

Sistemi multiantenna o MIMO

NON appena l'evoluzione tecnologica lo ha permesso i sistemi di trasmissione radio hanno iniziato a dotarsi di più di una antenna, presso una od entrambe le estremità del collegamento. Tale modifica architettonica ha reso possibile lo sviluppo di tecniche trasmissive basate sul *signal processing* del tutto nuove, capaci di migliorare le prestazioni sia nei termini di un maggior *SNR* in ricezione sia in quelli di una più elevata velocità di trasmissione, conseguendo così una *efficienza spettrale* prima non immaginabile. Ma la cosa più sorprendente è che i miglioramenti avvengono *grazie ai fenomeni di multipath*, che nello scenario ad antenna singola costituivano invece un aspetto peggiorativo.

In condizioni di *fading piatto* ciascun canale radio tra una coppia di antenne è descritto da un unico coefficiente complesso h_{ij} , l'insieme dei quali individua una matrice \mathbf{H} di dimensione $n_T \cdot n_R$ la cui conoscenza da parte del ricevitore, del trasmettitore, o di entrambi, costituisce un fattore determinante per la realizzazione di quasi tutte le soluzioni che andremo ad esaminare. Innanzitutto vedremo come molteplici antenne in ricezione consentano di combinare *in modo coerente* i segnali di ciascuna, mentre una molteplicità in trasmissione permette di inviare *più copie* dello stesso segnale. Quindi studieremo come la presenza di n_T antenne in trasmissione ed n_R in ricezione determini la possibilità di trasmettere *fino a* $\min(n_T, n_R)$ diverse comunicazioni indipendenti, il cui numero massimo dipende dal *rango* della matrice \mathbf{H} , funzione dal grado di indipendenza tra i suoi elementi, a sua volta funzione dalla spaziatura tra antenne e dalla natura dell'ambiente circostante.

Dopo aver ridefinito il concetto di capacità di canale per adattarlo al nuovo caso di canale *vettoriale* affetto da fading variabile, vengono descritte tecniche lineari subottime che permettono la *separazione* dei diversi flussi ricevuti senza incorrere in una eccessiva complessità computazionale. Tecniche duali possono essere altresì impiegate al lato trasmettente, aprendo la possibilità di adottare la soluzione multiantenna anche per realizzare schemi di trasmissione *uno a molti* tipici delle reti di telefonia mobile e di accesso WIFI e WIMAX. In tali contesti si è affermato l'abbinamento delle tecniche di trasmissione MIMO con quelle di modulazione a portante multipla OFDM, determinando la *moltiplicazione* del numero di canali frutto della molteplicità delle antenne, per il

numero delle sotto-portanti su cui si basa l'OFDM. Una soluzione del tutto simile si è affermata anche per sistemi di diffusione radio televisiva digitali o DVB, in cui un unico dispositivo riceve in modo sincrono lo stesso segnale OFDM trasmesso da molteplici antenne dislocate sul territorio.

21.1 Lo scenario delle possibilità

Iniziamo con un riassunto sommario dei molteplici aspetti da affrontare. Una prima discriminazione terminologica riguarda la *localizzazione* della molteplicità di antenne, specificando che mentre adottarne più di una sia in trasmissione che in ricezione viene indicato con l'acronimo MIMO ovvero *multiple input multiple output*, ed il caso *tradizionale* di una singola antenna da entrambi i lati è detto SISO (*single input single output*), l'utilizzo di antenne multiple da un solo lato del collegamento, in trasmissione od in ricezione, viene rispettivamente indicato con i termini MISO (*multiple input single output*) e SIMO (*single input multiple output*), vedi fig. 21.1a.

Il secondo punto importante è che la presenza di più antenne ad uno o entrambi i lati del collegamento consente l'attuazione di differenti principi genericamente descritti come *diversità*, *multiplazione spaziale* e *beamforming* (vedi fig. 21.1b), con lo scopo di perseguire tre obiettivi complementari e che in generale *non* possono essere conseguiti congiuntamente, ma di cui esiste la possibilità di individuare soluzioni di compromesso. Vediamo di che si tratta.

Diversità spaziale È la più diretta conseguenza della possibilità di poter trasmettere e/o ricevere una stessa trasmissione mediante antenne diverse. Come riferito al § 20.4.4 se le antenne sono abbastanza distanziate tra loro i *cammini multipli* su cui si sviluppa il collegamento sono differenti, così come la risposta in frequenza $H(f)$ vista dalle singole antenne, cosicché la distorsione lineare subita da ciascuna di esse diviene una v.a. statisticamente indipendente. Ciò rende molto improbabile che le antenne subiscano *tutte assieme* una forte attenuazione, cosicché la ricezione risente molto meno della variabilità del canale; come introdotto al (§ 20.3.3.1), questo risultato è frutto dello

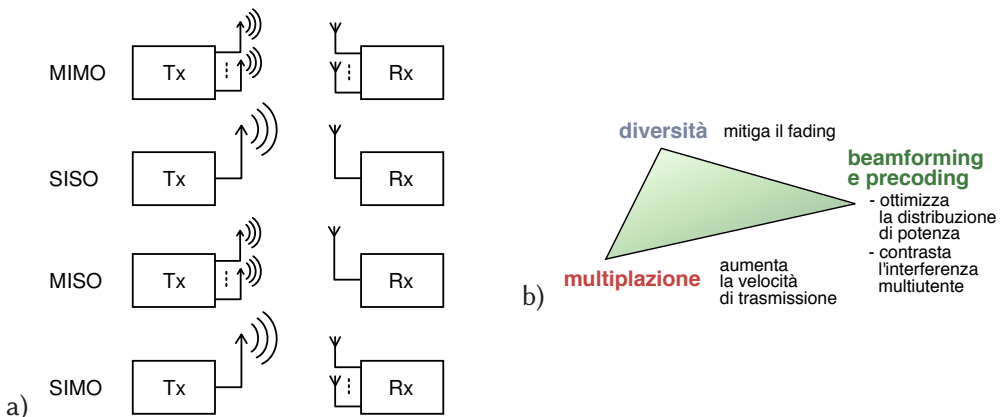


Figura 21.1: MIMO: a) - configurazioni di antenna; b) - tecniche applicabili

sfruttamento della *diversità spaziale*. In tale scenario la velocità di trasmissione non cambia, dato che ogni antenna trasmette lo stesso simbolo, o come approfondiremo al § 21.3.2.2, una versione codificata della stessa sequenza. Due particolarità della tecnica sono che in generale non è necessario conoscere le caratteristiche del canale per le diverse antenne, e che una qualche sua variante può essere attuata anche se solo una delle parti in comunicazione è dotata di più antenne. Il *guadagno di diversità* conseguito può essere impiegato a favore di una maggiore affidabilità del collegamento (ossia per ottenere un miglior E_b/N_0 e dunque P_e), oppure per preservare la stessa affidabilità di un sistema SISO ma diminuire la potenza totale irradiata, a vantaggio della durata delle batterie dei terminali di utente.

Multiplazione spaziale In presenza di più antenne da entrambi i lati del collegamento la multiplazione consiste nel trasmettere flussi numerici *diversi* da parte delle differenti antenne, contemporaneamente, ed alla stessa frequenza. Ciò è possibile a patto che il ricevitore conosca la matrice \mathbf{H} che descrive la risposta in frequenza tra coppie di antenne; quando tali valori risultano sufficientemente incorrelati la matrice diviene invertibile, rendendo possibile risalire ai singoli segnali a partire dalle loro diverse combinazioni lineari, ricevute da ciascuna antenna. Il risultato è che la velocità aggregata è *la somma*¹ di quella dei diversi flussi, mentre la banda occupata è quella di un flusso singolo, conseguendo pertanto una elevata *efficienza spettrale* (pag. 494). Qualora anche il trasmettitore conosca \mathbf{H} può combinare la multiplazione con il *pre-coding*, rendendo possibile trasmettere i diversi flussi verso *destinatari differenti*, anche ad antenna singola, realizzando in tal modo un sistema *multiutente* detto MU-MIMO basato sullo *space division multiple access* o SDMA.

Compromesso diversità - multiplazione Non è possibile conseguire allo stesso tempo sia un pieno guadagno di diversità che un altrettanto pieno guadagno di multiplazione, semplicemente perché le antenne o trasmettono gli stessi dati, oppure dati differenti. Nei sistemi reali si ricerca una condizione operativa di compromesso, ed in grado di adattarsi in funzione delle condizioni del collegamento: mentre in presenza di un buon *SNR* può non essere necessario ricorrere alla diversità e dunque è possibile sfruttare al massimo la multiplazione (conseguendo una velocità elevata), per valori di *SNR* ridotti è opportuno rinunciare a buona parte di multiplazione, a favore della diversità.

Beamforming Il senso più letterale di questo termine si riferisce alla consolidata tecnica di *orientare* verso una determinata direzione l'onda elettromagnetica emessa da una schiera di antenne², in modo da aumentare la potenza ricevuta in tale direzione e/o riducendo la potenza interferente per soggetti in posizioni diverse. Nato nel contesto di una comunicazione in visibilità, nel caso di fading di Rayleigh ovvero di prevalenza

¹In altre parole la velocità aggregata aumenta linearmente con il numero delle antenne, inteso come il numero *minimo* tra quelle di trasmissione e quelle di ricezione.

²Tecnica nota anche come *smart antenna* o *phased array*, vedi ad es. https://en.wikipedia.org/wiki/Phased_array

dei cammini riflessi tale approccio perde di applicabilità, ma lo stesso termine è usato per intendere anche la tecnica del...

Precoding Nel caso di una molteplicità di antenne di tipo MIMO e della perfetta conoscenza di \mathbf{H} da parte del trasmettitore, studieremo come tecniche di signal processing consentano di realizzare il cosiddetto *eigen-beamforming*, ovvero utilizzare il collegamento ripartendo l'informazione da trasmettere nelle *direzioni* individuate dagli autovettori della matrice di canale. Come per il multiplexing anche con il precoding si ottiene un aumento della velocità di trasmissione, sia per quanto riguarda un collegamento punto-punto, sia per la possibilità di servire più utenti in un sistema MU-MIMO. Ma in questo secondo caso la tecnica da adottare si modifica allo scopo di minimizzare l'effetto interferente nei confronti degli utenti non destinatari.

Esaurita l'esposizione delle tecniche principali, il capitolo prosegue illustrando le problematiche correlate, riassunte appresso.

Modalità duplex e canale di ritorno Le due direzioni di trasmissione possono essere distinte nel tempo o in frequenza: mentre nel primo caso la proprietà di reciprocità del canale radio permette di usare la stima di canale svolta in fase di ricezione anche durante quella di trasmissione, nel secondo caso ciò non è possibile, determinando l'esigenza di rendere nota al trasmettitore la matrice \mathbf{H} stimata in ricezione. La trasmissione di \mathbf{H} inoltre avviene per forza di cose in modo approssimato (quantizzato), comportando errori ed imprecisioni nel precoding, ed impegnando tante più risorse di trasmissione quanto più precisa si desidera la codifica della \mathbf{H} da trasmettere.

Trasmissione multiportante Gran parte degli approcci fin qui descritti si basano sul presupposto di avere a che fare con segnali a banda stretta ovvero per i quali si verificano le condizioni di *fading piatto*, impedendo di aumentare troppo la velocità di trasmissione ovvero l'occupazione di banda, a causa dell'aumento di complessità della procedura di equalizzazione che si renderebbe necessaria. Ciò non è più vero nel caso della trasmissione OFDM, per la quale la banda a disposizione è suddivisa in P sotto-canali di banda ridotta su ciascuno dei quali trasmettere una frazione del flusso informativo complessivo, rendendo l'ipotesi di $H(f)$ costante verificata per ciascun sotto-canale. In pratica è come avere a disposizione P diverse trasmissioni MIMO, su ciascuna delle quali applicare le tecniche discusse, con valutazioni di compromesso potenzialmente differenti per ciascuna sottoportante. A questo aspetto sono da aggiungere due ulteriori vantaggi di una trasmissione MIMO-OFDM: il primo deriva dalla semplicità con cui l'OFDM consente di equalizzare la distorsione lineare subita da ciascuna sottoportante, mentre la seconda è la possibilità di assegnare sottoportanti diverse ad utenti differenti per realizzare con facilità un sistema multiutente.

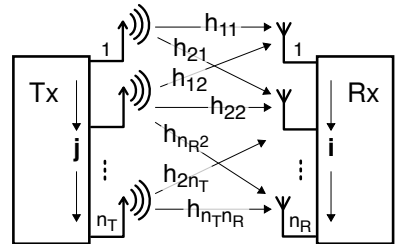
Rete a frequenza singola La trasmissione broadcast attuata dai sistemi di radio e TV digitale si basa su molteplici antenne disseminate sul territorio che trasmettono tutte il medesimo segnale OFDM, allo stesso tempo e nella stessa regione di frequenza,

secondo un approccio³ noto come *single frequency network* o SFN⁴. Ciò richiede una stringente sincronizzazione sia nel tempo che in frequenza (derivata dal sistema GPS) in modo che il ritardo relativo dei segnali che pervengono ad un medesimo destinatario ricada tutto entro i limiti del tempo di guardia del simbolo OFDM.

Come evidente la natura delle implicazioni della presenza di più antenne investe i più disparati contesti, che in questo capitolo tentiamo di affrontare uno alla volta, a partire dalla definizione formale del *canale MIMO*.

21.2 Il canale MIMO

In presenza di n_T antenne in trasmissione ed n_R in ricezione ogni coppia i, j di esse ($i = 1, 2, \dots, n_R$; $j = 1, 2, \dots, n_T$) vede un differente canale radio, che in condizioni di fading piatto (pag. 678) ovvero di segnale a banda stretta (§ 13.1.2.4) ossia con banda inferiore alla banda di coerenza (pag. 684), viene descritto da un unico coefficiente *complesso* h_{ij} , pari alla $H(f)$ valutata alla $f = f_0$ di modulazione.



Il valore esatto di h_{ij} dipende dai fenomeni di cammini multipli tipici delle comunicazioni wireless, sia rurali che urbane che indoor, ed in tal senso può variare nel tempo qualora la trasmissione sia da associare ad un contesto di mobilità. Il valore h_{ij} della risposta in frequenza tra le antenne j ed i ad un dato istante viene quindi messo in corrispondenza all'elemento i, j di una matrice \mathbf{H} di dimensioni $n_R \times n_T$, indicata anche come CSI ovvero *channel state information*.

La trasmissione da parte della antenna j di un segnale $x(t)$ ottenuto mediante modulazione numerica PSK o QAM a partire da una sequenza simbolica⁵ $\{s_j^k\}$ produce agli istanti di simbolo $t = kT_s$ la ricezione presso l'antenna i di un inviluppo complesso⁶

$$r_i^k = h_{ij}s_j^k + n_i \quad (21.1)$$

con n_i v.a. gaussiana complessa a media nulla, parti reale ed immaginaria incorrelate, e varianza⁷ $\sigma_n^2 = E\{nn^*\}$, condizione indicata anche come $n_i \in \mathcal{CN}\{0, \sigma_n^2\}$. Indicando

³Al contrario, le tecniche di broadcast analogico (cap. 25) su scala nazionale prevedono l'uso di regioni di frequenza diverse per lo stesso canale trasmesso in bacini di propagazione differenti, in cui questi ultimi sono definiti dalle condizioni di visibilità legate alla conformazione del territorio.

⁴Vedi ad es. https://en.wikipedia.org/wiki/Single-frequency_network

⁵I valori reale e immaginario di s^k rappresentano le coordinate nel piano dell'inviluppo complesso di un punto appartenente alla costellazione \mathcal{A} scelta per la trasmissione, vedi ad es. la figura a pag. 497.

⁶In virtù della (13.3) scriviamo $\underline{y}(t) = \frac{1}{2}\underline{h}(t) * \underline{x}(t) = \frac{1}{2}\underline{h} \cdot \underline{x}(t)$ dato che per fading piatto si ha $\underline{h}(t) = \underline{h} \cdot \delta(t)$; a sua volta (eq. (11.21)) $\underline{h} = \underline{H}(f) = 2H^+(f+f_0) = 2H(f_0)$ e dunque $\underline{y}(t) = \frac{1}{2} \cdot 2H(f_0) \cdot \underline{x}(t)$ ovvero, ponendo $\underline{x}(t_k) = s^k$ si ottiene $r^k = \underline{y}(t_k) = H(f_0) \cdot s^k$.

⁷Essendo n una v.a. complessa $n = n_c + jn_s$, otteniamo $\sigma_n^2 = E\{nn^*\} = E\{n_c^2 + n_s^2\} = 2\sigma_{n_c}^2 = 2\sigma_{n_s}^2$, in accordo con i risultati del § 14.1.3.

ora con \mathbf{s} il vettore n_T -dimensionale trasmesso ad un generico istante di simbolo⁸ da tutta la schiera di n_T antenne, presso le n_R antenne in ricezione otterremo al medesimo istante il vettore complesso di dimensione n_R

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{n} \quad (21.2)$$

con \mathbf{n} vettore gaussiano complesso ad n_R elementi. Come discusso al § 20.4.4 se ciascun valore $h_{ij} = \rho e^{j\varphi}$ è una v.a. gaussiana complessa a media nulla con parti reale ed immaginaria incorrelate e varianza σ_h^2 (⁹) il valore $\rho = |h_{ij}|$ del modulo di h_{ij} ha d.d.p. di Rayleigh

$$p(\rho) = \frac{\rho}{\sigma_h^2} \exp\left(-\frac{\rho^2}{2\sigma_h^2}\right)$$

mentre φ si distribuisce uniformemente tra $\pm\pi$, come avviene per collegamenti in assenza di visibilità e con molteplici cammini multipli statisticamente indipendenti.

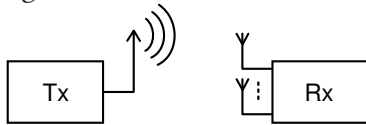
Come scopriremo strada facendo, molti risultati si basano sulla conoscenza dei valori di \mathbf{H} , stimati al ricevitore (§ 21.7.3.2) e la cui stima si ritiene valida per un periodo comparabile con il *tempo di coerenza* (pag. 688) del collegamento. In base a tali valori scopriremo come le proprietà di \mathbf{H} possano condizionare il livello di prestazione conseguibile. Ma iniziamo la trattazione con un caso *relativamente* semplice.

21.3 Diversità spaziale

L'esistenza contemporanea (e sulla stessa banda di frequenze) di più canali radio, differenziati per gli elementi di \mathbf{H} (21.2), offre innanzitutto l'opportunità di analizzare come sfruttare la *diversità spaziale* disponibile allo scopo di migliorare le prestazioni del collegamento, ovvero risparmiare potenza o migliorare la P_e , od un compromesso tra i due obiettivi. Discutiamo innanzitutto cosa si può fare quando solo uno dei due estremi del collegamento dispone di più di una antenna.

21.3.1 Ricevitore multi-antenna

Si tratta del caso indicato come SIMO ($n_T = 1$) che studiamo nel contesto di una trasmissione punto-punto oppure broadcast, in cui una unica antenna trasmette un segnale che viene ricevuto mediante $n_R > 1$ antenne, e dunque \mathbf{H} è un vettore ad



una colonna ed n_R righe, ovvero $\mathbf{H} = (h_1, h_2, \dots, h_{n_R})^T$. Nelle condizioni di fading di Rayleigh *piatto* se la separazione tra le antenne è sufficiente¹⁰ il valo-

⁸Di qui in poi sarà possibile riferirsi agli istanti di simbolo come all'“uso del canale”, ovvero pensando alla trasmissione di un simbolo \mathbf{s} n_T -dimensionale come ad *un singolo uso* del canale.

⁹ σ_h^2 dipende dalla somma delle intensità con cui le repliche del segnale trasmesso dall'antenna j giungono all'antenna i .

¹⁰Nel caso di un telefono *cellulare* sono presenti numerosi riflettori nelle vicinanze del ricevitore, producendo nei *downlink* fading incorrelati per distanze tra le antenne *dei mobili* di circa mezza lunghezza d'onda. Viceversa nel caso della *base station* fissa con cui il cellulare comunica, i cammini multipli *dell'uplink* hanno quasi tutti origine nei pressi del mobile, riducendo la gamma di angoli di incidenza dei raggi ricevuti, che iniziano ad essere indipendenti per distanze di decine di lunghezze d'onda: pertanto alla *base station* sono necessarie antenne molto più lontane tra loro.

re h_i relativo a ciascuna antenna è *incorrelato* da quello associato alle altre ovvero $E\{h_i h_j\} = 0$, e dunque *statisticamente indipendente* in virtù della gaussianità delle h_i . Pertanto se per una antenna si manifesta una forte attenuazione, ciò probabilmente non accade per le altre.

Sussistono tre possibili modi di sfruttare le n_R antenne di ricezione: preferire quella da cui si riceve più potenza, mediare equamente tra tutti i segnali ricevuti, oppure effettuare una somma *pesata*.

21.3.1.1 Selezione di diversità

Mostriamo innanzitutto come anche solo limitandosi a scegliere quale antenna utilizzare per la ricezione si riescano a conseguire risultati molto soddisfacenti. Consideriamo quindi un ricevitore per cui siano disponibili n_R *rami di diversità* indipendenti ed affetti da fading di Rayleigh, su ognuno dei quali (per ogni utilizzo del canale) si riceve un segnale con inviluppo complesso (vedi (21.1))

$$r_i = h_i s + n_i \quad (21.3)$$

con $i = 1, 2, \dots, n_R$, in cui s è il valore complesso che individua un punto della costellazione adottata, trasmesso con potenza¹¹ $\mathcal{P}_T = \mathcal{E}_s f_s$. I valori $\rho_i = |h_i|$ sono v.a. di Rayleigh con d.d.p. $p(\rho) = \frac{\rho}{\sigma_h^2} e^{-\rho^2/2\sigma_h^2}$, uguale per tutti i rami i ma con valori incorrelati ovvero $E\{\rho_i \rho_j\} = 0$ con $i \neq j$, mentre $n_i \in \mathcal{CN}\{0, \sigma_n^2\}$ è un campione dell'inviluppo complesso di un processo gaussiano bianco passabanda.

A differenza di un canale AWGN, il rapporto *SNR* che compete alla (21.3) non è un valore deterministico, ma una v.a. che indichiamo come *SNR istantaneo, per simbolo*¹²

$$\gamma_i = \frac{|h_i|^2 E\{s^2\}}{\sigma_n^2} = |h_i|^2 \frac{\mathcal{E}_s}{\sigma_n^2} \quad (21.4)$$

e che si distribuisce con la medesima d.d.p. di $|h_i|^2$, che è (vedi eq. (20.13)) una *esponenziale*; pertanto

$$p(\gamma_i) = \frac{1}{\Gamma} e^{-\gamma_i/\Gamma}$$

(con $\gamma_i \geq 0$), dove Γ è l'*SNR medio* valutabile come

$$\Gamma = E\{\gamma_i\} = E\{|h_i|^2\} \frac{\mathcal{E}_s}{\sigma_n^2} = 2\sigma_h^2 \frac{\mathcal{E}_s}{\sigma_n^2} \quad (21.5)$$

(vedi § 20.5.1) uguale per tutti i rami i nelle ipotesi poste. A questo punto abbiamo tutte le relazioni necessarie a valutare la probabilità che *un singolo* ramo abbia un γ_i inferiore ad un valore δ , pari a (vedi eq. (22.3) a pag. 767)

$$Pr\{\gamma_i \leq \delta\} = \int_0^\delta p(\gamma_i) d\gamma_i = 1 - \int_\delta^\infty \frac{1}{\Gamma} e^{-\gamma_i/\Gamma} d\gamma_i = 1 - e^{-\delta/\Gamma}$$

mentre la probabilità che *tutti* gli n_R rami indipendenti presentino *contemporaneamente*

¹¹Il valore \mathcal{E}_s individua l'*energia per simbolo*, e misura il valore di $E\{s^2\}$.

¹²*Istantaneo* perché dipende da h_i che in linea di principio può variare da istante ad istante; *per simbolo* perché $\frac{\mathcal{E}_s}{\sigma_n^2} = \frac{\mathcal{P}_s}{\mathcal{P}_n} \frac{1}{f_s}$.

$\gamma_i < \delta$ vale

$$Pr \{ \gamma_1, \gamma_2, \dots, \gamma_M \leq \delta \} = (1 - e^{-\delta/\Gamma})^{n_R}$$

che indichiamo come $P_{n_R}(\delta)$, da cui otteniamo la probabilità che *almeno uno* dei rami consegua $\gamma_i \geq \delta$ come

$$Pr \{ \gamma_i \geq \delta, \forall i \} = 1 - P_{n_R}(\delta) = 1 - (1 - e^{-\delta/\Gamma})^{n_R}$$

Esempio Consideriamo un ricevitore con quattro rami di diversità, ognuno affetto da fading di Rayleigh, e con un medesimo SNR medio Γ . Determinare la probabilità che l'SNR istantaneo γ_i di ciascun ramo si riduca contemporaneamente di 10 dB sotto il valor medio Γ , ossia che tutti i γ_i siano inferiori ad un valore δ tale che $\delta/\Gamma = 0.1$, e confrontare il risultato con il caso di un ricevitore senza diversità.

Risulta che $Pr \{ \gamma_1, \gamma_2, \gamma_3, \gamma_4 \leq \delta \} = P_4(\delta) = (1 - e^{-0.1})^4 = 8.2 \cdot 10^{-5}$, mentre per un sistema siso avremmo avuto $Pr \{ \gamma_i \leq \delta \} = P_1(\delta) = 1 - e^{-0.1} = 9.5 \cdot 10^{-2}$. Considerando i 10 dB di differenza come il valore di un margine oltre il quale il collegamento diviene troppo rumoroso, l'uso di quattro rami di diversità corrisponde ad un miglioramento della probabilità di fuori servizio di più di mille volte!

L'approccio della *selezione di diversità* è facilmente realizzabile in quanto coinvolge solamente il ricevitore, dove viene comparata la potenza del segnale in arrivo sulle diverse antenne, e quindi il segnale più forte è inviato al ricevitore, ne più ne meno come nello schema anticipato al § 20.3.3.1. Dato che la stima del livello di potenza di ricezione per ciascun ramo si basa su di una media temporale, la decisione non avviene in modo prettamente istantaneo; d'altra parte, è sufficiente che avvenga con tempi inferiori al *tempo di coerenza* eq. (20.29).

E' possibile mostrare¹³ che l'SNR medio Γ_{SD} per il ramo di volta in volta selezionato *migliora* all'aumentare del numero di antenne, anche se per ogni antenna aggiuntiva il miglioramento è sempre minore, risultando

$$\Gamma_{SD} = \Gamma \sum_{k=1}^{n_R} \frac{1}{k} \quad (21.6)$$

ovvero con un numero di 2, 3, 4 antenne si ottiene un fattore di miglioramento di 1.5, 1.83, 2.03... ma si può fare di meglio se vengono utilizzati *tutti* i rami in contemporanea, anziché uno solamente.

21.3.1.2 Combinazione di massimo rapporto - MRC

Riscriviamo la (21.3) espandendo il coefficiente h_i come $h_i = |h_i| e^{j\phi_i}$, ovvero

$$r_i = |h_i| e^{j\phi_i} s + n_i \quad (21.7)$$

con $|h_i|$ v.a. di Rayleigh e ϕ_i uniforme. La tecnica che stiamo per affrontare si basa sul calcolo di una combinazione lineare

$$\hat{r} = \sum_{i=1}^{n_R} |w_i| e^{j\phi_i} r_i \quad (21.8)$$

¹³Vedi D.G. Brennan, *Linear Diversity Combining Techniques*. Proc. IEEE, Vol. 91, N. 2, Feb 2003.

dei segnali ricevuti r_i mediante coefficienti complessi $w_i = |w_i|e^{-j\phi_i}$ ottenuti a partire da una stima¹⁴ dei valori h_i . Per quanto riguarda la fase ϕ_i questa viene scelta pari a $-\phi_i$, in modo da *annullare* il termine di fase in r_i e porsi nelle condizioni di *somma coerente*, dato che inserendo la (21.7) in (21.8) si ottiene

$$\begin{aligned}\hat{r} &= \sum_i |w_i| e^{-j\phi_i} r_i = \sum_i |w_i| e^{-j\phi_i} (|h_i| e^{j\phi_i} s + n_i) = \\ &= \sum_i |w_i| |h_i| s + \sum_i |w_i| e^{-j\phi_i} n_i = \sum_i |w_i| |h_i| s + \sum_i |w_i| n_i\end{aligned}\quad (21.9)$$

visto che la rotazione ϕ_i della v.a. complessa n_i non ne cambia la natura¹⁵.

La scelta dei valori $|w_i|$ avviene secondo il criterio di rendere massimo l'*SNR momentaneo*¹⁶ γ_{MR} di (21.9), da cui il nome di *maximal ratio combining* (MRC) della tecnica, in cui per *ratio* si intende l'*SNR*. A tale scopo osserviamo come alla componente di segnale di (21.9) compete una energia media (rispetto alla variabilità di s) $E\{(\sum_i |w_i| |h_i| s)^2\} = \mathcal{E}_s (\sum_i |w_i| |h_i|)^2$, mentre per la varianza del rumore si ottiene $E\{(\sum_i |w_i| n_i)^2\} = \sigma_n^2 \sum_i w_i^2$, essendo i campioni di rumore n_i statisticamente indipendenti: pertanto l'*SNR* di \hat{r} (21.9) risulta pari a

$$\gamma_{MR} = \frac{\mathcal{E}_s (\sum_{i=1}^{n_R} |w_i| |h_i|)^2}{\sigma_n^2 \sum_{i=1}^{n_R} w_i^2} = \frac{\mathcal{E}_s |\mathbf{w}_m^T \cdot \mathbf{h}_m|^2}{\sigma_n^2 \mathbf{w}_m^T \cdot \mathbf{w}_m}\quad (21.10)$$

dove con $\mathbf{w}_m = (|w_1|, \dots, |w_{n_R}|)^T$ si è indicato il vettore del *modulo* dei coefficienti \mathbf{w} e con $\mathbf{h}_m = (|h_1|, \dots, |h_{n_R}|)^T$ quello del guadagno di ampiezza dei rami, in modo che la (21.10) sia espressa nei termini di *prodotto scalare* tra vettori. Con tale formalismo il massimo valore per γ_{MR} si ottiene applicando la disuguaglianza di Schwartz (§ 2.4.3) che asserisce

$$|\mathbf{w}_m^T \cdot \mathbf{h}_m|^2 \leq (\mathbf{h}_m^T \cdot \mathbf{h}_m) \cdot (\mathbf{w}_m^T \cdot \mathbf{w}_m)$$

con il segno di uguale solo quando $\mathbf{w}_m = \alpha \mathbf{h}_m$, ovvero i vettori sono *paralleli*. Con tale scelta la (21.10) diviene

$$\gamma_{MR} = \frac{\mathcal{E}_s |\alpha \mathbf{h}_m^T \cdot \mathbf{h}_m|^2}{\sigma_n^2 \alpha^2 \mathbf{h}_m^T \cdot \mathbf{h}_m} = \frac{\mathcal{E}_s (\sum_{i=1}^{n_R} |h_i|^2)^2}{\sigma_n^2 \sum_{i=1}^{n_R} |h_i|^2} = \frac{\mathcal{E}_s}{\sigma_n^2} \sum_{i=1}^{n_R} |h_i|^2 = \sum_{i=1}^{n_R} \gamma_i\quad (21.11)$$

in cui l'ultima eguaglianza si basa sulla (21.4). L'*SNR* complessivo è dunque pari alla *somma* degli *SNR* dei singoli rami, e può così conseguire valori *accettabili* anche se nessuno dei rami lo ottiene individualmente.

In virtù dell'ipotesi che i valori h_i siano statisticamente indipendenti, dalla (21.11) otteniamo l'*SNR medio* Γ_{MR} a partire dai Γ_i dei singoli rami (eq. (21.5)) come

$$\Gamma_{MR} = \sum_{i=1}^{n_R} \Gamma_i\quad (21.12)$$

¹⁴Valida per un intervallo temporale minore del tempo di coerenza del canale.

¹⁵E' proprio in base a questa considerazione che il vettore gaussiano complesso costituito da campioni dell'involuppo complesso di un processo di rumore passa banda prende il nome di *processo circolare*.

¹⁶Nel senso che dato che (come stiamo per vedere) γ_{MR} dipende dagli h_i che sono v.a., è una v.a. anch'esso. Ma allo stesso tempo gli h_i sono considerati costanti per tutto il tempo di coerenza, ed altrettanto accade a γ_{MR} .

