

Cap 1 PRELUCRAREA NUMERICĂ A SEMNALELOR

Introducere

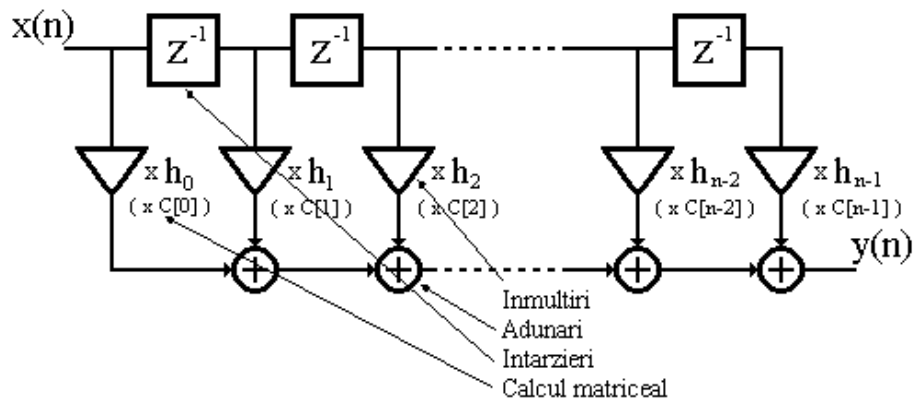
DSP este un acronim pentru Digital Signal Processing (procesare de semnal digital) și reprezintă o familie de aplicații care tratează o gamă largă de semnale (audio, video etc.) prin procesarea lor într-un procesor digital specializat. Acesta din urmă este cunoscut tot sub numele de DSP *Digital Signal Processor*, sau procesor de semnal digital.

Procesoarele de semnal digital preiau în timp real informația de mare viteză și o manipulează în multiple moduri și scopuri. Cu DSP putem restaura vechi înregistrări pe suport de vinyl la claritatea originală, putem șterge zgomotul static din liniile telefonice și putem "vedea" din sateliți detalii ale unor obiecte de dimensiunea unei mingi de golf. În autoturisme, DSP-urile creează, în linia audio, sunetul spațial (surround), în timp ce răspund de sistemele de suspensii active, frânele ABS sau injecția electronică. În telefonie mobilă, prin procesarea digitală de semnal putem "înghesui" mai multe conversații pe undele aeriene, simultan cu codificarea lor pentru prevenirea interceptării. Procesoarele de semnal digital pot fi întâlnite în aparatele de monitorizare medicale, în aparatele auditive, supercalculatoare sau modemuri. Sunt folosite în detectoare radar, în televizoare digitale sau combine audio, instalații de amplificare și laser pentru spectacole. DSP-urile sunt, de fapt, o parte componentă a existenței noastre moderne, translatând semnalele în inteligență și permițându-ne să luăm decizii mai precise, într-un timp mult mai scurt.

1.1 Arhitectura DSP

Deși există o multitudine de procesoare DSP, ele conțin un set comun de caracteristici care pot fi clasificate în trei categorii: o aritmetică specializată de mare viteză, o capacitate de transfer de date din și către procesul continuu și o arhitectură de accesare multiplă a memoriei. Funcționarea unui DSP constă din câteva operații specifice: adunări și înmulțiri, întârzieri (delays) și calcul matricial. Fiecare dintre aceste operații necesită câteva condiții specifice: adunările și înmulțirile presupun citirea simultană a doi operanzi, executarea operațiilor (de obicei în același timp), salvarea rezultatului sau reținerea lui pentru repetarea operației; întârzierile presupun reținerea unei valori pentru utilizarea sa ulterioară; calculul matricial constă în preluarea datelor din locații consecutive de memorie sau copierea datelor între zone diferite de memorie.

Pentru a putea obține aceste performanțe obligatorii, procesoarele de semnal trebuie să ofere posibilitatea efectuării în paralel a înmulțirii și adunării, accesării multiple a memoriei și generării eficiente a adreselor, să conțină o multitudine de regiștri pentru reținere temporară a datelor, precum și facilități speciale cum ar fi întârzierile și adresarea circulară.

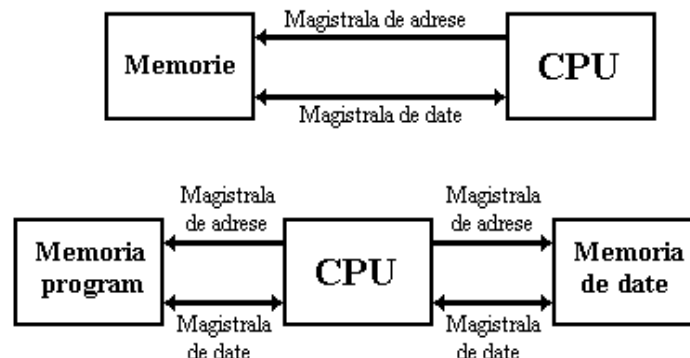


**Matematica unui DSP se rezuma la trei tipuri de operatii:
adunari, inmultiri, intarzieri**

Fig. 1-1

Cele mai multe operații pe care le efectuează un DSP constau în multiple adunări și înmulțiri. Pentru a citi doi operanzi într-un singur ciclu de instrucțiune va trebui să avem simultan două accesări ale memoriei. De fapt, gândindu-ne că va trebui să salvăm rezultatul și să citim instrucțiunea însăși, vom dori să avem mai mult de două accesări ale memoriei într-un același ciclu de instrucțiune. Se știe că nu se pot accesa două adrese de memorie diferite pe aceeași magistrală de date. Accesarea multiplă este posibilă în cazul DSP folosind două tehnologii derivate din modelele arhitecturale adoptate: arhitectura Harvard sau arhitectura Von Neumann modificată.

În arhitectura Harvard, DSP are două magistrale separate de memorie. Acestea permit accesarea simultană a două locații de memorie. În procesoarele de uz general, una dintre magistrale este dedicată citirii codului instrucțiunii, iar cealaltă aducerii operanzilor. Arhitectura Harvard-DSP permite ca magistrala de program să poată fi folosită pentru citirea operanzilor. Deoarece avem simultan nevoie de trei informații, Harvard-DSP include o memorie cache care poate stoca instrucțiunile comune ce vor fi reutilizate, lăsând ambele magistrale libere pentru operanzi. Această extensie poartă denumirea de *Super Harvard ARChitecture* (SHARC).



Arhitecturile Von Neumann si Harvard

Fig. 1-2

Există și un dezavantaj al arhitecturii Harvard. Un DSP care procesează cuvinte de 32 de biți, la un spațiu de memorie cu adrese de 32 de biți, necesită cel puțin 64 de pini pentru fiecare magistrală de date, deci un total de 128 de pini pentru conectarea cu exteriorul. Rezultatul este un cip foarte mare, greu de amplasat în circuite dedicate.

1.2 Contactul cu realitatea

Pe lângă calcule matematice, în practică procesorul de semnal digital trebuie să comunice cu lumea reală. Într-o aplicație DSP tipică, procesorul va trebui să prelucreze date provenite de la surse multiple. El va trebui să poată primi și transmite date în timp real, fără a-și întrerupe calculele matematice interne. Majoritatea aplicațiilor DSP implică tratarea semnalelor analogice. Acestea trebuie convertite în formă digitală (cu ajutorul unor convertoare A/D), într-o secvență discretă de valori măsurabile. O bună parte din informație este pierdută la conversia analog-digitală datorită impreciziei de măsurare, inacurateții în sincronizare, limitări ale duratei de efectuare a măsurării. Aceste efecte sunt numite erori de cuantificare.

Semnalul continuu, analog, trebuie să fie reținut înainte de a fi eșantionat. Altfel s-ar putea modifica în timpul măsurării. După ce a fost reținut, semnalul va fi măsurat și apoi convertit într-o valoare digitală. Prin eșantionare obținem o mulțime de numere în format binar ce reprezintă mărimi ale semnalului, luate de obicei la intervale constante de timp.

Există trei moduri de interacțiune a DSP-ului cu lumea reală: prin semnale care sunt recepționate și transmise direct, prin comunicarea cu un sistem de control de tip diferit sau prin comunicarea cu alte procesoare DSP de același tip. Aceste căi de comunicație multiple marchează cea mai importantă distincție dintre procesoarele DSP și procesoarele obișnuite.

Sistemele care conțin DSP trebuie să funcționeze în timp real, capturând și procesând informația pe măsură ce se schimbă. Mai întâi convertoarele analog-digitale trebuie să țină pasul cu modificările continue ale datelor reale. Dacă rămân în urmă, informația se pierde, iar semnalul este distorsionat. Convertorul A/D, "fotografiază" semnalul la intervale suficient de mici pentru a observa fluctuațiile relevante, iar DSP efectuează toate calculele necesare până în momentul în care o nouă "fotografie" apare pentru a fi procesată. Semnalul audio stereo trebuie prelucrat la frecvențe de aproximativ 44 kHz (44.000 de cicluri pe secundă), obligând DSP-ul să realizeze sute de milioane de operații pe secundă. Alte semnale, cum ar fi transmisiile prin satelit, sunt și mai rapide, ajungând la ordinul giga-hertzilor (miliarde de cicluri pe secundă).

Avantajele proiectării cu DSP sunt multiple:

- Cost - Cu un hardware foarte simplu, în absența unor elemente electronice sofisticate, putem construi dispozitive cu o putere de procesare extrem de mare.

- Fiabilitate - Orice circuit electronic este supus modificărilor de performanță, datorate variațiilor termice, îmbătrânirii, care pot duce la o funcționare incorectă.

- Flexibilitate - Prin folosirea tehnologiei DSP putem concepe un produs cu o anumită funcție, care, printr-o modificare de software, se transformă într-un cu totul alt produs, cu noi funcționalități.

CAP 2. Sisteme Liniare

Majoritatea tehnicilor DSP se bazează pe metoda "divide et impera" numită și metoda *superpoziției*. Semnalul procesat este împărțit în componente simple, fiecare componentă este procesată individual, rezultatele finale fiind adunate. Această metodă este foarte puternică în sensul că împarte o singură problemă mai complicată într-o serie de probleme mai simple și ușor de rezolvat. Superpoziția poate fi folosită doar pentru *sistemele liniare*, ceea ce înseamnă că numai anumite reguli matematice se pot aplica. Din fericire, majoritatea aplicațiilor întâlnite în știință și inginerie intră în această categorie. Acest capitol prezintă principiile fundamentale ale DSP: ce reprezintă un sistem liniar, diferite moduri de a împărți semnalele în componente mai simple și felul în care metoda superpoziției generează o serie de tehnici de procesare de semnal.

