETSI si è articolato secondo il grafo di flusso riportato nella figura A [6] e da considerarsi una novità significativa rispetto i precedenti lavori in 3GPP.

Nel dettaglio, il processo complessivo si è articolato in 4 fasi, di cui l'ultima ancora in corso e di seguito descritte; cruciale il ruolo dell'ETSI, sia per la gestione logistica che a garanzia di tutto il processo per quanto riguarda la raccolta dei dati di misura in chiaro e successiva gestione in forma anonimizzata:

Fase 1

- Selezione di un parco terminali da misurare con buona presenza commerciale in rete nel periodo 2017-1H2018 (finestra commerciale di 18 mesi), diversità di brand e varietà di modelli (dall'entry level al top gamma) [4].
- Scelta dei laboratori accreditati ISO 17025 su base candidature volontarie.
- Definizione di un criterio di allineamento misure tra i vari laboratori secondo un principio di round robin: 2 terminali, dei 21 DUT, sono stati misurati da tutti i laboratori così da mettere a confronto affidabilità e precisione di ciascun laboratorio [5].

Fase 2

- ETSI si è fatta carico della raccolta dei dispositivi, acquistati volontariamente da alcune delle compagnie attive in MSG-TFES, tra cui TIM.
- Successiva distribuzione e suddivisione, su base sorteggio, dei terminali da misurare ai 5 laboratori individuati (circa 4 DUT per ciascun lab).

Fase 3

- Raccolta dei risultati in chiaro provenienti dai laboratori e presentazione in forma anonimizzata al gruppo MSG-TFES (sotto la responsabilità ETSI) [7].
- Post-elaborazioni preliminari: definizioni per ciascuna banda di valor medio, deviazione standard e distribuzioni cumulative (CDF) dei risultati anonimizzati.

Fase 4 (attualmente in corso):

- Definizione dei percentili target di ciascuna CDF per la definizione dei minimum requirements.
- Accordo sui requisiti minimi OTA per LTE e successiva inclusione negli Harmonized Standards.

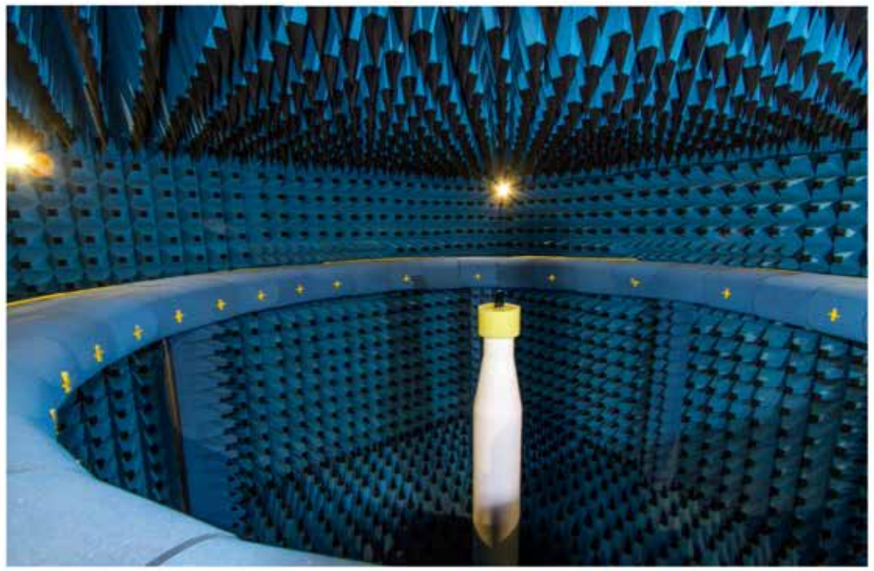
Il gruppo è dunque attualmente a lavoro sul primo punto della "fase 4", ovvero la scelta del percentile più appropriato, per ciascuna banda operativa, che individui nella pratica, la percentuale della popolazione di terminali che soddisfi il valore del parametro radio misurato: ad es. fissando il 10° percentile della distribuzione cumulativa di potenza tramessa dal DUT, la corrispondente misura è soddisfatta dal 90% della popolazione considerata che raggiunge o supera quella soglia di potenza così individuata come rappresentato nella figura B.