Se ogni ramo presenta il medesimo SNR medio Γ_i si può confrontare (21.12) con (21.6), ed osservare che ora Γ_{MR} aumenta *linearmente* con il numero di antenne, ovvero migliora di 3 dB ad ogni raddoppio di n_R . Se invece (essendo gli $|h_i|$ v.a. indipendenti) i Γ_i sono differenti tra loro il miglioramento è inferiore, ma Γ_{MR} risulta comunque sempre *migliore del miglior* Γ_i .

Per poter essere operativi non resta che effettuare una scelta ragionata per il coefficiente di proporzionalità α introdotto dalla disuguaglianza di Schwartz, e dal punto di vista teorico di nessun impatto. Includendo anche il termine di rifasamento, nella pratica conviene scegliere dei pesi

$$w_i = \frac{h_i^*}{\sum_{i=1}^{n_R} |h_i|^2} \quad (21.13)$$

in modo da rendere \hat{r} uno stimatore non polarizzato (§ 6.6.3): essendo infatti i termini n_i a media nulla, il valore atteso della (21.9) assume la forma

$$E\{\hat{r}\} = \sum_i w_i h_i s = \sum_i \frac{h_i^*}{\sum_i |h_i|^2} h_i s = \frac{\sum_i |h_i|^2}{\sum_i |h_i|^2} s = s \quad (21.14)$$

e dunque la tecnica MRC non altera *la dimensione* della costellazione \mathcal{A} a cui s appartiene, e la decisione sul simbolo trasmesso può basarsi sul criterio di massima verosimiglianza

$$\hat{s} = \arg \max_{s \in \mathcal{A}} \text{Prob}\{\hat{r}/s\} \quad \text{ovvero} \quad \hat{s} = \arg \min_{s \in \mathcal{A}} (s - \hat{r})^2$$

dove la seconda espressione individua una minima distanza euclidea¹⁷ conseguenza del passaggio ai logaritmi e della gaussianità di n .

Probabilità di errore Con una procedura di media statistica analoga¹⁸ a quella illustrata al § 20.5.1 si può arrivare a valutare la P_e^{bit} nel caso di costellazione BPSK ed in funzione del numero di ricevitori n_R e dell' SNR medio ricevuta Γ (lo stesso per tutte le antenne), che per valori di $E_b/N_0 \geq 10$ è bene approssimata come

$$P_{e, MRC, BPSK}^{bit} \simeq \binom{2n_R - 1}{n_R} \left[\frac{1}{2} \left(1 - \sqrt{\frac{\Gamma}{1 + \Gamma}} \right) \right]^{n_R} \quad (21.15)$$

da confrontare con l'espressione (20.35) ottenuta per un sistema di trasmissione SISO, confronto possibile anche mediante la figura a pagina seguente, in cui oltre al caso del fading SISO compare anche la curva di P_e per il caso AWGN in cui non si verifica fading. Evidentemente il miglioramento è notevole!

Svolgiamo ora due ultime considerazioni a partire dalla (21.13). La prima è che nel caso di una potenza di rumore σ_n^2 uguale su tutti i rami, si può dire che la combinazione lineare attuata dalla (21.8) dà maggior peso ai rami con l' SNR più elevato, dato che questo dipende da $|h_i|^2$. La seconda considerazione è che la scelta (21.13) rivela una analogia

¹⁷Con l'accortezza che per grandezze complesse il quadrato si valuta come prodotto per il coniugato, ossia $z^2 = z \cdot z^* = (\Re\{z\})^2 + (\Im\{z\})^2$.

¹⁸Nel senso di pervenire ad una espressione di P_e (γ_{MR}) e poi calcolarne il valore atteso rispetto alla v.a. γ_{MR} , che ora non è più esponenziale, bensì chi quadro (§ 6.6.5) con $2n_R$ gradi di libertà, essendo le h_i v.a. gaussiane complesse.

tra questa tecnica e quella del filtro adattato (§ 7.6) che pure trova motivazione nella massimizzazione dell' SNR , ottenuta in quel caso grazie ad una risposta in frequenza che è *il coniugato* del segnale da rivelare. Il paragone non è casuale dato che come per il filtro adattato, a supporto della scelta $\mathbf{w} \propto \mathbf{h}^*$ sussiste una argomentazione basata sulla disuguaglianza di Schwartz (§ 2.4.3)¹⁹; in altre parole, ciò che il filtro adattato compie nel dominio della frequenza, viene attuato dal ricevitore MRC nel dominio spaziale.

21.3.1.3 Combinazione equal gain

Torniamo all'ipotesi che rende massima la (21.12) ovvero quando i diversi rami subiscono un fading di intensità simile, cioè $|h_i| \approx |h| \forall i$. In tal caso i pesi w_i ottimi per ottenere \hat{r} tramite la (21.8) hanno tutti lo stesso modulo²⁰ pari a $|w_i| = \frac{1}{n_R |h_i|}$, in modo che la (21.14) ora è espressa come

$$E\{\hat{r}\} = \sum_i |w_i| |h_i| s = \sum_i \frac{1}{n_R |h_i|} |h_i| s = \frac{n_R}{n_R} s = s \quad (21.16)$$

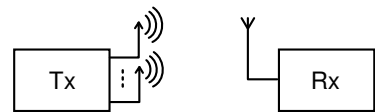
e quindi ogni antenna riceve lo stesso SNR medio Γ , conseguendo un SNR complessivo

$$\Gamma_{EG} = n_R \Gamma \quad (21.17)$$

ovvero n_R volte quello di ogni singolo ramo. In tal modo il ricevitore risulta molto semplificato, al punto da preferire a volte l'uso di pesi tutti uguali²¹ anche in presenza di coefficienti h_i diversi tra loro, nel qual caso il valore (21.17) non viene raggiunto. Tale scelta viene indicata come metodo *equal gain*, ed offre risultati solo di poco inferiori a quelli di *massimo rapporto*, e comunque migliori del metodo *a selezione*.

21.3.2 Trasmettitore multiantenna

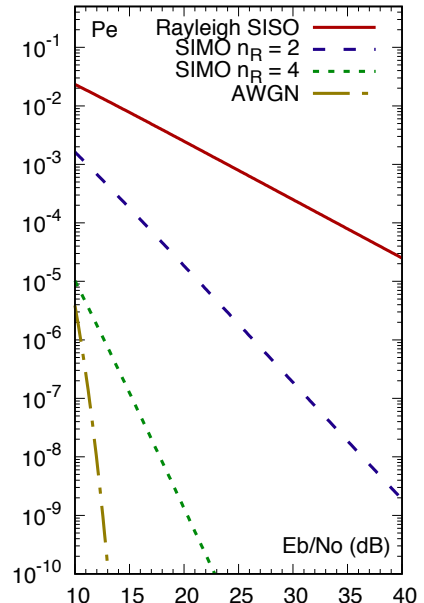
Ovviamente la presenza di una unica antenna ricevente impedisce di attuare strategie di scelta o di combinazione delle fonti di *diversità spaziale* disponibili. Con un numero $n_T > 1$ di trasmettenti i



¹⁹Vedi ad es. il riferimento della nota 13; per lo stesso risultato sussiste inoltre anche un'argomentazione basata sugli autovettori della matrice $\mathbf{h} \cdot \mathbf{h}^\dagger$ (l'apice \dagger indica il trasposto coniugato), vedi B. Holter, G.E. Oien, *The Optimal Weights of a Maximum Ratio Combiner using an Eigenfilter Approach*, Proc. 5th IEEE Nordic Signal Processing Symposium, Hurtigruten, 4-6 October 2002, reperibile presso <https://www.ux.uis.no/norsig/norsig2002/Proceedings/papers/cr1099.pdf>, e che esamina anche il caso in cui il rumore si presenti con potenze differenti sui diversi rami.

²⁰Rimane infatti la necessità di *rifasare* i rami per ottenere una somma *coerente*.

²¹Dove per uguali si intende $|w_i| = 1/n_R$ senza riguardo per $|h_i|$, causando nella (21.16) l'insorgenza di un fattore moltiplicativo, che può essere ignorato nel caso di una modulazione PSK ovvero con costellazione circolare.



simboli che giungono all'unico ricevente

$$r = \sum_{i=1}^{n_T} h_i s + n_i \quad (21.18)$$

risultano infatti *già combinati*. Ma come vedremo la conoscenza da parte *del ricevente* dei coefficienti complessi²² h_i (che costituiscono l'unica riga della matrice \mathbf{H} definita al § 21.2) permette comunque di sfruttare la diversità spaziale. Il modo con cui ciò avviene si basa su di una particolare tecnica di codifica di canale, nota come codifica *spazio-tempo* (STC) in quanto l'informazione da trasmettere viene *diluata* (ridondata) oltre che *nel tempo* (la sola dimensione possibile per un caso SISO²³) anche *sulle antenne* (ovvero nello spazio), riuscendo a ottenere un *guadagno di diversità*.

Una seconda cosa degna di nota in questo passaggio è che in un canale bidirezionale²⁴ trasmettitore e ricevitore si scambiano continuamente di ruolo, dunque se anche in modalità MISO si riesce ad ottenere un miglioramento comparabile al caso SIMO, è sufficiente che solo una delle parti in comunicazione - tipicamente la stazione radio base o l'access point WiFi²⁵ - sia dotata di più antenne.

21.3.2.1 Codice a traliccio spazio - tempo

Un primo approccio alla realizzazione di un STC si è basato sulla *modulazione a traliccio* (pag. 537) e per questo indicato come *space time trellis code* o STTC. In questa tecnica i simboli codificati sono ripartiti tra le antenne, mentre dal lato ricevente la decodifica avviene eseguendo un algoritmo di Viterbi *vettoriale*. Oltre ad un guadagno *di diversità* proporzionale al numero delle antenne, il metodo offre anche un guadagno *di codifica* legato al numero di stati del codice a traliccio sottostante (a spese della complessità di decodifica), cosa che non avviene con la tecnica STBC descritta appresso, ma che è attualmente favorita nonostante questa carenza, più che compensata dal vantaggio di poter svolgere la decodifica mediante operazioni assai più semplici. Scegliamo pertanto di non addentrarci nello studio degli STTC.

21.3.2.2 Codice a blocco spazio - tempo

A differenza dei codici di canale esaminati al § 17.4, in cui le quantità in ingresso ed in uscita erano semplici bit, uno STBC opera *a valle* della codifica di simbolo (indicata anche come *symbol mapper*), che ad un blocco di M bit fa corrispondere un valore complesso s_i che individua un punto di costellazione \mathcal{A} tra $L = 2^M$. Lo STBC osserva un numero k di tali simboli, e fa corrispondere ad essi una codeword che consiste in una matrice \mathbf{C} i cui elementi $c_{ij} \in \mathcal{A}$ identificano i simboli trasmessi dall'antenna

²² Aleatori a componenti gaussiane a media nulla etc etc...

²³ In realtà ad es. nella modulazione COFDM (§ 16.8.10) la ridondanza viene distribuita anche sulle diverse sottoportanti, dunque *in frequenza*. Ma ci torniamo al § 21.7.2.

²⁴ Qualora la trasmissione nelle due direzioni avvenga sulla stessa portante è necessario che le parti si alternino nei ruoli (*time-duplex*), mentre invece possono trasmettere e ricevere allo stesso tempo se si adottano frequenze differenti (*frequency-duplex*).

²⁵ Anche perché le ridotte dimensioni dei device mobili rendono problematico realizzare antenne sufficientemente distanziate.

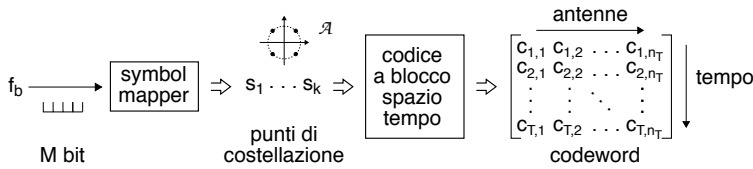


Figura 21.2: Generazione di un codice a blocco spazio-tempo

$j = 1, \dots, n_T$ all'istante $i = 1, 2, \dots, T$. Il tasso di codifica in questo caso è calcolato come il rapporto tra il numero k di simboli s_i ed il numero di istanti T impiegati per trasmetterli, ovvero

$$R_c = k/T \tag{21.19}$$

21.3.2.3 Codice di Alamouti

Esponiamo il metodo partendo dalla sua definizione iniziale dovuta ad *Alamouti*²⁶, relativa al caso di $n_T = 2$ antenne di trasmissione, che in due istanti di simbolo consecutivi trasmettono un totale di quattro simboli di canale c_{ij} legati ai due *di sorgente* s_1 ed s_2 secondo lo schema²⁷ illustrato in tabella 21.1. Le righe corrispondono ai due istanti di simbolo: durante il primo istante le antenne trasmettono rispettivamente s_1 la prima ed s_1 la seconda, mentre al secondo istante la prima trasmette $-s_2^*$ e la seconda s_1^* . Pertanto il codice agisce attraverso le due dimensioni della tabella, in cui le righe individuano la diversità *temporale*, mentre le colonne quella *spaziale*. Il calcolo del tasso di codifica fornisce $R_c = k/T = 2/2 = 1$, e dunque a differenza dei codici a blocco *tradizionali* non si verifica nessun aumento di banda. Vedremo che il codice di Alamouti è l'unico STBC a godere di tale proprietà, risultando in generale $R_c < 1$.

	antenna 1	antenna 2
tempo t	s_1	s_2
tempo $t + T_s$	$-s_2^*$	s_1^*

Tabella 21.1: Codice di Alamouti

Ricezione Applicando la (21.18) al segnale ricevuto nei due istanti, otteniamo

$$\begin{aligned} r_1 &= h_1 s_1 + h_2 s_2 + n_1 \\ r_2 &= -h_1 s_2^* + h_2 s_1^* + n_2 \end{aligned} \tag{21.20}$$

dove n_i è il solito rumore gaussiano complesso. Dalla seconda relazione si calcola

$$r_2^* = -h_1^* s_2 + h_2^* s_1 + n_2^* \tag{21.21}$$

e quindi, purché il ricevitore conosca i coefficienti h_1 ed h_2 , i due valori r_1 ed r_2^* sono *combinati* in modo *simile* a quanto visto per l'MRC, questa volta allo scopo di ottenere

²⁶S.M. Alamouti, *A simple transmit diversity technique for wireless communications*, IEEE Journal on Selected Areas in Communications, Oct 1998. Reperibile presso

https://mast.queensu.ca/~fady/Math800/papers/Alamouti_JSAC98.pdf

²⁷Dunque le codeword C sono costituite dai simboli di sorgente stessi, una sorta di codice a ripetizione, se non fosse che ora ci sono anche i coniugati, e quel segno cambiato... in effetti la scelta di tab. 21.1 è un caso particolare di una regola più generale, ovvero costruire gli elementi di C come combinazione lineare dei simboli s_i , e dei loro coniugati.

una stima \hat{s}_1 e \hat{s}_2 per *entrambi* i simboli di sorgente, calcolando

$$\begin{aligned}\hat{s}_1 &= h_1^* r_1 + h_2 r_2^* \\ \hat{s}_2 &= h_2^* r_1 - h_1 r_2^*\end{aligned}\quad (21.22)$$

Sostituendo infatti le (21.20) e (21.21) in (21.22) si ottiene

$$\begin{aligned}\hat{s}_1 &= h_1^* (h_1 s_1 + h_2 s_2 + n_1) + h_2 (-h_1^* s_2 + h_2^* s_1 + n_2^*) \\ \hat{s}_2 &= h_2^* (h_1 s_1 + h_2 s_2 + n_1) - h_1 (-h_1^* s_2 + h_2^* s_1 + n_2^*)\end{aligned}$$

da cui

$$\begin{aligned}\hat{s}_1 &= (|h_1|^2 + |h_2|^2) s_1 + h_1^* n_1 + h_2 n_2^* \\ \hat{s}_2 &= (|h_1|^2 + |h_2|^2) s_2 + h_2^* n_1 - h_1 n_2^*\end{aligned}\quad (21.23)$$

dove i termini di interferenza tra simboli sono *scomparsi*, e si manifesta un *rinforzo* delle ampiezze dei simboli di sorgente pari alla somma dei quadrati dei guadagni di Rayleigh, come anche avveniva per l'MRC, e che indichiamo con $\alpha = |h_1|^2 + |h_2|^2$.

Prestazioni Se dividiamo le (21.23) per α otteniamo anche in questo caso delle stime non polarizzate dei simboli di sorgente

$$\begin{aligned}\tilde{s}_1 &= \frac{1}{\alpha} \hat{s}_1 = s_1 + \tilde{n}_1 \\ \tilde{s}_2 &= \frac{1}{\alpha} \hat{s}_2 = s_2 + \tilde{n}_2\end{aligned}\quad (21.24)$$

in cui \tilde{n}_1 e \tilde{n}_2 sono v.a. gaussiane complesse con parti reale ed immaginaria incorrelate, a media nulla e varianza²⁸ $\sigma_{\tilde{n}}^2 = E \{ \tilde{n}_i \tilde{n}_i^* \} = \sigma_n^2 / (|h_1|^2 + |h_2|^2)$ in cui $\sigma_n^2 = E \{ n_i n_i^* \}$ è la potenza del rumore in ingresso al ricevitore. Il criterio di decodifica è come di consueto quello di massima verosimiglianza, che si riduce alla regola di decisione

$$s_i^\diamond = \arg \min_{s \in \mathcal{A}} (s - \tilde{s}_i)^2 \quad \text{con} \quad i = 1, 2 \quad (21.25)$$

ovvero ad ognuno dei due istanti i si decide per la ricezione del simbolo s_i^\diamond (che figura nella costellazione \mathcal{A}) *più vicino*²⁹ al simbolo (valore complesso) normalizzato \tilde{s}_i espresso dalla (21.24), che possiamo riguardare come un valore di decodifica *soft*. In alternativa, se i simboli trasmessi (di sorgente) provengono da uno stadio di codifica di canale *esterno* per il quale è prevista una decodifica *soft*, i valori \tilde{s}_i possono essere passati a quest'ultima *così come sono*.

La probabilità con cui la regola (21.25) dà esito errato è naturalmente legata in modo inverso al rapporto SNR *istantaneo* γ per le variabili \tilde{s}_i (21.24) su cui si effettua

²⁸Svolgendo infatti i calcoli per una di esse, ad esempio $\tilde{n}_1 = \frac{1}{\alpha} (h_1^* n_1 + h_2 n_2^*)$, in virtù dell'incorrelazione tra n_1 ed n_2 si ottiene

$$\begin{aligned}\sigma_{\tilde{n}_1}^2 &= E \{ \tilde{n}_1 \tilde{n}_1^* \} = \frac{1}{\alpha^2} E \{ (h_1^* n_1 + h_2 n_2^*) (h_1 n_1^* + h_2^* n_2) \} = \\ &= \frac{1}{\alpha^2} [|h_1|^2 E \{ n_1 n_1^* \} + |h_2|^2 E \{ n_2 n_2^* \} + h_1^* h_2^* E \{ n_1 n_2 \} + h_1 h_2 E \{ n_2^* n_1^* \}] = \\ &= \frac{1}{\alpha^2} [|h_1|^2 \sigma_{n_1}^2 + |h_2|^2 \sigma_{n_2}^2] = \frac{1}{\alpha^2} [\alpha \sigma_n^2] = \sigma_n^2 \frac{1}{|h_1|^2 + |h_2|^2}\end{aligned}$$

²⁹Si ribadisce che per grandezze complesse il modulo quadro si calcola come prodotto per il coniugato, dunque la (21.25) diviene $(s - \tilde{s}_i)^2 = (s - \tilde{s}_i) (s^* - \tilde{s}_i^*) = s s^* + \tilde{s}_i \tilde{s}_i^* - s \tilde{s}_i^* - \tilde{s}_i s^*$, da valutare per ogni $s \in \mathcal{A}$.

la decisione, che risulta pari a

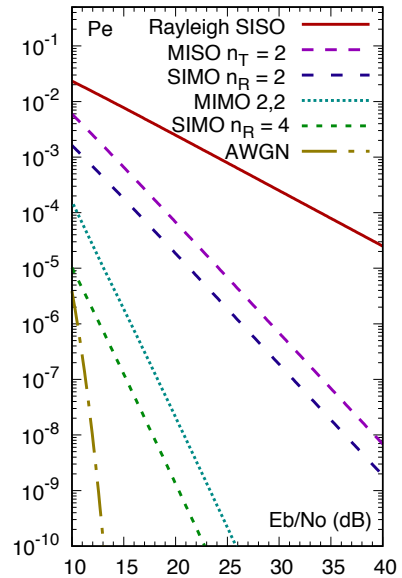
$$\gamma_{Ala} = \frac{E\{s_i s_i^*\}}{\sigma_n^2} = \frac{\mathcal{E}_s}{\sigma_n^2} (|h_1|^2 + |h_2|^2)$$

fornendo un *guadagno di diversità* identico a quello di un ricevitore SIMO-MRC (eq. (21.11)), tranne che... molto spesso le specifiche relative alla massima potenza irradiata impongono di suddividere la stessa in parti uguali tra le antenne di trasmissione. Ciò significa che anche l'energia per simbolo \mathcal{E}_s della costellazione adottata risulta (con $n_T = 2$) dimezzata, cosicché il sistema MISO-STBC con $n_T = 2$ subisce una penalizzazione di 3 dB³⁰ rispetto alle prestazioni di SIMO-MRC con $n_R = 2$.

Il legame esatto tra la P_e legata all'uso della (21.25) e l'SNR medio

$$\Gamma = E\{\gamma_{Ala}\} = \frac{\mathcal{E}_s}{\sigma_n^2} E\{|h_1|^2 + |h_2|^2\}$$

dipende dal tipo di modulazione, ma per il caso BPSK la figura a lato riporta il risultato di simulazioni messe a confronto con le curve (pag. 707) dei casi SISO, SIMO ed AWGN, in modo da poter apprezzare l'effetto della perdita di 3 dB prima indicata sul risultato finale, poca cosa rispetto al miglioramento comunque conseguito. La curva indicata in figura come MIMO 2,2 si riferisce al caso che affrontiamo appresso.



21.3.2.4 Ricezione multiantenna di un codice di Alamouti

Eccoci dunque arrivati alla prima configurazione propriamente MIMO: vediamo come ricevere una trasmissione MISO-STBC mediante una schiera di n_R antenne di ricezione, conseguendo un ordine di diversità *massimo* pari a $2n_R$, ossia quello ottenuto con due antenne di trasmissione moltiplicato per n_R .

Il trasmettitore non modifica il suo operato, ed invia la codifica espressa in tabella 21.1 mediante le sue due antenne. Facciamo dapprima il caso di adottare $n_R = 2$, ricevendo così un totale di *quattro* diversi segnali

$$\begin{aligned} r_{11} &= h_{11}s_1 + h_{12}s_2 + n_1 \\ r_{21} &= h_{21}s_1 + h_{22}s_2 + n_2 \\ r_{12} &= -h_{11}s_2^* + h_{12}s_1^* + n_3 \\ r_{22} &= -h_{21}s_2^* + h_{22}s_1^* + n_4 \end{aligned} \quad (21.26)$$

in cui r_{ik} è il segnale ricevuto dall'antenna $i = 1, 2$ all'istante $k = 1, 2$, ed h_{ij} è il

³⁰Questi 3 dB di differenza sono da interpretare come un guadagno di *array* legato al disporre di due antenne di ricezione, per cui in pratica viene ricevuta *il doppio* della potenza che si riceverebbe con una sola antenna, mentre in trasmissione ciò non si verifica, per la limitazione sulla potenza massima trasmessa.

coefficiente complesso del canale radio tra le antenne j e i . I valori r_{ik} delle (21.26) vengono quindi combinati per formare le grandezze di decisione

$$\begin{aligned}\hat{s}_1 &= h_{11}^* r_{11} + h_{12} r_{12}^* + h_{21}^* r_{21} + h_{22} r_{22}^* \\ \hat{s}_2 &= h_{12}^* r_{11} - h_{11} r_{12}^* + h_{22}^* r_{21} - h_{21} r_{22}^*\end{aligned}\quad (21.27)$$

e sostituendo le (21.26) nelle (21.27) dopo alcuni passaggi si ottiene

$$\begin{aligned}\hat{s}_1 &= (|h_{11}|^2 + |h_{12}|^2 + |h_{21}|^2 + |h_{22}|^2) s_1 + h_{11}^* n_1 + h_{12} n_3^* + h_{21}^* n_2 + h_{22} n_4^* \\ \hat{s}_2 &= (|h_{11}|^2 + |h_{12}|^2 + |h_{21}|^2 + |h_{22}|^2) s_2 + h_{12}^* n_1 - h_{11} n_3^* + h_{22}^* n_2 - h_{21} n_4^*\end{aligned}\quad (21.28)$$

riproducendo quindi la situazione già osservata alle (21.23), con la differenza che ora il *rinforzo* per le ampiezze dei simboli di sorgente vale $\alpha = |h_{11}|^2 + |h_{12}|^2 + |h_{21}|^2 + |h_{22}|^2$, ottenendo un'ordine di diversità complessivo che è *il prodotto* degli ordini ai due estremi multiantenna.

Gli stessi risultati sono facilmente estensibili ad un numero $n_R > 2$ di antenne riceventi, dato che ponendo le (21.26) nella forma

$$\begin{aligned}r_{i1} &= h_{i1} s_1 + h_{i2} s_2 + n_{i1} \\ r_{i2} &= -h_{i1} s_2^* + h_{i2} s_1^* + n_{i2}\end{aligned}\quad (21.29)$$

con $i = 1, 2, \dots, n_R$, le (21.27) si estendono come

$$\begin{aligned}\hat{s}_1 &= \sum_{i=1}^{n_R} (h_{i1}^* r_{i1} + h_{i2} r_{i2}^*) \\ \hat{s}_2 &= \sum_{i=1}^{n_R} (h_{i2}^* r_{i1} - h_{i1} r_{i2}^*)\end{aligned}\quad (21.30)$$

ed al posto delle (21.28) otteniamo

$$\begin{aligned}\hat{s}_1 &= \sum_{i=1}^{n_R} \left[(|h_{i1}|^2 + |h_{i2}|^2) s_1 + h_{i1}^* n_{i1} + h_{i2} n_{i2}^* \right] \\ \hat{s}_2 &= \sum_{i=1}^{n_R} \left[(|h_{i1}|^2 + |h_{i2}|^2) s_2 + h_{i2}^* n_{i1} - h_{i1} n_{i2}^* \right]\end{aligned}\quad (21.31)$$

Ortogonalità Mostriamo come i risultati fin qui ottenuti siano una diretta conseguenza del fatto che le righe della matrice delle codeword $\mathbf{C} = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix}$ di Alamouti sono *ortogonali* tra loro, ossia il prodotto scalare tra due diverse righe è nullo, e dunque

$$\begin{aligned}\mathbf{C}\mathbf{C}^\dagger &= \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix} \begin{bmatrix} s_1^* & -s_2 \\ s_2^* & s_1 \end{bmatrix} = \begin{bmatrix} s_1 s_1^* + s_1 s_2^* & -s_1 s_2 + s_1 s_2 \\ -s_1^* s_2^* + s_1^* s_2^* & s_2 s_2^* + s_1 s_1^* \end{bmatrix} = \\ &= \begin{bmatrix} |s_1|^2 + |s_2|^2 & 0 \\ 0 & |s_1|^2 + |s_2|^2 \end{bmatrix}\end{aligned}$$

in cui l'operatore † è noto come *Hermitiano* ed esegue il coniugato della trasposta della matrice³¹, equivalente per le matrici complesse dell'operazione di coniugazione valida per gli scalari. E' stato dimostrato che il codice di Alamouti è l'unico STBC a simboli complessi con codeword ortogonali.

Esprimendo ora la (21.29) in forma matriciale come

$$\begin{bmatrix} r_{i1} \\ r_{i2}^* \end{bmatrix} = \begin{bmatrix} h_{i1} & h_{i2} \\ h_{i2}^* & -h_{i1}^* \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} n_{i1} \\ n_{i2}^* \end{bmatrix} \quad \text{ovvero} \quad \mathbf{r}_i = \mathbf{G}_i \mathbf{s} + \mathbf{n}_i$$

³¹Indicato anche come operatore *aggiunto*, mentre \mathbf{A}^\dagger è detta *matrice aggiunta* di \mathbf{A} .

osserviamo come anche la matrice $\mathbf{G}_i = \begin{bmatrix} h_{i1} & h_{i2} \\ h_{i2}^* & -h_{i1}^* \end{bmatrix}$ sia *ortogonale*, per qualunque valore dei coefficienti h , e dunque $\mathbf{G}_i \mathbf{G}_i^\dagger = \alpha_i \mathbf{I}_2$ in cui \mathbf{I}_2 è la matrice identità 2×2 ed $\alpha_i = |h_{i1}|^2 + |h_{i2}|^2$. Pre-moltiplicando il vettore ricevuto dalla i -esima antenna \mathbf{r}_i per $\mathbf{G}_i^\dagger = \begin{bmatrix} h_{i1}^* & h_{i2} \\ h_{i2}^* & -h_{i1} \end{bmatrix}$, come avviene per le (21.22), otteniamo quindi

$$\tilde{\mathbf{r}}_i = \mathbf{G}_i^\dagger \mathbf{r}_i = \mathbf{G}_i^\dagger \mathbf{G}_i \mathbf{s} + \mathbf{G}_i^\dagger \mathbf{n}_i = \alpha_i \mathbf{s} + \tilde{\mathbf{n}}_i$$

dove $\tilde{\mathbf{n}}_i$ è ancora gaussiano complesso a media nulla e componenti incorrelate. A questo punto la decisione di massima verosimiglianza per l'antenna i assume la forma

$$\mathbf{s}^\diamond = \arg \min_{\mathbf{s} \in \mathcal{A}^2} (\tilde{\mathbf{r}}_i - \alpha_i \mathbf{s}) (\tilde{\mathbf{r}}_i - \alpha_i \mathbf{s})^* \quad (21.32)$$

ma essendo come osservato $\tilde{\mathbf{n}}_i = \tilde{\mathbf{r}}_i - \alpha_i \mathbf{s}$ a componenti incorrelate, la (21.32) si scompone in due minimizzazioni indipendenti, come espresso dalle (21.25) e (21.31).

21.3.3 Prestazioni limite

Abbandoniamo gli sviluppi ottenuti per il codice di Alamouti per tornare al caso più generale, e definire uno STBC come una procedura per ottenere una codeword definita da una matrice ad elementi complessi

$$\mathbf{C}^j = \begin{bmatrix} c_{1,1}^j & c_{1,2}^j & \cdots & c_{1,n_T}^j \\ c_{2,1}^j & c_{2,2}^j & \cdots & c_{2,n_T}^j \\ \vdots & \vdots & \ddots & \vdots \\ c_{T,1}^j & c_{T,2}^j & \cdots & c_{T,n_T}^j \end{bmatrix}$$

con $j = 1, 2, \dots, 2^{kM}$ in corrispondenza di ciascuna delle altrettante possibili sequenze di k simboli complessi a 2^M valori, come descritto al § 21.3.2.2, da trasmettere da parte di n_T antenne in $T \geq n_T$ istanti e ricevere mediante n_R antenne, cercando di individuare quali siano i fattori principali che concorrono alle prestazioni del codice, in termini di probabilità di errore.

A fronte di sviluppi analitici che tralasciamo³², si dimostra che la probabilità di decidere erroneamente per la codeword \mathbf{C}^i quando viene trasmessa \mathbf{C}^j per valori elevati di SNR medio Γ non supera un valore

$$\text{Prob} \{ \mathbf{C}^j \rightarrow \mathbf{C}^i \} \leq \frac{1}{(G_c \Gamma \cdot 1/4)^{G_d}} \quad (21.33)$$

in cui $G_d = r \cdot n_R$ esprime il *guadagno di diversità* che dipende oltre che da n_R anche dal *rango*³³ r della matrice $\mathbf{A} = (\mathbf{C}^i - \mathbf{C}^j)^\dagger (\mathbf{C}^i - \mathbf{C}^j)$, mentre $G_c = \sqrt{\prod_{n=1}^r \lambda_n}$ misura il *guadagno del codice* e dipende dagli autovalori λ_n diversi da zero della matrice \mathbf{A} .