2.1 Semnale și sisteme

Un **semnal** este o descriere a modului în care un parametru variază în funcție de alt parametru. Spre exemplu, tensiunea se modifică în timp într-un circuit electronic sau luminozitatea variază cu distanța într-o imagine. Un **sistem** este un proces oarecare ce produce un *semnal de ieșire* ca răspuns la un *semnal de intrare* (figura 2-1). Sistemele continue au ca intrare și ieșire semnale continue cum ar fi circuitele analogice. Sistemele discrete au ca intrare semnale discrete, cum ar fi programele software ce manipulează vectori de valori.

Sunt mai multe reguli de numire a acestor semnale. Prima, semnalele *continue* folosesc paranteze rotunde, cum ar fi: $x(t)$ și $y(t)$, în timp ce semnalele *discrete* folosesc paranteze drepte, ca: $x[n]$ și $y[n]$. A doua, semnalele folosesc litere mici. Literele mari sunt rezervate pentru domeniul frecvență. A treia, numele dat semnalelor descrie, de obicei, parametrul care-l reprezintă.

La prima vedere poate părea imposibilă înțelegerea tuturor sistemelor existente. Din fericire majoritatea acestora intră într-o categorie numită **sisteme liniare**. Acest fapt este extrem de important. Fără conceptul de sistem liniar, suntem nevoiți să examinăm caracteristicile individuale a multor sisteme neprecizate. În schimb, cu acest concept, ne putem concentra asupra tratării categoriei sistemelor liniare ca un întreg. Prima cerință este să identificăm ce proprietăți fac un sistem liniar și dacă ele aparțin sistemelor electronice, software și altor sisteme de procesare.

2.2 Cerințe pentru liniaritate

Un sistem este numit *liniar* dacă prezintă două proprietăți matematice: **omogenitate** și **aditivitate**. Dacă putem demonstra că un sistem prezintă ambele proprietăți, atunci vom demonstra că sistemul este liniar. De asemenea dacă demonstrăm că sistemul nu prezintă una sau ambele proprietăți, am demonstrat, că sistemul este neliniar. O a treia proprietate, **invarianța la deplasare** nu este o cerință strictă a liniarității dar este o necesitate pentru majoritatea tehnicilor DSP. Dacă avem un sistem liniar folosit în DSP, obligatoriu el va include și proprietatea de invarianță la deplasare. Aceste trei proprietăți alcătuiesc teoria matematică de definire și utilizare a sistemelor liniare.

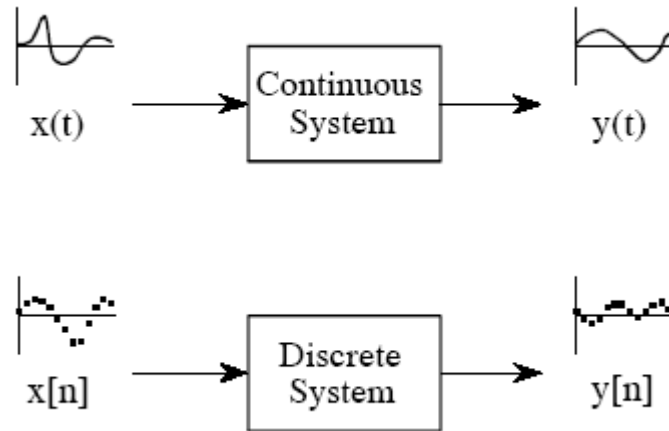


Fig. 2-1

Cum s-a ilustrat în figura 2-2, omogenitatea înseamnă că o modificare în amplitudinea semnalului de intrare, rezultă într-o modificare corespunzătoare în amplitudinea semnalului de ieșire. În termeni matematici, dacă un semnal de intrare $x[n]$ rezultă într-un semnal de ieșire $y[n]$, o intrare $k \times x[n]$ rezultă într-o ieșire $k \times y[n]$, pentru oricare ar fi semnalul de intrare și constanta k .

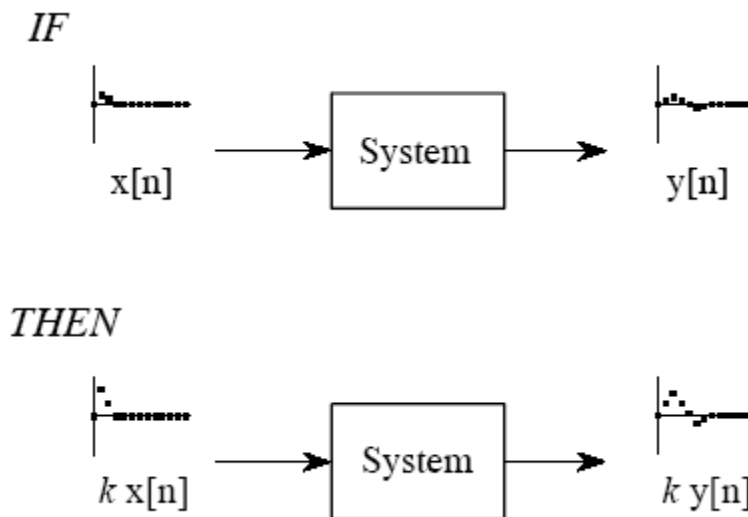


Fig. 2-2

Proprietatea de aditivitate este ilustrată în figura 2-3. Se consideră un sistem în care o intrare, $x_1[n]$, produce o ieșire $y_1[n]$. Mai departe, presupunem că o intrare diferită, $x_2[n]$, generează o altă ieșire, $y_2[n]$. Sistemul este aditiv dacă o intrare $x_1[n] + x_2[n]$ rezultă într-o ieșire $y_1[n] + y_2[n]$ pentru toate semnalele de intrare posibile.

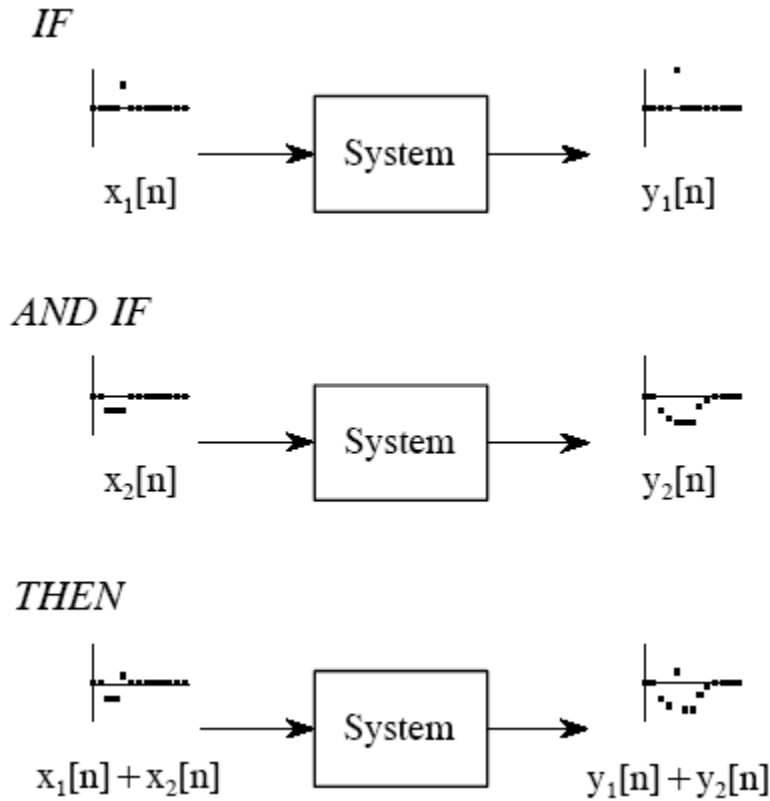


Fig. 2-3

Așa cum este arătat în figura 2-4, invarianța la deplasare înseamnă că o deplasare a semnalului de intrare va rezulta într-o deplasare a semnalului de ieșire. Cu alte cuvinte, dacă un semnal de intrare $x[n]$ rezultă într-un semnal de ieșire $y[n]$, atunci un semnal de intrare $x[n+s]$ rezultă într-un semnal de ieșire $y[n+s]$, pentru oricare ar fi semnalul de intrare și constanta s . Prin adunarea unei constante s la variabila independentă n forma de undă poate fi deplasată la dreapta sau la stânga pe direcția orizontală. Spre exemplu, când $s = 2$ semnalul se deplasează la *stânga* cu două eșantioane; când $s = -2$ semnalul se deplasează la *dreapta* cu două eșantioane.

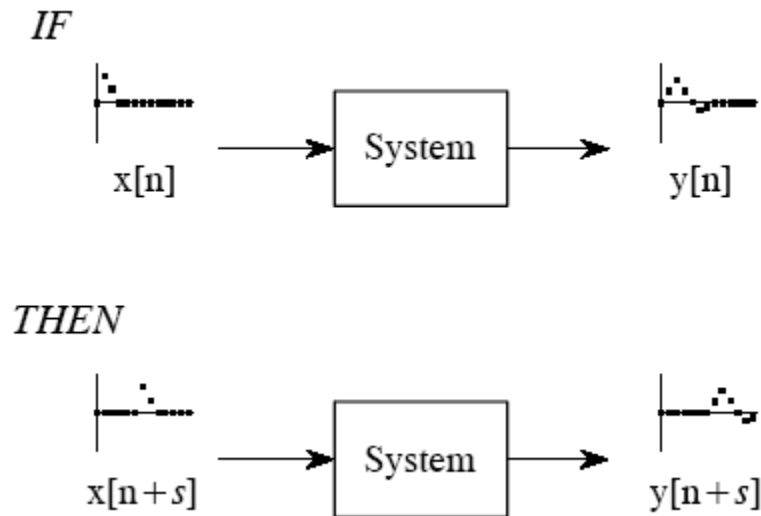


Fig. 2-4

2.3 Superpoziția, Fundament al DSP.

Când avem de-a face cu sisteme liniare, singurul mod în care pot fi combinate este *scalarea* (înmulțirea semnalelor cu constante) urmată de *adunare*. Figura 2-5 prezintă un exemplu: trei semnale $x_0[n]$, $x_1[n]$, $x_2[n]$, sunt adunate pentru a forma un al patrulea semnal $x[n]$. Acest proces de combinare a semnalelor prin scalare și adunare se numește **sinteză**.

Descompunerea este operația inversă sintezei în care un singur semnal este împărțit în două sau mai multe componente aditive. Aceasta este mult mai complicată decât sinteza deoarece există o infinitate de descompuneri pentru oricare semnal dat. Spre exemplu, numerele 15 și 25 pot fi sintetizate numai în numărul 40. În schimb numărul 40 poate fi descompus în: $1 + 39$ sau $2 + 38$ sau $-30,5 + 60 + 10,5$ etc.