- le prestazioni radio (sensibilità in ricezione e potenza in trasmissione) sono più significativamente misurate considerando il comportamento dello smartphone "Over The Air", cioè includendo gli effetti di accoppiamento dell'RF Transceiver con le antenne, così come avviene nell'utilizzo quotidiano dello smartphone. Tale accoppiamento, nel SoC, si realizza attraverso i componenti di RF Front-End, quali gli amplificatori di potenza ed i filtri RF. In particolare, i filtri, sempre più evoluti nei moderni smartphone 4G e 5G, permettono di gestire contemporaneamente differenti bande di frequenza, minimizzando eventuali effetti di intermodulazione intra-device. L'efficace gestione contemporanea di differenti bande di frequenza è alla base del supporto di funzionalità quali Carrier Aggregation 4G e 5G NR, nonché Dual Connectivity NSA tra 4G e 5G NR.

Smartphone e Metodologie di test

Gli smartphone moderni vengono sottoposti in tutte le loro fasi (sviluppo, integrazione delle componenti, produzione, realizzazione dei primi prototipi, verifica dei terminali commerciali e pre-commerciali) ad una serie quasi infinita di test che ad oggi possiamo categorizzare a grandi linee nelle seguenti tipologie:

1. Development phase Testing: verifica la corretta implementazione durante le fasi di progetto e produzione.
2. Integration Testing: verifica che tutte le componenti hardware di uno smartphone inter-lavorino correttamente tra di loro.
3. Conformance Testing: verifica che lo smartphone sia conforme alla normativa 3GPP [2].
4. Laboratory Carrier Acceptance Testing: verifica la conformità ad ulteriori requisiti definiti dall'operatore.
5. Interoperability Testing: verifica l'inter-lavoro di apparati di rete e di terminali di produttori diversi.



2
Ambiente di misura MIMO OTA

Molti di questi test vengono eseguiti in laboratorio e non su rete reale. I test di laboratorio hanno generalmente bisogno di una strumentazione dedicata alla quale viene connesso lo smartphone per poter essere verificato.

In particolare, dal punto di vista del terminale, il simulatore di rete opera a tutti gli effetti come se fosse una stazione radio base reale.
A seconda del tipo di collegamento con il simulatore di rete sono definite 2 categorie di test:

1. Test In Condotta: lo smartphone è collegato con cavo coassiale al simulatore, escludendo l'antenna.
2. Test in irradiata (o OTA "Over The Air"): comunicazione radio tramite sistema di antenna.

I test in irradiata sono più complessi rispetto ai test in condotta, in quanto la comunicazione avviene via radio, con un'attenuazione più alta rispetto a quella su cavo coassiale.
Un altro problema è il fatto che operando in ambienti chiusi (laboratorio) il segnale è soggetto inevitabilmente a riflessioni che introducono delle interferenze nella misura spesso non facilmente prevedibili e calcolabili.

Di conseguenza i test OTA sono effettuati in ambienti speciali (Camere Anecoiche), che grazie a speciali coni assorbitori, minimizzano il problema delle riflessioni.

Le camere sono anche schermate per evitare interferenza da e verso l'esterno a garantire la riproducibilità della misura.

Le prove OTA introducono una incertezza sulla misura più alta rispetto ai test in condotta.
L'incertezza di misura è inevitabile in un sistema reale di misura e

3
Misura "Beside The Head" in Camera Anecoica

apre scenari di discussione infiniti (specialmente negli enti di normativa e regolatori) tra produttori di smartphone e gli operatori quando la si confronta con un limite di accettazione.

Nello standard si introducono di conseguenza delle Tolleranze di Test per evitare che terminali siano dichiarati non conformi a causa delle incertezze di misura.
In generale per tutte le tipologie di test che non intendono specificatamente includere le prestazioni dell'antenna si preferisce utilizzare l'ambiente in condotta (esempio verifica dei protocolli [2]).

L'utilizzo della connessione in condotta non è più possibile per range di frequenza superiori ai 6 GHz (cosiddette onde millimetriche utilizzate dal 5G) a causa dell'utilizzo di antenne patch, senza connettori d'antenna.

L'ambiente OTA ha l'obiettivo di valutare le prestazioni dell'antenna del terminale e prevede due tipologie:

1. SISO OTA: focalizzato a misurare la Sensibilità Del Ricevitore e la Potenza Massima emessa dal trasmettitore come se si trovasse a bordo cella.
2. MIMO OTA: simula un ambiente più reale che tiene conto delle caratteristiche del canale (attenuazioni, riflessioni e fading) in presenza di antenne multiple, V. Figura 2.