³²Che ho trovato accennati, con i dovuti rimandi, su E. KROUK, S. SEMENOV, *Modulation and coding techniques in wireless communications*, 2011 John Wiley & Sons Ltd.

³³Il rango di una matrice quadrata $\mathbf{A} n \times n$ è definito come il numero delle sue righe (o colonne) linearmente indipendenti, ma è anche uguale al numero di autovalori diversi da zero, dove gli autovalori sono gli zeri del polinomio caratteristico definito come $\det(\mathbf{A} - \lambda \mathbf{I}_n)$

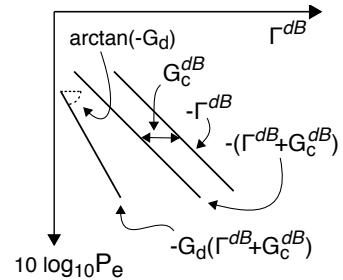
Pertanto il limite superiore (21.33) per la P_e può essere ridotto aumentando sia G_c che G_d , e se per tutte le coppie di codeword i, j la matrice A è a rango pieno (pari ad n_T), allora l'ordine di diversità conseguito dal codice è pari a $n_T \cdot n_R$.

Esempio Il simbolo vettoriale $\mathbf{s} = [s_1, s_2]^T$ è codificato secondo Alamouti come $C = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix}$, mentre un diverso input $\mathbf{s}' = [s'_1, s'_2]^T$ come $C' = \begin{bmatrix} s'_1 & s'_2 \\ -s_2'^* & s_1'^* \end{bmatrix}$, dunque $C - C' = \begin{bmatrix} s_1 - s'_1 & s_2 - s'_2 \\ -s_2^* + s_2'^* & s_1^* - s_1'^* \end{bmatrix}$ il cui determinante vale $|s_1 - s'_1|^2 + |s_2 - s'_2|^2 \neq 0$ per qualsiasi $\mathbf{s} \neq \mathbf{s}'$. Pertanto la matrice A ha sempre rango pieno³⁴, ed il codice offre un ordine di diversità pari a $n_T \cdot n_R$.

Prestazioni asintotiche Prima di procedere svolgiamo una interessante considerazione sulla legge di dipendenza della P_e a simbolo da G_c e G_d , come espressa dalla (21.33), che per SNR elevato possiamo generalizzare come $P_e \propto (G_c \Gamma)^{-G_d}$ o, esprimendo le grandezze in dB

$$10 \log_{10} P_e \propto 10 \log_{10} (G_c \Gamma)^{-G_d} = -G_d (G_c^{dB} + \Gamma^{dB})$$

che rappresentiamo alla figura a lato³⁵ in modo da meglio apprezzare come, mentre il guadagno di codifica G_c^{dB} determina una semplice *traslazione a sinistra* della curva di prestazione, il guadagno di diversità G_d *modifica l'inclinazione*, e dunque ha un effetto sempre più pronunciato all'aumentare dell'SNR.



21.3.4 Codici sub ottimi

Concludiamo la sezione sulla diversità spaziale citando le alternative che, utilizzando più di due antenne in trasmissione, conseguono un ordine di diversità pari a $n_T \cdot n_R$ a spese di un tasso di codifica $R_c = k/T < 1$, oppure conseguono $R_c = 1$ ma con un ordine di diversità ridotto.

Codice ortogonale reale quadrato Un primo risultato da citare è che limitando il campo di applicazione a simboli *reali* (ovvero ad una modulazione L-ASK) possono essere definiti STBC *ortogonali*³⁶ con codeword *quadrato* ovvero $n_T = T$ pari a 2, 4 ed 8, ed in grado di offrire $R_c = 1$ e diversità $n_T \cdot n_R$, prestazioni che per simboli complessi sono invece conseguite esclusivamente dal codice di Alamouti con $n_T = 2$. Un esempio di STBC che usa $n_T = 4$ antenne per trasmettere $k = 4$ simboli *reali* in $T = 4$ istanti è indicato come C_4 nello schema che segue:

³⁴Infatti il determinante è anche pari al prodotto degli autovalori, e dunque il suo essere $\neq 0$ implica che non vi siano autovalori nulli.

³⁵Anziché mostrare la P_e in scala logaritmica come si fa di solito, per coerenza con l'espressione $10 \log_{10} P_e \propto -G_d (G_c^{dB} + \Gamma^{dB})$ sulle ordinate è mostrato il log di P_e , ed essendo $P_e < 1$, il suo log è negativo.

³⁶L'ortogonalità delle codeword è la proprietà che consente di conseguire la piena diversità spaziale e che permette la decodifica semplificata discussa precedentemente.

$$C_4 = \begin{bmatrix} s_1 & s_2 & s_3 & s_4 \\ -s_1 & s_2 & -s_3 & s_4 \\ -s_1 & s_2 & s_3 & -s_4 \\ -s_1 & -s_2 & s_3 & s_4 \end{bmatrix} \quad C_{8 \times 4}^T = \begin{bmatrix} s_1 & -s_1 & -s_1 & -s_1 & s_1^* & -s_1^* & -s_1^* & -s_1^* \\ s_2 & s_2 & s_2 & -s_2 & s_2^* & s_2^* & s_2^* & -s_2^* \\ s_3 & -s_3 & s_3 & s_3 & s_3^* & -s_3^* & s_3^* & s_3^* \\ s_4 & s_4 & -s_4 & s_4 & s_4^* & s_4^* & -s_4^* & s_4^* \end{bmatrix}$$

Codice complesso associato Un secondo risultato è che è possibile ottenere un codice ortogonale *complesso* a partire da uno reale trasmettendo la stessa struttura di codeword ma ad elementi complessi, fatta seguire da altrettanti T istanti in cui viene trasmessa la codeword coniugata, come esemplificato dalla matrice $C_{8 \times 4} = \begin{bmatrix} C_4 \\ C_4^* \end{bmatrix}$, sopra rappresentata *trasposta* per ragioni di spazio. Come evidente l'adozione di $C_{8 \times 4}$ comporta un raddoppio del numero di istanti necessari alla trasmissione, e dunque un dimezzamento di R_c ; d'altra parte lo stesso raddoppio di T determina anche un miglioramento di 3 dB per l' SNR di decisione, in quanto uno stesso simbolo s_i viene ricevuto il doppio delle volte. D'altra parte le esigenze di *tempo reale* della trasmissione comportano che il raddoppio del numero di istanti T richieda il dimezzamento della loro durata, e quindi la ricezione di metà dell'energia per simbolo: tale peggioramento è compensato dal guadagno di 3 dB legato alla doppia trasmissione.

Codice reale non quadrato E' anche possibile realizzare *STBC reali* con codeword *non quadrate* ovvero con $n_T \neq T$, purché il numero di elementi $T \cdot n_T$ nella matrice del codice sia un multiplo intero del numero k di simboli da codificare; tali soluzioni esistono per qualunque n_T , conseguono massima diversità e decodifica lineare, e quando utilizzano un numero di istanti T pari a quello k dei simboli da rappresentare esibiscono anche un tasso di codifica $R_c = 1$

Esempio Citiamo il caso di un *SBTC reale* con $n_T = 3$ e $k = T = 4$ mostrato a lato, che ripete ciascun simbolo tre volte e consegue $R_c = 1$; quando da questo se ne ricava un *SBTC complesso* concatenando la matrice con la coniugata, il tasso si riduce ad $1/2$.

$$C_{4 \times 3} = \begin{bmatrix} s_1 & s_2 & s_3 \\ -s_2 & s_1 & -s_4 \\ -s_3 & s_4 & s_1 \\ -s_4 & -s_3 & s_2 \end{bmatrix}$$

Ritardo minimo Finché ci si limita a raddoppiare un codice reale per farne uno complesso come nell'esempio precedente, il tasso non può essere maggiore di $1/2$; sono però stati scoperti codici complessi in grado di conseguire $R_c = 3/4$ che, a parità del numero n_T di antenne, riducono il numero T di istanti necessari, conseguendo anche un minor *ritardo* di codifica.

Codici quasi-ortogonali Anziché sacrificare il tasso R_c per l'ortogonalità, è possibile fare l'opposto e mantenere $R_c = 1$ assieme alla piena diversità (con $n_T > 2$) mediante *STBC quasi-ortogonali*³⁷, penalizzati da una modesta perdita di prestazioni (P_e) ed una maggior complessità di decodifica.

³⁷Ovvero per i quali il prodotto $C \cdot C^\dagger$ non assume la forma aI_T .

21.4 Capacità di canale con fading di Rayleigh

Le tecniche di *valorizzazione* della diversità spaziale fin qui discusse apportano un beneficio alla comunicazione che è stato quantificato come un aumento dell'*SNR* per la variabile di decisione. Tale aumento può essere *capitalizzato* aumentando il numero M di bit (e di livelli $L = 2^M$) rappresentati da ciascun simbolo s_i del messaggio da trasmettere, aumentando così la velocità di trasmissione, a parità di periodo di simbolo e di banda occupata. Il legame tra *SNR* di un canale AWGN e la conseguente massima velocità di trasmissione è stato analizzato al § 17.3 ed espresso nei termini della *capacità* C del canale; affrontiamo ora l'adattamento di tale formulazione al caso di canale affetto da fading di Rayleigh ed in presenza di più antenne trasmittenti e/o riceventi, descritto dalla relazione

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{n} \quad (21.34)$$

come illustrato al § 21.2, con il vincolo di mantenere la potenza $\mathcal{P}_s = E\{\mathbf{s}^2\} \cdot f_s$ emessa *collettivamente* da tutte le n_T antenne trasmittenti limitata a \mathcal{P}_T .

Canale SISO Per avere un termine di paragone rispetto al quale confrontare i risultati valutiamo per prima l'espressione di C con $n_T = n_R = 1$, nel cui caso la (21.34) è una semplice relazione scalare (a valori complessi) $r = hs + n$. L'espressione (eq. (17.18)) della capacità AWGN $C = B \cdot \log_2\left(1 + \frac{\mathcal{P}_T}{N_0B}\right)$ [bit/sec] viene riscritta come³⁸

$$C_{SISO} = \log_2(1 + \rho |h|^2) \quad \text{bit/secondo/Hertz} \quad (21.35)$$

in cui ρ è l'*SNR* che si sarebbe ricevuto in assenza di fading, $|h|$ ha una d.d.p. di Rayleigh dovuta al fading, ed il risultato viene espresso *per Hertz* di banda occupata anziché in forma cumulativa, e dunque la (21.35) ha il significato di massima *efficienza spettrale* (pag. 494) conseguibile. La stessa (21.35) può essere equivalentemente misurata in bit/simbolo o bit/uso del canale.

³⁸Ripercorriamo lo sviluppo svolto al § 17.3 per un canale a valori *reali*, che definisce $C = \max_{p(s)} \{I(S; R)\}$ in cui $I(S; R) = h(R) - h(R/S)$ è l'informazione mutua media tra simbolo s trasmesso e valore ricevuto r . Il termine $h(R/S)$ è dovuto al solo rumore, e per esso rimane valido il risultato (17.16) ovvero $h(R/S) = \frac{1}{2} \log_2(2\pi e \sigma_n^2)$; il massimo di C si ottiene quindi massimizzando $h(R)$, che come noto (§ 9.6.2) avviene con r gaussiano, fornendo $h(R) = \frac{1}{2} \log_2(2\pi e \sigma_r^2)$ in cui per σ_r^2 si ottiene

$$\sigma_r^2 = E\{r^2\} = E\{(hs + n)(hs + n)\} = |h|^2 E\{s^2\} + E\{n^2\} = |h|^2 \sigma_s^2 + \sigma_n^2$$

in virtù dell'incorrelazione tra s ed n , entrambi a media nulla. Procedendo come al § 17.3 si ha

$$\begin{aligned} C &= \max_{p(s)} \{I(S; R)\} = \max_{p(s)} \{h(R) - h(R/S)\} = \frac{1}{2} \log_2 \left(2\pi e \left(\sigma_n^2 + \sigma_s^2 |h|^2 \right) \right) - \frac{1}{2} \log_2 \left(2\pi e \sigma_n^2 \right) = \\ &= \frac{1}{2} \log_2 \frac{\sigma_n^2 + \sigma_s^2 |h|^2}{\sigma_n^2} = \frac{1}{2} \log_2 \left(1 + \frac{\sigma_s^2}{\sigma_n^2} |h|^2 \right) \end{aligned}$$

in cui definiamo $\frac{\sigma_s^2}{\sigma_n^2} = \rho$ pari all'*SNR* in assenza di fading. Per arrivare alla (21.35) osserviamo come un canale a valori *complessi* equivalga a due canali reali indipendenti (basti pensare al mo-demodulatore in fase e quadratura), essenzialmente in virtù della indipendenza delle parti reale ed immaginaria del coefficiente complesso h e del rumore n . Pertanto la capacità del canale complesso è il *doppio* di quanto ora calcolato, come espresso dalla (21.35). Qualcuno può giustamente chiedersi ora se l'*SNR* ρ sia da riferirsi al singolo ramo (I o Q), oppure al segnale complesso. In realtà i due valori sono *equivalenti*, perché per il canale complesso sia σ_s^2 che σ_n^2 *raddoppiano*.

Capacità ergodica e garantita Osserviamo che il valore (21.35), come quelli che seguiranno, è in realtà una v.a., in quanto h lo è. A partire dall'espressione (21.35) si possono ricavare due *numeri*: il primo si ottiene eseguendo il valore atteso di C rispetto alla variabilità di h , ottenendo un valore *medio* indicato anche come capacità *ergodica*. Il secondo è invece riferito ad un grado di servizio g ed esprime il valore di capacità *superato* nel $g\%$ del tempo, ed in questo senso detta capacità *garantita* (per il $g\%$ del tempo). Ma ai fini di quel che segue sono presi in considerazione casi in cui h non varia per tutta la durata del collegamento, in modo da permettere il confronto tra risultati.

Canale SIMO In presenza di n_R antenne riceventi al § 21.3.1.2 si è analizzato come un ricevitore MRC determini un aumento dell' SNR di una quantità pari a $\sum_{i=1}^{n_R} |h_i|^2$, e difatti anche il valore teorico di capacità risulta pari a

$$C_{SIMO} = \log_2 \left(1 + \rho \sum_{i=1}^{n_R} |h_i|^2 \right) \quad \text{bit/secondo/Hertz} \quad (21.36)$$

in cui h_i è il guadagno complesso per l' i -esima antenna. Notiamo che la dipendenza tra n_R e C_{SIMO} è di tipo *logaritmico*, e dunque l'incremento di C_{SIMO} è via via minore all'aumentare di n_R .

Canale MISO Nel caso di n_T antenne trasmettenti, ma in assenza di conoscenza da parte del trasmettitore dei valori che compaiono nella matrice \mathbf{H} , ogni antenna utilizza un n_T -esimo della potenza totale di trasmissione, e la capacità risulta pari a

$$C_{MISO} = \log_2 \left(1 + \frac{\rho}{n_T} \sum_{i=1}^{n_T} |h_i|^2 \right) \quad \text{bit/secondo/Hertz} \quad (21.37)$$

ovvero oltre all'incremento solamente logaritmico con n_T , si assiste anche alla progressiva riduzione dell' SNR di ricezione, in virtù del vincolo sulla potenza totale di trasmissione che va suddivisa su tutte le n_T antenne.

21.4.1 Capacità del canale MIMO

Il calcolo della capacità in presenza di più antenne sia in trasmissione (n_T) che in ricezione (n_R) segue le linee guida indicate alla nota 38, con l'obiettivo di rendere massima l'informazione mutua media

$$I(\mathbf{s}; \mathbf{r}) = h(\mathbf{r}) - h(\mathbf{r}/\mathbf{s})$$

tra i vettori dei simboli trasmessi \mathbf{s} e di quelli ricevuti \mathbf{r} in modo da ottenere $C = \max_{\rho(\mathbf{s})} \{I(\mathbf{s}; \mathbf{r})\}$. Dato che si suppone la matrice di canale \mathbf{H} perfettamente nota in ricezione la sua incertezza condizionata ad \mathbf{s} è nulla, dunque

$$h(\mathbf{r}/\mathbf{s}) = h(\mathbf{H}\mathbf{s} + \mathbf{n}/\mathbf{s}) = h(\mathbf{n}/\mathbf{s}) = h(\mathbf{n})$$

in quanto \mathbf{n} ed \mathbf{s} sono statisticamente indipendenti, e quindi

$$I(\mathbf{s}; \mathbf{r}) = h(\mathbf{r}) - h(\mathbf{n}) \quad (21.38)$$

L'entropia differenziale $h(\mathbf{n})$ del vettore gaussiano complesso \mathbf{n} è calcolata al § 21.9.1 come $h(\mathbf{n}) = \log(\det(\pi\epsilon \Sigma_n))$, dove $\Sigma_n = E\{\mathbf{nn}^\dagger\}$ è la matrice di covarianza del rumore \mathbf{n} . Per massimizzare la (21.38) occorre quindi massimizzare $h(\mathbf{r})$, il che avviene

quando anche \mathbf{r} è un vettore gaussiano complesso, per il quale nuovamente $h(\mathbf{r}) = \log(\det(\pi e \Sigma_r))$, in cui la matrice di covarianza Σ_r del vettore (a media nulla) ricevuto risulta pari a

$$\begin{aligned}\Sigma_r &= E\{\mathbf{r}\mathbf{r}^\dagger\} = E\{(\mathbf{H}\mathbf{s} + \mathbf{n})(\mathbf{H}\mathbf{s} + \mathbf{n})^\dagger\} = E\{\mathbf{H}\mathbf{s}(\mathbf{H}\mathbf{s})^\dagger\} + E\{\mathbf{n}\mathbf{n}^\dagger\} = \\ &= \mathbf{H}E\{\mathbf{s}\mathbf{s}^\dagger\}\mathbf{H}^\dagger + \Sigma_n = \mathbf{H}\Sigma_s\mathbf{H}^\dagger + \Sigma_n\end{aligned}$$

data l'incorrelazione tra \mathbf{s} ed \mathbf{n} , ed avendo sostituito $(\mathbf{H}\mathbf{s})^\dagger = \mathbf{s}^\dagger\mathbf{H}^\dagger$. Mettendo tutto assieme otteniamo dunque

$$\begin{aligned}C_{MIMO} &= \max_{p(\mathbf{s})}\{I(\mathbf{s}; \mathbf{r})\} = \max_{p(\mathbf{s})}\{h(\mathbf{r}) - h(\mathbf{n})\} = \\ &= \log_2(\det[\pi e \Sigma_r]) - \log_2(\det[\pi e \Sigma_n]) = \log_2\left(\frac{\det[\pi e \Sigma_r]}{\det[\pi e \Sigma_n]}\right) = \\ &= \log_2(\det[\Sigma_r \Sigma_n^{-1}]) = \log_2(\det[(\mathbf{H}\Sigma_s\mathbf{H}^\dagger + \Sigma_n)\Sigma_n^{-1}]) = \\ &= \log_2(\det[\mathbf{H}\Sigma_s\mathbf{H}^\dagger \Sigma_n^{-1} + \Sigma_n \Sigma_n^{-1}]) = \\ &= \log_2(\det[\mathbf{I}_{n_R} + \Sigma_n^{-1}\mathbf{H}\Sigma_s\mathbf{H}^\dagger]) \quad \text{bit/sec/Hz}\end{aligned}\quad (21.39)$$

in cui la capacità MIMO viene a dipendere dalla covarianza di sorgente Σ_s , da quella di rumore Σ_n , e dalla matrice di canale \mathbf{H} . Vediamo ora come diverse ipotesi su tali grandezze permettano di arrivare a risultati adatti ai singoli casi.

Rumore incorrelato tra antenne La circostanza di avere un rumore *spazialmente bianco*, che si verifica quando per le componenti n_i del vettore \mathbf{n} risulta $E\{n_i n_j^*\} = 0$ con $i \neq j$ oppure σ_n^2 se $i = j$, può essere interpretata come una assenza di interferenti *in comune* tra le antenne di ricezione. A ciò consegue una covarianza *del rumore* nella forma $\Sigma_n = \sigma_n^2 \mathbf{I}_{n_R}$, ossia una matrice nulla tranne che sulla diagonale, dove vale σ_n^2 . In tal caso la (21.39) diviene

$$C_{MIMO-WN} = \log_2\left(\det\left[\mathbf{I}_{n_R} + \frac{1}{\sigma_n^2}\mathbf{H}\Sigma_s\mathbf{H}^\dagger\right]\right) \quad \text{bit/sec/Hz}\quad (21.40)$$

Potenza trasmessa uniforme e simboli incorrelati In generale il trasmettitore non è a conoscenza dei valori di \mathbf{H} , e l'unica cosa che può fare è trasmettere con la stessa potenza su tutte le n_T antenne. Allo stesso tempo è lecito considerare i simboli trasmessi incorrelati, ossia per due elementi s_i, s_j di \mathbf{s} si ha $E\{s_i s_j\} = 0$ se $i \neq j$. Sotto tali condizioni la covarianza del segnale è pari a $\Sigma_s = \frac{\mathcal{E}_s}{n_T} \mathbf{I}_{n_T}$ ossia è tutta nulla tranne che sulla diagonale, dove vale un n_T -esimo dell'energia per simbolo $\mathcal{E}_s = E\{\mathbf{s}^\dagger \mathbf{s}\}$. In queste circostanze la (21.40) diviene

$$C_{MIMO-EP} = \log_2\left(\det\left[\mathbf{I}_{n_R} + \frac{\rho}{n_T}\mathbf{H}\mathbf{H}^\dagger\right]\right) \quad \text{bit/sec/Hz}\quad (21.41)$$

in cui $\rho = \frac{\mathcal{E}_s}{\sigma_n^2}$ è l'SNR *per simbolo* del segnale ricevuto da ciascuna delle n_R antenne a partire da *tutte* le n_T trasmettenti.

Derivazione della capacità MISO e SIMO Notiamo che la condizione di equipartizione di \mathcal{P}_T è in comune alla configurazione MISO che conduce alla (21.37), ma mentre

in quel caso la capacità aumenta con legge logaritmica rispetto ad n_T , si dimostra che la (21.41) aumenta *linearmente* con $m = \min(n_T, n_R)$. Osserviamo inoltre che (21.41) diviene esattamente pari a C_{MISO} (eq. (21.37)) quando $n_R = 1$, dato che in questa circostanza \mathbf{H} è un vettore riga, e dunque $\mathbf{H}\mathbf{H}^\dagger = \sum_{i=1}^{n_T} |h_i|^2$. Per lo stesso motivo (21.41) diviene pari a C_{SIMO} (eq. (21.36)) quando $n_T = 1$.

Il contributo di \mathbf{H} Quando entrambi i lati del collegamento sono equipaggiati con più antenne la (21.41) può essere riscritta in una forma che permette di individuare quali caratteristiche della matrice di canale MIMO \mathbf{H} di dimensioni $n_R \times n_T$ rendono il valore $C_{MIMO-EP}$ più o meno grande. Andiamo a mostrare che ciò dipende dagli *autovalori non nulli* $\lambda_1, \lambda_2, \dots, \lambda_m$ della matrice $n_R \times n_R$

$$\mathbf{W} = \mathbf{H}\mathbf{H}^\dagger$$

simmetrica, semidefinita positiva e di rango $m \leq \min(n_T, n_R)$ ³⁹, che può quindi essere espressa (§ 6.7.3) come $\mathbf{W} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^\dagger$ in cui $\mathbf{\Lambda}$ è una matrice quadrata tutta nulla tranne per gli autovalori λ_i sulla diagonale, e $\mathbf{\Gamma}$ è una matrice unitaria⁴⁰ con colonne pari agli autovettori di \mathbf{W} , posti nello stesso ordine con cui gli autovalori compaiono in $\mathbf{\Lambda}$. Partendo dalla (21.41) scriviamo pertanto

$$\begin{aligned} C_{MIMO-EP} &= \log_2(\det[\mathbf{I}_{n_R} + \frac{\rho}{n_T}\mathbf{H}\mathbf{H}^\dagger]) = \log_2(\det[\mathbf{I}_{n_R} + \frac{\rho}{n_T}\mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^\dagger]) = \\ &= \log_2(\det[\mathbf{I}_m + \frac{\rho}{n_T}\mathbf{\Gamma}^\dagger\mathbf{\Gamma}\mathbf{\Lambda}]) = \log_2(\det[\mathbf{I}_m + \frac{\rho}{n_T}\mathbf{\Lambda}]) = \\ &= \log_2\left(\left(1 + \frac{\rho}{n_T}\lambda_1\right)\left(1 + \frac{\rho}{n_T}\lambda_2\right) \cdots \left(1 + \frac{\rho}{n_T}\lambda_m\right)\right) = \\ &= \sum_{i=1}^m \log_2\left(1 + \frac{\rho}{n_T}\lambda_i\right) \quad \text{bit/sec/Hz} \end{aligned} \quad (21.42)$$

in cui la terza eguaglianza deriva dall'identità riportata alla nota 39, la quarta dall'essere $\mathbf{\Gamma}$ unitaria, mentre alla quinta si è sviluppato il determinante come prodotto dei termini sulla diagonale. Discutiamo del risultato non appena ottenuto:

- la (21.42) mostra come $C_{MIMO-EP}$ sia pari alla *somma* delle capacità di m canali *virtuali* di tipo SISO (eq. (21.35)) *indipendenti*, ognuno con SNR per simbolo pari a $\frac{\rho}{n_T} = \frac{1}{n_T} \frac{\mathcal{E}_s}{\sigma_n^2}$ e guadagno di Rayleigh $|h|^2 = \lambda_i$;
- essendo il determinante pari al prodotto degli autovalori, la matrice

$$\mathbf{I}_m + \frac{\rho}{n_T}\mathbf{\Lambda} = \begin{bmatrix} 1 + \frac{\rho}{n_T}\lambda_1 & 0 & \cdots & 0 \\ 0 & 1 + \frac{\rho}{n_T}\lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 + \frac{\rho}{n_T}\lambda_m \end{bmatrix} \quad \text{di dimensioni } m \times m \text{ ha auto-}$$

³⁹La definizione $\mathbf{W} = \mathbf{H}\mathbf{H}^\dagger$ è valida quando $n_R \leq n_T$, mentre se $n_T < n_R$ conviene scrivere $\mathbf{W} = \mathbf{H}^\dagger\mathbf{H}$, con dimensioni $n_R \times n_R$ e $n_T \times n_T$ rispettivamente. Da questo punto di vista è da notare che in base al teorema di *Weinstein-Aronszajn* sussiste l'identità

$$\det[\mathbf{I}_{n_R} + \frac{\rho}{n_T}\mathbf{H}\mathbf{H}^\dagger] = \det[\mathbf{I}_{n_T} + \frac{\rho}{n_T}\mathbf{H}^\dagger\mathbf{H}]$$

vedi ad es. https://en.wikipedia.org/wiki/Weinstein-Aronszajn_identity: date due matrici \mathbf{A} e \mathbf{B} di dimensioni $m \times n$ ed $n \times m$, risulta $\det(\mathbf{I}_m + \mathbf{A}\mathbf{B}) = \det(\mathbf{I}_n + \mathbf{B}\mathbf{A})$.

⁴⁰Ovvero per la quale $\mathbf{\Gamma}\mathbf{\Gamma}^\dagger = \mathbf{\Gamma}^\dagger\mathbf{\Gamma} = \mathbf{I}$, ossia la matrice identità.

valori $1 + \frac{\rho}{n_T} \lambda_i$, così come anche $\mathbf{I}_{n_R} + \frac{\rho}{n_T} \mathbf{H}\mathbf{H}^\dagger$;

- essendo \mathbf{W} semidefinita positiva⁴¹ i suoi autovalori λ_i sono tutti reali e non negativi, risultando inoltre⁴² $\det \mathbf{W} = \det \Lambda$. Dunque $\mathbf{I}_m + \frac{\rho}{n_T} \Lambda$ ha autovalori reali e positivi, così come $\mathbf{I}_{n_R} + \frac{\rho}{n_T} \mathbf{H}\mathbf{H}^\dagger$;
- la (21.42) è tanto più grande quanto maggiori sono gli autovalori λ_i , o meglio (dopo averli ordinati in ordine decrescente) quanto più il loro valore *non decade* troppo rapidamente. Autovalori piccoli o peggio ancora nulli (qualora il rango di \mathbf{W} sia minore di n_R) sono indicativi di antenne poco distanziate, ovvero di una matrice \mathbf{H} con righe (o colonne) circa pari alla combinazione lineare di altre righe (o colonne), ossia con valori h_{ij} correlati, da attribuirsi ad una ridotta variabilità del multipath, eventualmente anche dovuta ad un elevato *fattore di Rice* ossia una eccessiva visibilità tra le antenne (§ 20.3.1 e pag. 682);
- gli elementi della matrice *aleatoria* $\mathbf{W} = \mathbf{H}\mathbf{H}^\dagger$ sono pari a $w_{ij} = \sum_{k=1}^{n_T} h_{ik} h_{jk}^*$ ossia al *prodotto scalare* tra le righe i e j di \mathbf{H} , di valore tanto maggiore quanto più i guadagni complessi tra ciascuna delle due antenne riceventi i e j e tutte le n_T trasmettenti sono *paralleli*. Da questo punto di vista gli elementi w_{ij} sono proporzionali ad una *stima campionaria* della correlazione tra i valori ricevuti dalle antenne i e j .

21.4.1.1 Trasmissione a potenza differenziata

Rimuoviamo ora l'ipotesi $\Sigma_s = \frac{\mathcal{E}_s}{n_T} \mathbf{I}_{n_T}$ adottata per giungere alle (21.41) e (21.42) ed investighiamo se esista una diversa matrice Σ_s tale da rendere il termine $\det \left[\mathbf{I}_{n_R} + \frac{1}{\sigma_n^2} \mathbf{H} \Sigma_s \mathbf{H}^\dagger \right]$ che compare nella (21.40) maggiore di quanto ottenuto per la (21.42), mantenendo il vincolo⁴³

$$\text{tr}(\Sigma_s) = \sum_{i=1}^{n_T} E \{ s_i s_i^* \} = \mathcal{E}_s \leq \frac{\mathcal{P}_T}{f_s}$$

A questo scopo consideriamo per \mathbf{H} la sua *scomposizione ai valori singolari* (SVD⁴⁴)

$$\mathbf{H} = \mathbf{U} \mathbf{D} \mathbf{V}^\dagger$$

dove \mathbf{D} è una matrice $n_R \times n_T$ tutta nulla tranne che per gli elementi sulla diagonale, noti come *valori singolari* di \mathbf{H} , e pari alla *radice quadrata* degli autovalori non nulli λ_i ($i = 1, \dots, m$) di $\mathbf{H}\mathbf{H}^\dagger$ (ma anche di $\mathbf{H}^\dagger \mathbf{H}$), ovvero $(\mathbf{D})_{ii} = \lambda_i^{1/2}$, mentre \mathbf{U} e \mathbf{V} sono

⁴¹Ossia tale che $\Re \{ \mathbf{c}^\dagger \mathbf{W} \mathbf{c} \} \geq 0$ per $\forall \mathbf{c}$, vedi anche § 6.7.3. Per la precisione \mathbf{W} è una matrice *Hermitiana*, ossia tale che $\mathbf{W} = \mathbf{W}^\dagger$, e la condizione precedente diviene $\mathbf{c}^\dagger \mathbf{W} \mathbf{c} \geq 0$, vedi ad es. https://it.wikipedia.org/wiki/Matrice_definita_positiva. Anzi, ad essere ancora più precisi, per la sua natura aleatoria \mathbf{W} è nota come *matrice di Wishart*, e la legge di distribuzione probabilistica dei suoi valori è descritta dalla omonima distribuzione, vedi ad es. https://it.wikipedia.org/wiki/Distribuzione_di_Wishart.

⁴²In quanto \mathbf{W} e Λ sono matrici *simili*, vedi ad es.

https://it.wikipedia.org/wiki/Similitudine_tra_matrici.

⁴³Ricordiamo che la *traccia* di una matrice quadrata è la *somma* degli elementi sulla diagonale.

⁴⁴Vedi ad es. https://it.wikipedia.org/wiki/Decomposizione_ai_valori_singolari

matrici unitarie di dimensione $n_R \times n_R$ e $n_T \times n_T$ rispettivamente, le cui colonne sono (per \mathbf{U}) gli autovettori di $\mathbf{H}\mathbf{H}^\dagger$, e per \mathbf{V} gli autovettori di $\mathbf{H}^\dagger\mathbf{H}$.