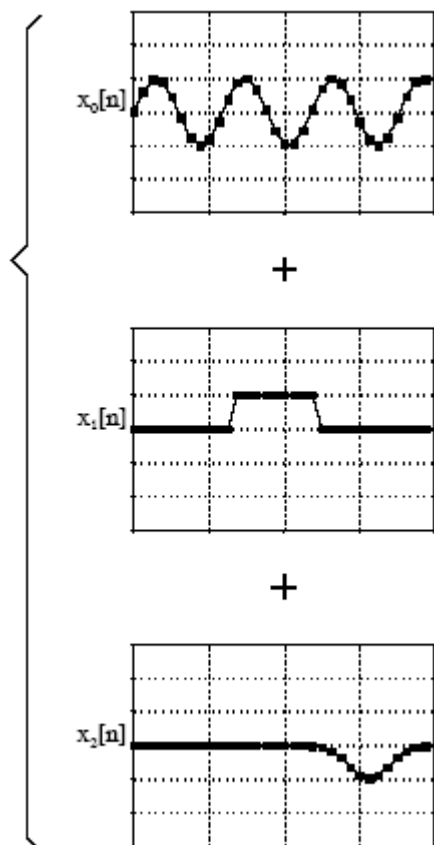
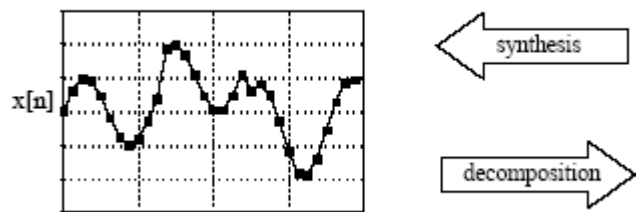


Fig. 2-5

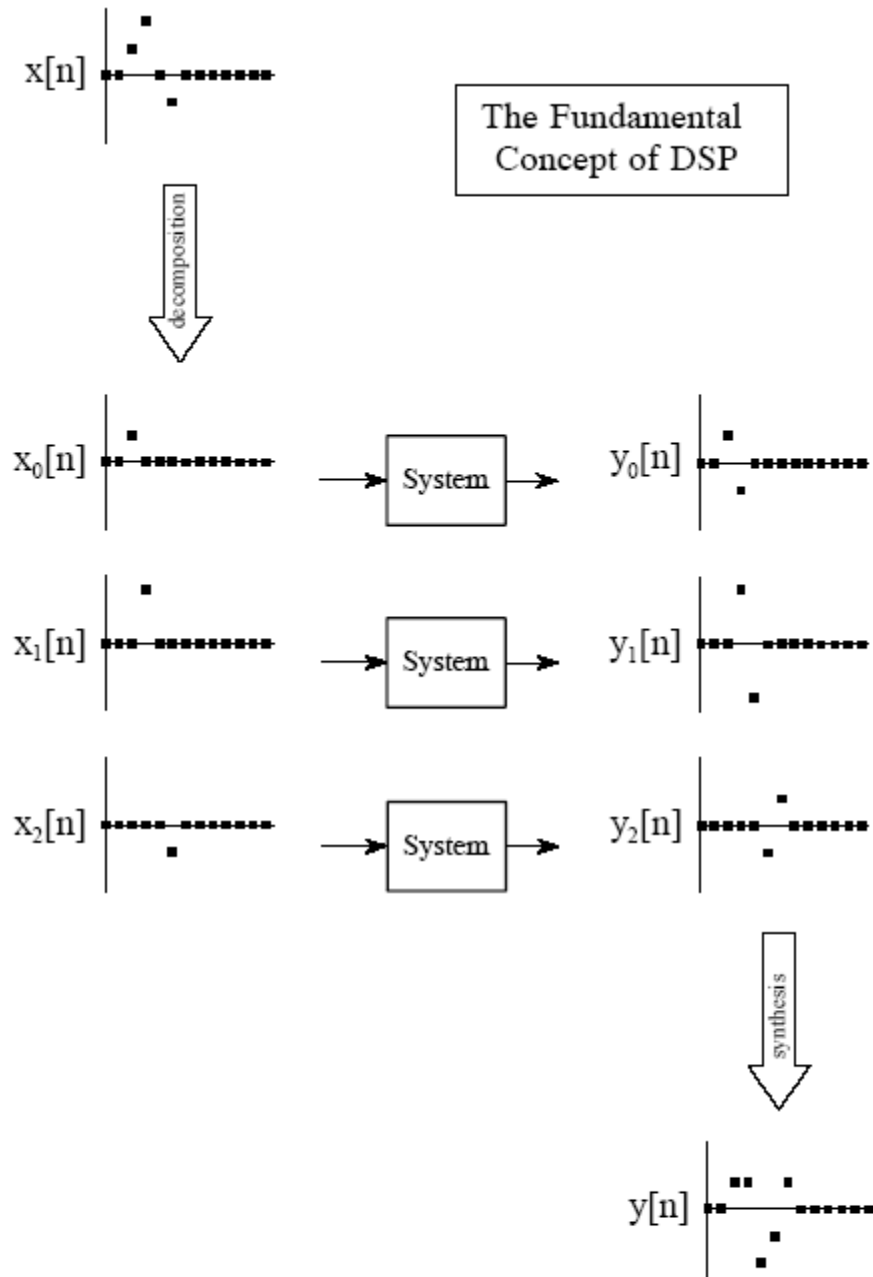


Fig. 2-6

Superpoziția, este strategia generală pentru înțelegerea modului cum pot fi analizate semnalele și sistemele. Considerăm un semnal de intrare notat cu $x[n]$, trecut printr-un sistem liniar, rezultând un semnal de ieșire, $y[n]$. Așa cum este prezentat în figura 2-6 semnalul de intrare poate fi descompus într-un grup de semnale mai simple (**componente ale semnalului de intrare**). Fiecare componentă a semnalului de intrare este trecută

individual prin sistem, rezultând un set de **componente ale semnalului de ieșire**: $y_0[n]$, $y_1[n]$, $y_2[n]$ etc. Aceste componente ale semnalului de ieșire sunt apoi sintetizate în semnalul de ieșire, $y[n]$.

Partea importantă constă în faptul că semnalul de ieșire obținut prin această metodă este *identic* cu cel produs prin trecerea directă a semnalului de intrare prin sistem. Acesta este un rezultat foarte important: în loc să încercăm să înțelegem cum sunt modificate de către un sistem semnalele *complicate*, tot ce trebuie să știm este cum sunt modificate semnalele *simple*.

2.4 Descompuneri uzuale

Scopul acestei metode este de a înlocui o problemă complicată cu mai multe probleme ușoare. Există mai multe moduri de descompunere a semnalelor: *descompunerea în impulsuri*, *descompunerea treaptă*, *descompunerea pară/impară*, *descompunerea intercalată* și *descompunerea Fourier*.

2.4.1 Descompunerea în Impulsuri

Cum este prezentată în figura 2-7, descompunerea în impulsuri desparte un semnal cu N eșantioane în N semnale componente, fiecare conținând N eșantioane. Fiecare dintre semnalele componente conține un punct din semnalul original, restul fiind zero. Un singur punct diferit de zero într-un șir de valori egale cu zero este numit *impuls*. Descompunerea în impulsuri este importantă deoarece ea permite semnalelor să fie examinat un eșantion la un moment dat. La fel, sistemele sunt caracterizate de răspunsul lor la impulsuri. Dacă cunoaștem răspunsul unui sistem la impuls, ieșirea sistemului poate fi calculată pentru orice intrare dată (*convoluție*).

2.4.2 Descompunerea Treaptă

Pentru descompunerea treaptă, prezentată în figura 2-8, se împarte de asemenea semnalul cu N eșantioane în N semnale componente, fiecare alcătuite din N eșantioane. Fiecare semnal component este o treaptă, asta înseamnă că primele eșantioane au valoarea zero în timp ce ultimele au o valoare constantă. Considerând descompunerea unui semnal de N puncte, $x[n]$, în componentele: $x_0[n]$, $x_1[n]$, $x_2[n]$, ..., $x_{N-1}[n]$ al i -lea semnal component $x_i[n]$, este compus din zerouri de la punctul 0 până la $i-1$, în timp ce restul punctelor au valoarea de $x[i] - x[i-1]$.

2.4.3 Descompunerea Pară/Impară

Este prezentată în Figura 2-9, și împarte un semnal în două semnale componente, unul având **simetrie pară** și celălalt **simetrie impară**. Un semnal de N puncte se spune că are simetrie pară dacă prezintă o imagine în oglindă în jurul punctului $N/2$. Asta înseamnă că eșantionul $x[N/2+1]$ trebuie să fie egal cu $x[N/2-1]$, eșantionul $x[N/2+2]$ trebuie să fie egal cu $x[N/2-2]$ etc. La fel, simetria impară apare atunci când punctele opuse au valori egale dar semne diferite: $x[N/2+1] = -x[N/2-1]$, $x[N/2+2] = -x[N/2-2]$ etc. Aceste

definiții obligă semnalul să fie alcătuit dintr-un număr par de valori iar indexul lor să fie de la 0 la $N - 1$. Descompunerea este calculată după relațiile:

$$x_E[n] = \frac{x[n] + x[N-n]}{2}$$

$$x_O[n] = \frac{x[n] - x[N-n]}{2}$$

Ecuția 2-1.

Motivul pentru care ultimul eșantion din semnal este văzut ca urmând după primul eșantion este o tehnică DSP numită *analiza Fourier*. Matematica analizei Fourier vede semnalul ca fiind circular, deși nu există nici o semnificație fizică pentru aceasta.