Le misure SISO OTA sul device si effettuano generalmente nelle seguenti modalità:

1. In Spazio Libero (Free space): posizionato come oggetto isolato.
2. Di Fianco alla Testa (Beside Head, BH): lo smartphone è posizionato è di fianco ad una testa "fantoccio" costituita da materiale che ha proprietà dielettriche, e dimensioni simili al contenuto organico della testa umana, V. figura 3.
3. Di Fianco alla Testa con Mano (Beside Head & Hand, BHH): insieme alla testa "fantoccio" viene utilizzata anche una "mano fantoccio" che sorregge il terminale.
4. Modalità di gioco: lo smartphone è impugnato da entrambi i lati da una doppia "mano fantoccio".
5. Modalità di navigazione: lo smartphone è impugnato da una "mano fantoccio" per riprodurre la navigazione web.

Data la complessità del sistema di misura ne consegue una elevata incertezza che introduce delle problematiche nel confronto tra misure di laboratori diversi. Nel caso di Acceptance Testing dell'operatore infatti occorre spesso confrontarsi con le misure effettuate dal costruttore.

Per garantire ufficialità e la confrontabilità dei risultati sono definite delle procedure di accreditamento internazionale.

Un laboratorio per essere accreditato deve compiere una serie rigorosa di procedure codificate di taratura, utilizzo e validazione, sia per gli ambienti (camera anecoica) sia per la strumentazione.

Per essere internazionalmente riconosciuto il laboratorio deve essere sottoposto ad una visita periodica da parte di un ente accreditatore esterno (es. ACCREDIA per l'Italia).

TIM investe ogni anno nella calibrazione, aggiornamento e accreditamento (es. ISO17025) delle camere anecoiche preposte per le varie tipologie di test.

Conclusioni

Il nostro telefonino si basa su architetture SoC realizzate con tecnologia produttiva inferiore a 10 nm, integrando quindi miliardi di transistor e facendoli rivaleggiare con i supercomputer degli anni '90 del secolo scorso.

Le diverse componenti del modulo di comunicazione permettono di raggiungere prestazioni molto elevate, con throughput di circa 2 Gbps nei primi smartphone commerciali 5G. Le caratteristiche tecniche di queste componenti impattano la qualità percepita dal cliente, in termini per esempio di raggiungibilità (ovvero di essere in copertura) e di throughput.

Una sensibilità in ricezione 3 dB inferiore rispetto ai requisiti target comporterebbe, a parità di condizioni di rete mobile, un livello di segnale del 50% inferiore a quello garantito da uno smartphone conforme ai requisiti.

Si rende quindi necessario un processo di verifica di conformità alle specifiche 3GPP ed ai requisiti di TIM.

Questo processo è tecnicamente molto complesso, ma è fondamentale per assicurare il buon funzionamento degli smartphone con brand TIM e di garantire un livello minimo di qualità di quelli che arrivano dal mercato retail ■

Bibliografia

[1] B. Melis, D. Rapone, G. Romano "Dentro lo smartphone: Banda base e protocolli radio", Notiziario Tecnico n.1-2020

[2] Approfondimento "Il 3GPP conformance testing", massimiliano.ubicini@telecomitalia.it Notiziario Tecnico n° 2 - 2015 Articolo 3 "L'evoluzione Dell'accesso Radio LTE", Andrea Buldorini, Maurizio Fodrini, Giovanni Romano

[3] TCAM WG (09) 16: Receiver sensitivity of mobile phones - NL administration (TFES(17)000016)

[4] TFES(19)000032r5: OTA testing - devices and information received up to date (pre-testing phase) – ETSI Secretariat.

[5] TFES(19)000019r4: LTE UE OTA Measurement Campaign Technical Guidance Document - Samsung Electronics R&D Institute.

[6] TFES(19)000021r3: Questionnaire for OEM's of the selected DUT's for MSG-TFES OTA test campaign – TIM, Orange, Vodafone, Telefonica.

[7] TFES(20)065022: Results of the LTE UE OTA Antenna Parameters Test Campaign – ETSI Secretariat.