Con tali posizioni per l'argomento di (21.41) otteniamo

$$\begin{aligned} \det\left[\mathbf{I}_{n_R} + \frac{1}{\sigma_n^2}\mathbf{H}\boldsymbol{\Sigma}_s\mathbf{H}^\dagger\right] &= \\ &= \det\left[\mathbf{I}_{n_R} + \frac{1}{\sigma_n^2}\mathbf{U}\mathbf{D}\mathbf{V}^\dagger\boldsymbol{\Sigma}_s\mathbf{V}\mathbf{D}^\dagger\mathbf{U}^\dagger\right] = \\ &= \det\left[\mathbf{I}_{n_R} + \frac{1}{\sigma_n^2}\mathbf{D}\mathbf{V}^\dagger\boldsymbol{\Sigma}_s\mathbf{V}\mathbf{D}^\dagger\right] \end{aligned} \quad (21.43)$$

dove all'ultima eguaglianza si è invocata l'identità di cui alla nota 39 e ricordato che $\mathbf{U}^\dagger\mathbf{U} = \mathbf{I}_{n_R}$. Evocando ora la *disuguaglianza di Hadamard*⁴⁵, che recita

per una matrice \mathbf{A} semidefinita positiva risulta $\det(\mathbf{A}) \leq \prod_i a_{ii}$, con il segno di uguale solo se \mathbf{A} è diagonale

osserviamo come per massimizzare la (21.43) occorra che $\mathbf{I}_{n_R} + 1/\sigma_n^2\mathbf{D}\mathbf{V}^\dagger\boldsymbol{\Sigma}_s\mathbf{V}\mathbf{D}^\dagger$ sia diagonale. A tale scopo eseguiamo il cambio di variabile⁴⁶ $\tilde{\mathbf{s}} = \mathbf{V}^\dagger\mathbf{s}$ in modo da poter scrivere

$$\boldsymbol{\Sigma}_{\tilde{\mathbf{s}}} = E\{\tilde{\mathbf{s}}\tilde{\mathbf{s}}^\dagger\} = E\{\mathbf{V}^\dagger\mathbf{s}\mathbf{s}^\dagger\mathbf{V}\} = \mathbf{V}^\dagger\boldsymbol{\Sigma}_s\mathbf{V}$$

ovvero $\boldsymbol{\Sigma}_s = \mathbf{V}\boldsymbol{\Sigma}_{\tilde{\mathbf{s}}}\mathbf{V}^\dagger$, e quindi riformulare la (21.43) come

$$\begin{aligned} \det\left[\mathbf{I}_{n_R} + \frac{1}{\sigma_n^2}\mathbf{D}\mathbf{V}^\dagger\boldsymbol{\Sigma}_s\mathbf{V}\mathbf{D}^\dagger\right] &= \det\left[\mathbf{I}_{n_R} + \frac{1}{\sigma_n^2}\mathbf{D}\mathbf{V}^\dagger\mathbf{V}\boldsymbol{\Sigma}_{\tilde{\mathbf{s}}}\mathbf{V}^\dagger\mathbf{V}\mathbf{D}^\dagger\right] = \\ &= \det\left[\mathbf{I}_{n_R} + \frac{1}{\sigma_n^2}\mathbf{D}\boldsymbol{\Sigma}_{\tilde{\mathbf{s}}}\mathbf{D}^\dagger\right] \end{aligned} \quad (21.44)$$

che è massimo quando $\mathbf{I}_{n_R} + \frac{1}{\sigma_n^2}\mathbf{D}\boldsymbol{\Sigma}_{\tilde{\mathbf{s}}}\mathbf{D}^\dagger$ è diagonale, ovvero quando $\boldsymbol{\Sigma}_{\tilde{\mathbf{s}}}$ lo è. Ma l'elemento i -esimo sulla diagonale di $\boldsymbol{\Sigma}_{\tilde{\mathbf{s}}}$ è pari all'energia per simbolo \mathcal{E}_{s_i} trasmesso dall' i -esima antenna⁴⁷: scriviamo dunque $\boldsymbol{\Sigma}_{\tilde{\mathbf{s}}} = \text{diag}(\mathcal{E}_{s_1}, \mathcal{E}_{s_2}, \dots, \mathcal{E}_{s_{n_T}})$ in modo da ottenere dalle (21.40), (21.43) e (21.44)

$$\begin{aligned} C_{MIMO-NEP} &= \log_2(\det[\mathbf{I}_{n_R} + \frac{1}{\sigma_n^2}\mathbf{D}\boldsymbol{\Sigma}_{\tilde{\mathbf{s}}}\mathbf{D}^\dagger]) = \\ &= \log_2(\prod_{i=1}^{n_T} (1 + \frac{1}{\sigma_n^2}\lambda_i\mathcal{E}_{s_i})) = \sum_{i=1}^{n_T} \log_2(1 + \frac{\mathcal{E}_{s_i}}{\sigma_n^2}\lambda_i) = \\ &= \sum_{i=1}^m \log_2(1 + \rho_i\lambda_i) \quad \text{bit/sec/Hz} \end{aligned} \quad (21.45)$$

del tutto simile alla (21.42), tranne che ora ogni antenna trasmittente può avere un diverso SNR per simbolo, pari a $\rho_i = \frac{\mathcal{E}_{s_i}}{\sigma_n^2}$. Notiamo che all'ultimo passaggio la sommatoria arriva fino al numero m di autovalori non nulli, in quanto per quelli nulli il contributo sarebbe stato pari a $\log_2(1) = 0$.

⁴⁵Vedi ad es. https://en.wikipedia.org/wiki/Hadamard's_inequality

⁴⁶Che, essendo \mathbf{V} unitaria ossia una rotazione, non modifica le caratteristiche informative ed energetiche delle quantità in gioco.

⁴⁷In quanto

$$(\boldsymbol{\Sigma}_{\tilde{\mathbf{s}}})_{i,i} = E\{\tilde{s}_i\tilde{s}_i^*\} = E\{(\mathbf{V}^\dagger)_i s_i s_i^* (\mathbf{V})_i\} = E\{s_i s_i^*\} (\mathbf{V}^\dagger)_i (\mathbf{V})_i = \mathcal{E}_{s_i}$$

ovvero l'energia per il simbolo trasmesso dall'antenna i . Con $(\mathbf{V}^\dagger)_i$ si è indicata l' i -esima riga di \mathbf{V}^\dagger e con $(\mathbf{V})_i$ l' i -esima colonna di \mathbf{V} , il cui prodotto scalare è uno, essendo la matrice unitaria. La trasformazione $\tilde{\mathbf{s}} = \mathbf{V}^\dagger\mathbf{s}$ è infatti una *rotazione* che non altera la norma di \mathbf{s} , e dunque $\boldsymbol{\Sigma}_s$ e $\boldsymbol{\Sigma}_{\tilde{\mathbf{s}}}$ hanno la stessa diagonale.

21.4.1.2 Codifica a riempimento d'acqua

A questo punto non rimane che individuare la distribuzione delle potenze di trasmissione $\mathcal{P}_i = \mathcal{E}_{s_i} f_s$ (e dunque dei corrispondenti SNR per simbolo $\rho_i = \mathcal{E}_{s_i} / \sigma_n^2$) tale da rendere massima (21.45), nel rispetto del vincolo che $\sum_{i=1}^{n_T} \rho_i - \frac{\mathcal{E}_s}{\sigma_n^2} = 0$ con $\mathcal{E}_s = E\{\mathbf{s}^\dagger \mathbf{s}\}$ e $\mathcal{P}_T = \mathcal{E}_s f_s$. Si tratta quindi di un classico problema di *ottimizzazione vincolata*, da affrontare con il metodo dei moltiplicatori di Lagrange (§ 9.6.1): a questo scopo definiamo la funzione *Lagrangiana* \mathcal{L} con moltiplicatore μ come

$$\mathcal{L} = \sum_{i=1}^m \log_2(1 + \rho_i \lambda_i) + \mu (\sum_{i=1}^m \rho_i - \rho_T)$$

avendo posto $\rho_T = \frac{\mathcal{E}_s}{\sigma_n^2}$, ed eguagliamone a zero le derivate rispetto ai ρ_i e rispetto a μ :

$$\begin{aligned} \frac{\partial}{\partial \rho_i} \mathcal{L} &= \frac{\lambda_i}{1 + \rho_i \lambda_i} + \mu = 0 \quad i = 1, 2, \dots, m \\ \frac{\partial}{\partial \mu} \mathcal{L} &= \sum_{i=1}^m \rho_i - \rho_T = 0 \end{aligned}$$

da cui

$$\frac{1}{\frac{1}{\lambda_i} + \rho_i} = -\mu \quad \text{ossia} \quad \frac{1}{\lambda_i} + \rho_i = -\frac{1}{\mu} \quad i = 1, 2, \dots, m \quad (21.46)$$

$$\sum_{i=1}^m \rho_i = \rho_T$$

dove la seconda riga è ancora il vincolo sulla potenza. Il valore del moltiplicatore μ si ottiene sommando (21.46) su tutti gli i e tenendo conto del vincolo, ovvero

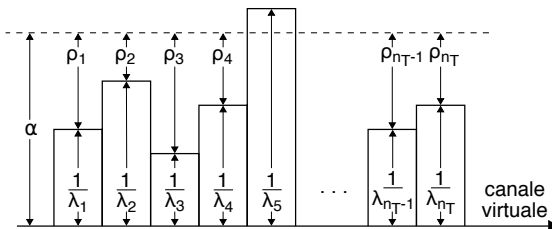
$$\sum_{i=1}^m \left(\frac{1}{\lambda_i} + \rho_i \right) = \sum_{i=1}^m \frac{1}{\lambda_i} + \rho_T = -\frac{m}{\mu} \quad \rightarrow \quad -\frac{1}{\mu} = \frac{1}{m} \left(\sum_{i=1}^m \frac{1}{\lambda_i} + \rho_T \right) = \alpha \quad (21.47)$$

in modo che dalla (21.46) si ottenga

$$\begin{cases} \rho_i = \alpha - \frac{1}{\lambda_i} & \text{se } \rho_i > 0 \\ \rho_i = 0 & \text{altrimenti} \end{cases} \quad (21.48)$$

che come è facile verificare⁴⁸ rispetta il vincolo, mentre la limitazione $\rho_i \geq 0$ ha un evidente significato fisico.

Ricordando la discussione svolta a riguardo della (21.42), i termini λ_i rappresentano i guadagni di Rayleigh dei *canali virtuali* di tipo SISO di cui risulta costituito il canale MIMO. La soluzione (21.48) è detta *a riempimento d'acqua* perché, potendo essere scritta anche come $\rho_i + \frac{1}{\lambda_i} = \alpha$, impone che la somma $\rho_i + \frac{1}{\lambda_i}$ sia *la stessa* per ogni canale, aumentando quindi la potenza (ossia l'SNR per simbolo ρ_i) laddove $\frac{1}{\lambda_i}$ è minore, ovvero



λ_i è maggiore, ossia quando il canale virtuale i è più *affidabile*. Il senso di tale strategia è quello di affidarsi in misura maggiore ai canali migliori, e non sprecare potenza su quelli peggiori.

⁴⁸Basta calcolare $\sum_{i=1}^m \rho_i = \sum_{i=1}^m \left(\alpha - \frac{1}{\lambda_i} \right) = \sum_{i=1}^m \frac{1}{\lambda_i} + \rho_T - \sum_{i=1}^m \frac{1}{\lambda_i} = \rho_T$

Sostituendo i valori forniti dalla (21.48) per ρ_i nella (21.45) otteniamo infine⁴⁹

$$C_{MIMO-WF} = \sum_{\substack{i=1 \\ i \notin \mathcal{B}}}^m \log_2(\alpha \lambda_i) \quad \text{bit/sec/Hz} \quad (21.49)$$

in cui \mathcal{B} è l'insieme degli indici i per cui $\frac{1}{\lambda_i} \geq \alpha = \frac{1}{m} \left(\sum_{i=1}^m \frac{1}{\lambda_i} + \rho_T \right)$.

Architettura di trasmissione Dopo questa analisi così particolare, mostriamo come la stessa SVD della matrice $\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^\dagger$ possa essere *di guida* per la realizzazione di un sistema di trasmissione MIMO in grado di *approcciare*⁵⁰ la velocità per simbolo espressa dalla (21.49). Per riuscirci è necessario che il lato trasmittente conosca gli autovalori non nulli λ_i di $\mathbf{H}\mathbf{H}^\dagger$, la matrice \mathbf{V} , e la potenza di rumore σ_n^2 in ricezione.

A partire dalla potenza \mathcal{P}_T a disposizione, il trasmettente valuta l'SNR per simbolo

$$\rho_T = \frac{1}{\sigma_n^2} \frac{\mathcal{P}_T}{f_s} = \frac{\mathcal{E}_s}{\sigma_n^2} = \frac{E\{\mathbf{s}^\dagger \mathbf{s}\}}{\sigma_n^2}$$

da utilizzare nella (21.47) per calcolare il valore α , e da questo ottiene i valori ρ_i come prescritto dalla (21.48). Le ampiezze dei simboli s_i da inviare alle antenne *non spente* sono quindi (più o meno) amplificati in modo da ottenere $\frac{E\{s_i s_i^*\}}{\sigma_n^2} = \rho_i$ come calcolato, ed il vettore di simboli \mathbf{s}^{WF} così ottenuto viene moltiplicato per la matrice \mathbf{V} fornendo il nuovo vettore $\tilde{\mathbf{s}} = \mathbf{V}\mathbf{s}^{WF}$, i cui elementi \tilde{s}_i sono usati per generare il segnale dati che modula la portante del segnale trasmesso *realmente* dalle antenne.

Dall'altro lato del collegamento (vedi fig. 21.3-a)) viene ora ricevuto il vettore $\tilde{\mathbf{r}} = \mathbf{H}\tilde{\mathbf{s}} + \mathbf{n}$, ma al suo posto il ricevitore prende in considerazione il prodotto di $\tilde{\mathbf{r}}$ per la matrice \mathbf{U}^\dagger , ottenendo così un nuovo vettore di simboli

$$\begin{aligned} \mathbf{r} &= \mathbf{U}^\dagger \tilde{\mathbf{r}} = \mathbf{U}^\dagger \mathbf{H} \tilde{\mathbf{s}} + \mathbf{U}^\dagger \mathbf{n} = \mathbf{U}^\dagger \mathbf{U} \mathbf{D} \mathbf{V}^\dagger \mathbf{V} \mathbf{s}^{WF} + \mathbf{U}^\dagger \mathbf{n} = \\ &= \mathbf{D} \mathbf{s}^{WF} + \tilde{\mathbf{n}} \end{aligned}$$

e dato che $\mathbf{D} = \text{diag}(\lambda_1^{1/2}, \lambda_2^{1/2}, \dots, \lambda_m^{1/2})$ ciò equivale a ricevere m segnali *indipendenti*

$$r_i = \lambda_i^{1/2} s_i^{WF} + \tilde{n}_i \quad i = 1, 2, \dots, m$$

in cui \tilde{n}_i ha esattamente le stesse caratteristiche statistiche (media nulla e varianza σ_n^2) del valore n_i , in virtù dell'essere \mathbf{U}^\dagger una matrice unitaria⁵¹. Pertanto in questo modo il ricevitore *vede* direttamente gli m canali virtuali, tutti di tipo SISO.

⁴⁹Infatti

$$\sum_{i=1}^m \log_2(1 + \rho_i \lambda_i) = \sum_{\substack{i=1 \\ i \notin \mathcal{B}}}^m \log_2(1 + (\alpha - \frac{1}{\lambda_i}) \lambda_i) = \sum_{\substack{i=1 \\ i \notin \mathcal{B}}}^m \log_2(1 + \alpha \lambda_i - 1)$$

dove i canali per i quali in base alla (21.48) si ottiene $\rho_i = 0$ (ovvero elementi di \mathcal{B}) avrebbero dato un contributo $\log_2(1 + \rho_i \lambda_i) = \log_2(1) = 0$.

⁵⁰Ricordiamo che il valore di capacità è un *limite massimo* di velocità per la trasmissione senza errori, rispetto al quale confrontare le prestazioni della codifica di canale in uso.

⁵¹Considerando l'intero vettore $\tilde{\mathbf{n}} = \mathbf{U}^\dagger \mathbf{n}$, le sue componenti \tilde{n}_i sono v.a. gaussiane complesse a media nulla in quanto combinazioni lineari di v.a. della stessa natura. Per la covarianza si ottiene

$$\Sigma_{\tilde{\mathbf{n}}} = E\{\tilde{\mathbf{n}}\tilde{\mathbf{n}}^\dagger\} = E\{\mathbf{U}^\dagger \mathbf{n}(\mathbf{U}^\dagger \mathbf{n})^\dagger\} = E\{\mathbf{U}^\dagger \mathbf{n} \mathbf{n}^\dagger \mathbf{U}\} = \mathbf{U}^\dagger \Sigma_{\mathbf{n}} \mathbf{U} = \sigma_n^2 \mathbf{U}^\dagger \mathbf{I} \mathbf{U} = \sigma_n^2 \mathbf{I}$$

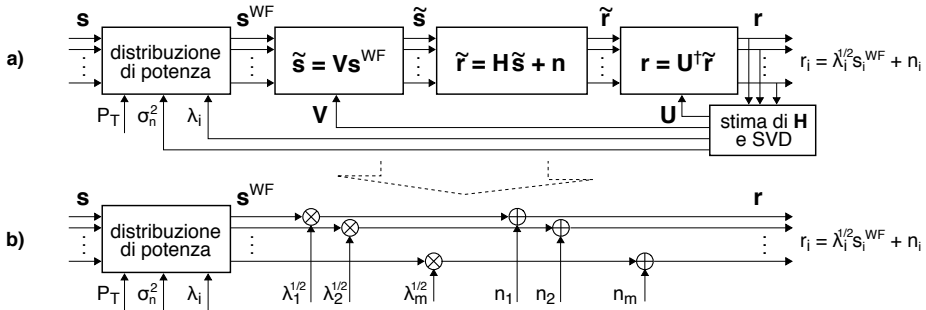


Figura 21.3: Trasmissione MIMO con CSI in trasmissione e *water-filling*: a) - schema di elaborazione, b) - canali virtuali equivalenti

Esempio Per un canale MIMO con $H = \begin{bmatrix} .1 & .3 & .7 \\ .5 & .4 & .1 \\ .2 & .6 & .8 \end{bmatrix}$ si ottiene una svd $H = UDV^\dagger$ pari a

$$H = \begin{bmatrix} -.555 & .3764 & -.7418 \\ -.3338 & -.9176 & -.2158 \\ -.7619 & .1278 & .6349 \end{bmatrix} \begin{bmatrix} 1.3333 & 0 & 0 \\ 0 & 0.5129 & \\ 0 & 0 & 0.0965 \end{bmatrix} \begin{bmatrix} -.2811 & -.7713 & -.5710 \\ -.5679 & -.3459 & .7469 \\ -.7736 & .5342 & -.3408 \end{bmatrix}$$

Notiamo che i canali virtuali presentano guadagni molto differenti, ed in particolare il terzo ($\lambda_3^{1/2} = 0.0965$, dunque $\lambda_3 \approx 0.009$) produrrà un contributo alla capacità (21.49) trascurabile, ed in base alla (21.48) si vedrà allocata una potenza di molto ridotta.

Considerazioni applicative La diseguale distribuzione della potenza disponibile tra le antenne di trasmissione permette di raggiungere velocità più elevate di quanto possibile per una distribuzione uniforme, ma richiede la conoscenza da parte del trasmettitore di informazioni relative al canale MIMO, dette *channel status information* (CSI), che devono essere *stimate* dal lato ricevente⁵². Nel caso di una trasmissione *half-duplex* in cui le parti si scambiano alternativamente di ruolo utilizzando la medesima frequenza portante può essere sfruttata la reciprocità⁵³ della matrice H , e dunque ognuna delle due parti effettua la stima per proprio conto. Ciò non è possibile qualora la trasmissione sia di tipo *full-duplex*⁵⁴, rendendo necessario comunicare la CSI da ricevente a trasmettitore, con due inconvenienti: il primo è che per evitare l'impegno di una eccessiva banda a ritroso tale informazione deve necessariamente essere quantizzata, introducendo errori che possono inficiare i benefici legati alla conoscenza della CSI da parte del trasmettitore; il secondo è che in caso di mobilità la velocità di variazione del canale MIMO può essere troppo elevata rispetto ai tempi necessari alla stima della CSI

⁵²Le tecniche usate a questo scopo sono generalmente basate sulla trasmissione preventiva di *sequenze di training* o di *toni pilota* concordati, in base alla ricezione (alterata) dei quali il ricevitore è in grado di ricostruire le alterazioni subite dal segnale trasmesso. In linea generale questo processo richiede un tempo proporzionale ad n_T , vedi ad es. https://en.wikipedia.org/wiki/Channel_state_information

⁵³Nel senso che la risposta in frequenza h_{ij} del canale radio tra l'antenna j di trasmissione e quella i di ricezione è la stessa di quella in senso inverso. Pertanto se un ricevitore con n_R antenne stima una H_{avanti} relativa ai segnali ricevuti, quando lo stesso si comporta da trasmettitore (con uguale numero di antenne) *trasponde* la matrice stimata e la usa come CSI per il canale MIMO in direzione opposta, ovvero $H_{indietro} = H_{avanti}^T$.

⁵⁴Infatti in tal caso occorre adottare due portanti differenti nelle due direzioni, per non incorrere in una forte auto-interferenza in ricezione.

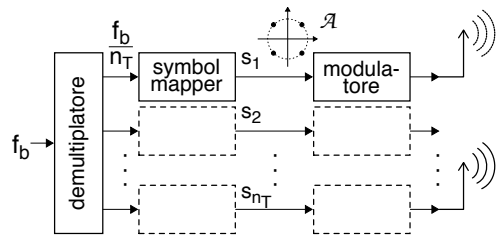
ed alla sua trasmissione, con il rischio di introdurre errori anche per questo motivo.

Rispetto al miglioramento di prestazioni, come già osservato esso dipende dalla particolare realizzazione della matrice \mathbf{H} , dal suo rango, e dalla distribuzione degli autovalori. Il miglioramento inoltre diminuisce all'aumentare dell' SNR di ricezione, sino a (di fatto) annullarsi per $SNR > 30$ dB⁵⁵. Per bassi valori di SNR il miglioramento è invece sensibile, fino a più che raddoppiare il valore di capacità per $SNR < 0$ dB.

21.5 Moltiplicazione spaziale

Mentre nei collegamenti con un basso SNR il principale beneficio del disporre di più antenne è quello di poter sfruttare la diversità spaziale (§ 21.3) per conseguire una P_e altrimenti insufficiente, se al contrario l' SNR è abbastanza elevato da garantire una comunicazione affidabile si possono sfruttare le antenne per trasmettere su ognuna di esse un diverso segnale dati, moltiplicando la velocità di trasmissione di un fattore pari a $g_M = \min(n_T, n_R)$, detto *guadagno di moltiplicazione*.

Dato che come osservato nel caso di un buon SNR il vantaggio derivante dalla scomposizione del collegamento in canali virtuali associati alla SVD di \mathbf{H} è ridotto, si preferisce che il flusso di simboli s_i convogliato dall'antenna i -esima sia frutto della suddivisione del flusso originario a velocità f_b operata *a monte* della trasmissione, da parte di un *demultiplatore* che assegna ad ogni antenna un flusso di f_b/n_T bit/sec. Ogni gruppo di M bit di ciascuno degli n_T flussi viene quindi *mappato* su di un simbolo a valori complessi s_i , $i = 1, 2, \dots, n_T$ appartenente ad un alfabeto \mathcal{A} definito dagli $L = 2^M$ punti di costellazione della modulazione adottata; l'insieme dei simboli trasmessi *contemporaneamente e sulla stessa portante* viene quindi descritto nei termini di un vettore $\mathbf{s} = (s_1, s_2, \dots, s_{n_T})^T$ a valori complessi.



Osserviamo ora che per ricevere correttamente tutti gli n_T diversi flussi dati occorre disporre di un numero n_R di antenne in ricezione almeno pari ad n_T ovvero deve essere $n_R \geq n_T$; le antenne *in più* in ricezione possono essere usate per conseguire anche un *guadagno di diversità*. In questo contesto l'equazione (21.2) del canale MIMO

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{n} \quad (21.50)$$

con $\mathbf{r} \in \mathbb{C}^{n_R}$, $\mathbf{s} \in \mathcal{A}^{n_T}$, $E\{\mathbf{s}\mathbf{s}^\dagger\} = \frac{E_s}{n_T} \mathbf{I}_{n_T}$, $\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$ e $\mathbf{n} \in \mathcal{CN}(0, \sigma_n^2)$, individua una *trasformazione lineare* affetta da rumore, formalmente assimilabile alla distorsione lineare subita da un segnale in transito su di un canale selettivo in frequenza e rumoroso, con la differenza che l'alterazione agisce ora su di un dominio spaziale anziché frequenziale. In tal senso il tentativo di risalire al vettore \mathbf{s} a partire da quello \mathbf{r} noto al ricevitore è simile ad una operazione di *equalizzazione* (§ 18.4) al punto che, con un abuso di terminologia, è lecito riferirsi al primo come ad una *equalizzazione spaziale*.

⁵⁵E' lecito pensare come per SNR elevati *l'acqua sia profonda*, e dunque l'allocatione di potenza sia circa la stessa per tutti i canali virtuali.

21.5.1 Ricevitore a massima verosimiglianza (ML)

Effettua la stima del vettore trasmesso \mathbf{s} come

$$\hat{\mathbf{s}} = \arg \max_{\mathbf{s} \in \mathcal{A}^{n_T}} Pr \{ \mathbf{r} / \mathbf{s} \}$$

ed essendo il vettore \mathbf{n} nella (21.50) costituito da v.a. gaussiane complesse circolari ed incorrelate, equivale ad un criterio di *minima distanza*, ovvero

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s} \in \mathcal{A}^{n_T}} | \mathbf{r} - \mathbf{H}\mathbf{s} |^2 \quad (21.51)$$

dove l'argomento minimizzato dalla (21.51) è la norma quadratica del vettore \mathbf{n} . La ricerca del minimo si estende a tutte le *disposizioni con ripetizione* sulle n_T antenne degli L possibili valori per il generico simbolo s_i , ovvero L^{n_T} diverse disposizioni, causando una complessità *esponenziale* nel numero di antenne: ad es. adottando una semplice modulazione 16-QAM su $n_T = 4$ antenne (corrispondenti a 16 bit/simbolo \mathbf{s}), si ottengono... $16^4 = 65536$ combinazioni, decisamente una quantità *sproporzionata*.

Soluzioni sub-ottime Sebbene la decodifica ML sia quella ottimale, la sua complessità ha spinto la ricerca di soluzioni sub-ottime ma avvicinabili. Una soluzione *algoritmica* è quella nota come *sphere decoding*⁵⁶, che non approfondiamo. Una *famiglia* di soluzioni alternative sono quelle *lineari*, che (in analogia al caso dell'equalizzazione) pervengono ad una matrice \mathbf{G} di dimensioni $n_T \times n_R$ tale da poter scrivere

$$\tilde{\mathbf{s}} = \mathbf{G}\mathbf{r}$$

in cui si è rimosso il vincolo di voler direttamente ottenere un vettore appartenente ad \mathcal{A}^{n_T} . Il valore $\tilde{\mathbf{s}}$ viene quindi indicato come vettore *soffice*, il cui corrispondente vettore *hard* $\hat{\mathbf{s}} \in \mathcal{A}^{n_T}$ si ottiene decidendo per ogni elemento s_i il valore *più vicino* ad uno di quelli ammessi, ossia

$$\hat{s}_i = \arg \min_{s \in \mathcal{A}} (\tilde{s}_i - s)^2 \quad (21.52)$$

Nel seguito descriviamo due criteri per scegliere \mathbf{G} , oltre ad un ulteriore metodo di soluzione che adotta una proiezione lineare nel contesto di una procedura di *cancelazione ordinata*, coniugando efficienza con precisione, la cui implementazione più diffusa è nota come V-BLAST.

21.5.2 Ricevitore zero-forcing

Individua il vettore $\tilde{\mathbf{s}}$ che (senza rispettare il vincolo $\tilde{\mathbf{s}} \in \mathcal{A}^{n_T}$ della (21.51)) risolve il problema

$$\tilde{\mathbf{s}} = \arg \min_{\mathbf{s}} | \mathbf{r} - \mathbf{H}\mathbf{s} |^2 \quad (21.53)$$

calcolando una matrice $\mathbf{G}_{n_T \times n_R}$ tale che $\tilde{\mathbf{s}} = \mathbf{G}\mathbf{r}$.

⁵⁶Vedi ad es. B. Hassibi, H. Vikalo, On the Sphere-Decoding Algorithm, Expected Complexity, ma anche Mathworks, comm.SphereDecoder. In estrema sintesi, questa tecnica riduce la complessità della ricerca del ricevitore ML limitandola ai vettori \mathbf{s} che ricadono (in qualche modo) all'interno di una sfera di raggio fisso e centrata sul vettore ricevuto \mathbf{r} . All'aumentare del raggio aumenta la complessità ed al limite si ottiene la soluzione ML, ma anche per complessità ridotte la soluzione trovata non si discosta molto da quella ottima.

Nel caso generale in cui $n_R \geq n_T$ si ottiene

$$\mathbf{G}_{ZF} = (\mathbf{H}^\dagger \mathbf{H})^{-1} \mathbf{H}^\dagger = \mathbf{H}_{PI} \quad (21.54)$$

nota anche come *pseudo-inversa*⁵⁷ \mathbf{H}_{PI} di \mathbf{H} , che qualora \mathbf{H} sia quadrata ($n_R = n_T$) ed invertibile si riduce a $\mathbf{G} = \mathbf{H}^{-1}$. In entrambi i casi la (21.53) fornisce

$$\tilde{\mathbf{s}} = \mathbf{G}_{ZF} \mathbf{r} = (\mathbf{H}^\dagger \mathbf{H})^{-1} \mathbf{H}^\dagger (\mathbf{H} \mathbf{s} + \mathbf{n}) = \mathbf{s} + (\mathbf{H}^\dagger \mathbf{H})^{-1} \mathbf{H}^\dagger \mathbf{n} \quad (21.55)$$

ottenendo l'effetto di invertire \mathbf{H} , come dire *annullare*, da cui il nome *forzante a zero*. Il costo computazionale associato è circa cubico in n_T per matrici quadrate, ma ottenere $\hat{\mathbf{s}} \in \mathcal{A}^{n_T}$ da $\tilde{\mathbf{s}}$ applicando la (21.52) è lineare in n_T .

Notiamo ora che, in perfetta analogia con quanto osservato a pag. 621 per l'omonimo equalizzatore, il valore $\tilde{\mathbf{s}}$ ottenuto mediante la (21.55) è gravato da un termine di rumore $\tilde{\mathbf{n}} = (\mathbf{H}^\dagger \mathbf{H})^{-1} \mathbf{H}^\dagger \mathbf{n}$ le cui componenti sono *correlate* tra antenne riceventi, e la cui entità può risultare assai rilevante qualora \mathbf{H} sia *mal condizionata*⁵⁸ ovvero $\mathbf{H}^\dagger \mathbf{H}$ presenti autovalori per cui $\frac{|\lambda_{\max}|}{|\lambda_{\min}|} \gg 1$, in quanto⁵⁹

$$\begin{aligned} \Sigma_{\tilde{\mathbf{n}}} &= E\{\tilde{\mathbf{n}}\tilde{\mathbf{n}}^\dagger\} = E\{(\mathbf{H}^\dagger \mathbf{H})^{-1} \mathbf{H}^\dagger \mathbf{n} \mathbf{n}^\dagger \mathbf{H} (\mathbf{H} \mathbf{H}^\dagger)^{-1}\} = \\ &= \sigma_n^2 (\mathbf{H}^\dagger \mathbf{H})^{-1} \mathbf{H}^\dagger \mathbf{H} (\mathbf{H}^\dagger \mathbf{H})^{-1} = \sigma_n^2 (\mathbf{H}^\dagger \mathbf{H})^{-1} \end{aligned}$$

e l'inversione di $\mathbf{H} \mathbf{H}^\dagger$ comporta la divisione per il suo determinante⁶⁰, che essendo pari al prodotto degli autovalori può nelle condizioni indicate risultare molto piccolo; in particolare, il rumore è particolarmente amplificato per le antenne trasmettenti che corrispondono alle colonne di \mathbf{H} associate ai valori singolari $\lambda_i^{1/2}$ più piccoli.