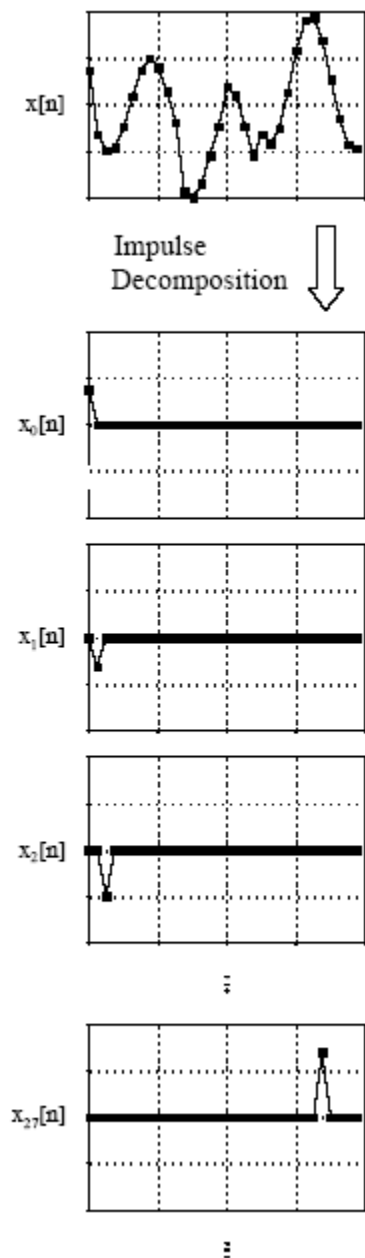


Fig. 2-7

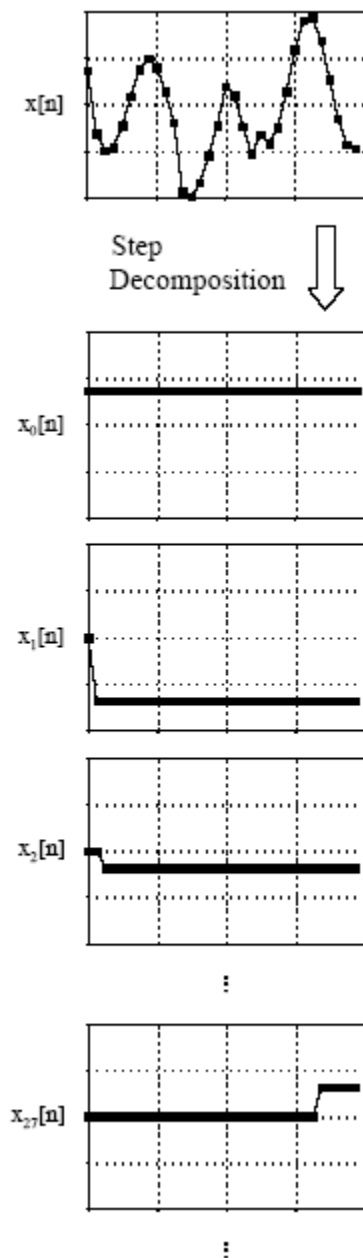


Fig. 2-8

2.4.4 Descompunerea Intercalată

Aceasta împarte semnalul în două semnale componente: *eșantioanele pare* și *eșantioanele impare* (Figura 2-10). Pentru a găsi semnalul cu eșantioanele pare, se începe cu semnalul original și se setează toate eșantioanele impare la zero. Pentru a găsi semnalul cu eșantioanele impare, se pornește de la semnalul original și se setează toate eșantioanele pare la zero.

Descompunerea intercalată este baza unui algoritm foarte important în DSP, Transformata Fourier Rapidă (FFT). Procedura pentru calcularea descompunerii Fourier se știa de acum câteva sute de ani. Din nefericire, este foarte învechită. Strategia discutată este un excelent exemplu al DSP: reduce semnalul la componente elementare prin folosirea repetată a transformatei de intercalare; calculează descompunerea Fourier a componentelor individuale; sintetizează rezultatul într-un răspuns final. Rezultatul este că viteza de calcul crește de *sute* sau chiar *mii* de ori.

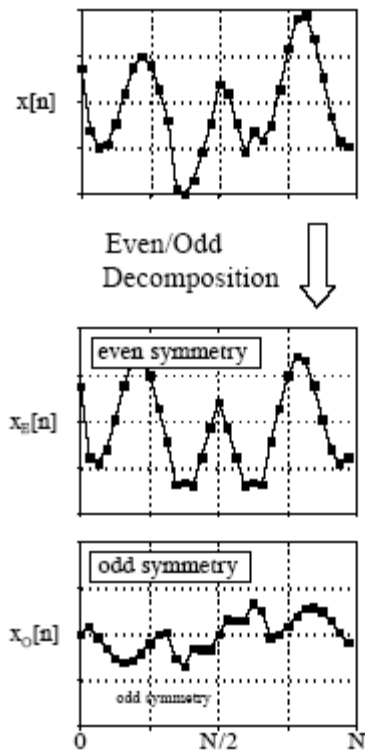


Fig. 2-9

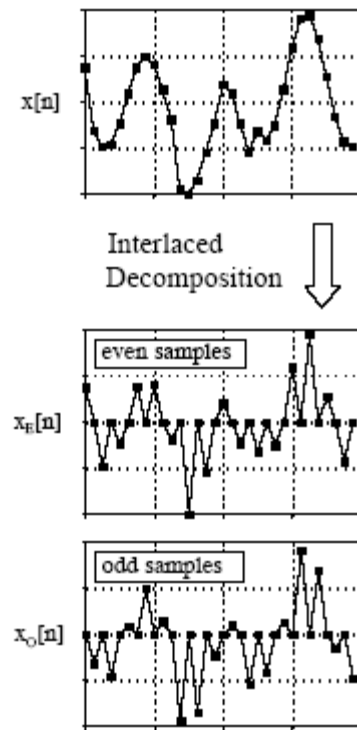


Fig. 2-10

2.4.5 Descompunerea Fourier

Figura 2-11 prezintă un exemplu din această tehnică. Orice semnal de N puncte poate fi descompus în $N+2$ semnale, jumătate sinusoide iar cealaltă jumătate cosinusoide. Cosinusoida cu frecvența cea mai mică ($x_{c0}[n]$), execută zero cicli compleți în N eșantioane, este un semnal DC. Următoarele cosinusoide: $x_{c1}[n]$, $x_{c2}[n]$ și $x_{c3}[n]$, execută 1, 2 sau 3 cicli compleți în N eșantioane. Acest lucru este valabil și pentru componentele sinusoidale. Din moment ce frecvența fiecărei componentă este fixă, singurul lucru care se modifică pentru diferitele semnale ce sunt descompuse, este *amplitudinea* fiecărei sinusoide sau cosinusoide.

Descompunerea Fourier este importantă din trei motive. Primul, o mare varietate de semnale sunt create prin suprapunerea mai multor sinusoide. Semnalul audio este un bun exemplu pentru aceasta. Descompunerea Fourier permite o analiză directă a informației conținută în aceste tipuri de semnale. Al doilea, sistemele liniare răspund la sinusoide

într-un mod unic: o intrare sinusoidală va rezulta întotdeauna într-o ieșire sinusoidală. Sistemele sunt caracterizate prin modul cum modifică amplitudinea și faza sinusoidelor ce trec prin ele. Din moment ce un semnal de intrare poate fi descompus în sinusoidale, știind cum reacționează sistemul la acestea, putem determina ieșirea sistemului. Al treilea, descompunerea Fourier reprezintă baza pentru un domeniu puternic și foarte răspândit al matematicii, numit *analiza Fourier*, pe lângă alte tipuri de transformate cum ar fi *Z* și *Laplace*. Mulți algoritmi DSP se bazează pe aceste tehnici.

2.5 Alternative la Liniaritate

Pentru a aprecia importanța sistemelor liniare, să considerăm că există o *singură* strategie majoră pentru analiza sistemelor neliniare. Strategia este de a face sistemul neliniar să *semine* cu un sistem liniar. Există trei moduri pentru a face asta:

Primul, ignorarea neliniarității. Dacă neliniaritatea este destul de mică, sistemul poate fi aproximat ca fiind liniar. Erorile rezultate sunt ignorate.

Al doilea, păstrarea unor semnale mici. Multe sisteme neliniare apar ca fiind liniare dacă semnalele au o amplitudine foarte mică. Spre exemplu, tranzistoarele sunt foarte neliniare, dar prezintă o amplificare liniară bună dacă semnalele sunt menținute sub câțiva milivolți. Amplificatoarele operaționale duc această idee la extrem.

Al treilea, aplicarea unei transformări de liniarizare. Spre exemplu, se consideră două semnale ce se înmulțesc formând un al treilea: $a[n] = b[n] \times c[n]$. Luând logaritmul semnalelor, schimbăm procesul neliniar al înmulțirii într-un proces liniar de adunare: $\log(a[n]) = \log(b[n]) + \log(c[n])$.

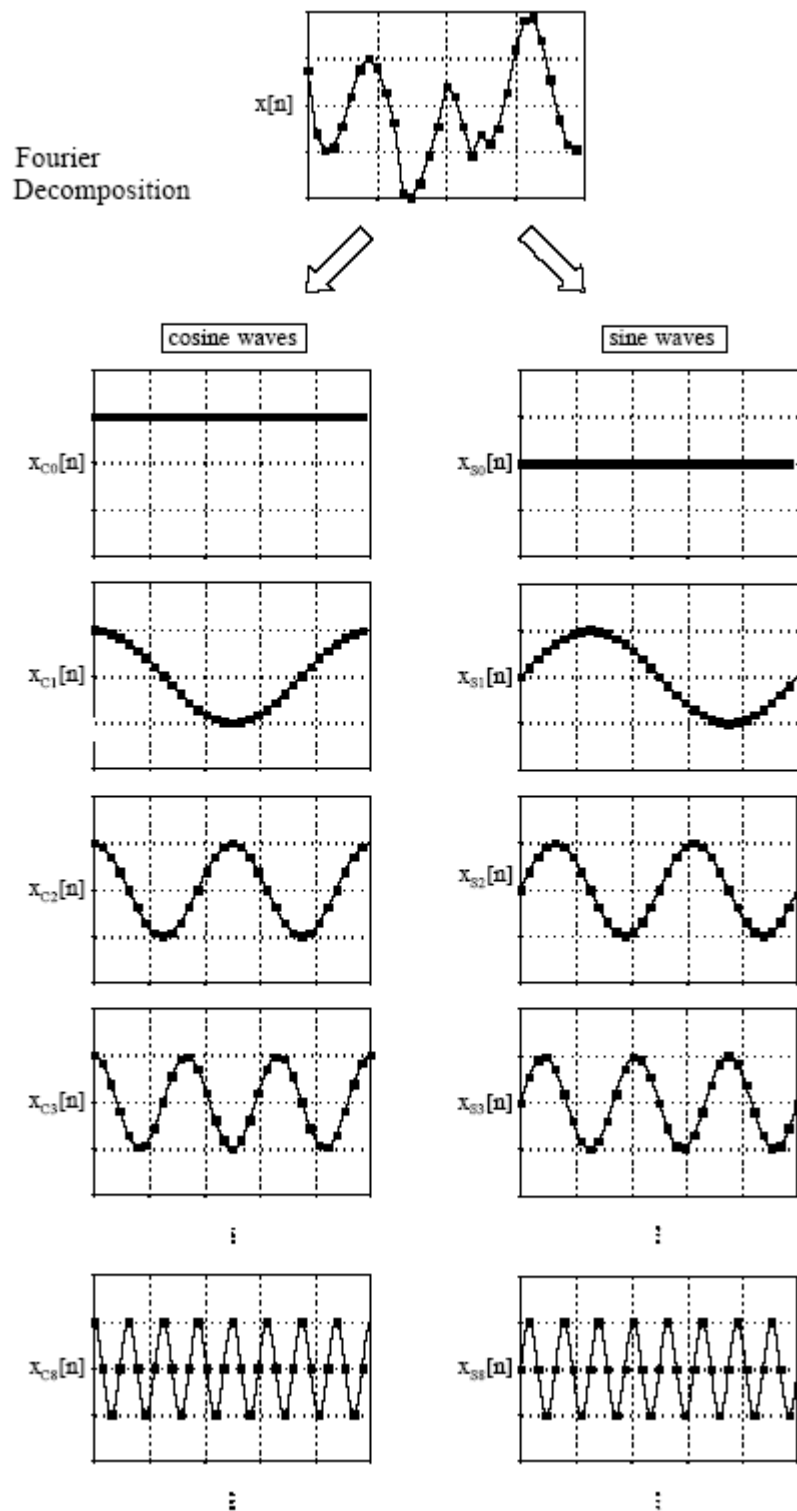


Fig. 2-11

CAP 3. Convoluția

Este o metodă matematică care combină două semnale pentru a forma un al treilea. Este cea mai importantă tehnică în procesarea de semnal. Folosind metoda descompunerii în impulsuri sistemele sunt descrise de un semnal numit răspunsul la impuls. Convoluția face legătura între cele trei semnale de interes: semnalul de intrare, semnalul de ieșire, răspunsul la impuls. Se va prezenta convoluția din două puncte de vedere: a) al vectorului de intrare și b) al vectorului de ieșire.