Acronimi

256QAM	256 Quadrature Amplitude Modulation
3GPP	Third Generation Partnership Project
BH	Beside Head
BHH	Beside Head & Hand
CDF	Cumulative Density Function
CEN	Comité Européen de Normalisation
CENELEC	Comité Européen de Normalisation Électrotechnique
DUT	Device Under Test
ETSI	European Telecommunication Standard Institute
FDD	Frequency Division Duplex
GCF	Global Certification Forum
GSM	Global System for Mobile Communications
HW	Hardware
LTE	Long Term Evolution
MIMO	Multiple Input Multiple Output
MNO	Mobile Network Operator
MSG-TFES	Mobile Standards Group - Task Force for the production of Harmonised Standards under the RED for the IMT family
NR	New Radio
NSA	Non-Stand Alone
OEM	Original Equipment Manufacturer
OFDM	Orthogonal Frequency Division Multiplexing
OTA	Over The Air
SISO	Single Input, Single Output
SoC	System on Chip
RAN	Radio Access Network
RAN1	RAN Working Group 1
RAN2	RAN Working Group 2
RAN4	RAN Working Group 4
RAN5	RAN Working Group 5
RF	Radio Frequency
UE	User Equipment

Domenico Arena *domenico.arena@telecomitalia.it*

Ingegnere delle Telecomunicazioni (Università di Pisa - 1999), entra in Azienda (CSELT) nel 2001 occupandosi della progettazione di componenti satellitari in guida d'onda, nell'ambito dei programmi dell'Agenzia Spaziale Europea (ESA). Da metà anni 2000 si sposta nell'ambito dell'Accesso Radio Mobile presidiando tematiche interferenziali e di coesistenza tra reti radiomobili e tra reti eterogenee a supporto della pianificazione radio (spectrum optimization) e delle aste frequenziali. Dal 2006 al 2009, delegato TIM presso 3GPP RAN WG4. Negli ultimi anni ha fa focalizzato la sua attività sulla modellizzazione del canale di propagazione radio mobile, le onde millimetriche e profili di canali "aerei" a supporto delle nuove applicazioni verticali basate sui "droni connessi". Dal 2018 presidia il gruppo ETSI MSG-TFES nell'ambito dei cosiddetti Harmonized Standards europei. È coautore di numerose pubblicazioni tematiche oltre ad un brevetto sul monitoraggio della latenza di rete ■

Camillo Carlini *camillo.carlini@telecomitalia.it*

Laureato in Ingegneria Elettronica, sono entrato in Azienda nel 2006, dopo una prima esperienza nel settore dei semiconduttori. Mi sono occupato da subito di ingegnerizzazione dei Devices Cliente, sia Fissi che Mobili, focalizzandomi progressivamente sulla gestione dell'evoluzione dei Terminali LTE e 5G, contribuendo lato Devices al lancio commerciale dei servizi 4.5G e 5G. In ambito internazionale, oltre ad essere coautore di diverse pubblicazioni IEEE su tecnologie Wireless, ho partecipato ai gruppi standard Small Cell Forum, Open Mobile Alliance e 3GPP RAN WG5 e WG2, essendo oggi il delegato TIM in GCF Steering Group. Attualmente sono responsabile del progetto "Mobile Devices Engineering and Specifications" e della valorizzazione del relativo gruppo di lavoro ■

Massimiliano Ubicini *massimiliano.ubicini@telecomitalia.it*

Laureato al politecnico di Torino in ingegneria elettronica in Telecomunicazioni, entra nei laboratori Telecom Italia dal 1999 dove ha modo di approfondire tutte le tematiche relative al testing del terminale radiomobile attraverso le tecnologie che si sono susseguite negli anni successivi (GERAN, UMTS, IMS, LTE e attualmente 5G). L'attività presso i laboratori ha permesso di sviluppare la proprie esperienza su strumenti di misura, procedure, ambienti di test, automazione e standard. Dal 2003 ha iniziato a seguire, in sinergia con le attività in laboratorio, la standardizzazione in 3GPP RAN WG5 su tematiche relative al test di conformità dei terminali mobili dove ha potuto seguire attivamente la standardizzazione del testing di tutte le tecnologie di accesso mobile a partire dal 3G. Attualmente il principale focus di attività è l'evoluzione tecnologica di banchi di misura verso il 5G e relativa standardizzazione in ambito 3GPP RAN5 e ETSI MSG TFES come delegato TIM ■