21.5.3 Ricevitore lineare a minimo errore medio quadratico L-MMSE

Per ovviare al problema evidenziato si può adottare il diverso criterio di trovare la matrice \mathbf{G} (con cui calcolare $\tilde{\mathbf{s}} = \mathbf{G} \mathbf{r}$) che rende minimo l'*errore quadratico medio* (MSE), ovvero

$$\mathbf{G}^\diamond = \arg \min_{\mathbf{G}} E\{|\mathbf{s} - \mathbf{G} \mathbf{r}|^2\} \quad (21.56)$$

Si dimostra⁶¹ che la matrice \mathbf{G} che verifica la (21.56) deve necessariamente soddisfare

⁵⁷In questo testo la definizione di pseudo inversa è scaturita all'eq. (7.39) nel contesto della *regressione lineare multipla* (§ 7.7.1), con l'analogia che diviene evidente qualora si confronti la (21.53) con la (7.38) e si sostituisca $\mathbf{r}, \mathbf{H}, \mathbf{s}$ con $\mathbf{y}, \mathbf{X}, \beta$. La discussione ivi svolta nel nostro caso significa che se fosse stato trasmesso $\tilde{\mathbf{s}}$ che soddisfa (21.53) avremmo ricevuto $\tilde{\mathbf{r}} = \mathbf{H} \tilde{\mathbf{s}} + \mathbf{n}$ vincolato (a parte per il rumore) a giacere nello spazio esplorato dalle colonne di \mathbf{H} , e dunque la differenza $\boldsymbol{\epsilon} = \tilde{\mathbf{r}} - \mathbf{r} = \mathbf{H}(\tilde{\mathbf{s}} - \mathbf{s})$ deve essere *ortogonale* alle stesse colonne, e quindi $\mathbf{H}^\dagger \cdot \boldsymbol{\epsilon} = \mathbf{0}$, dopodiché i passaggi sono gli stessi utilizzando l'hermitiano \dagger anziché il trasposto \top .

⁵⁸Vedi ad es. https://en.wikipedia.org/wiki/Condition_number

⁵⁹Alla seconda eguaglianza manca il passaggio

$$\tilde{\mathbf{n}}^\dagger = ((\mathbf{H}^\dagger \mathbf{H})^{-1} \mathbf{H}^\dagger \mathbf{n})^\dagger = \mathbf{n}^\dagger ((\mathbf{H}^\dagger \mathbf{H})^{-1} \mathbf{H}^\dagger)^\dagger = \mathbf{n}^\dagger \mathbf{H} ((\mathbf{H}^\dagger \mathbf{H})^{-1})^\dagger = \mathbf{n}^\dagger \mathbf{H} (\mathbf{H}^\dagger \mathbf{H})^{-1}$$

dato che $(\mathbf{A}^{-1})^\dagger = (\mathbf{A}^\dagger)^{-1}$

⁶⁰Vedi ad es. https://it.wikipedia.org/wiki/Matrice_invertibile

⁶¹Indicando l'argomento interno di (21.56) come l'errore $\boldsymbol{\epsilon}(\mathbf{G}) = \mathbf{s} - \mathbf{G} \mathbf{r}$ ed il suo valore atteso quadratico come $J(\mathbf{G}) = E\{\boldsymbol{\epsilon}^2\} = E\{\boldsymbol{\epsilon}^\dagger \boldsymbol{\epsilon}\}$, quando J è minimo tutte le componenti del suo gradiente devono annullarsi, ovvero $\nabla_{\mathbf{G}} J(\mathbf{G})|_{\mathbf{G}=\mathbf{G}^\diamond} = \mathbf{0}_{n_T \times n_R}$ (cioè si annullano le derivate di J rispetto a tutti gli

anche la relazione

$$E \{ (\mathbf{s} - \mathbf{G}^\circ \mathbf{r}) \mathbf{r}^\dagger \} = \mathbf{0}_{n_T \times n_R} \quad (21.57)$$

nota come principio di *ortogonalità*⁶², sviluppando la quale (dopo un po' di conti) si perviene al risultato⁶³

$$\mathbf{G}_{MMSE} = \mathbf{G}^\circ = \left(\mathbf{H}\mathbf{H}^\dagger + \frac{1}{\rho} \mathbf{I} \right)^{-1} \mathbf{H}^\dagger \quad (21.58)$$

in cui

$$\rho = \frac{\mathcal{E}_s}{\sigma_n^2} = \frac{\mathcal{P}_T}{f_s n_T} \frac{n_R}{E\{\mathbf{n}^\dagger \mathbf{n}\}}$$

è l'*SNR* per simbolo e per antenna ricevente, comprensivo dell'eventuale guadagno di diversità n_R/n_T . Notiamo quindi che per ρ elevato la (21.58) equivale alla (21.54), cancellando gli interferenti ma senza considerare il rumore; qualora invece ρ tenda a zero la (21.58) tende ad \mathbf{H}^\dagger , un risultato simile al *MRC*⁶⁴ (§ 21.3.1.2), e che non considera gli interferenti. Dopodiché il valore di $\hat{\mathbf{s}} \in \mathcal{A}^{n_T}$ si ottiene applicando anche per questo caso la (21.52).

21.5.4 Ricevitore a cancellazioni successive - **vBLAST**

Nei casi precedenti la decisione da parte del ricevitore in merito al simbolo s_i trasmesso dall' i -esima antenna viene presa *congiuntamente* a tutti gli altri, calcolando il vettore

elementi di \mathbf{G}). Possiamo quindi scrivere

$$\frac{\partial J}{\partial \mathbf{G}} = \frac{\partial}{\partial \mathbf{G}} E\{\boldsymbol{\epsilon}^2\} = 2E\{\boldsymbol{\epsilon} \cdot \frac{\partial \boldsymbol{\epsilon}^\dagger}{\partial \mathbf{G}}\} = 0$$

e dato che $\frac{\partial \boldsymbol{\epsilon}^\dagger}{\partial \mathbf{G}} = \frac{\partial}{\partial \mathbf{G}} (\mathbf{s} - \mathbf{G}\mathbf{r})^\dagger = -\mathbf{r}^\dagger$ si ottiene J minimo quando

$$E\{\boldsymbol{\epsilon}_{MMSE} \cdot \frac{\partial \boldsymbol{\epsilon}^\dagger}{\partial \mathbf{G}}\} = E\{(\mathbf{s} - \mathbf{G}^\circ \mathbf{r}) \mathbf{r}^\dagger\} = \mathbf{0}_{n_T \times n_R}$$

⁶²In effetti ciò che la (21.57) afferma è l'incorrelazione tra ogni componente del vettore di errore minimo rispetto ad ogni componente di quello ricevuto, ma vedi anche https://en.wikipedia.org/wiki/Orthogonality_principle.

⁶³A partire dalla (21.57) otteniamo $E\{(\mathbf{s} - \mathbf{G}^\circ \mathbf{r}) \mathbf{r}^\dagger\} = E\{\mathbf{s}\mathbf{r}^\dagger - \mathbf{G}^\circ \mathbf{r}\mathbf{r}^\dagger\} = \boldsymbol{\Sigma}_{SR} - \mathbf{G}^\circ \boldsymbol{\Sigma}_{RR} = 0$ per cui deve risultare $\mathbf{G}^\circ = \boldsymbol{\Sigma}_{SR}/\boldsymbol{\Sigma}_{RR}$, dove le matrici di covarianza $\boldsymbol{\Sigma}$ coincidono con quelle di correlazione, dato il valore atteso nullo di \mathbf{s} e di \mathbf{r} . Calcoliamone il valore:

$$\begin{aligned} \boldsymbol{\Sigma}_{SR} &= E\{\mathbf{s}\mathbf{r}^\dagger\} = E\{\mathbf{s}(\mathbf{H}\mathbf{s} + \mathbf{n})^\dagger\} = E\{\mathbf{s}\mathbf{s}^\dagger \mathbf{H}^\dagger\} + E\{\mathbf{s}\mathbf{n}^\dagger\} = \boldsymbol{\Sigma}_S \mathbf{H}^\dagger = \mathcal{E}_s \mathbf{H}^\dagger \\ \boldsymbol{\Sigma}_{RR} &= E\{(\mathbf{H}\mathbf{s} + \mathbf{n})(\mathbf{H}\mathbf{s} + \mathbf{n})^\dagger\} = E\{\mathbf{H}\mathbf{s}\mathbf{s}^\dagger \mathbf{H}^\dagger + \mathbf{H}\mathbf{s}\mathbf{n}^\dagger + \mathbf{n}\mathbf{s}^\dagger \mathbf{H}^\dagger + \mathbf{n}\mathbf{n}^\dagger\} = \mathbf{H}\boldsymbol{\Sigma}_S \mathbf{H}^\dagger + \boldsymbol{\Sigma}_N = \mathcal{E}_s \mathbf{H}\mathbf{H}^\dagger + \sigma_n^2 \mathbf{I} \end{aligned}$$

dato che $E\{\mathbf{s}\mathbf{n}^\dagger\}$, $E\{\mathbf{H}\mathbf{s}\mathbf{n}^\dagger\}$ e $E\{\mathbf{n}\mathbf{s}^\dagger \mathbf{H}^\dagger\}$ sono nulli in virtù dell'incorrelazione tra \mathbf{s} ed \mathbf{n} , e si è posto $\boldsymbol{\Sigma}_S = \mathcal{E}_s \mathbf{I}$ e $\boldsymbol{\Sigma}_N = \sigma_n^2 \mathbf{I}$ in virtù dell'incorrelazione tra simboli delle diverse antenne, nonché tra campioni di rumore, in cui $\mathcal{E}_s = E\{s_i^* s_i\} = \mathcal{P}_T/f_s n_T$ è pari all'energia per simbolo e per antenna, uguale per tutte, mentre $\sigma_n^2 = E\{n_i^* n_i\} = 1/n_R E\{\mathbf{n}^\dagger \mathbf{n}\}$ è la potenza del campione di rumore in ingresso all' i -esima antenna ricevente. Sostituendo ora $\boldsymbol{\Sigma}_{SR}$ e $\boldsymbol{\Sigma}_{RR}$ nella relazione $\mathbf{G}^\circ = \boldsymbol{\Sigma}_{SR}/\boldsymbol{\Sigma}_{RR}$ otteniamo il risultato cercato

$$\mathbf{G}^\circ = \frac{\boldsymbol{\Sigma}_{SR}}{\boldsymbol{\Sigma}_{RR}} = \frac{\mathcal{E}_s \mathbf{H}^\dagger}{\mathcal{E}_s \mathbf{H}\mathbf{H}^\dagger + \sigma_n^2 \mathbf{I}} = \frac{\mathbf{H}^\dagger}{\mathbf{H}\mathbf{H}^\dagger + \frac{\sigma_n^2}{\mathcal{E}_s} \mathbf{I}}$$

⁶⁴Nel senso che il vettore di pesi \mathbf{w} con coefficienti (21.13) è il coniugato dell'unica colonna di \mathbf{H} presente nel caso SIMO, come ora la riga i -esima di \mathbf{H}^\dagger è coniugata dell' i -esima colonna di \mathbf{H} - ma vedi anche la discussione seguente.

$\tilde{\mathbf{s}} = \mathbf{G}\mathbf{r}$ mediante una unica operazione. Al contrario, l'approccio che stiamo per illustrare⁶⁵ decodifica i simboli s_i antenna (trasmittente) per antenna, *cancellando* ogni volta il contributo ad \mathbf{r} dovuto ai simboli già decodificati.

Prima di procedere torniamo all'espressione $\tilde{\mathbf{s}} = \mathbf{G}\mathbf{r}$ in cui \mathbf{G} può essere ottenuta con il metodo ZF oppure MMSE, ma specializzando il ragionamento al caso ZF osserviamo che il valore per l'elemento \tilde{s}_i si ottiene da \mathbf{r} come

$$\tilde{s}_i = \mathbf{w}_i^T \mathbf{r} \quad (21.59)$$

in cui \mathbf{w}_i è il vettore con elementi pari alla i -esima riga di \mathbf{G} , che indichiamo come $(\mathbf{G})_i$, ovvero $\mathbf{w}_i = (\mathbf{G})_i^T$. Ebbene, accade che \mathbf{w}_i è *ortogonale* a tutte le colonne di \mathbf{H} tranne l' i -esima: ciò è evidente nel caso di matrice \mathbf{H} quadrata, per la quale si ottiene $\mathbf{G}_{ZF} = \mathbf{H}^{-1}$ e dunque ogni sua riga i è ortogonale alle colonne $j \neq i$ di \mathbf{H} in quanto $\mathbf{H}^{-1}\mathbf{H} = \mathbf{I}$. Ma anche se \mathbf{H} non è quadrata, la sua pseudo-inversa $\mathbf{G}_{ZF} = \mathbf{H}_{PI}$ data dalla (21.54) ha righe ortogonali alle colonne di \mathbf{H} non omologhe, risultando⁶⁶ $\mathbf{H}_{PI}\mathbf{H} = \mathbf{I}$. Tale proprietà di ortogonalità è alla base dell'effetto di inversione del sistema ottenuto con la (21.55).

Consideriamo ora che dopo la decodifica del primo simbolo ($i = 1$)

$$\tilde{s}_i = \mathbf{w}_i^T \mathbf{r}$$

ed alla sua quantizzazione in $\hat{s}_i \in \mathcal{A}$, nel caso in cui l'SNR sia sufficiente ad avere una P_e abbastanza piccola si può procedere alla *cancellazione* dell'effetto di \hat{s}_i su \mathbf{r} , semplicemente valutando⁶⁷ un nuovo vettore

$$\mathbf{r}_{i+1} = \mathbf{r} - \hat{s}_i (\mathbf{H})_i$$

in cui $(\mathbf{H})_i$ è l' i -esima colonna di \mathbf{H} : per il nuovo \mathbf{r}_{i+1} è quindi come se l'antenna i -esima fosse *spenta*, ovvero come se la colonna $(\mathbf{H})_i$ fosse tutta nulla: indichiamo quindi con \mathbf{H}_0^i la matrice \mathbf{H} a cui sono state *azzerate* le colonne $j = 1, 2, \dots, i$. Il vettore \mathbf{w}_{i+1} con cui decodificare il successivo simbolo

$$\tilde{s}_{i+1} = \mathbf{w}_{i+1}^T \mathbf{r}_{i+1}$$

deve quindi essere ortogonale *non a tutte* le colonne di \mathbf{H} , ma solo a quelle non nulle di \mathbf{H}_0^i , coerentemente con il ridotto numero di interferenti ancora presenti nel vettore \mathbf{r}_{i+1} . Il vettore \mathbf{w}_{i+1}^T viene quindi posto pari alla $(i+1)$ -esima *riga della pseudo-inversa* di \mathbf{H}_0^i .

Procedendo in questo modo i simboli decodificati per ultimi godono di un ordine di diversità maggiore di quelli decodificati per primi, in quanto il numero di antenne trasmittenti *si riduce* via via. D'altra parte nel caso in cui la P_e non sia trascurabile si

⁶⁵Noto come V-BLAST, vedi P.W. WOLNIANSKY ET AL, *V-BLAST: An Architecture for Realizing Very High Data Rates Over the Rich-Scattering Wireless Channel*, 1998 URSI Int. Symposium Conference Proceedings, reperibile presso <https://www.ee.columbia.edu/~jiantan/E6909/wolnianskyandfoschini.pdf>

⁶⁶Vedi ad es. <https://it.wikipedia.org/wiki/Pseudo-inversa>

⁶⁷Infatti (a parte il rumore) è possibile scrivere $\mathbf{r} = \sum_{k=1}^{n_T} s_k (\mathbf{H})_k$, dunque

$$\mathbf{r}' = \mathbf{r} - \hat{s}_i (\mathbf{H})_i = \sum_{\substack{k=1 \\ k \neq i}}^{n_T} s_k (\mathbf{H})_k$$

può verificare un effetto di propagazione degli errori, specialmente se questi avvengono nei simboli decodificati per primi. Occorre quindi stabilire un criterio di ordinamento in grado di garantire le migliori prestazioni: si dimostra che l'ordinamento ottimale si ottiene scegliendo ogni volta l'indice i_M dell'antenna che trasmette il simbolo s_{i_M} per il quale l'SNR dopo decodifica è il più grande, ovvero $i_M = \arg \max_i \text{SNR}(i)$. Si può altresì dimostrare⁶⁸ che

$$\text{SNR}(i) = \frac{E\{|s_i|^2\}}{\sigma_n^2 |\mathbf{w}_i|^2} \quad (21.60)$$

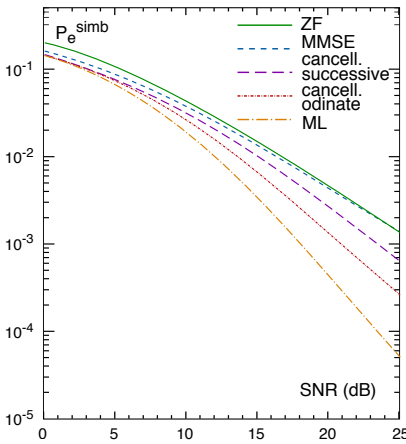


Figura 21.4: Prestazioni della multiplazione spaziale

e quindi se ogni antenna adotta la stessa costellazione \mathcal{A} la massimizzazione di (21.60) si ottiene scegliendo $i_M = \arg \min_i |\mathbf{w}_i|^2$; dato che \mathbf{w}_i è pari alla i -esima riga della pseudo-inversa di \mathbf{H}_i^0 , quest'ultima dopo essere stata ricalcolata ad ogni iterazione, viene esplorata per individuare la riga con norma minore.

Qualora anziché la pseudo inversa dello ZF si adotti la soluzione MMSE data dalla (21.58), il metodo è ancora applicabile, con un lieve vantaggio per SNR basso, come mostrato nella figura a fianco⁶⁹, che confronta le prestazioni per diverse opportunità di decodifica di un sistema MIMO 2×2 con modulazione QPSK per una tipica matrice \mathbf{H} .

21.5.5 Compromesso diversità - multiplazione

Quando in ricezione si dispone di un numero di antenne maggiore del minimo (ossia $n_R > n_T$) le antenne *in più* non contribuiscono al guadagno di multiplazione (al massimo pari a $\min(n_T, n_R)$), ma consentono di conseguire un ordine di diversità *almeno* pari a $n_R - n_T + 1$ (con il metodo di cancellazione ordinata anche di più, dato che in pratica n_T *diminuisce* man mano). Ma anche se $n_R = n_T$ è comunque possibile (e necessario se $n_R < n_T$) sacrificare parte della velocità di trasmissione per migliorare le prestazioni nei confronti del rumore, sfruttando la diversità spaziale frutto del fading indipendente. Si può infatti ad esempio scegliere di dimezzare il numero di antenne usate per la multiplazione, ed impiegare la restante metà per trasmettere la ridondanza associata ad un codice spazio-tempo (§ 21.3.2.2). Ma anziché provare le

⁶⁸Sostituendo in $\tilde{s}_i = \mathbf{w}_i^T \mathbf{r}$ l'espressione del vettore ricevuto $\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{n}$ otteniamo

$$\tilde{s}_i = \mathbf{w}_i^T \sum_{k=1}^{n_T} s_k (\mathbf{H})_k + \mathbf{w}_i^T \mathbf{n} = s_i \mathbf{w}_i^T (\mathbf{H})_i + \mathbf{w}_i^T \mathbf{n} = s_i + \mathbf{w}_i^T \mathbf{n}$$

in cui si tiene conto dell'ortogonalità tra \mathbf{w}_i e le colonne $(\mathbf{H})_j$ con $j \neq i$, e del fatto che $\mathbf{w}_i^T (\mathbf{H})_i = 1$ essendo \mathbf{w}_i^T pari all' i -esima riga della pseudoinversa di \mathbf{H} . Pertanto l'SNR risulta pari a

$$\rho = \frac{E\{|s_i|^2\}}{E\{|\mathbf{w}_i^T \mathbf{n}|^2\}} = \frac{E\{|s_i|^2\}}{E\{\mathbf{w}_i^T \mathbf{n} \mathbf{n}^T \mathbf{w}_i\}} = \frac{E\{|s_i|^2\}}{\sigma_n^2 \mathbf{w}_i^T \mathbf{w}_i}$$

⁶⁹In realtà la figura è frutto di un *missaggio* di due prelevate da lavori differenti, ed i valori mostrati sono da ritenersi indicativi e non esatti, oltre che frutto della \mathbf{H} adottata.

diverse combinazioni possibili, riferiamo di un risultato generale⁷⁰. La relazione tra guadagno di moltiplicazione spaziale r e di diversità d può essere ottenuta dopo avere definito queste due grandezze nei termini della rispettiva legge di dipendenza *asintotica* da SNR , ovvero

$$r = \lim_{SNR \rightarrow \infty} \frac{R(SNR)}{\log_2 SNR}; \quad d = - \lim_{SNR \rightarrow \infty} \frac{\log P_e(SNR)}{\log SNR} \quad (21.61)$$

rappresentando così il risultato che per SNR elevato la velocità di trasmissione R aumenta come $r \log_2 SNR$ ⁽⁷¹⁾, mentre la probabilità di errore decade come $1/SNR^d$ ⁽⁷²⁾.

In questo modo può essere tracciata la curva $d(r)$ mostrata in fig. 21.5 e che *racconta* i valori estremi di massima diversità $d_{\max} = n_T \cdot n_R$ e massima moltiplicazione $r_{\max} = \min(n_T, n_R)$ mediante la relazione

$$d_{opt}(r) = (n_T - r)(n_R - r) \quad \text{con} \quad r = 0, 1, \dots, \min(n_T, n_R) \quad (21.62)$$

che individua il *massimo* ordine di diversità conseguibile per un guadagno di moltiplicazione assegnato.

Esempio Desiderando ottenere una velocità R di 15 bps/Hz da un sistema MIMO con $n_T = n_R = 8$ ed $SNR = 15$ dB, qual'è il massimo guadagno di diversità d_{opt} che il sistema può offrire? **R:** dalla relazione $R = r \log_2 SNR$ con $SNR = 15$ dB otteniamo $15 = r \log_2 10^{1.5}$ e dunque $r = 3.01 \approx 3$. Pertanto in linea teorica 3 delle 8 antenne sono usate per la moltiplicazione, e 5 per la diversità, con un guadagno di diversità che *al massimo* può arrivare a $d_{opt}(r) = (n_T - r)(n_R - r) = (8 - 3)(8 - 3) = 25$. Nella pratica *naif*, se si adotta un codice di Alamouti che utilizza 6 antenne per trasmettere 3 flussi codificati, ognuno dei quali è ricevuto da due antenne, l'ordine di diversità per ogni flusso risulta pari a $2n_R = 2 \cdot 2 = 4$ (vedi § 21.3.2.4), e dunque complessivamente pari solamente a $3 \cdot 4 = 12$. In compenso, la potenza non usata dalle 2 antenne spente in trasmissione può essere ridistribuita sulle 6 attive, così come le due antenne *in più* di ricezione possono essere impiegate nella ricezione di 2 dei 3 flussi, portando l'ordine di diversità complessivo a $6 + 6 + 4 = 16$. Alla nota⁷³ alcuni approcci in grado di avvicinarsi di più al risultato (21.62).

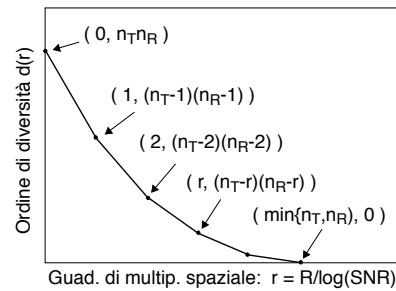


Figura 21.5: Compromesso diversità - moltiplicazione

⁷⁰Vedi L.ZHENG, D.N.C.TSE, *Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels*, IEEE Trans. on Inf. Theory, May 2003, dove sono riportate le considerazioni e le ipotesi che determinano il risultato; trovo una copia *libera* presso l'Univ. di Stanford.

⁷¹Dato che la capacità per simbolo $C = \log_2(1 + SNR)$ di un canale siso tende a $\log_2 SNR$ con $SNR \rightarrow \infty$, il valore r nella (21.61) esprime il numero di canali siso equivalenti.

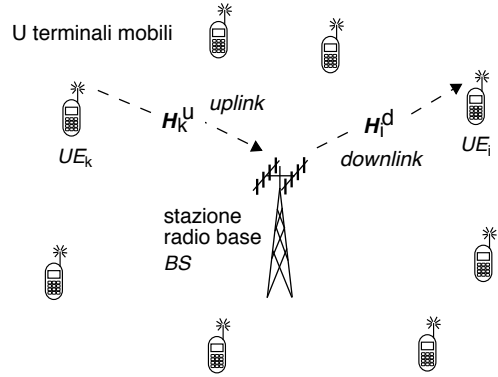
⁷²Probabilmente ci vorrebbe qualche passaggio in più, ma la relazione $P_e \propto 1/SNR^d$ deriva dal confronto tra la (21.15) del ricevitore MRC che sfrutta la diversità, e la (20.35) valida per un canale siso.

⁷³S.SFAR, L.DAI, K.B.LETAIEF, *Optimal Diversity-Multiplexing Tradeoff With Group Detection for MIMO Systems*, IEEE Tr. on Comm. 2005; P. ELIA ed al, *Explicit Space-Time Codes Achieving The Diversity-Multiplexing Gain Tradeoff*, IEEE Tr. on Inf. Th. 2006

21.6 Trasmissione multiutente o MU - MIMO

Fino ad ora abbiamo discusso di collegamenti MIMO punto-punto, ma in generale esiste una forte asimmetria nel numero antenne presenti ai due lati di un collegamento *radiomobile*, che sono molte di più presso una *stazione radio base*⁷⁴ o BS (il cui numero, decine o anche centinaia, indichiamo ora come n_{BS}) rispetto a quelle (che ora chiamiamo n_k) di cui è equipaggiato il k -esimo *User Equipment* UE_k (o terminale, o dispositivo) ad essa *associato* - che in gran parte dei casi, dispone di una sola antenna.

Come poter sfruttare in questo caso l'elevata velocità di trasmissione della BS, resa possibile dalla multiplexazione spaziale di cui è capace grazie alle sue n_{BS} antenne? Molto semplice: dedicando ad ogni diverso terminale mobile UE_k una (o più) differente antenna di trasmissione della BS, realizzando una sorta di MIMO *distribuito*, anche se il passaggio non è così *indolore*, come andiamo ad approfondire dopo aver distinto tra le due direzioni di trasmissione, indicando come *uplink* il verso tra i terminali e la BS, e come *downlink* la direzione opposta, il cui insieme dei singoli collegamenti è indicato anche come *canale broadcast*.



Uplink In questa direzione le cose non sono molto diverse da quanto già approfondito. La BS riceve un segnale⁷⁵

$$\mathbf{r}_{BS} = \sum_{k=1}^U \mathbf{H}_k^u \mathbf{s}_k + \mathbf{n} = \mathbf{H}^u \mathbf{s} + \mathbf{n}$$

in cui \mathbf{H}_k^u è la matrice ($N \times n_k$) di canale MIMO di *uplink* tra UE_k e BS, \mathbf{s}_k è il vettore ($n_k \times 1$) dei simboli inviati da UE_k , $\mathbf{H}^u = [\mathbf{H}_1^u \ \mathbf{H}_2^u \ \dots \ \mathbf{H}_U^u]$ è la matrice ($n_{BS} \times \sum_{k=1}^U n_k$) di *uplink* da *tutti* gli U terminali UE_k , ed $\mathbf{s} = [\mathbf{s}_1^T \ \mathbf{s}_2^T \ \dots \ \mathbf{s}_U^T]^T$ è il vettore dei simboli complessivamente trasmessi.

Ogni UE_k con più di una antenna può trasmettere più flussi dati ricorrendo alla multiplexazione spaziale, oppure se $n_k = 1$ il singolo collegamento è di tipo SIMO. In entrambi i casi la BS deve *separare* tra loro i contributi ricevuti dagli U utenti adottando le tecniche illustrate al § 21.5 e basate sulla conoscenza della \mathbf{H}^u di *uplink*, che la BS può stimare nei confronti degli UE_k attivi. Fortunatamente la dispersione spaziale degli utenti fa sì che \mathbf{H}^u sia ben condizionata⁷⁶, senza dunque porre problemi per la sua

⁷⁴Il punto accesso di una rete di telefonia mobile, o di una rete WiMax, o l'access point di una rete WiFi casalinga, vedi ad es. https://it.wikipedia.org/wiki/Stazione_radio_base, <https://it.wikipedia.org/wiki/WiMAX>, https://it.wikipedia.org/wiki/Access_point

⁷⁵Supponiamo qui che i terminali mobili siano sincronizzati temporalmente, e che il sistema sia in grado di compensare i diversi ritardi di propagazione per la corretta determinazione degli istanti di campionamento dei simboli.

⁷⁶Infatti ogni canale utente-BS sperimenta cammini multipli completamente differenti ed incorrelati, rendendo le colonne di \mathbf{H} indipendenti, al punto da definirle come una *firma spaziale* di ciascun utente, dando al sistema di trasmissione la denominazione di *space division multiple access* o SDMA.

inversione.

Downlink In questa direzione il generico dispositivo UE_k con n_k antenne riceve un vettore \mathbf{r}_k di dimensione $n_k \times 1$ pari a

$$\mathbf{r}_k = \mathbf{H}_k^d \mathbf{s} + \mathbf{n} \quad (21.63)$$

in cui \mathbf{H}_k^d ($n_k \times n_{BS}$) individua il canale MIMO di broadcast tra tutte le antenne di BS e le n_k di UE_k, e \mathbf{s} ($n_{BS} \times 1$) è il vettore di simboli destinati a tutti gli U utenti. Purtroppo essendo $n_k < n_{BS}$, il dispositivo mobile non può adottare le soluzioni previste al § 21.5 per cancellare i termini di interferenza, più evidenti particolarizzando la (21.63) al caso di un terminale ad antenna singola, che riceve un valore

$$r_k = \mathbf{h}_k^T \mathbf{s} + n = h_k s_k + \sum_{\substack{h=1 \\ h \neq k}}^U h_h s_h + n \quad (21.64)$$

in cui \mathbf{h}_k è il vettore⁷⁷ ($n_{BS}, 1$) dei guadagni complessi tra le n_{BS} antenne di trasmissione e l'unica di ricezione, h_k è l'elemento di \mathbf{h}_k corrispondente all'antenna della BS che trasmette il simbolo s_k diretto all'utente k , ed il termine $\sum_{h=1, h \neq k}^U h_h s_h$ individua l'interferenza da parte degli altri e che si aggiunge al contributo del rumore⁷⁸, in modo che si possa parlare di rapporto segnale / interferenti più rumore o SINR.

21.6.1 Precodifica

Nell'impossibilità per i terminali mobili con antenna singola di sopprimere gli interferenti, occorre trovare una soluzione semplice⁷⁹ che possa essere attuata presso la BS: fortunatamente questa esiste ed è la duale di quella esposta ai § 21.5.2-21.5.3, necessitando come preconditione la conoscenza da parte della BS di tutti i vettori \mathbf{h}_k visti dagli utenti, in modo da poter ricostruire la matrice

$$\mathbf{H}^d = [\mathbf{h}_1 \mathbf{h}_2 \cdots \mathbf{h}_U]^T \quad (21.65)$$

di dimensioni $U \times n_{BS}$ e che caratterizza il canale broadcast del downlink tra le antenne della BS e quelle (singole) degli U utenti.

Zero forcing precoding Una volta nota \mathbf{H}^d , la BS ne può calcolare la pseudo inversa (eq. (21.54))

$$\mathbf{P} = (\mathbf{H}^{d \dagger} \mathbf{H}^d)^{-1} \mathbf{H}^{d \dagger} \quad (21.66)$$

e sostituire al vettore di simboli \mathbf{s} da inviare agli U terminali un nuovo vettore precodificato

$$\mathbf{s}_P = \mathbf{P} \mathbf{s} = \sum_{h=1}^U \mathbf{p}_h s_h$$

dove \mathbf{p}_h è pari alla h -esima colonna di \mathbf{P} , di fatto spargendo ogni simbolo s_h destinato a ciascun utente h su tutte le n_{BS} antenne. Osserviamo ora che in virtù della definizione

⁷⁷Corrispondente all'unica riga di \mathbf{H}_k^d pertinente all'utente k qualora $n_k = 1$. Il vettore \mathbf{h}_k può essere pensato nel senso letterale di indicare una direzione (a coordinate complesse) nello spazio generato dalle antenne della BS, dunque non la direzione fisica dell'utente, ma che lo distingue comunque dagli altri.