3.1 Funcția delta și răspunsul la impuls.

Figura 3-1 definește toți termenii importanți folosiți în convoluția DSP. Primul este funcția **delta** simbolizată de litera grecească $\delta[n]$. Funcția delta este un impuls normalizat, asta însemnând că eșantionul cu numărul zero are valoarea unu în timp ce restul eșantioanelor au valoarea zero. Din această cauză funcția delta este frecvent numită **impuls unitate**.

Al doilea termen este **răspunsul la impuls**. Așa cum sugerează numele, răspunsul la impuls este semnalul care iese din sistem când o funcție delta reprezintă intrarea. Dacă două sisteme diferă în vreun fel, ele vor avea un răspuns la impuls diferit. Așa cum semnalele de intrare și ieșire sunt adesea numite: $x[n]$ și $y[n]$, răspunsului la impuls i se dă de obicei numele de $h[n]$.

Orice impuls poate fi reprezentat ca o funcție delta deplasată și scalată. Considerăm un semnal $a[n]$ compus din zerouri, cu excepția eșantionului 8, care are valoarea -3 . În formă matematică: $a[n] = -3\delta[n-8]$.

Dacă intrarea într-un sistem este un impuls, cum ar fi $-3\delta[n-8]$, care este ieșirea sistemului? Aici se folosește proprietatea de omogenitate și invarianță la deplasare. Scalând și deplasând intrarea rezultă o ieșire identică, scalată și deplasată. Dacă $\delta[n]$ rezultă în $h[n]$, atunci $-3\delta[n-8]$ rezultă $-3h[n-8]$. Cu alte cuvinte ieșirea este o versiune a răspunsului la impuls care este *deplasată* și *scalată* cu aceeași cantitate ca și funcția delta de pe intrare. Dacă știm răspunsul la impuls a sistemului, vom ști imediat cum va reacționa acesta la *orice* impuls.

3.2 Convoluția

Este o operație matematică obișnuită, cum ar fi înmulțirea, adunarea sau integrarea. Adunarea primește două *numere* și generează un al treilea număr, în timp ce convoluția primește două *semnale* și generează un al treilea *semnal*. Este folosită în multe domenii ale matematicii cum ar fi probabilitatea sau statistică.

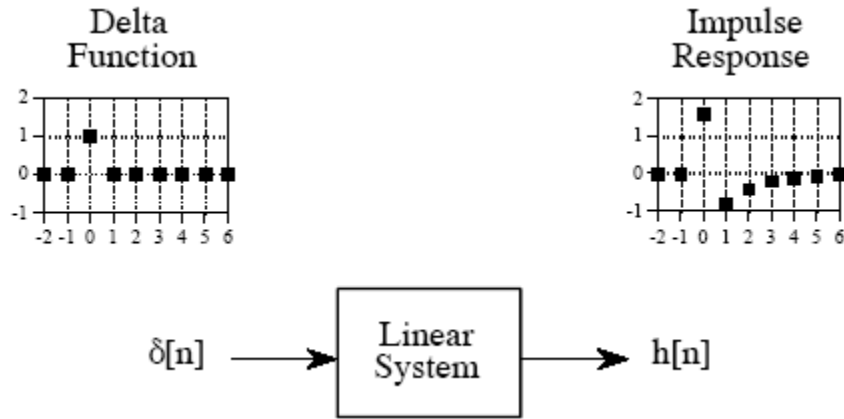


Fig. 3-1

Figura 3-2 arată modul de reprezentare a convoluției în sistemele liniare. Un semnal de intrare $x[n]$, intră într-un sistem liniar cu un răspuns la impuls, $h[n]$ rezultând un semnal de ieșire, $y[n]$ ($x[n] * h[n] = y[n]$).

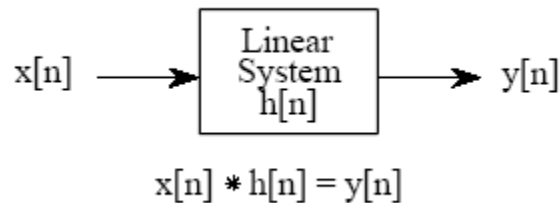
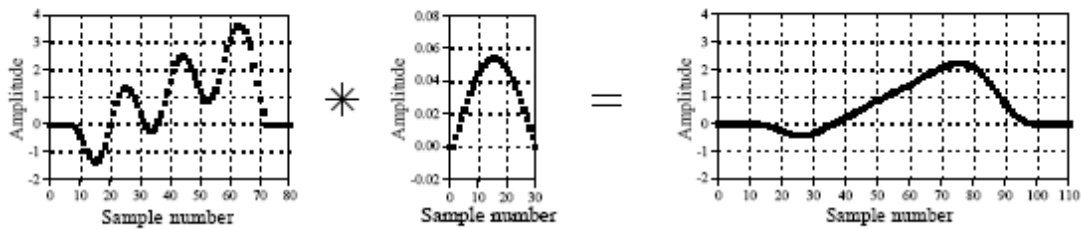


Fig. 3-2

Figura 3-3 prezintă convoluția folosită la o filtrare trece-jos și o filtrare trece-sus. Semnalul de intrare este suma a două componente: trei cicli dintr-o sinusoidă reprezentând frecvențe înaltă și un semnal pantă reprezentând frecvența joasă. În (a), răspunsul la impuls pentru filtrul trece-jos este o pantă nu foarte abruptă, rezultând căre ieșire doar formele de undă a căror variație nu este foarte bruscă. La fel, filtrul trece-sus (b), permite trecerea numai a formelor de undă a căror variație este rapidă.

a. Low-pass Filter



b. High-pass Filter

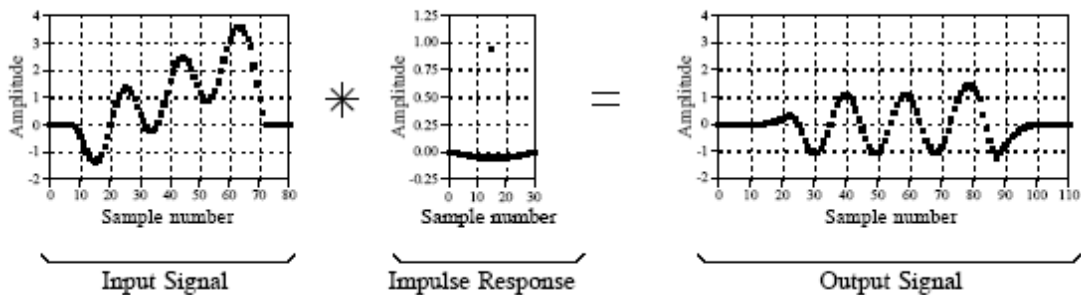


Fig. 3-3

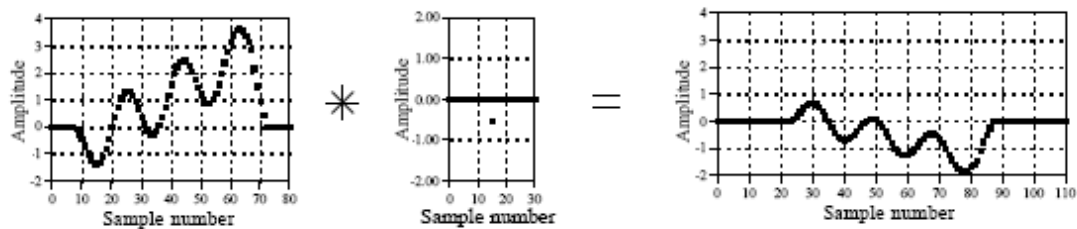
Figura 3-4 prezintă alte două exemple de folosire a convoluției. Atenuatorul cu inversare (a), inversează semnalul de sus în jos și-i reduce amplitudinea. Derivata discretă este prezentată în (b).

Semnalul de intrare conține 81 de eșantioane, în timp ce fiecare răspuns la impuls este alcătuit din 31 de eșantioane. În majoritatea aplicațiilor DSP semnalul de intrare prezintă sute, mii sau chiar milioane de eșantioane lungime. Răspunsul la impuls este de obicei mult mai scurt, de la câteva puncte, până la câteva sute de puncte. Nu există restricții asupra lungimii acestor semnale. Ele specifică lungimea semnalului de ieșire care este egală cu lungimea semnalului de intrare, plus lungimea răspunsului la impuls, minus unu. Pentru semnalele din figurile 3-3 și 3-4, fiecare semnal de ieșire este: $81 + 31 - 1 = 111$ eșantioane.

Detalierea matematică a convoluției poate fi înțeleasă din două puncte de vedere. Primul, vede convoluția din **punctul de vedere al semnalului de intrare**. Acesta implică analiza modului cum fiecare eșantion din semnalul de intrare *contribuie* la formarea mai multor puncte din semnalul de ieșire. Al doilea, tratează convoluția din **punctul de vedere la semnalului de ieșire** și redă modul cum fiecare eșantion din semnalul de ieșire *primește* informație de la mai multe puncte ale semnalului de intrare.