⁷⁸A ben vedere, per un numero U elevato di utenti si applica il teorema centrale del limite, ed il termine $\sum_{h=1, h \neq k}^U h_h s_h$ tende ad una v.a. gaussiana.

⁷⁹La soluzione ottima è nota come dirty paper coding (DPC), di difficile implementazione, vedi ad es. https://it.wikipedia.org/wiki/Dirty_paper_coding

di pseudoinversa si ha $\mathbf{H}^d \mathbf{P} = \mathbf{I}$, e dunque ogni colonna \mathbf{p}_k di \mathbf{P} è ortogonale a tutte le righe di \mathbf{H}^d tranne che alla k -esima, ovvero $\mathbf{h}_k^T \mathbf{p}_h = \delta_{hk}$, in modo che la (21.64) diviene

$$r_k = \mathbf{h}_k^T \mathbf{s} \mathbf{P} + n = \mathbf{h}_k^T \sum_{h=1}^U \mathbf{p}_h s_h + n = \mathbf{h}_k^T \mathbf{p}_k s_k + n = s_k + n \quad (21.67)$$

da cui *scompare* il termine di interferenza da parte dei simboli s_h destinati agli altri utenti, ed il simbolo ricevuto può essere stimato come $\hat{s}_k = \arg \min_{s \in \mathcal{A}} |s - s_k|^2$.

MMSE precoding Tornando all'espressione (21.64), occorre distinguere il caso in cui le prestazioni del sistema sono *limitate dal rumore* e cioè quando $E\{|n|^2\} > E\{|\sum_{h=1, h \neq k}^U \mathbf{h}_h s_h|^2\}$, dal caso in cui sono invece limitate *dagli interferenti*, dove la disuguaglianza cambia verso. La precodifica (21.66) si comporta bene solo nel secondo caso, mentre ponendo come obiettivo della matrice \mathbf{P} quello di rendere minimo l'errore medio quadratico $E\{|s - s_k|^2\}$ si ottiene una soluzione che pur non cancellando completamente gli interferenti permette di bilanciare la riduzione della loro potenza con la riduzione della potenza di rumore; tale soluzione è indicata come versione *regolarizzata* di (21.66) e fornisce

$$\mathbf{P}_{MMSE} = (\mathbf{H}^{d \dagger} \mathbf{H}^d + \frac{1}{\rho} \mathbf{I})^{-1} \mathbf{H}^{d \dagger} \quad (21.68)$$

in cui $\rho = E\{|\mathbf{P}\mathbf{s}|^2\}/E\{|n|^2\}$ è una misura dell'*SNR*.

21.6.2 Controllo di potenza

I valori ottenuti per le colonne della matrice di precodifica \mathbf{P} devono essere alterati, essenzialmente per due motivi:

- rispettare il vincolo sulla potenza trasmessa $P_T = E\{|\mathbf{P}\mathbf{s}|^2\} \cdot f_s$;
- rendere massimo un criterio di qualità complessiva come ad es. la velocità aggregata, cioè la somma di quella per ciascun utente.

Il secondo obiettivo è quello più *intrigante*, dato che aumentare la potenza per un utente significa aumentare la capacità del suo canale, a discapito di quella degli altri per i quali si verifica invece un aumento della potenza interferente, e (per il vincolo sulla potenza complessiva) la riduzione della propria. Inoltre la potenza allocata ad ogni utente (ossia al simbolo a lui diretto) deve essere determinata in base all'*SNR* di ricezione, altra informazione che deve essere comunicata alla BS, in modo che quest'ultima possa attuare tecniche di *water filling*, ed assegnare più potenza a chi ha un *SNR* migliore. Non ci addentriamo nelle strategie di ottimizzazione congiunta, alcune delle quali iterative: questo è solamente un testo introduttivo.

21.6.3 Prioritizzazione degli utenti

Fino ad ora si è implicitamente supposto che $U < n_{BS}$, ma gli utenti che intendono comunicare (tra quelli registrati presso una BS) possono essere ovviamente più delle antenne a disposizione. Tutti i terminali che richiedono il servizio comunicano alla BS le informazioni sullo stato del proprio canale o CSI (\mathbf{h}_k^d, ρ_k) in modo da permetterle di calcolare la \mathbf{P} di precodifica. La BS deve quindi selezionare un sottoinsieme di

$U_M \leq n_{BS}$ utenti scegliendoli in modo da massimizzare la velocità somma; ma la velocità di ciascun utente ha come limite superiore la capacità del suo canale, che a sua volta dipende da relativo ρ_k . La ricerca esaustiva del migliore gruppo di $U_M < U$ utenti verso cui trasmettere ha complessità esponenziale, per cui è necessario ricorrere a strategie euristiche. Una tecnica (cosiddetta *greedy* o *vorace*) consiste nello scegliere per primo l'utente che può dare il maggior contributo di velocità, poi il secondo migliore, e così via. Ovviamente l'obiettivo da massimizzare, anziché tenere conto solamente della velocità-somma, può prendere in considerazione anche altri fattori, come la lunghezza di coda (da minimizzare) od altro ancora, vedi appresso.

21.6.4 Precodifica con feedback limitato

Il calcolo della matrice \mathbf{P} (eq. (21.66) o (21.68)) da parte della BS necessita della conoscenza da parte di quest'ultima della matrice \mathbf{H}^d (21.65), la cui riga \mathbf{h}_k^d descrive il canale tra BS e UE_k , stimato presso quest'ultimo. Nel caso di un sistema *full duplex* basato su portanti differenti per le due direzioni del collegamento il canale *non è reciproco* (nota 53), e quindi occorre un canale di uplink *di ritorno* attraverso il quale i terminali mobili possano trasmettere alla BS la loro stima di \mathbf{h}_k^d , codificata mediante B bit frutto della quantizzazione dei valori di \mathbf{h}_k^d e di ρ_k . Come noto dalla teoria dell'informazione (§ 9.5.2), una distorsione nulla si ottiene solamente per $B \rightarrow \infty$, dunque per B finito è inevitabile la presenza di errori negli \mathbf{h}_k^d quantizzati, che a loro volta causano errori nella determinazione delle colonne \mathbf{p}_k della matrice di precodifica. Scrivendo quindi $\tilde{\mathbf{p}}_k = \mathbf{p}_k + \mathbf{e}_k$ in cui \mathbf{p}_k e $\tilde{\mathbf{p}}_k$ sono i vettori di precodifica calcolati per una CSI esatta oppure quantizzata, la (21.67) diviene

$$\begin{aligned} r_k &= \mathbf{h}_k^T \sum_{h=1}^U \tilde{\mathbf{p}}_h s_h + n = \mathbf{h}_k^T \sum_{h=1}^U \mathbf{p}_h s_h + \mathbf{h}_k^T \sum_{h=1}^U \mathbf{e}_h s_h + n = \\ &= \mathbf{h}_k^T \mathbf{p}_k s_k + \mathbf{h}_k^T \sum_{h=1}^U \mathbf{e}_h s_h + n = s_k + n_I + n \end{aligned} \quad (21.69)$$

in cui $n_I = \mathbf{h}_k^T \sum_{h=1}^U \mathbf{e}_h s_h$ costituisce un termine di rumore *interferente*.

Legame feedback - SNR Il termine n_I nella (21.69) è tanto maggiore quanto peggiore è la risoluzione della quantizzazione di \mathbf{h}_k^d , ovvero quanto minore è il numero B di bit di feedback inviati da ciascun terminale. Mantenendo B fisso si produce un effetto *piattaforma* (da *error floor*) nelle prestazioni (sia di velocità che di errore) in funzione dell'*SNR*, dato che n_I non cambia anche se il rumore n si riduce. Il rimedio è quello di richiedere una maggiore accuratezza (ossia un maggior numero di bit B_k) nei valori trasmessi \mathbf{h}_k^d in forma quantizzata per quei terminali UE_k che sperimentano un rumore termico n minore⁸⁰, in modo da mantenere i due tipi di rumore (n e n_I) alla stessa potenza. Inoltre essendo i vettori \mathbf{h}_k e \mathbf{p}_k di dimensione n_{BS} , l'espressione $n_I = \mathbf{h}_k^T \sum_{h=1}^U \mathbf{e}_h s_h$ implica una somma di n_{BS} termini, cosicché anche il numero di antenne di trasmissione contribuisce in modo diretto ad n_I . Qualora l'*SNR* sia uguale per tutti, è stato trovato⁸¹ che la BS può conseguire un guadagno di multiploazione massimo

⁸⁰Ovvero un *SNR* ρ_k maggiore.

⁸¹N. JINDAL, *MIMO Broadcast Channels With Finite-Rate Feedback*, IEEE Trans. on Inf. Th. Nov. 2006, ne trovo una copia libera presso Univ. of Minnesota

(n_{BS}) quando ogni terminale invia

$$B = (n_{BS} - 1) \log_2 SNR \approx \frac{n_{BS} - 1}{3} SNR_{dB} \quad \text{bit} \quad (21.70)$$

soffrendo una penalizzazione di soli 3 dB di prestazione rispetto alla conoscenza perfetta della CSI.

Quantizzazione vettoriale del canale di downlink Approfondiamo un minimo come vengono prodotti i B bit di feedback. La QV, introdotta al § 10.1.2.4, è la tecnica per eseguire la codifica di sorgente $\tilde{\mathbf{h}}_k = Q(\mathbf{h}_k)$ della grandezza vettoriale \mathbf{h}_k . Consiste in un codebook⁸² $\mathcal{H} = \{\mathbf{h}_{q1}, \mathbf{h}_{q2}, \dots, \mathbf{h}_{q2^B}\}$ di 2^B vettori di canale a norma unitaria, tra cui si sceglie il vettore \mathbf{h}_{opt} più vicino a quello misurato \mathbf{h}_k (reso a norma unitaria anch'esso) usando come criterio di distorsione il *modulo quadro del prodotto scalare*⁸³. Si sceglie pertanto l'indice⁸⁴

$$n_{opt} = \arg \max_{\mathbf{h}_n \in \mathcal{H}} |\mathbf{h}_n^\dagger \mathbf{h}_k|^2 = \arg \max_{\mathbf{h}_n \in \mathcal{H}} \cos^2 [\angle(\mathbf{h}_n^\dagger, \mathbf{h}_k)]$$

decidendo quindi per $\tilde{\mathbf{h}}_k = Q(\mathbf{h}_k) = \mathbf{h}_{opt}$, e si trasmettono i B bit della codifica binaria di n_{opt} . Anche la BS è a conoscenza dello stesso CB \mathcal{H} , da cui preleva $\mathbf{h}_{n_{opt}}$ e lo usa per costruire \mathbf{H}^d da cui ricavare \mathbf{P} . Anche se i vettori del codebook sono a norma unitaria, si ottengono vettori di precoding (ad es., mediante *zero forcing*) comunque in grado di cancellare gli interferenti.

Qualora il terminale rilevi una variazione di SNR per cui la (21.70) indica la necessità di variare B , l'adozione di un CB organizzato *gerarchicamente* semplifica l'aumento (o la riduzione) della sua cardinalità, che raddoppia per ogni bit aggiunto a B . Infine, qualora il numero di utenti attivi sia il risultato di una strategia di prioritizzazione tra un insieme più ampio, si possono ottenere risultati migliori scegliendo quei terminali che abbiano comunicato vettori $\tilde{\mathbf{h}}_k$ il più possibile ortogonali tra loro.

Quantizzazione del precoder Osserviamo ora che l'errore $\mathbf{e}_h = \mathbf{p}_h - \tilde{\mathbf{p}}_h$ che compare nell'espressione (21.69) del segnale ricevuto r_k come un termine di *rumore interferente* $n_I = \mathbf{h}_k^T \sum_{h=1}^U \mathbf{e}_h s_h$, essendo causato dal calcolo delle colonne della matrice \mathbf{P} di pre-

⁸²La scelta degli elementi del codebook può avvenire in accordo ad un modello di d.d.p. $p(\mathbf{h}^d)$ nel contesto dell'algoritmo di *Lloyd-Max* (nota 26 a pag. 100), oppure a partire da una base dati di vettori osservati su cui eseguire un algoritmo di clusterizzazione, vedi ad es.

https://en.wikipedia.org/wiki/K-means_clustering. Infine è da citare la possibilità di popolare il CB in modo *casuale*, a partire da un seme iniziale, trasmettendo il quale il CB stesso può essere autonomamente *ri-generato* alla BS.

⁸³Vedi § 2.4.3. In questo modo due vettori sono tanto più simili quanto più condividono lo stesso orientamento; la normalizzazione ne ha portato *la punta* sulla superficie di una sfera di raggio unitario, e qui si va a finire negli *spazi Grassmaniani*. Qualcuno può chiedersi: come mai la distorsione è nulla anche per vettori *opposti*? Una possibile risposta è che i vettori quantizzati $\tilde{\mathbf{h}}_h$ servono alla BS per calcolare vettori \mathbf{p}_k di precoding *ortogonali* ad $\mathbf{h}_{h \neq k}$. L'ortogonalità dunque permane anche nei confronti di un vettore con il segno cambiato.

⁸⁴Per evitare che più terminali possano scegliere lo stesso vettore del CB, ognuno di essi può utilizzare un diverso CB. Dato che questi devono essere tutti noti presso la BS, la questione può essere semplificata nel caso di CB generati casualmente a partire da un seme.

codifica a partire dalla \mathbf{H}^d quantizzata, è un errore *causato da un altro errore*. Tanto vale allora cercare di quantizzare direttamente le colonne di \mathbf{P} , e lasciare che siano i terminali stessi a decidere quale vettore di precodifica $\tilde{\mathbf{p}}_k$ è il più idoneo moltiplicatore dei simboli destinati a ciascuno di essi, senza necessità di comunicare la CSI.

A questo fine si adotta un codebook $\mathcal{P} = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{2^B}\}$ casuale con vettori complessi \mathbf{p}_h ($n_{BS} \times 1$) a norma unitaria ed uniformemente distribuiti in *tutte le direzioni*⁸⁵, noto sia alla BS che a tutti i terminali UE_k . Ogni dispositivo, dopo aver stimato il proprio vettore di canale \mathbf{h}_k^d , può individuare il vettore $\mathbf{p}_{opt} \in \mathcal{P}$ più parallelo ad \mathbf{h}_k^d e dunque in grado di massimizzare il proprio $SINR$ ⁸⁶, ovvero

$$SINR_{k,m} = \frac{|\mathbf{h}_k^T \mathbf{p}_m|^2}{1/\rho + \sum_{h=1, h \neq m}^U |\mathbf{h}_k^T \mathbf{p}_h|^2} \quad \mathbf{p}_{opt} = \arg \max_{\mathbf{p}_m \in \mathcal{P}} SINR_{k,m} \quad (21.71)$$

dove $SINR_{k,m}$ è l' $SINR$ per UE_k qualora i simboli ad esso destinati siano precodificati con il vettore \mathbf{p}_m . Il terminale quindi confronta $SINR_{k,opt}$ con un valore soglia γ che gli è stato comunicato dalla BS, e se lo supera può inviare sia $SINR_{k,opt}$ che i B bit della codifica binaria dell'indice opt , indicando così il desiderio che i simboli s_k a lui diretti siano precodificati mediante \mathbf{p}_{opt} . Dal canto suo dopo aver ricevuto il feedback di ciascuno, la BS individua per ogni vettore \mathbf{p}_h il terminale k (ai cui dati s_k verrà applicato \mathbf{p}_h) che ha comunicato il valore $SINR_{k,h}$ più grande, ovvero per il quale \mathbf{p}_h porta più beneficio. Con tali vettori \mathbf{p}_h costruisce la matrice \mathbf{P} , ed effettua la trasmissione dedicata ai terminali selezionati.

Notiamo infine che adottando un codebook \mathcal{P} costituito dalle n_{BS} colonne di una matrice di precoding *identità*, ovvero $\mathbf{P} = \mathbf{I}_{n_{BS}}$, si ottengono vettori \mathbf{p}_h tutti nulli tranne che per l' h -esimo elemento pari ad uno. Con la scelta (21.71) ciascun terminale indica in tal caso l'antenna della BS rispetto alla quale sperimenta le migliori condizioni di propagazione, come avviene per la selezione di diversità (§ 21.3.1.1). Specialmente nel caso di un ridotto numero di utenti, ciò permette di organizzare le trasmissioni di downlink in modalità *a divisione di tempo*, trasmettendo di volta in volta con *tutta* la

⁸⁵Ovvero tali da rendere *massima* la separazione angolare *minima* d (\mathcal{P}) tra due elementi di \mathcal{P} , avendo definito $d(\mathcal{P}) = \min_{1 \leq i \leq j \leq 2^B} \sin(\theta_{i,j})$ e $\theta_{i,j} = \arccos(\mathbf{p}_i^T \mathbf{p}_j)$, e quindi $\mathcal{P} : d(\mathcal{P}) = \max$. Di nuovo, è un problema noto, vedi J.H. CONWAY, R.H. HARDIN, N.J.A. SLOANE, *Packings in Grassmannian Spaces*, Experimental Mathematics 5:2, 1996, vedi <http://www2.stat.duke.edu/~sayan/SAMSI/lec/conway.pdf>. Notiamo che qualora $2^B > n_{BS}$ non è possibile trovare 2^B vettori mutuamente ortogonali.

⁸⁶Consideriamo la (ipotetica) trasmissione da parte della BS del segnale $\mathbf{s}_P = \sum_{h=1}^{2^B} \mathbf{p}_h s_h$, che determina presso UE_k la ricezione (vedi eq. (21.67)) del segnale $\mathbf{r}_k = \mathbf{h}_k^T \mathbf{s}_P + n = \sum_{h=1}^{2^B} \mathbf{h}_k^T \mathbf{p}_h s_h + n$. Si intende trovare l'indice h associato al vettore \mathbf{p}_h che determina le migliori condizioni di ricezione per UE_k . Se consideriamo come componente di segnale quella relativa al generico indice m ovvero $\mathbf{h}_k^T \mathbf{p}_m s_m$, il termine di rumore è $\sum_{h=1, h \neq m}^{2^B} \mathbf{h}_k^T \mathbf{p}_h s_h + n$, e dunque il valore di $SINR_{k,m}$ si calcola come

$$SINR_{k,m} = \frac{E\{|\mathbf{h}_k^T \mathbf{p}_m s_m|^2\}}{\sigma_n^2 + E\{|\sum_{h=1, h \neq m}^{2^B} \mathbf{h}_k^T \mathbf{p}_h s_h|^2\}} = \frac{\mathcal{E}_s \cdot |\mathbf{h}_k^T \mathbf{p}_m|^2}{\sigma_n^2 + \mathcal{E}_s \cdot \sum_{h=1, h \neq m}^{2^B} |\mathbf{h}_k^T \mathbf{p}_h|^2} = \frac{|\mathbf{h}_k^T \mathbf{p}_m|^2}{1/\rho + \sum_{h=1, h \neq m}^{2^B} |\mathbf{h}_k^T \mathbf{p}_h|^2}$$

in cui $\rho = \mathcal{E}_s/\sigma_n^2$ e $\mathcal{E}_s = E\{|s_h|^2\}$, ed il passaggio a denominatore è possibile considerando i \mathbf{p}_h *ortonormali*. $SINR_{k,m}$ è dunque massimo quando lo è il numeratore, ossia quando lo è il prodotto scalare tra \mathbf{h}_k e \mathbf{p}_m .

potenza disponibile ed alla *massima* velocità, mediante una unica antenna m a beneficio del terminale k che ha dichiarato un miglior $SINR_{k,m}$.

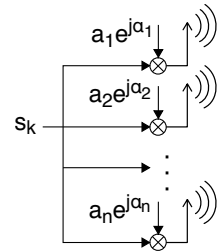
21.6.5 Beamforming

L'effetto di uno specifico vettore di precodifica \mathbf{p}_k a norma unitaria è quello di moltiplicare il simbolo s_k per un valore complesso $a_{k,i}e^{j\alpha_{k,i}}$ diverso per ciascuna antenna $i = 1, 2, \dots, n_{BS}$ della BS. Di fatto questo è il modo di operare tipico delle *antenne intelligenti*⁸⁷ che realizzano un diagramma di radiazione *direttivo ed orientabile* senza dove *muovere* nulla, ma adottando una schiera di antenne e *sfasando* con valori α_i crescenti il segnale diretto all' i -esima antenna.

Al diagramma di radiazione ottenuto è stato dato il nome di *beam* ed alla tecnica quello di *beamforming*, in cui beam può essere tradotto come *trave, raggio, fascio*; data la somiglianza formale con l'operazione di precodifica l'uso del termine di *beamforming* si è esteso ad indicare non solo le operazioni di precoding, ma anche quelle (§ 21.4.1.1) basate sulla SVD ed il water-filling per ottenere un numero di canali *indipendenti* pari al rango di \mathbf{H} , modalità operativa indicata anche come *eigen-beamforming*.

Il vantaggio più evidente del beamforming *propriamente detto* è che, trasmettendo in direzione (geometrica) del destinatario, si riduce la potenza interferente per gli altri utenti dislocati altrove e che *non rientrano* nel fascio. Per ottenere tale risultato è però necessario che la BS conosca la posizione geografica dei terminali mobili. In alternativa si possono generare *più beam* in direzioni casuali diverse⁸⁸ e lasciare che siano i terminali ad indicare per quale beam si verifica una ricezione migliore, un po' come illustrato per il caso della quantizzazione del precoder, che viene infatti descritto anche con il termine di beamforming *opportunistico*.

Nel caso di propagazione affetta da fading di Rayleigh (e dunque priva di un cammino diretto prevalente) il concetto di direttività delle antenne perde un po' di significato, mentre continua a rivestire importanza per celle di raggio più ampio, e in cui antenne (della BS) siano situate su di una torre sufficientemente elevata rispetto agli ostacoli presenti. In tal caso si determina una forte componente di onda diretta ed un fattore di Rice non trascurabile, riducendo di fatto il grado di diversità spaziale: in questo caso l'adozione del beamforming *geografico* non è solo vantaggiosa, ma necessaria. Al contrario, per un ambiente privo di percorsi diretti e ricco di scattering è preferibile attuare il precoding atto a massimizzare l' $SINR$ (eq. (21.71)) e capace di far combinare in modo coerente le repliche che viaggiano su cammini multipli differenti, come esemplificato nella animazione presso <https://youtu.be/XBb481RNqGw>.



⁸⁷O *phased array*, vedi ad es. https://en.wikipedia.org/wiki/Phased_array ma anche <https://www.vialattea.net/content/1875/>.

⁸⁸Cioè pilotare i modulatori collegati alle antenne con un vettore $\mathbf{s}_B = \sum_{h=1}^{n_B} \mathbf{b}_h s_h$ in cui il simbolo s_h destinato all' h -esimo utente viene *affasciato* nella direzione stabilita dal beam \mathbf{b}_h , ne più ne meno come avviene nel precoding.

21.7 Trasmissione MIMO - OFDM

La necessità di subire un fading *piatto*, ovvero che il segnale modulato occupi una banda inferiore alla *banda di coerenza*, determina un *limite* alla massima velocità binaria della trasmissione MIMO *a portante singola*. Infatti per contenere la banda si dovrebbe ricorrere ad una costellazione con un numero molto elevato di livelli, incappando in un rapido degrado delle prestazioni; lasciare invece che le banda aumenti in proporzione alla velocità di simbolo significa dover equipaggiare ogni antenna di ricezione con uno stadio di equalizzazione adattativa, con evidente aumento di complessità, tanto maggiore con l'aumento della velocità. D'altra parte l'adozione di una modulazione a spettro espanso che affronta l'equalizzazione mediante una architettura di ricezione *Rake* (§ 20.5.2) determina una complessità che cresce in modo quadratico con la velocità di trasmissione⁸⁹.

Al contrario, se il segnale trasmesso da ciascuna antenna adotta una modulazione OFDM (§ 16.8) i problemi appena esposti semplicemente *svaniscono*, ed al tempo stesso si aprono possibilità di ulteriore miglioramento della qualità e delle prestazioni del collegamento. Tutto nasce dal fatto che, ripartendo il flusso binario da trasmettere su P canali ortogonali (in frequenza) con una occupazione di banda ridotta, per ciascuno di essi risulta verificata l'ipotesi di fading piatto. Inoltre l'adozione del *prefisso ciclico* consente di annullare gli effetti dell'ISI senza dovere ricorrere ad uno stadio di equalizzazione, se non per la componente di stima di canale alla frequenza delle sottoportanti, che è comunque necessaria per la ricezione di trasmissioni MIMO. Infine l'utilizzo di una FFT per realizzare i processi di modulazione e demodulazione fa sì che la complessità dei dispositivi aumenti *solo* con legge $P \log_2 P$ al crescere del numero di sottoportanti P , e dunque della velocità di trasmissione.

Vantaggio di diversità La trasmissione MIMO-OFDM consente inoltre di sfruttare, oltre a quella *spaziale* dovuta alla molteplicità delle antenne, anche la diversità *in frequenza*⁹⁰. Pur se il canale associato a ciascuna sottoportante è considerato sede di fading *piatto*, il valore h del suo guadagno complesso è una v.a. che può assumere una realizzazione *diversa* per sottoportanti differenti⁹¹, aspetto che nelle trasmissioni SISO viene affrontato con la tecnica COFDM (§ 16.8.10). In particolare si dimostra⁹² che un canale per il quale sono individuabili L cammini multipli (a cui si deve la selettività in frequenza, § 20.4.5) ed equipaggiato da n_T, n_R antenne ai lati del collegamento, offre

⁸⁹La fonte di questa affermazione (Wikipedia) cita la tesi di dottorato G.RALEIGH, *On Multivariate Communication Theory and Data Rate Multiplying Techniques for Multipath Channels*, 1998, ma il link a cui puntava non risulta più attivo.

⁹⁰La diversità in frequenza è esattamente quella discussa al § 20.3.3.1 e dovuta ai cammini multipli che rendono il canale selettivo in frequenza (§ 20.4.5).

⁹¹Possiamo notare che come le antenne riescono ad offrire diversità spaziale solo qualora le stesse siano sufficientemente distanziate rispetto alla lunghezza d'onda λ della trasmissione, così le sottoportanti offrono diversità in frequenza solo su canali separati da un intervallo di frequenza maggiore della banda di coerenza.

⁹²H. BOLCSKEI, A.J. PAULRAJ, *Space-frequency coded broadband OFDM systems*, 2000 IEEE Wireless Comm. and Networking Conference

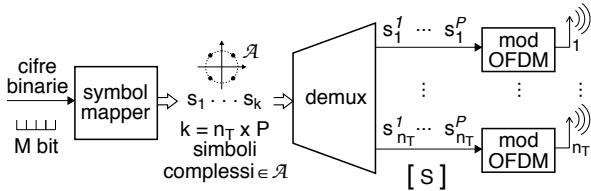
un ordine di diversità *massimo*⁹³ pari al prodotto di queste quantità, ovvero pari a

$$L \cdot n_T \cdot n_R \quad (21.72)$$

Ma procediamo con ordine.

21.7.1 Modello di canale MIMO-OFDM

Partiamo dal caso più generale in figura, in cui $n_T \times P$ simboli complessi vengono sistematicamente⁹⁴ ripartiti su n_T antenne mediante ciascuna delle quali viene trasmesso un simbolo



OFDM costituito da P sottoportanti. Anziché un semplice vettore \mathbf{s} di n_T simboli, per ogni utilizzo del canale viene ora trasmessa una matrice \mathbf{S} di $n_T \times P$ elementi complessi s_{jp} ; similmente, in ricezione si ottiene la matrice \mathbf{R} di $n_R \times P$ elementi r_{ip} , secondo la relazione

$$r_{ip} = \sum_{j=1}^{n_T} h_{ijp} \cdot s_{jp} + n_{ip} \quad (21.73)$$

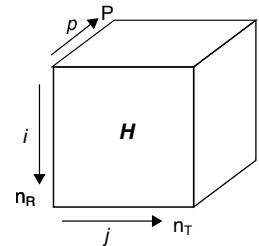
con $i = 1 \dots n_R$ e $p = 1 \dots P$, in cui h_{ijp} è la risposta in frequenza tra le antenne j e i alla frequenza p (supposta costante almeno per la durata del simbolo OFDM), la sommatoria prende in considerazione tutte le antenne di trasmissione, e n_{ip} è il corrispondente campione complesso di rumore gaussiano circolare.

La classica matrice \mathbf{H} dei guadagni complessi tra le n_R antenne di ricezione (le righe i) e la n_T di trasmissione (colonne j) che compare nella (21.2) si è dunque arricchita di un ulteriore indice, quello che individua il *piano* p corrispondente ad ognuna delle sottoportanti usate dall'OFDM, in modo che tutti i $n_R \cdot n_T \cdot P$ valori possano essere individuati come $(\mathbf{H})_{i,j,p} = h_{ijp}$. La relazione (21.73) può quindi essere riscritta in forma matriciale considerando le matrici \mathbf{S} ed \mathbf{R} dei simboli trasmessi e ricevuti lette per colonne, ottenendo rispettivamente i vettori \mathbf{s}_p degli n_T valori trasmessi alla frequenza p , e quelli \mathbf{r}_p degli n_R valori ricevuti alla medesima frequenza, in modo da poter scrivere

$$\mathbf{r}_p = \mathbf{H}_p \cdot \mathbf{s}_p + \mathbf{n}_p \quad \text{con } p = 1, 2, \dots, P$$

in cui \mathbf{H}_p di dimensione $n_R \times n_T$ è il *piano* p della matrice tridimensionale \mathbf{H} e corrisponde alla tradizionale relazione (21.2) tra le antenne, specializzata per la frequenza p , ed \mathbf{n}_p di dimensione n_R rappresenta il rumore alla stessa frequenza. La costruzione della matrice \mathbf{R} completa si realizza quindi concatenando le sue colonne \mathbf{r}_p , ovvero

$$\mathbf{R} = [\mathbf{r}_1 \mathbf{r}_2 \dots \mathbf{r}_P] = [\mathbf{H}_1 \mathbf{s}_1 \mathbf{H}_2 \mathbf{s}_2 \dots \mathbf{H}_P \mathbf{s}_P]$$



⁹³Con il massimo conseguito solo per il caso di indipendenza di tutti i percorsi alternativi.

⁹⁴In alternativa la ripartizione può contemplare elaborazioni più complesse come quelle descritte al § 21.7.2, o prevedere anche uno stadio di precoding, od anche di codifica di canale e/o interleaving.

La variazione di \mathbf{H}_p con l'indice di portante p è dunque l'elemento che aggiunge al MIMO-OFDM anche la componente di diversità *in frequenza* oltre che spaziale.

21.7.2 Codice spazio-tempo-frequenza

Per beneficiare del guadagno di diversità spaziale la trasmissione MIMO-OFDM si deve *affidare* ai codici spazio-tempo affrontati al § 21.3.2.2. A tale scopo lo schema di pag. 709 si modifica come indicato in fig. 21.6, che mostra come ad una sequenza di k simboli s_i ad $L = 2^M$ valori venga fatta corrispondere una codeword

$$\mathbf{C} = \begin{matrix} \rightarrow \text{antenne} \rightarrow \\ \begin{bmatrix} \mathbf{c}_1^1 & \mathbf{c}_2^1 & \cdots & \mathbf{c}_{n_T}^1 \\ \mathbf{c}_1^2 & \mathbf{c}_2^2 & \cdots & \mathbf{c}_{n_T}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{c}_1^T & \mathbf{c}_2^T & \cdots & \mathbf{c}_{n_T}^T \end{bmatrix} \\ \downarrow \\ \text{istanti} \\ \downarrow \end{matrix}$$

i cui elementi \mathbf{c}_t^j sono ora vettori di P elementi che individuano i punti di costellazione L -aria con cui modulare le sottoportanti del simbolo OFDM trasmesso all'istante t dall'antenna j . In questo contesto il tasso di codifica (eq. (21.19)) prende in considerazione anche il numero di portanti, ed è definito come $R_c = \frac{k}{PT}$.

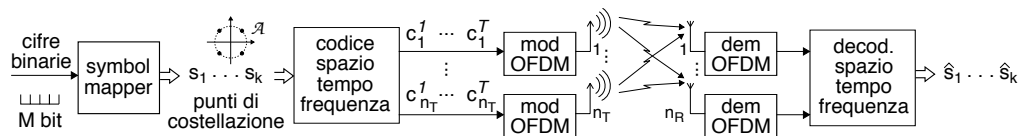


Figura 21.6: Codifica spazio-tempo-frequenza per un collegamento MIMO-OFDM

Si aprono ora diverse possibilità: in fig.21.7-a viene mostrata l'applicazione di un codice STBC di Alamouti su di una unica sottoportante, trasmesso da due antenne in due istanti temporali: si ottiene $R_c = 1$ in quanto si trasmettono due simboli in due istanti, e lo stesso può essere fatto su tutte le P portanti. La fig. 21.7-b mostra invece lo stesso codice trasmesso sempre da due antenne, ma applicato su due portanti *dello stesso simbolo*, conseguendo nuovamente $R_c = 1$: questo caso viene indicato come *space-frequency block code* (SFBC). E' evidente che mentre nel primo caso possiamo trarre vantaggio solo sul fronte della diversità spaziale, nel secondo vorremmo sfruttare anche la diversità frequenziale, sempre nei limiti di quanto evidenziato alla nota 91; in entrambi i casi non si riesce però a conseguire⁹⁵ tutta la diversità offerta sia nello spazio che in frequenza, vedi eq. (21.72). Qualcosa di meglio si riesce a fare secondo gli approcci di fig. 21.7-c e -d, che raddoppiano il numero di portanti su cui si sviluppa il codice, con un dimezzamento di R_c ; per entrambi se il numero di repliche del codice eguaglia quello (L) dei cammini multipli, lo schema può offrire il massimo guadagno (21.72) con un rate $R_c = 1/L$. Tuttavia, l'implementazione di fig. 21.7-d presenta una maggiore complessità di decodifica.

⁹⁵Una buona sintesi storica di questo filone di studio si trova in W ZHANG, XG XIA, KB LETAIEF, *Space-time/frequency coding for MIMO-OFDM in next generation broadband wireless systems*, IEEE Wireless Comm. June 2007, di cui trovo una copia presso <https://www.eecis.udel.edu/~xxia/WeiZhang1.pdf>

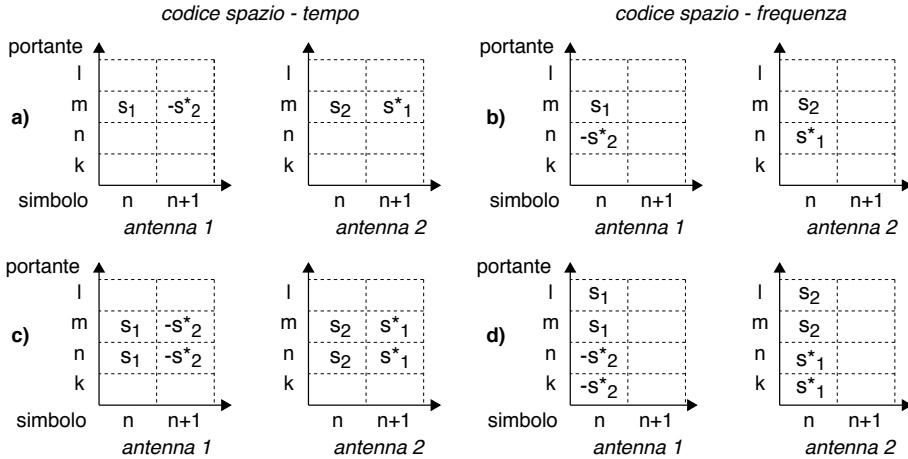


Figura 21.7: Schemi di codifica spazio-tempo e spazio-frequenza

Rimandando alla nota (95) per gli approfondimenti, qui citiamo solamente che sono stati individuati codici SFBC in grado di sfruttare il massimo guadagno di diversità (21.72) con $R_c = 1$, ed anche (impiegando codici *algebrici*⁹⁶) con $R_c = n_T$, ovvero capaci di trasmettere un diverso flusso per antenna. Ovviamente, tutto ciò al prezzo di una complessità di decodifica ancora maggiore.

Ma non è finita! Nel caso in cui il canale presenti una accentuata variabilità temporale, con valori di \mathbf{H} da considerare costanti nell'ambito di un blocco (in cui entrano uno o pochi simboli OFDM) ma variabili da un blocco all'altro, il canale viene definito come soggetto a *block-fading*, ed una ulteriore categoria di codici STFC⁹⁷ che si estende su di n_B blocchi con fading *indipendente* può conseguire un ordine di diversità ancora maggiore, e pari a $L \cdot n_T \cdot n_R \cdot n_B$, con un tasso $R_C = n_T$.

21.7.3 Sistema multiutente MU-MIMO-OFDM

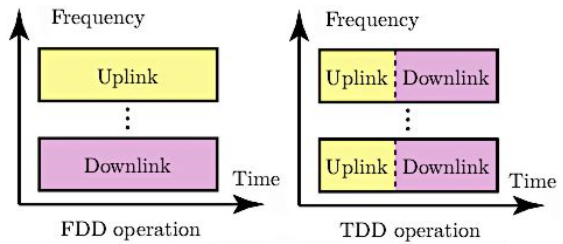
La trasmissione MIMO-OFDM si è rivelata particolarmente idonea a realizzare un sistema ad accesso multiplo come introdotto al § 21.6, in cui una stazione radio base BS con n_{BS} antenne comunica con U terminali UE_k ad antenna singola, sia in direzione downlink (DL o canale broadcast) che in uplink (UL o accesso multiplo). L'unica antenna di ricezione dei terminali impedisce l'uso di codici spazio-tempo, ma non di quelli spazio-frequenza.

Alternanza temporale Nel caso OFDM si rendono possibili considerazioni che portano a preferire uno schema di condivisione delle risorse radio in cui la trasmissione nei due versi *si alterna* nel tempo secondo una modalità nota come *time division duplex* o TDD, ed occupa la medesima banda nelle due direzioni, banda che per l'OFDM è di

⁹⁶Vedi H. EL GAMAL, M. O. DAMEN, *Universal Space-Time Coding*, IEEE Trans. on Inf. Th., May 2003, reperibile presso <http://www2.ece.ohio-state.edu/~elgamal/print12.pdf>

⁹⁷Vedi W. ZHANG, X. XIA, P. C. CHING, *High-Rate Full-Diversity Space-Time-Frequency Codes for Broadband MIMO Block-Fading Channels*, IEEE Trans. on Comm., January 2007, di cui trovo una copia su CiteseerX

estensione ben maggiore rispetto al caso a portante singola, coprendo un intervallo pari a diversi multipli della banda di coerenza B_c . Lo schema TDD consente una serie di semplificazioni rispetto al *frequency division duplex* (FDD), prima tra tutte la possibilità per gli utenti mobili di evitare di dover comunicare alla BS la stima di canale da essi effettuata (§ 21.7.3.2), in virtù della reciprocità del canale (vedi nota 53).



Uplink La BS alla portante p riceve un segnale

$$\mathbf{r}_{BS}^p = \sum_{k=1}^U \mathbf{h}_k^p s_k^p + \mathbf{n} = \mathbf{H}_p^u \mathbf{s}_p + \mathbf{n} \quad (21.74)$$

in cui \mathbf{h}_k^p è il vettore ($n_{BS} \times 1$) del canale MIMO di *uplink* per la portante p tra UE $_k$ e le antenne della BS, s_k^p è il simbolo inviato dall'unica antenna di UE $_k$ sulla portante p con energia $\mathcal{E}_k = E\{|s_k^p|^2\}$, $\mathbf{H}_p^u = [\mathbf{h}_1^p \mathbf{h}_2^p \dots \mathbf{h}_U^p]$ è la matrice ($n_{BS} \times U$) di *uplink* da tutti gli U dispositivi UE $_k$, ed \mathbf{s}_p è il vettore ($U \times 1$) ad elementi s_k^p ; qualora uno UE non trasmetta sulla portante p , pone il valore s_k^p a zero. Infine, \mathbf{n} è un vettore aleatorio gaussiano complesso a media nulla e covarianza $\sigma_{UL}^2 \cdot \mathbf{I}_{n_{BS} \times n_{BS}}$.

Per effettuare la detezione del simbolo s_j^p la BS calcola la matrice \mathbf{G}_p a partire da \mathbf{H}_p con uno dei metodi discussi al § 21.5 e valuta il prodotto tra la j -esima riga \mathbf{g}_j^p di \mathbf{G}_p e la (21.74), ottenendo

$$\tilde{s}_j^p = \mathbf{g}_j^p \mathbf{r}_{BS}^p = \mathbf{g}_j^p \mathbf{h}_j^p s_j^p + \sum_{\substack{k=1 \\ k \neq j}}^U \mathbf{g}_j^p \mathbf{h}_k^p s_k^p + \mathbf{g}_j^p \mathbf{n} \quad (21.75)$$

in cui il primo termine è quello desiderato, il secondo sono gli interferenti (che se in numero elevato possono essere ritenuti a somma gaussiana) e l'ultimo è il nuovo termine di rumore; dopodiché si valuta $\hat{s}_j = \arg \min_{s \in \mathcal{A}} (\hat{s}_j - s)^2$.

Notiamo che qualora \mathbf{G}_p sia ottenuta con l'approccio *zero forcing* (§ 21.5.2), il termine di interferenza si annulla; notiamo inoltre che il calcolo (21.75) può essere svolto in contemporanea per tutti gli utenti, come $\tilde{\mathbf{s}} = \mathbf{G}_p \mathbf{r}_{BS}^p = \mathbf{G}_p \mathbf{H}_p^u \mathbf{s}_p + \mathbf{G}_p \mathbf{n}$.

Downlink In questa direzione le n_{BS} antenne inviano (sulla portante p) il segnale

$$\mathbf{x}_p = \sum_{k=1}^U \mathbf{p}_k^p s_k^p = \mathbf{P}_p \mathbf{s}_p \quad (21.76)$$

dove \mathbf{p}_k^p è il vettore di *precodifica* (§ 21.6.1) a norma unitaria relativo al terminale k , ovvero la k -esima colonna dell'omonima matrice \mathbf{P}_p ($n_{BS} \times U$) ottenuta a partire dalla conoscenza della matrice ($U \times n_{BS}$) di DL $\mathbf{H}_p^d = (\mathbf{H}_p^u)^T$, trasposta di quella di UL per la reciprocità del canale MIMO alla stessa frequenza p , e che determina la direttività *spaziale* (e non *geografica* come nel beamforming) della trasmissione del simbolo s_k^p destinato al terminale k , mentre \mathbf{s}_p è il vettore di tutti i simboli trasmessi. Il terminale j -esimo riceve dunque (alla portante p)

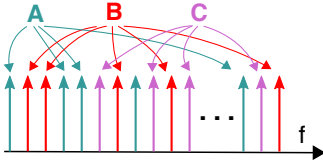
$$\mathbf{r}_j^p = (\mathbf{h}_j^p)^T \mathbf{x}_p = (\mathbf{h}_j^p)^T \sum_{k=1}^U \mathbf{p}_k^p s_k^p + \mathbf{n} = (\mathbf{h}_j^p)^T \mathbf{p}_j^p s_j^p + \sum_{\substack{k=1 \\ k \neq j}}^U (\mathbf{h}_j^p)^T \mathbf{p}_k^p s_k^p + \mathbf{n} \quad (21.77)$$

in cui \mathbf{h}_j^p è il vettore delle risposte in frequenza tra UE $_k$ e le antenne della BS, che viene trasposto per la reciprocità del canale di DL; come per la (21.75) inoltre, il primo termine di (21.77) è quello desiderato, il secondo esprime l'interferenza dei simboli destinati agli altri utenti, ed il terzo è il rumore. Anche qui, una matrice di precodifica *zero forcing* ha l'effetto di annullare del tutto gli interferenti, fornendo $r_j = s_j + n = \tilde{s}_j$ da cui ottenere $\hat{s}_j = \arg \min_{s \in \mathcal{A}} (\hat{s}_j - s)^2$.

21.7.3.1 Ripartizione delle risorse

Oltre alla suddivisione dell'asse dei tempi per le due direzioni di trasmissione, nella trasmissione multiutente viene ripreso anche l'approccio dell'OFDMA (§ 16.8.12) di assegnare sotto-gruppi di portanti OFDM ad utenti differenti.

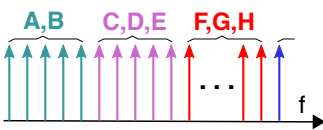
Una o più portanti per utente Qualora i dispositivi collegati alla BS non siano soggetti a spostamenti frequenti e non siano in numero troppo elevato si può procedere in modo *distribuito* come esemplificato in figura, in cui i singoli utenti sono rappresentati da colori diversi, utilizzando un numero di portanti possibilmente distanziate da un intervallo di frequenza maggiore della banda di coerenza in modo da poter sfruttare la relativa diversità



in frequenza offerta dalla trasmissione OFDM. Il numero di portanti per utente dipende dalle sue esigenze trasmissive, mentre il relativo posizionamento in frequenza può dipendere dalla stima di canale effettuata dalla BS, in modo da usare frequenze per le quali si hanno buone condizioni di ricezione. Le portanti OFDM non utilizzate da un singolo utente vengono da queste poste a zero.

in frequenza offerta dalla trasmissione OFDM. Il numero di portanti per utente dipende dalle sue esigenze trasmissive, mentre il relativo posizionamento in frequenza può dipendere dalla stima di canale effettuata dalla BS, in modo da usare frequenze per le quali si hanno buone condizioni di ricezione. Le portanti OFDM non utilizzate da un singolo utente vengono da queste poste a zero.

Uno o più utenti per blocco di coerenza Se invece si ha a che fare con uno scenario di utenza mobile, che determina la variabilità temporale delle condizioni di ricezione, le sottoportanti OFDM da assegnare ad uno (o più⁹⁸) utenti possono essere *contigue in frequenza* per la durata di un ciclo UL-DL; il numero di portanti e la durata del ciclo dipendono dai valori di banda e tempo di coerenza B_c e T_c (§ 20.4.7) del collegamento. Anche se questi possono differire tra



essere *contigue in frequenza* per la durata di un ciclo UL-DL; il numero di portanti e la durata del ciclo dipendono dai valori di banda e tempo di coerenza B_c e T_c (§ 20.4.7) del collegamento. Anche se questi possono differire tra

utenti diversi, prendendo *il caso peggiore* di entrambi si può suddividere il piano tempo-frequenza in unità denominate *Coherence Block* (CB) che raggruppano portanti e simboli contigui in intervalli entro i quali il canale può essere considerato stazionario e privo di distorsione lineare, e riservare i CB per la trasmissione (alternata) tra la BS ed i singoli UE.

Cosa si intende per blocco di coerenza Tralasciando per il momento l'aspetto MIMO, la parte superiore di fig. 21.8 rappresenta come il guadagno (in dB) della risposta in frequenza tra una antenna della BS e quelle di una coppia di utenti in movimento

⁹⁸Distinguibili grazie ai vettori di precoding \mathbf{p}_k e di combinazione \mathbf{g}_k .

possa variare sia rispetto alla frequenza che al tempo, guadagno rappresentato da una superficie di diverso colore per i due utenti. La parte inferiore mostra invece come i diversi CB vengano assegnati ora all'uno ora all'altro utente, in base a chi dei due subisce una minore attenuazione. La presenza di più antenne presso la BS *appiattisce* le superfici mostrate per quanto riguarda l'UL grazie alla possibilità di combinare in modo coerente i segnali ricevuti (eq. (21.75)), mentre il precoding (eq. (21.76)) attuato dalla BS in trasmissione ha il solo scopo di combattere l'interferenza tra gli utenti assegnati allo stesso CB⁹⁹, e dunque non modifica l'attenuazione subita nel DL.

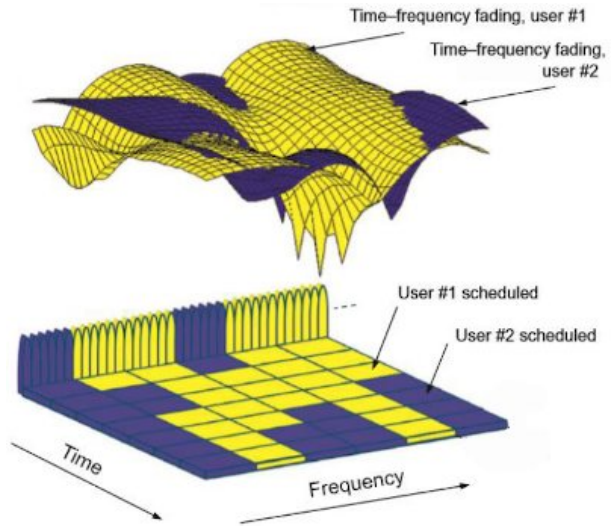


Figura 21.8: Assegnazione dei Coherence Block

La fig. 21.9 mostra¹⁰⁰ la suddivisione del piano tempo - frequenza in CB, evidenziando in giallo gli intervalli temporali dedicati alla trasmissione dell'uplink ed in rosa quelli per il downlink, mentre all'interno di un CB evidenzia la suddivisione in sottoportanti, ed i campioni complessi che si possono ottenere per ogni sottoportante, in numero totale (per ogni CB) τ_c proporzionale¹⁰¹ al prodotto $B_c T_c$, per cui scriviamo

$$\tau_c = \alpha B_c T_c \quad \alpha < 1 \tag{21.78}$$

⁹⁹Verso i quali si trasmette *in contemporanea* realizzando lo SDMA.

¹⁰⁰Tratta dall'ottimo testo di E. BjÖRNSON, J. HOYDIS, L. SANGUINETTI, *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency*, 2017, accessibile presso <https://massivemimobook.com>

¹⁰¹Considerando che un involuppo complesso che occupa una banda *bilatera* B_c viene campionato a frequenza doppia di $B_c/2$, in T_c secondi si ottengono appunto $B_c T_c$ campioni. Una volta eliminato il prefisso ciclico (fig. 21.10) tale numero si riduce, ed il ricevitore OFDM ne effettua la FFT ottenendo altrettanti campioni complessi in frequenza.

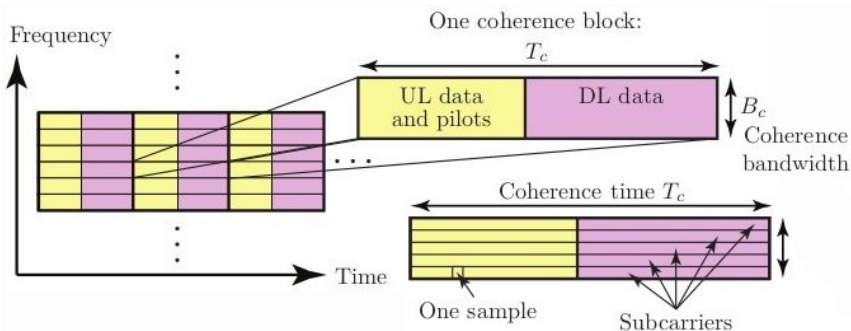


Figura 21.9: Suddivisione del piano tempo-frequenza in blocchi di coerenza e relativi intervalli temporali di trasmissione, sottoportanti, e campioni complessi

suddivisi tra tutte le portanti e tutti i simboli OFDM trasmessi nelle due direzioni nell'ambito di uno stesso CB.

Esempio In prima approssimazione valutiamo il tempo di coerenza T_c , legato alla velocità v del mobile, come quello necessario a spostarsi di $1/4$ di lunghezza d'onda λ , ovvero $T_c \simeq \lambda/4v$: quindi T_c è inversamente proporzionale alla frequenza portante $f_0 = c/\lambda$ centrale della trasmissione. D'altra parte la banda di coerenza può essere approssimata come $B_c \simeq 1/2\Delta\tau$ dove $\Delta\tau$ è la dispersione temporale tra la prima e l'ultima replica del multipath. Valutiamo ora due scenari per una frequenza $f_0 = 2$ GHz ovvero $\lambda = 15$ cm: il primo è un sistema *outdoor* con velocità $v = 37.5$ m/sec ossia 135 Km/h, ed un $\Delta\tau = 2.5$ μ sec (cioè una differenza tra i percorsi di 750 metri). In tal caso si ottiene $T_c = 1$ msec e $B_c = 200$ KHz, e dunque $T_c B_c = 200$. Il secondo scenario è una trasmissione *indoor* caratterizzata da una mobilità $v = 0.75$ m/s (2.7 Km/h) ed un $\Delta\tau = 0.5$ μ sec (differenza tra percorsi di 150 metri), a cui corrispondono $T_c = 50$ msec e $B_c = 1$ MHz, fornendo $T_c B_c = 50000$.

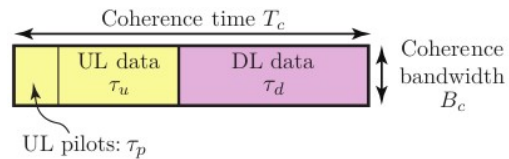
21.7.3.2 Stima di canale

Tutte le operazioni di ricezione delle trasmissioni sia MIMO che OFDM prevedono la conoscenza della risposta in frequenza del canale, che si suppone vari lentamente nel tempo rispetto alle costanti di tempo del sistema di trasmissione¹⁰², e che fino ad ora abbiamo assunto nota senza chiederci come venirne a conoscenza: nel caso dell'OFDM la risposta in frequenza viene stimata in base alle condizioni di ricezione delle *portanti pilota* (§ 16.8.11, fig. 21.10) inserite in trasmissione.

Suddivisione del blocco di coerenza Lo schema di alternanza TDD che utilizza la stessa banda di frequenze per i due versi di trasmissione permette come già accennato di inserire le p. pilota nella sola direzione di UL ossia nella trasmissione da UE a BS, all'inizio di ogni blocco di coerenza come mostrato nella figura seguente, riservando per le stesse un numero di campioni complessi τ_p sul totale di τ_c (trasmessi e ricevuti, eq. (21.78)) in accordo alla ripartizione

$$\tau_c = \tau_p + \tau_u + \tau_d \tag{21.79}$$

in cui i pedici u e d individuano rispettivamente i campioni riservati ai dati di uplink ed a quelli di downlink, come esemplificato in figura.



Ripartizione degli utenti La stima di canale dunque avviene per ciascun CB_h ($h = 1 \dots n_{CB}$) in cui è suddiviso l'asse della frequenza ed a cui sono assegnati U_h utenti sul totale di U , ovvero $\sum_h U_h = U$, ad opera dalla sola BS, che ne usa il risultato sia per ricevere i dati di UL da parte di tutti gli UE_k , $k = 1 \dots U_h$ assegnati allo stesso CB_h

¹⁰²Per la modalità TDD su cui stiamo basando l'esposizione l'espressione *costante di tempo* individua la durata di un CB, mentre più in generale ci si riferisce non tanto al periodo di simbolo OFDM quanto ad un *periodo di trama*, che comprende un *preambolo* composto da qualche simbolo OFDM *accorciato* in cui trovano posto le portanti pilota, a cui fa seguito la sotto-trama dei simboli con i dati, per la cui ricezione si fa uso del risultato della stima di canale. Spesso alcune pilota sono presenti anche nei simboli dati, per favorire il mantenimento delle condizioni di sincronizzazione.

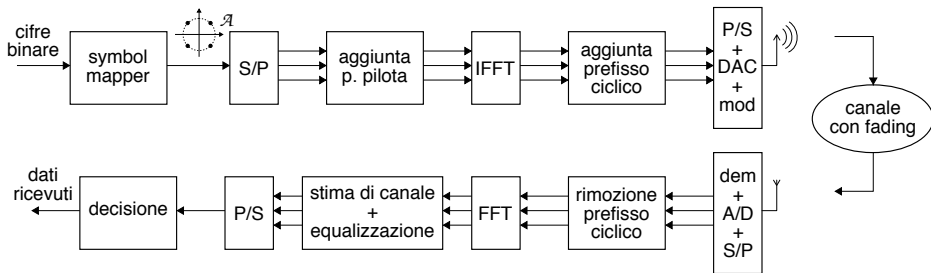


Figura 21.10: Schema simbolico di un *modem* OFDM per un collegamento siso

(eq. (21.75)), sia per la trasmissione in DL verso gli stessi UE_k (eq. (21.77)). La relazione (21.79) pone un limite al numero massimo di campioni dedicati alle p. pilota, dato che tipicamente si adotta il vincolo $\tau_p < \tau_c/2$.

Sequenze pilota Per permettere alla BS di stimare il canale verso gli utenti, all'inizio del CB ogni UE_k trasmette una sequenza-pilota ϕ_k di τ_p campioni complessi di *ampiezza unitaria*, ovvero tali che $\phi_k^\dagger \phi_k = \tau_p$, ed ortogonale¹⁰³ alle sequenze ϕ_l trasmesse dagli altri utenti UE_l ($l \neq k$) assegnati allo stesso CB, ovvero $\phi_k^\dagger \phi_l = 0$. La sequenza ϕ_k viene assegnata a ciascun utente k da parte della BS durante la fase di *richiesta di accesso* dell'utente alla rete, ed è scelta tra le colonne di una matrice Φ di dimensione $\tau_p \times \tau_p$ detta *pilot-book* e che soddisfa la relazione $\Phi^\dagger \Phi = \tau_p I_{\tau_p}$. Per Φ può essere adottata una matrice di *Walsh-Hadamard*¹⁰⁴ che pone il vincolo su τ_p di essere una potenza di due, e che essendo composta da elementi ± 1 implica la trasmissione di simboli BPSK nella eq. (21.80); in alternativa per un valore τ_p qualunque la matrice Φ può essere ottenuta come quella che definisce la DFT (eq. (4.12) a pag. 105) i cui elementi sono valori complessi equispaziati sul cerchio unitario e che dunque danno luogo a simboli L-PSK.

Ricezione delle pilota e stima di canale Gli elementi di ϕ_k sono moltiplicati dai dispositivi UE_k per la costante $\sqrt{\mathcal{E}_k}$ in modo da presentare la stessa energia dei simboli che trasportano informazione, e sono quindi trasmessi come i simboli s_j in (21.74), con la differenza che ora la trasmissione è distribuita su tutte le portanti che fanno parte del CB e ripartita su più simboli OFDM. Per la definizione di CB la risposta in frequenza del canale è la stessa sia per le diverse portanti che per i diversi simboli, portando a descrivere il segnale ricevuto per il CB dalla BS in questa fase nella forma

$$\mathbf{Y}_h = \sum_{k=1}^{U_h} \sqrt{\mathcal{E}_k} \mathbf{h}_k^h \phi_k^T + \mathbf{N}_h \quad (21.80)$$

in cui \mathbf{h}_k^h è il vettore delle n_{BS} risposte in frequenza tra l'antenna dell'utente k assegnato al CB h e tutte le antenne della BS, \mathbf{Y}_h è la matrice complessa di dimensione $n_{BS} \times \tau_p$

¹⁰³Notiamo che per uno spazio descritto da una base vettoriale τ_p -dimensionale si possono individuare non più di τ_p diversi vettori *ortogonali* ϕ_k .

¹⁰⁴Vedi ad es. https://en.wikipedia.org/wiki/Walsh_matrix

utilizzata per stimare il canale, e \mathbf{N}_h è la matrice di eguali dimensioni dei campioni complessi ed indipendenti di rumore, a media nulla e varianza σ_{UL}^2 .

Dato che la BS conosce la sequenza ϕ_k utilizzata da ciascun utente k , può effettuare la stima del canale \mathbf{h}_j^h relativo ad uno specifico utente j moltiplicando (correlando) la (21.80) per la sequenza coniugata ϕ_j^* di quella assegnata all'utente j , ottenendo

$$\mathbf{y}_j^h = \mathbf{Y}_h \phi_j^* = \sum_{k=1}^{U_h} \sqrt{\mathcal{E}_k} \mathbf{h}_k^h \phi_k^T \phi_j^* + \mathbf{N}_h \phi_j^* = \sqrt{\mathcal{E}_k} \tau_p \mathbf{h}_j^h + \mathbf{N}_h \phi_j^*$$

in quanto per l'ortogonalità delle sequenze pilota assegnate ad utenti differenti, il prodotto $\phi_k^T \phi_j^*$ è nullo per $k \neq j$ e pari a τ_p per $k = j$. Dato che il prodotto $\mathbf{N}_h \phi_j^*$ è un vettore complesso gaussiano circolare a media nulla e covarianza $\sigma_{UL}^2 \tau_p \mathbf{I}_{n_{BS}}$, la stima di massima verosimiglianza (e dunque di minima distanza) per \mathbf{h}_j^h si ottiene come

$$\hat{\mathbf{h}}_j^h = \frac{\mathbf{y}_j^h}{\sqrt{\mathcal{E}_k} \tau_p} = \mathbf{h}_j^h + \frac{\mathbf{N}_h \phi_j^*}{\sqrt{\mathcal{E}_k} \tau_p} \quad (21.81)$$

e dunque le componenti \hat{h}_{ij} di $\hat{\mathbf{h}}_j^h$ sono v.a. gaussiane a media h_{ij} e varianza $\frac{\sigma_{UL}^2}{\sqrt{\mathcal{E}_k} \tau_p}$: in altre parole, l'utilizzo di una sequenza-pilota di lunghezza τ_p determina un *guadagno di processo* pari a τ_p ovvero un miglioramento dell'SNR di τ_p volte quello che si otterrebbe per una pilota singola.

Numero di utenti per blocco di coerenza Per garantire l'ortogonalità delle sequenze pilota assegnate ai diversi utenti che ricadono nello stesso CB_h il loro numero U_h non può superare il valore della lunghezza τ_p delle sequenze. Dato che aumentare τ_p determina una riduzione dei campioni utili alla trasmissione dati (eq. (21.79)) si manifesta un compromesso tra il numero di utenti e capacità trasmissiva.

21.7.3.3 Rete cellulare

Con questo termine si indica una suddivisione del territorio del tipo di fig. 21.11, in cui più BS sono contemporaneamente attive. In tale contesto per ogni CB le BS ricevono le trasmissioni, oltre che degli UE ad esse associati, anche degli UE nelle celle vicine ed assegnati al CB nella stessa regione di frequenza, così come i diversi UE ricevono il segnale delle altre BS oltre a quello della propria. Ciò comporta che nelle espressioni (21.74) e (21.77) dei segnali ricevuti rispettivamente dalla BS e dagli UE di una singola cella compaiano termini di interferenza che originano dalle celle limitrofe, la cui entità può essere ridotta mediante operazioni di combinazione lineare e precoding estendendo la procedura di stima di canale da parte di ciascuna BS anche alle risposte in frequenza relative agli UE associati ad altre BS ossia *residenti* in altre celle.

E' ovvio che per quanto grande possa essere la dimensione del pilot book Φ , è impossibile assegnare sequenze differenti a tutti gli utenti che possono interferire tra loro, sicché la stima di canale svolta da parte delle BS è gravata da un errore detto *pilot contamination*, e che consiste in un termine di errore nella stima (21.81) che origina da

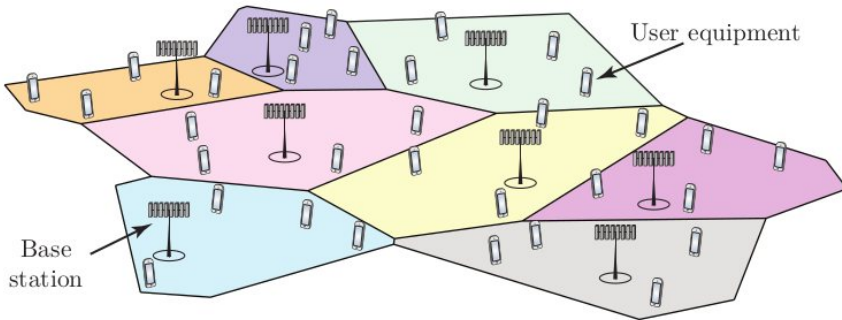


Figura 21.11: Suddivisione del territorio in celle in cui gli utenti (UE) si associano alla base station (BS) il cui segnale è più forte

tutti gli UE nelle altre celle che condividono la stessa sequenza pilota ϕ dello UE di cui si sta stimando la risposta in frequenza.

Senza avere la minima pretesa di approfondire a sufficienza l'argomento, al cui riguardo si rimanda il lettore al testo indicato alla nota 100, ci si limita a citare che in questo caso la problematica viene affrontata tenendo conto della diversa correlazione spaziale \mathcal{R}_h^k che lega statisticamente le componenti del vettore delle risposte in frequenza \mathbf{h}_k di ciascun generico utente k nei confronti delle antenne di una BS.