Acestea reprezintă două moduri de a vedea aceeași operație matematică. Primul este important deoarece se leagă de implementarea convoluției în DSP. Al doilea, descrie matematica convoluției. Unul din cele mai importante lucruri în DSP este de a face înțeleasă *conceptual* teoria matematică și implementată în sistemele de calcul.

a. Inverting Attenuator



b. Discrete Derivative

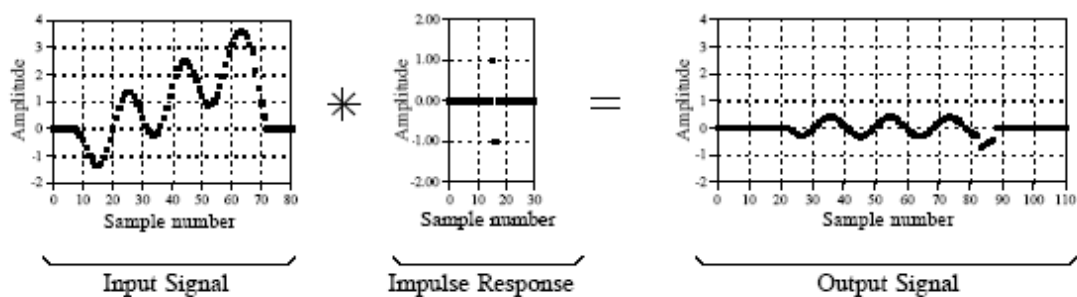


Fig. 3-4

3.3 Convoluția văzută din punct de vedere al vectorului de intrare

Figura 3-5 prezintă o problemă simplă de convoluție: un semnal de intrare de 9 puncte, $x[n]$ este trecut printr-un sistem având un răspuns la impuls de 4 puncte, $h[n]$ rezultând un semnal de ieșire $y[n]$ de $9+4-1=12$ puncte. În acest exemplu, fiecare din cele nouă eșantioane a semnalului de intrare va genera o componentă scalată și deplasată a răspunsului la impuls pentru semnalul de ieșire (fig. 3-6). Prin adunarea acestor semnale se generează semnalul de ieșire, $y[n]$.

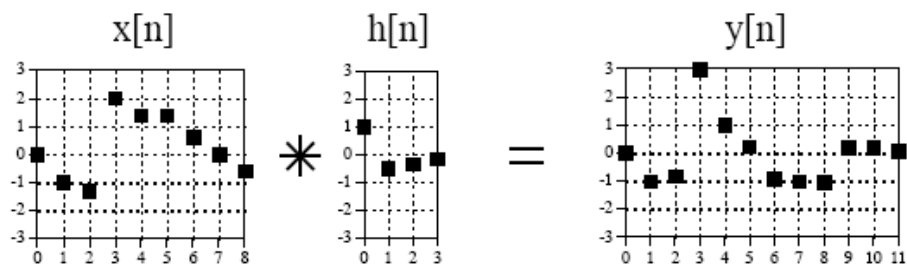


Fig. 3-5

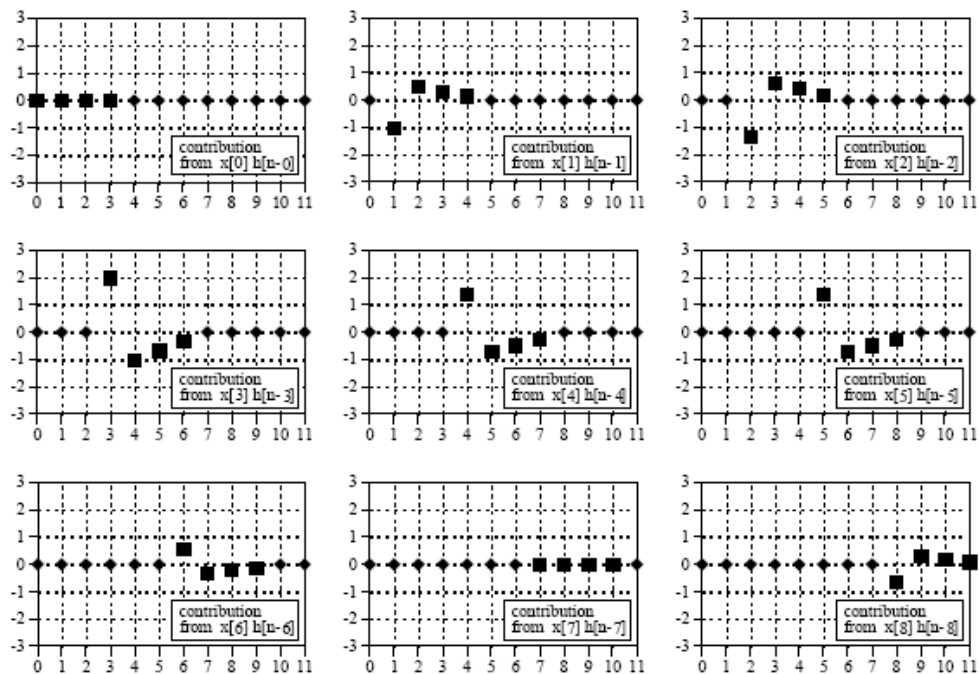


Fig. 3-6

În acest exemplu, $x[n]$ este un semnal de nouă puncte și $h[n]$ este de patru puncte. În următorul exemplu, prezentat în figura 3-7, vom schimba situația făcând $x[n]$ de patru puncte și $h[n]$ de nouă puncte. Sunt folosite aceleași forme de undă, acestea fiind doar schimbate între ele.

După cum se observă, semnalul rezultat este identic cu semnalul de ieșire din figura 3-5.

De aici rezultă o proprietate importantă. Convoluția este *comutativă*:

$$a[n] * b[n] = b[n] * a[n].$$

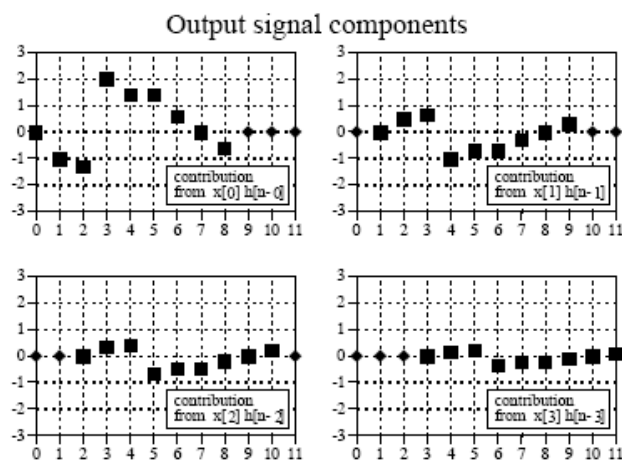
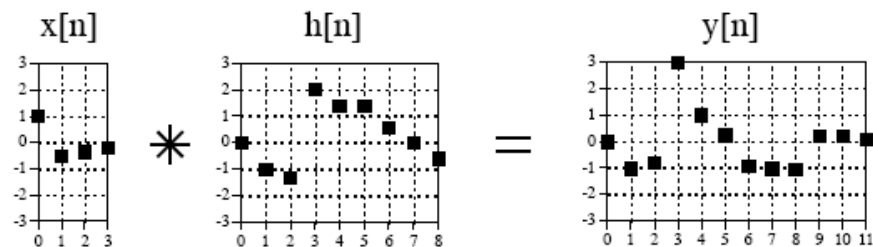


Fig. 3-7

3.4 Convoluția văzută din punct de vedere al vectorului de ieșire

Să considerăm exemplul în care un singur punct al semnalului de ieșire este influențat de mai multe puncte din semnalul de intrare. Vom lua ca exemplu punctul $y[6]$ în figura 3-5. Acest punct este egal cu suma a 6 puncte din cele nouă componente prezentate în figura 3-6. Să verificăm care din cele 9 componente influențează $y[6]$. Asta înseamnă să căutăm care din aceste 9 semnale conțin o valoare diferită de zero pe poziția a șasea: $x[3]$, $x[4]$, $x[5]$, și $x[6]$. Prin adunarea eșantionului șase din fiecare dintre aceste componente, rezultă: $y[6] = x[3]h[3] + x[4]h[2] + x[5]h[1] + x[6]h[0]$. Figura 3-8 prezintă acest algoritm ca o mașină de convoluție, o schemă a procesului de convoluție. Valoarea rezultată este plasată în locația corespunzătoare din vectorul de ieșire.

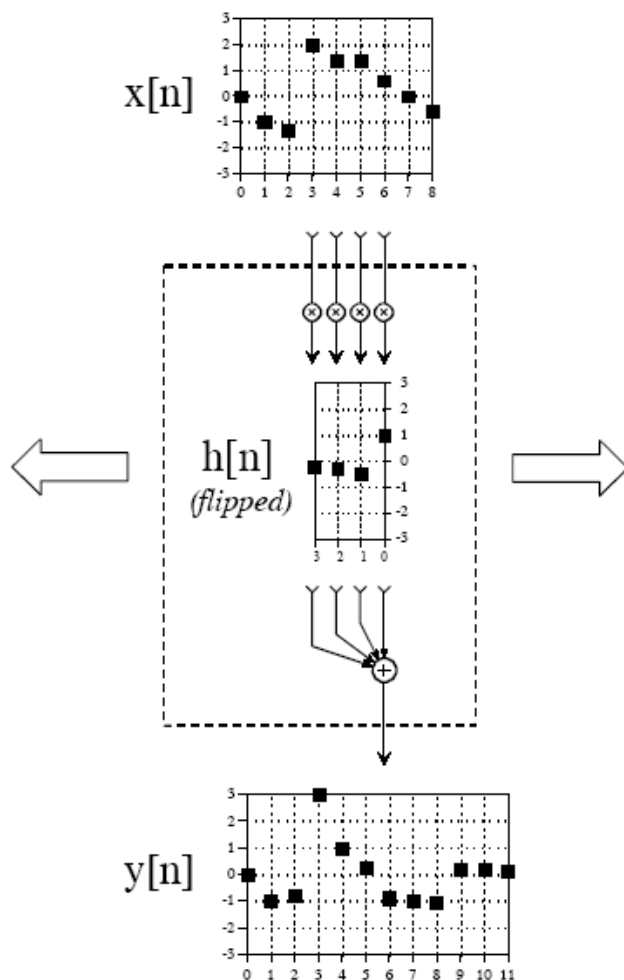
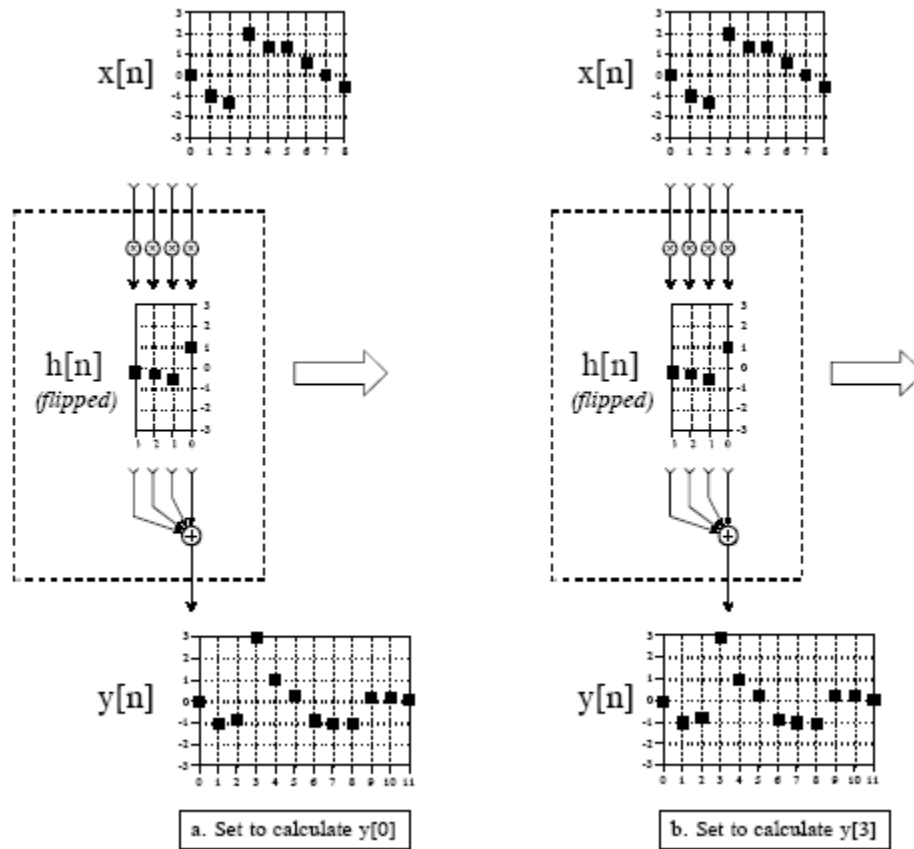