A differenza di un collegamento MIMO punto-punto, in cui tra le diverse antenne da ambo i lati si manifestano vettori \mathbf{h} descritti da matrici di correlazione \mathcal{R}_h sostanzialmente simili, nel caso multiutente ogni utente k è in genere distante da tutti gli altri per un buon multiplo di lunghezze d'onda λ , e dunque manifesta una diversa matrice \mathcal{R}_h^k , ovvero ogni vettore \mathbf{h}_k risulta essere una realizzazione di una diversa d.d.p. gaussiana multivariata a media nulla e covarianza \mathcal{R}_h^k . Ciò consente di impostare la stima di canale secondo un approccio MMSE, capace di sfruttare l'informazione legata alla conoscenza delle matrici \mathcal{R}_h dei diversi utenti, e in grado di subire la *pilot contamination* solo nel caso di utenti che, oltre alla stessa sequenza ϕ , hanno in comune anche il medesimo orientamento rispetto alla BS nei cui confronti si sta stimando il vettore \mathbf{h} . Per una esemplificazione di quanto illustrato, si veda la fig. 21.12.

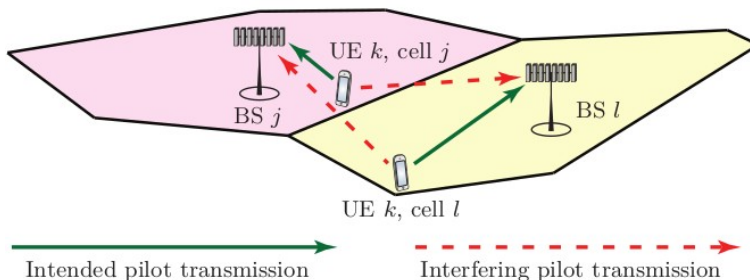


Figura 21.12: Mentre lo UE nella cella l causa una forte interferenza nella ricezione presso la BS $_j$ della sequenza pilota ϕ che ha in comune con lo UE nella cella j , lo stesso non avviene tra la cella j e quella l in virtù della diversa matrice di correlazione \mathcal{R}_h , che a sua volta dipende dagli elementi riflettenti prossimi alla BS e che determinano i cammini multipli, simili per i due UE per la BS j , e differenti per quella l

Per quanto riguarda le matrici \mathcal{R}_h , esse vengono stimate a partire dalle stime dei vettori \mathbf{h} collezionate nel tempo, e mantenute aggiornate nel caso di mobilità adattandone la stima a partire dai dati via via raccolti.

21.8 Single frequency network - SFN

Il caso della diffusione televisiva digitale terrestre DVB-T¹⁰⁵ rientra nella casistica di un sistema MISO in quanto prevede la ricezione da parte dell'apparecchio televisivo di segnali trasmessi da più antenne dislocate sul territorio, ognuna delle quali effettua una trasmissione OFDM

1. centrata sulla stessa frequenza portante, i cui simboli
2. contengono esattamente gli stessi bit e
3. sono trasmessi esattamente allo stesso istante.

Notiamo come i requisiti 1) e 3) siano comuni a quelli necessari a realizzare il collegamento di uplink di un sistema MU-MIMO (§ 21.7.3), conseguiti in tale contesto per mezzo dei protocolli di controllo peculiari dei relativi sistemi di accesso multiplo, che non abbiamo approfondito¹⁰⁶. In questa sezione illustriamo invece come sia affrontato il problema per i sistemi di diffusione DVB-T, non prima però di aver discusso le differenze rispetto alla tecnica *multiple frequency network* (MFN) adottata dalla diffusione televisiva *analogica*.

Rete a frequenza multipla o MFN La televisione analogica (§ 25.1) si basa sulla assegnazione di una diversa portante (o canale) a ciascuna emittente (o broadcaster), su cui trasmettere il proprio segnale con modulazione AM-BLR, tipicamente ricevuto dall'apparecchio TV mediante una antenna direttiva posta sul tetto ed orientata verso il ripetitore più vicino, dunque in condizione di visibilità. Nel caso di antenna indoor si verifica(va) di frequente la comparsa di immagini sdoppiate (dette *ghost* o fantasma¹⁰⁷) dovute all'insorgenza di cammini multipli, con conseguente ricezione di copie ritardate dello stesso segnale. La presenza di più ripetitori, necessari a coprire differenti regioni del territorio per le emittenti nazionali, richiede(va) quindi l'adozione di frequenze portanti differenti per ciascuna regione, pena la comparsa di *ghost* legati alla ricezione del segnale trasmesso (per la stessa emittente) dai diversi ripetitori, ognuno posto ad una differente distanza dalla TV, qualora utilizzanti la stessa portante.

Fantasma televisivi ma non radiofonici Cogliamo l'occasione per investigare sul meccanismo di formazione dei *ghost* televisivi in una SFN: sono causati dalla *sincronizzazione di quadro* (§ 25.1.2) che l'apparecchio analogico consegue con riferimento al segnale *più forte* ricevuto. In virtù della linearità della AM, la demodulazione della somma delle

¹⁰⁵Vedi ad es. <https://en.wikipedia.org/wiki/DVB-T>

¹⁰⁶In effetti nella trasmissione di *uplink* di un sistema MU-MIMO il requisito 3) è ancora più stringente, dato che in tal caso è richiesto che i simboli trasmessi dagli UE *arrivino* alla BS allo stesso istante, cosa resa possibile dalla conoscenza della distanza tra ciascuno di essi e la BS; ciò non è possibile in un sistema *broadcast*, data la presenza di *molteplici* ricevitori ad antenna singola, al posto di una singola BS.

¹⁰⁷Vedi ad es. [https://en.wikipedia.org/wiki/Ghosting_\(television\)](https://en.wikipedia.org/wiki/Ghosting_(television))

repliche produce un segnale di immagine (luminanza e cromaticità) che è la somma delle relative immagini, sfalsate dei corrispondenti ritardi nel tempo di arrivo. Nel caso della radio AM-BLD-PI (§ 12.1.1.2) manca invece la necessità di sincronizzazione, la frequenza portante ridotta rende trascurabile la differenza di fase tra le repliche ritardate, e la minore occupazione spettrale rende trascurabile l'entità della distorsione lineare. Infine nella radio FM il PLL del ricevitore (§ 12.3.2.1) si aggancia alla portante del segnale più forte.

Vantaggi di una SFN Il vantaggio più evidente rispetto ad una MFN è una migliore efficienza spettrale, in quanto lo stesso programma diffuso via MFN in regioni limitrofe necessita di altrettanti canali radio, contro l'unico canale necessario con una architettura SFN. A questa considerazione si può aggiungere che mentre nel caso di TV analogica basata su MFN ogni emittente necessita di un proprio canale, nel DVB i canali radio sono usati per trasmettere interi *transport stream* MPEG o TS (§ 10.3.2.1), ognuno dei quali ospita un multiplex di *elementary stream* (§ 10.3.2) provenienti da più emittenti, che così condividono la medesima SFN e dunque un unico canale radio.

Un secondo vantaggio è indicato come *network gain* e consiste nel vantaggio tipico della trasmissione MISO di poter sommare in modo coerente più repliche dello stesso messaggio, purché giunte nei limiti del tempo di guardia per come descritto appresso. A questo corrisponde una migliore qualità del segnale in termini di SNR, ovvero una minore criticità nei confronti della figura di rumore del ricevitore. I dispositivi che ne traggono maggior vantaggio sono i ricevitori mobili o nomadici, che non utilizzando antenne direttive possono facilmente trovarsi nelle condizioni di ricevere segnali da più fonti.

Tempo di guardia per una SFN Come anticipato, la diffusione in una rete SFN si avvale di una trasmissione OFDM *sincronizzata* da parte dei diversi trasmettitori Tx_i , che grazie all'inserimento del *prefisso ciclico* di durata T_g (o *tempo di guardia*, pag. 516) rende il ricevitore *esente* da interferenza intersimbolica qualora T_g sia maggiore del massimo ritardo legato al multipath (che indichiamo con $\delta\tau$), sommato (vedi fig. 21.13) alla massima differenza dei ritardi $\Delta\tau$ con cui viene ricevuto il segnale trasmesso dalle antenne Tx_i .

Dalla formula $\tau = d/c$ che fornisce il ritardo τ con cui si riceve un segnale ra-

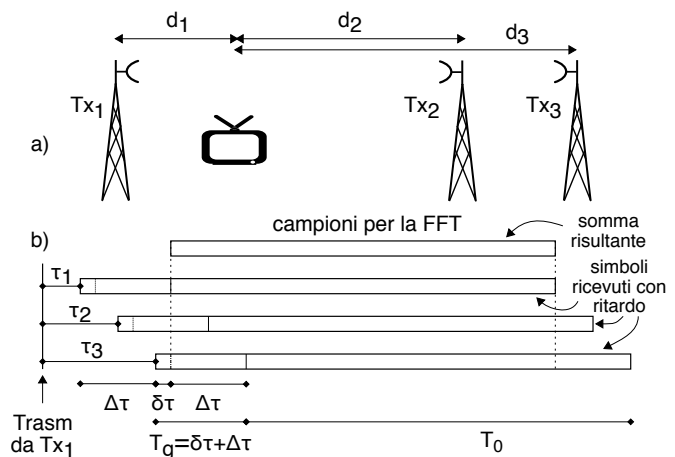


Figura 21.13: Ricezione SFN: a) - tre trasmettitori a diversa distanza dal ricevitore b) - tempo di guardia T_g somma di due ritardi *massimi*: $\delta\tau$ del mutipath, e $\Delta\tau$ tra prima e ultima replica

dio trasmesso da una distanza d e che si propaga alla velocità $c = 3 \cdot 10^8$ metri/sec si ottiene che per $d = 100$ metri τ ha un valore di 330 nsec, mentre per 100 Km si ottiene $\tau = 330 \mu\text{sec}$: la scelta del valore di T_g impone quindi quello del *massimo* distanziamento tra antenne che prendono parte allo stesso sistema SFN.

Valori di $T_g \approx 200 \mu\text{sec}$ sono adeguati per le reti a diffusione nazionale con ripetitori distanziati di una cinquantina di Km, mentre reti di ambito regionale o locale possono adottare valori inferiori.

Legame tra T_g ed efficienza spettrale Come illustrato in fig. 21.13, dalla somma delle repliche ricevute si individua una finestra di durata T_0 secondi (il periodo principale) i cui campioni di involuppo complesso vengono inviati (vedi fig. 21.10) al computo di una FFT in grado di ri-ottenere i punti di costellazione con cui sono modulate le sottoportanti. A questo riguardo il sistema DVB-T prevede di poter operare nelle modalità dette $2K$ ed $8K^{108}$, corrispondenti ad una FFT di dimensione N pari a 2048 e 8192 (vedi tab. 21.2) di cui solamente un sottoinsieme di portanti \tilde{N} sono *attive* ovvero effettivamente modulate (§ 16.8.2), in parte con dati di servizio¹⁰⁹, ed in larga misura con i dati relativi alla trasmissione¹¹⁰. Notiamo inoltre che i parametri di tab. 21.2 si riferiscono ad un distanziamento tra canali SFN (ovvero TS differenti) pari ad 8 MHz, mantenendo quindi una separazione (o banda di guardia) di 580 KHz tra due canali SFN.

Rimarchiamo ora il fatto che solamente la porzione di simbolo OFDM relativa al periodo principale (di durata T_0) trasporta informazione, e dunque una volta determinato il valore T_g necessario a supportare le caratteristiche territoriali della SFN che si intende sviluppare è opportuno far sì che la somma $T_g + T_0$ sia la massima possibile, in quanto minore è il rapporto T_g/T_0 e maggiore sarà l'efficienza del collegamento¹¹¹, e dunque la velocità binaria effettivamente trasferita.

Sempre con riferimento alla tab. 21.2 osserviamo che, fissata la durata di T_0 per le due modalità $8K$ e $2K$, il numero delle possibili scelte per il rapporto T_g/T_0 è ristretto a 4 alternative, ovvero $1/4$, $1/8$, $1/16$ ed $1/32$; in particolare, i valori più elevati di T_g (associati a ripetitori più distanti) sono possibili solo in modalità $8K$. Per lo stesso

¹⁰⁸Le cui sigle non hanno nulla a che vedere con la qualità dell'immagine!

¹⁰⁹Come *dati di servizio* si intendono le portanti pilota di sincronizzazione in frequenza e nel tempo, di ausilio alla stima di canale, e che trasportano dati di segnalazione a riguardo (tra l'altro) del tipo di modulazione (M-QAM con $M = 4, 16$ o 64) e di codifica di canale, parametri questi ultimi che determinano la velocità binaria conseguita dal flusso dati trasmesso, variabile in un intervallo da 5 a 31 Mbps.

¹¹⁰Che consistono nel *transport stream* MPEG che multiplexa le emittenti che prendono parte alla SFN, a cui sono stati applicati stadi di codifica di canale e di scrambling, vedi ad es. Lo standard DVB-T, Centro Ricerche RAI, da cui è anche tratta la tab. 21.2

¹¹¹Indicando con \tilde{N}_d il numero di portanti che trasportano informazione, con R_c il tasso di codifica FEC, e con M il numero di bit/portante della costellazione adottata, se il tempo di guardia fosse nullo si otterrebbe una velocità $f_b^{max} = \tilde{N}_d \cdot R_c \cdot M / T_0$. Dato però che il periodo di simbolo è pari a $T_s = T_g + T_0$, la velocità si riduce a $f_b = f_b^{max} \frac{T_0}{T_g + T_0}$ e dunque l'efficienza è pari a $\eta = \frac{T_0}{T_g + T_0} = \frac{1}{1 + T_g/T_0}$. Con i valori di T_g/T_0 in fig. 21.2 si ottiene $\eta = 0.8, 0.88, 0.94$ e 0.97 .

	modo 8k				modo 2k			
dimensione FFT N	8192				2048			
portanti attive \tilde{N} (di cui dati)	6817 (6048)				1705 (1512)			
periodo principale T_0	896 μsec				224 μsec			
intervallo tra portanti $\Delta = 1/T_0$	1116 Hz				4464 Hz			
banda occupata $B = \tilde{N}\Delta$	7.61 MHz							
rapporto T_g/T_0	1/4	1/8	1/16	1/32	1/4	1/8	1/16	1/32
durata T_g (μsec)	224	112	56	28	56	28	14	7

Tabella 21.2: Parametri di una trasmissione DVB-T per una canalizzazione di 8 MHz

motivo la modalità 2k, che impone l'uso di valori di T_g inferiori, può essere adottata solamente in regioni in cui la densità delle antenne è particolarmente elevata.

Rete di distribuzione e sincronizzazione in frequenza Descriviamo per ultima la soluzione che si è adottata per soddisfare le condizioni elencate a pag. 750 e che permettono la realizzazione di una SFN, ovvero come fare in modo che le diverse antenne trasmettano allo stesso istante e frequenza simboli OFDM che contengono gli stessi bit. Questi ultimi fanno parte di un flusso di trasporto MPEG o TS (§ 10.3.2.1)¹¹², che deve poi essere ulteriormente elaborato aggiungendo una protezione FEC mediante due stadi di codifica di canale associati ad interleaving (§ 17.4.2.6). La trasmissione OFDM del risultato viene anche indicata come trasmissione COFDM (§ 16.8.10).

La soluzione cercata è ottenuta grazie all'adozione della architettura di fig. 21.14, che centralizza la creazione del flusso di bit da trasmettere in un unico luogo e lo inoltra ai ripetitori che realizzano la SFN per mezzo di una differente rete di distribuzione, tipicamente in fibra ottica, oppure realizzata con ponti radio, o via satellite.

¹¹²Costituito da una serie di pacchetti di 188 byte il cui contenuto è descritto dal campo PID presente nell'header, che può anche specificare un pacchetto *nullo* inserito per adattare la velocità variabile dei contenuti multimediali con quella fissa offerta dal canale a disposizione.

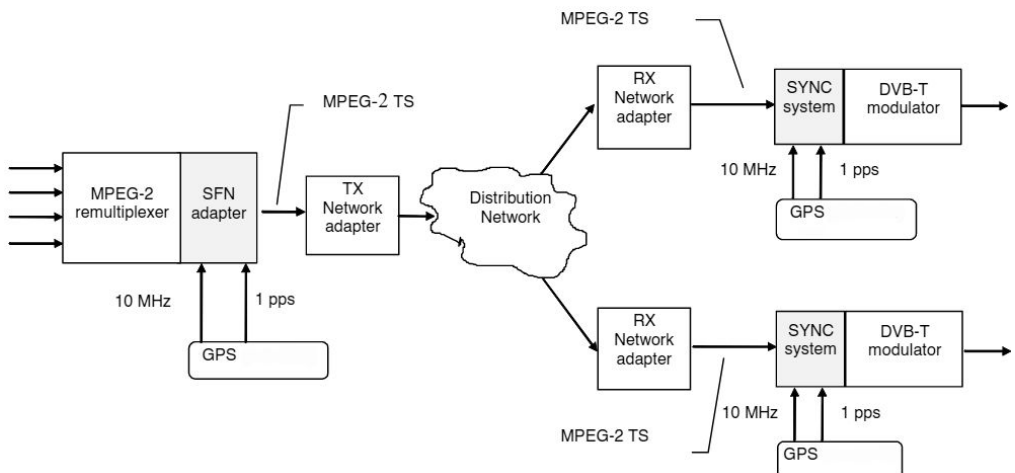


Figura 21.14: Distribuzione ai ripetitori del segnale DVB-T mediante adattatori SFN

La sincronizzazione *in frequenza* tra le diverse antenne è conseguita grazie al comune utilizzo da parte delle antenne di una medesima *sorgente esterna* di sincronismo, come quella offerta dal segnale GPS¹¹³ che provvede a fornire due riferimenti temporali, uno (detto PPS o *pulse per second*) che scandisce intervalli di 1 secondo, e l'altro a 10 MHz che suddivide lo stesso secondo in 10^7 parti ovvero individua intervalli di 100 nanosecondi. Lo stadio di modulazione finale di ciascuna antenna può pertanto lavorare ad una frequenza ottenuta per sintesi digitale (§ 12.4.3) a partire dal comune riferimento a 10 Mhz.

Pacchetto MIP e megafame Lo stesso segnale GPS, ricevuto oltre che dalle antenne anche dal nodo di controllo, permette a quest'ultimo di orchestrare la *sincronizzazione temporale* dei simboli OFDM trasmessi (e del loro contenuto) grazie all'inserimento nel flusso MPEG di uno speciale pacchetto TS (con PID 0x15) denominato *mega frame initialization packet* o MIP, contenente oltre ad un *timbro temporale* ottenuto con riferimento al PPS e con risoluzione 100 nsec, anche le direttive a riguardo dei parametri per lo stadio FEC e per il tipo di costellazione M-QAM da applicare alle portanti.

L'inserimento del pacchetto MIP avviene in un punto qualunque di una struttura sintattica denominata *mega frame* e che viene definita dallo *SFN adapter* del multiplexer, delimitando una sequenza di pacchetti TS che sono a loro volta contenuti in 8 *frame*¹¹⁴ DVB-T in modalità 8K (oppure in 32 frames in modalità 2K). La durata temporale di un megafame dipende solamente dal rapporto T_s/T_0 e dalla banda del canale, ma non da T_0 né dalla scelta della costellazione o dal tipo di FEC, variando tra circa 0.5 e 0.6 secondi per un canale di 8 MHz.

Sincronizzazione temporale Dato che la rete di distribuzione presenta ritardi di attraversamento differenti per raggiungere le diverse antenne, la sincronizzazione temporale è ottenuta inserendo in un punto qualunque del megafame m il pacchetto MIP contenente (tra le altre) le informazioni

- un *puntatore* che individua l'inizio del megafame seguente $m + 1$, espresso come il numero di pacchetti TS che separano il MIP dal megafame successivo;
- un *synchronization time stamp* (STS) che misura il tempo che intercorre¹¹⁵ tra l'impulso PPS che *precede* il megafame $m + 1$, e la trasmissione del primo bit del primo pacchetto di tale megafame, evento indicato come **A** in figura;
- il valore *maximum delay*¹¹⁶ che deve essere maggiore del massimo ritardo di attraversamento della rete di distribuzione, per un qualsiasi percorso.

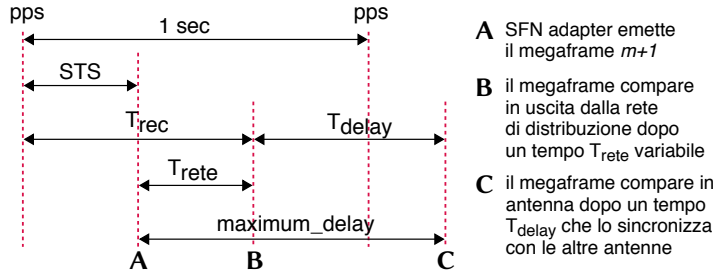
¹¹³Vedi https://it.wikipedia.org/wiki/Sistema_di_posizionamento_globale

¹¹⁴La definizione esatta di *frame* DVB-T trascende lo scopo di questa sezione, e verrà (forse) affrontata in una prossima edizione. Il lettore interessato può affrontare la lettura di ETSI EN 300 744 V1.6 - Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television

¹¹⁵Espresso come numero di colpi di clock da 10 MHz, ovvero come multiplo di 0.1 μ sec.

¹¹⁶sempre misurato in multipli di 0.1 μ sec

L'evento indicato come **B** alla figura seguente rappresenta l'arrivo del megaframe presso una delle antenne che prendono parte alla SFN, ed il valore T_{rec} individua il ritardo con cui si verifica **B** rispetto al pps che precede l'inizio del megaframe. Il compito del SYNC system (mostrato in fig. 21.14) di ciascuna antenna della SFN è dunque quello di



- localizzare l'inizio del megaframe $m+1$ a partire dall'informazione fornita dal campo *puntatore* presente nel MIP contenuto nel megaframe m ;
- calcolare il valore

$$T_{delay} = (STS + maximum\ delay - T_{rec}) \text{ modulo } 10^7$$

di cui ritardare la trasmissione del megaframe $m+1$ (evento **C**) rispetto all'istante **B** in cui lo si è ricevuto per mezzo della rete di distribuzione.

In tal modo l'inizio della trasmissione del megaframe sarà lo stesso per tutte le antenne che partecipano alla SFN. Come anticipato, il MIP convoglia anche altre informazioni, come il tipo di codifica FEC e la costellazione da adottare: questi (eventualmente nuovi) parametri vengono inseriti nelle portanti pilota dei simboli OFDM utilizzati per trasmettere il contenuto del megaframe $m+1$, e resi operativi a partire dal megaframe $m+2^{117}$, in modo da permettere ai dispositivi riceventi di adattare a loro volta lo stadio di demodulazione alla nuova configurazione del modem OFDM.

Il sistema di diffusione DVB-T presenta molteplici aspetti qui solamente accennati o del tutto trascurati; il lettore interessato può approfondirli presso ETSI TR 101 190 V1.3 - Digital Video Broadcasting (DVB); Implementation guidelines for DVB terrestrial services; Transmission aspects.

21.9 Appendice

21.9.1 Entropia di variabile gaussiana complessa multivariata

Estendiamo il risultato trovato al § 9.3.2 per l'entropia differenziale $h_G(X) = \frac{1}{2} \log_2(2\pi e \sigma^2)$ di una v.a. gaussiana reale X con varianza σ^2 a due altri casi, quello in cui X sia una v.a. *vettoriale* a componenti gaussiane (§ 6.5) con matrice di covarianza Σ , e quello in cui oltre a ciò, le v.a. marginali assumano valori *complessi*.

Caso reale La v.a. X (vettore colonna ad n elementi) è descritta dalla d.d.p. n -dimensionale

$$p_X(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \Sigma^{-1}(\mathbf{x} - \mathbf{m})\right\} \quad (21.82)$$

¹¹⁷Vedi ETSI TS 101 191 V1.4 - Digital Video Broadcasting (DVB); DVB mega-frame for Single Frequency Network (SFN) synchronization

alla quale applichiamo la definizione di entropia $H_X = E \{-\log_2 p_X(\mathbf{x})\}$, ponendo $\mathbf{m} = 0$ in quanto l'entropia differenziale è indipendente dalle traslazioni (pag. 266), e tenendo conto che $\log_2 p(\mathbf{x}) = \frac{\ln p(\mathbf{x})}{\ln 2}$, in modo da ottenere

$$\begin{aligned} h_{GM} &= \frac{1}{\ln 2} E \left\{ \frac{n}{2} \ln(2\pi) + \frac{1}{2} \ln(\det(\Sigma)) + \frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x} \right\} = \\ &= \frac{1}{\ln 2} \left(\frac{n}{2} \ln(2\pi) + \frac{1}{2} \ln(\det(\Sigma)) + \frac{1}{2} E \{ \mathbf{x}^T \Sigma^{-1} \mathbf{x} \} \right) \end{aligned} \quad (21.83)$$

All'ultimo termine si applica ora *il trucco della traccia*¹¹⁸ che permette due sostituzioni: la prima è¹¹⁹ $\mathbf{x}^T \Sigma^{-1} \mathbf{x} = \text{tr}(\mathbf{x}^T \Sigma^{-1} \mathbf{x})$, e la seconda che si può scrivere¹²⁰ $\text{tr}(UVW) = \text{tr}(WUV)$, dunque

$$E \{ \mathbf{x}^T \Sigma^{-1} \mathbf{x} \} = E \{ \text{tr}(\mathbf{x}^T \Sigma^{-1} \mathbf{x}) \} = \text{tr}(E \{ \mathbf{x} \mathbf{x}^T \} \Sigma^{-1}) = \text{tr}(\Sigma \Sigma^{-1}) = n$$

dato che $E \{ \mathbf{x} \mathbf{x}^T \}$ è una matrice con elementi $E \{ x_i x_j \} = \sigma_{ij}$ ossia esattamente gli elementi della covarianza Σ , il cui prodotto per Σ^{-1} è dunque la matrice identità di dimensione n . Pertanto risulta

$$h_{GM} = \frac{1}{\ln 2} \left(\frac{n}{2} \ln(2\pi) + \frac{1}{2} \ln(\det(\Sigma)) + \frac{n}{2} \right) \quad (21.84)$$

e dato che $\ln(e) = 1$ scriviamo

$$\begin{aligned} h_{GM} &= \frac{1}{\ln 2} \left(\frac{n}{2} \ln(2\pi) + \frac{1}{2} \ln(\det(\Sigma)) + \frac{n}{2} \ln e \right) = \\ &= \frac{1}{2 \ln 2} \ln((2\pi e)^n \det(\Sigma)) = \frac{1}{2} \log_2((2\pi e)^n \det(\Sigma)) \end{aligned}$$

che può anche essere scritta come $h_{GM} = \frac{1}{2} \log_2(\det(2\pi e \Sigma))$.

Caso complesso In questo caso \mathbf{Z} è un vettore ad n componenti complesse $z_k = x_k + jy_k$ note come v.a. *gaussiana centrata a simmetria circolare*¹²¹, le cui parti reale ed immaginaria x_k, y_k sono v.a. gaussiane indipendenti (in quanto $E \{ x_k y_k^* \} = 0$) a media nulla e varianza $\frac{1}{2} \sigma_k^2$ (in modo che $E \{ z_k z_k^* \} = E \{ x_k x_k^* \} + E \{ y_k y_k^* \} = \sigma_k^2$), con matrice di covarianza Σ *Hermitiana*¹²² ad elementi $\sigma_{i,j} = E \{ z_i z_j^* \}$. Per calcolare l'entropia h_{GC} di questo tipo di sorgente informativa¹²³ si può procedere in (almeno) due modi: il primo prende in considerazione la d.d.p. di \mathbf{Z} che ora si scrive

$$p_X(\mathbf{z}) = \frac{1}{\pi^n \det(\Sigma)} \exp \{ -\mathbf{z}^\dagger \Sigma^{-1} \mathbf{z} \} \quad (21.85)$$

¹¹⁸Vedi ad es. <https://sgfin.github.io/>

¹¹⁹L'operatore matriciale $\text{tr}(A)$ restituisce la somma degli elementi sulla diagonale della matrice quadrata A . Dal momento che $\mathbf{x} \Sigma \mathbf{x}^T$ è un numero, ovvero una matrice 1×1 , la sua traccia è pari al numero stesso.

¹²⁰Nota come proprietà *ciclica*, vedi [https://en.wikipedia.org/wiki/Trace_\(linear_algebra\)](https://en.wikipedia.org/wiki/Trace_(linear_algebra))

¹²¹Vedi ad es. https://en.wikipedia.org/wiki/Complex_normal_distribution

¹²²Ossia Σ coincide con la propria trasposta coniugata (o aggiunta); dunque se la matrice è ad elementi reali, è simmetrica. Vedi ad es. https://it.wikipedia.org/wiki/Matrice_hermitiana

¹²³O molto più banalmente, il campionamento dell'involuppo complesso del rumore osservato su n canali equivalenti di bassa frequenza.

in cui (a parte il resto) al posto del trasposto T che compare nella (21.82) ora si usa il trasposto coniugato \dagger o *Hermitiano*. A partire dalla (21.85), applicando sviluppi analoghi alle (21.83) - (21.84) si può ottenere

$$h_{GC}(\mathbf{Z}) = \log(\det(\pi e \Sigma))$$

Un secondo modo di procedere prende in considerazione il fatto che l'entropia $h_{GC}(\mathbf{Z})$ è pari a quella $h(\mathbf{X}, \mathbf{Y})$ di un vettore $[\mathbf{X}, \mathbf{Y}] = [\Re\{\mathbf{Z}\}, \Im\{\mathbf{Z}\}]$ ottenuto concatenando i vettori delle parti reale ed immaginaria di \mathbf{Z} . Dato che gli elementi di \mathbf{X} ed \mathbf{Y} sono statisticamente indipendenti a coppie perché incorrelati, si ha $h(\mathbf{X}, \mathbf{Y}) = h(\mathbf{X}) + h(\mathbf{Y})$. Dato che inoltre sia \mathbf{X} che \mathbf{Y} sono vettori gaussiani a media nulla e covarianza $\frac{1}{2}\Sigma_z$ l'entropia differenziale di ciascuno di essi è pari a $\frac{1}{2} \log_2(\det(2\pi e \frac{1}{2}\Sigma_z))$. Mettendo tutto assieme si ottiene

$$\begin{aligned} h_{GC}(\mathbf{Z}) &= h(\mathbf{X}, \mathbf{Y}) = h(\mathbf{X}) + h(\mathbf{Y}) = \\ &= \frac{1}{2} \log_2\left(\det(2\pi e \frac{1}{2}\Sigma_z)\right) + \frac{1}{2} \log_2\left(\det(2\pi e \frac{1}{2}\Sigma_z)\right) = \\ &= \log_2(\det(\pi e \Sigma_z)) \end{aligned}$$

L'opera

Trasmissione dei Segnali e Sistemi di Telecomunicazione

è il risultato di un progetto ventennale di cultura libera, aggiornato di continuo ed evolutosi fino alla forma attuale. La sua disponibilità pubblica è regolata dalle norme di licenza CREATIVE COMMONS

*Attribuzione - Non commerciale -
Condividi allo stesso modo*



<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.it>

e tutte le risorse relative al testo sono accessibili presso

<https://teoriadeisignali.it/libro/>

Puoi contribuire al suo successo promuovendone la diffusione e supportarne lo sviluppo attraverso una donazione, in buona parte devoluta ai progetti *open source*¹ che ne hanno resa possibile realizzazione e divulgazione. Ai donatori viene accordato un accesso *vitalizio* al formato PDF *navigabile* di tutte le edizioni presenti *e future*.

1

- . Lyx - <http://www.lyx.org/>
- . L^AT_EX - <https://www.latex-project.org/>
- . TeX Users Group - <https://tug.org/>
- . Inkscape - <http://www.inkscape.org/>
- . Gnuplot - <http://www.gnuplot.info/>
- . Octave - <http://www.gnu.org/software/octave/>
- . Geany - <https://www.geany.org/>
- . Linux - <https://www.linux.it/>
- . Free Software Foundation - <https://shop.fsf.org/>
- . GNOME Foundation - <https://www.gnome.org/>
- . Mozilla Foundation - <https://www.mozilla.org/it/>
- . Wikipedia - <https://it.wikipedia.org>
- . Internet Archive - <https://archive.org/about/>
- . Creative Commons - <https://creativecommons.it/chapterIT/>
- . WordPress - <https://it.wordpress.org/>
- . Phplist - <https://www.phplist.org/>