Fig. 3-8

Aranjarea vectorului răspuns la impuls în *interiorul* mașinii de convoluție este foarte importantă. Răspunsul la impuls este *inversat de la dreapta la stânga*. Această rotire rezultă din calculul matematic.

Figura 3-9 prezintă mașina de convoluție utilizată pentru a calcula mai multe eșantioane în semnalul de ieșire. Această diagramă mai prezintă și realele probleme care apar în convoluție. În (a), mașina de convoluție este plasată în extremitatea stângă cu ieșirea pe $y[0]$. În această poziție, ea încearcă să primească informații de la punctele: $x[-3]$, $x[-2]$, $x[-1]$ și $x[0]$. Problema este că eșantioanele $x[-3]$, $x[-2]$ și $x[-1]$ nu există. La fel și în (d), în care mașina de convoluție încearcă să citească valorile $x[9]$, $x[10]$ și $x[11]$.

Un mod de a rezolva această problemă este de a adăuga eșantioane la capetele semnalului de intrare, având valoarea zero. Această “umplere” nu va afecta rezultatul final.



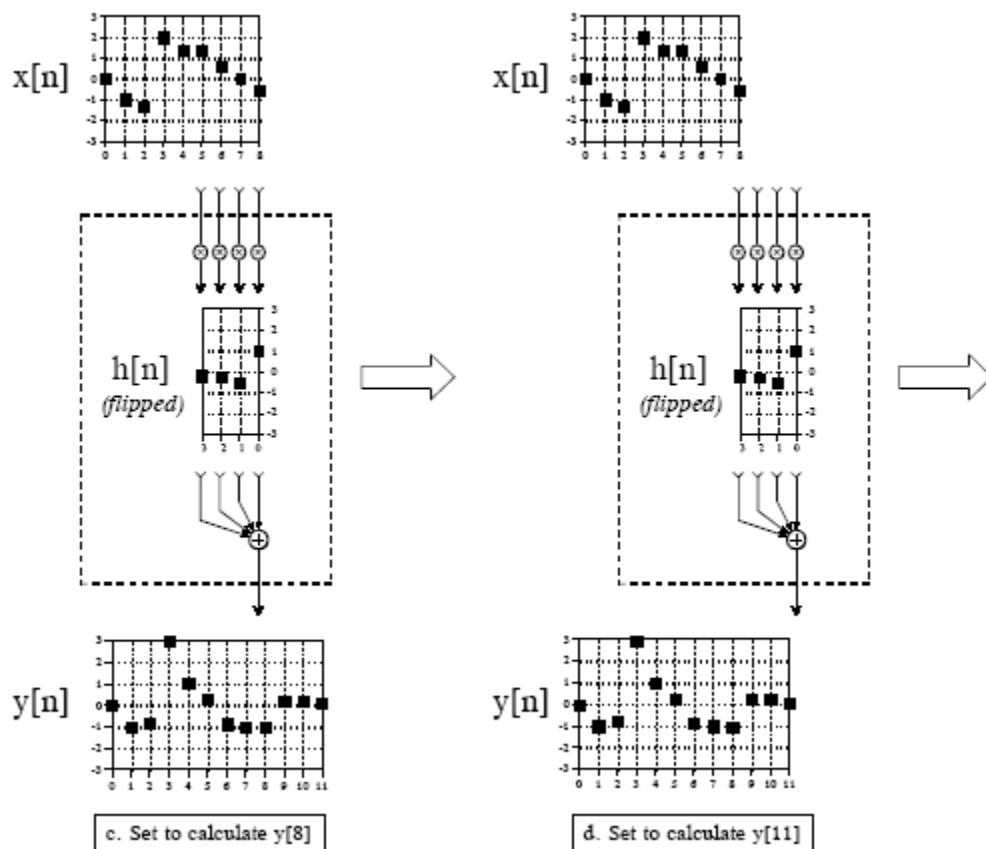


Fig. 3-9

Partea importantă a acestei probleme este că extremitățile semnalului de ieșire sunt construite dintr-o informație incompletă. Acest lucru este asemănător cu un circuit electronic care necesită un anumit timp după pornire, pentru a-și stabili parametri. Diferența este că această tranziție este ușor de ignorat în electronică, dar foarte dăunătoare în DSP.

Figura 3-10 prezintă un exemplu în care apar aceste efecte de capăt. Semnalul de intrare este o sinusoidă plus o componentă DC. Dorim să eliminăm din acest semnal componenta DC, lăsând sinusoida intactă. Acesta necesită un filtru trece-sus, având răspunsul la impuls prezentat în figură. Problema este că primele și ultimele 30 de puncte sunt "false". Forma acestor capete, poate fi înțeleasă imaginându-ne semnalul de intrare având 30 de puncte egale cu zero în partea stângă, de la $x[-1]$ până la $x[-30]$ și 30 de zerouri în partea dreaptă, de la $x[81]$ până la $x[110]$. Ieșirea reprezintă semnalul filtrat al acestei forme de undă mai lungi. Aceste probleme ale "efectelor de capăt" sunt foarte răspândite în DSP. Ca regulă generală, începutul și sfârșitul semnalelor digitale în procesarea de semnal, vor fi neutilizabile.

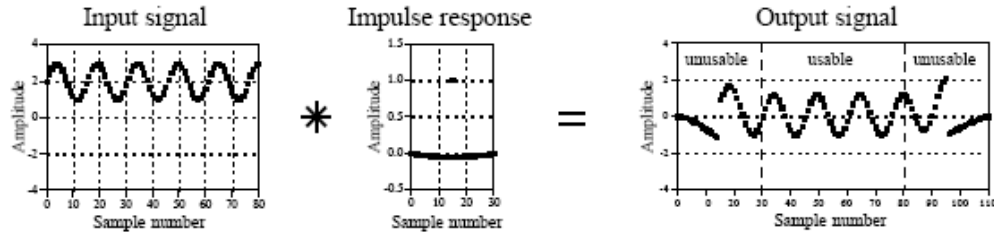


Fig. 3-10

Folosind mașina de convoluție, putem scrie ecuația standard a convoluției. Dacă $x[n]$ este un semnal de N puncte de la 0 la $N-1$ și $h[n]$ are M puncte, de la 0 la $M-1$, convoluția celor două: $y[n] = x[n] * h[n]$, este un semnal de lungime $N+M-1$ de la 0 la $N+M-2$, dat de:

$$y[i] = \sum_{j=0}^{M-1} h[j]x[i-j]$$

Ecuația 3-1

Această ecuație este numită **sumă de convoluție**. Formula permite fiecărui punct din semnalul de ieșire să fie calculat independent de celelalte puncte ale acestuia. Indexul i , determină care eșantion din semnalul de ieșire este calculat. Majoritatea aplicațiilor DSP se bazează pe această ecuație.

CAP 4. Proprietățile convoluției

Se prezintă proprietățile și utilizarea convoluției în diferite domenii. În prima parte sunt prezentate câteva răspunsuri la impuls uzuale, apoi modurile de conectare în serie sau paralel a sistemelor liniare. În final se face o introducere a tehnicii de corelație.

4.1 Răspunsuri la impuls uzuale

Cel mai simplu răspuns la impuls este funcția delta, prezentată în figura 4-1a. Un impuls la intrare generează un impuls identic la ieșire, toate semnalele trec prin sistem fără a fi modificate. Convoluționând orice semnal cu o funcție delta rezultatul va fi același semnal. Ele sunt ideale pentru memorarea datelor, comunicații și măsurători.

$$x[n] * \delta[n] = x[n]$$

Ecuația 4-1

Figura 4-1b prezintă o funcție delta răspuns la impuls modificată. Dacă funcției delta i se mărește sau i se micșorează amplitudinea, sistemul rezultat este un amplificator sau un

atenuator. Dacă k este mai mare ca 1, atunci se poate vorbi de amplificare. Dacă k este mai mic ca 1, atunci se obține o atenuare a semnalului.

$$x[n] * k\delta[n] = kx[n] \quad \text{Ecuația 4-2}$$

În figura 4-1c este prezentată o funcție delta cu o deplasare. Aceasta va avea ca efect o deplasare identică între semnalul de ieșire și cel de intrare. Poate fi înțeleasă ca o întârziere sau o avansare a semnalului, depinzând de direcția deplasării.

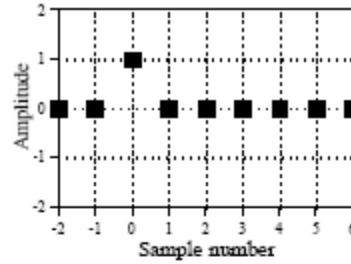
$$x[n] * \delta[n+s] = x[n+s] \quad \text{Ecuația 4-3}$$

Există o mulțime de exemple în care un semnal este o variantă deplasată a altui semnal: un semnal radio transmis și recepționat către și de la un obiect în spațiu. Timpul scurs între emisia și recepția acestui semnal cauzează apariția unei întârzieri între semnalul transmis și cel recepționat.

Figura 4-1d prezintă un răspuns la impuls compus dintr-o funcție delta plus o funcție delta deplasată și scalată. Prin superpoziție, ieșirea acestui sistem este semnalul de intrare plus o variantă atenuantă și întârziată a aceluiași semnal de intrare, un ecou. Ecourile sunt importante în multe aplicații DSP. Suprapunerea ecourilor face ca înregistrările sunetelor din natură să fie mai plăcute și mai naturale. Radarele și sonarele analizează ecourile pentru a detecta avioanele sau submarinele. Geofizicienii folosesc ecourile pentru a descoperi zăcămintele de petrol. Ecourile sunt foarte importante în rețeaua telefonică, deoarece dorim să le evităm.

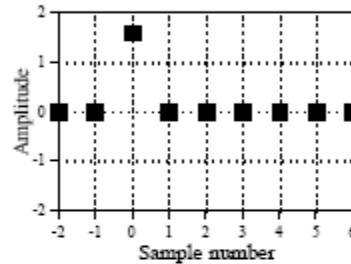
a. Identity

The delta function is the *identity* for convolution. Convoluting a signal with the delta function leaves the signal unchanged. This is the goal of systems that transmit or store signals.



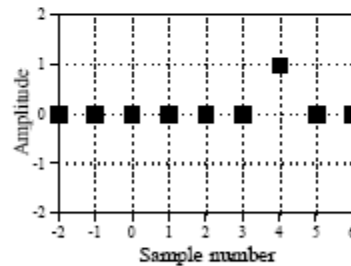
b. Amplification & Attenuation

Increasing or decreasing the amplitude of the delta function forms an impulse response that *amplifies* or *attenuates*, respectively. This impulse response will amplify the signal by 1.6.



c. Shift

Shifting the delta function produces a corresponding shift between the input and output signals. Depending on the direction, this can be called a *delay* or an *advance*. This impulse response delays the signal by four samples.



d. Echo

A delta function plus a shifted and scaled delta function results in an *echo* being added to the original signal. In this example, the echo is delayed by four samples and has an amplitude of 60% of the original signal.

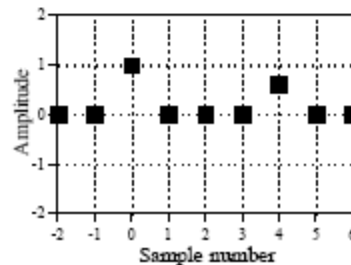


Fig. 4-1

4.2 Operații asupra semnalelor.

Convoluția poate acționa ca integrator sau derivator de semnale. Pentru semnalele analogice se folosesc termenii de “derivată” și “integrală”. La semnalele discrete apare denumirea de “derivată discretă” sau “integrală discretă”.

Figura 4-2 prezintă răspunsul la impuls care implementează derivata discretă și integrala discretă. Figura 4-3 prezintă un exemplu în care se utilizează aceste operații. În figura 4-3a, semnalul original este compus din mai multe componente cu pante diferite. Convoluționând acest semnal cu răspunsul la impuls derivata discretă se generează

semnalul din figura 4-3b. Amplitudinea fiecărui punct din semnalul derivata discretă este egală cu panta corespunzătoare locației din semnalul original. Integrarea discretă este operația inversă derivatei discrete. Asta înseamnă că, convoluționând semnalul din (b) cu răspunsul la impuls la a integralei discrete se obține semnalul (a).

De obicei nu este necesară implementarea unui program de convoluție pentru obținerea acestor funcții. Spre exemplu derivata discretă se poate calcula ca fiind:

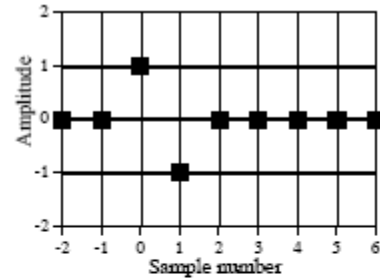
$$y[n] = x[n] - x[n-1] \quad \text{Ecuația 4-4a}$$

Fiecare eșantion din semnalul de ieșire este egal cu diferența dintre două puncte adiacente ale semnalului de intrare. O altă metodă uzuală este aceea de a calcula panta în jurul punctului examinat:

$$y[n] = (x[n+1] - x[n-1]) / 2 \quad \text{Ecuația 4-4b}$$

a. First Difference

This is the discrete version of the *first derivative*. Each sample in the output signal is equal to the *difference* between adjacent samples in the input signal. In other words, the output signal is the *slope* of the input signal.



b. Running Sum

The running sum is the discrete version of the *integral*. Each sample in the output signal is equal to the sum of all samples in the input signal to the *left*. Note that the impulse response extends to infinity, a rather nasty feature.

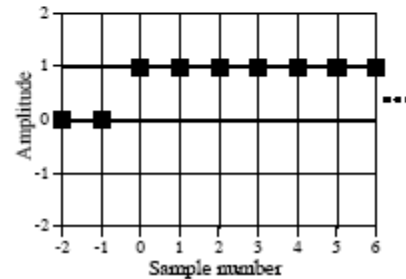


Fig. 4-2

Folosind aceeași aproximare, fiecare eșantion din integrala discretă poate fi calculat prin adunare tuturor punctelor din semnalul original aflate la stânga locației eșantionului curent:

$$y[n] = x[n] + y[n-1] \quad \text{Ecuația 4-5}$$

Relațiile de acest tip se numesc ecuații recursive sau ecuații cu diferențe finite. Ideea este de a înțelege că aceste relații sunt identice cu folosirea convoluției și răspunsului la impuls din figura 4-2.

Example of calculus-like operations. The signal in (b) is the *first difference* of the signal in (a). Correspondingly, the signal in (a) is the *running sum* of the signal in (b). These processing methods are used with discrete signals the same as differentiation and integration are used with continuous signals.

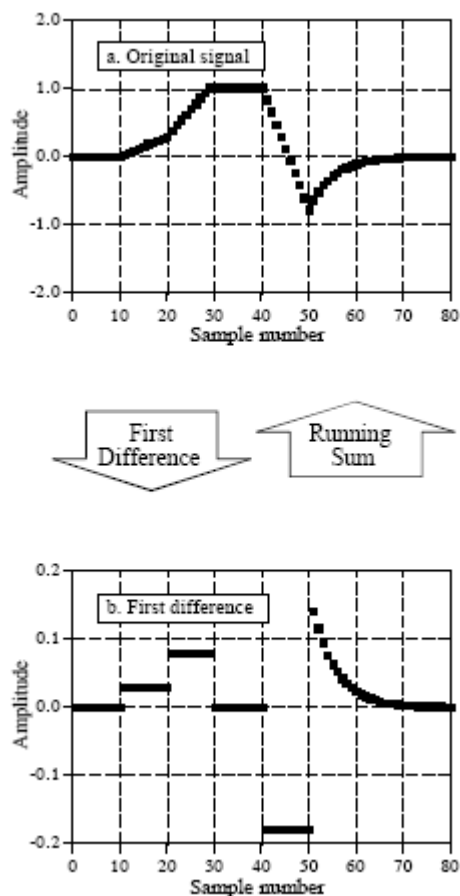


Fig. 4-3

4.3 Filtre trece-jos și trece-sus.

Pentru a înțelege forma generală a unui filtru trece-jos sau a unui filtru trece-sus să privim figurile 4-4 sau 4-5. În general, filtrele trece-jos sunt compuse din grupuri de puncte pozitive adiacente. Acesta are ca rezultat în semnalul de ieșire o medie ponderată a mai multor puncte adiacente a semnalului de intrare. Această medie “netezește” semnalul eliminând componentele cu frecvență înaltă. Așa cum este prezentată funcția sinc (c), câteva nuclee de filtre trece-jos includ în cozi și valori negative ale eșantioanelor.

Frecvența de tăiere a filtrelor este modificată prin construirea filtrelor mai largi sau mai înguste. Dacă un filtru trece-jos prezintă o amplificare de unu la frecvența zero (DC), atunci suma tuturor punctelor din răspunsul la impuls trebuie să fie egală cu unu. Așa cum este prezentat în figurile (a) (b) și (c), teoretic, filtrele pot fi extinse la infinit fără a atinge punctul zero. În practică, ele sunt trunchiate după un număr de eșantioane, pentru a putea fi memorate într-un vector.

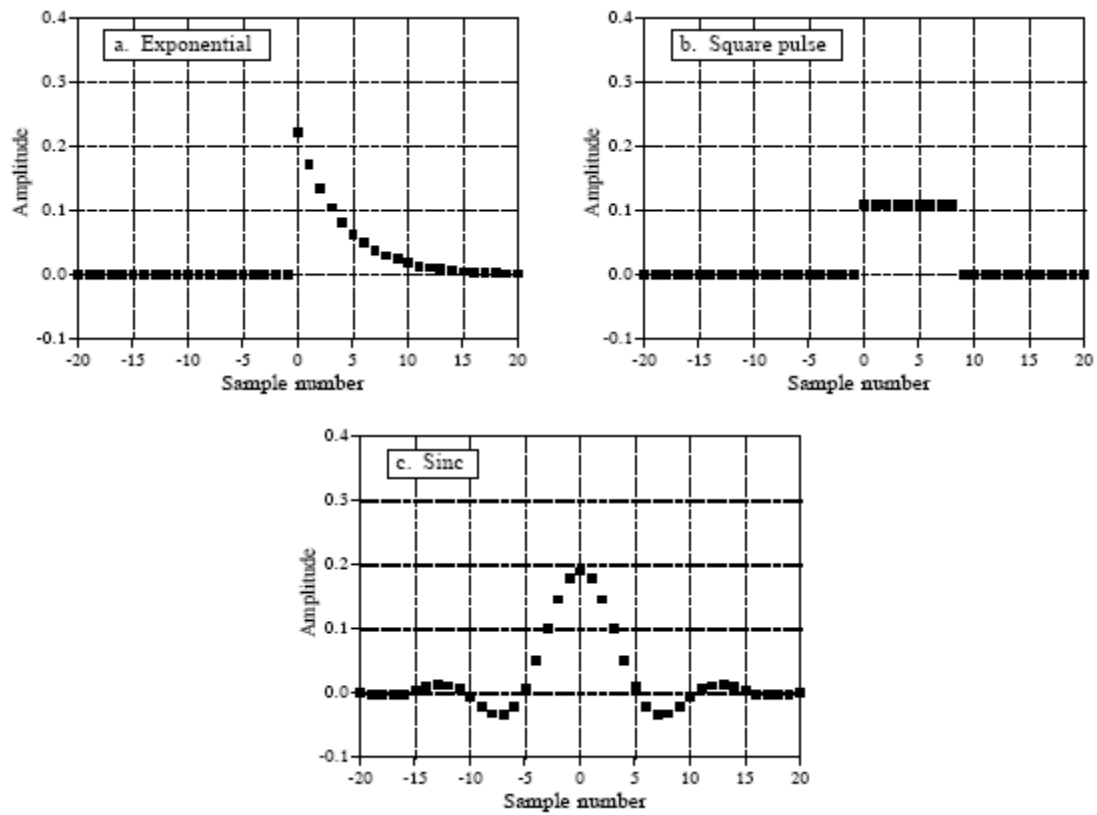


Fig. 4-4

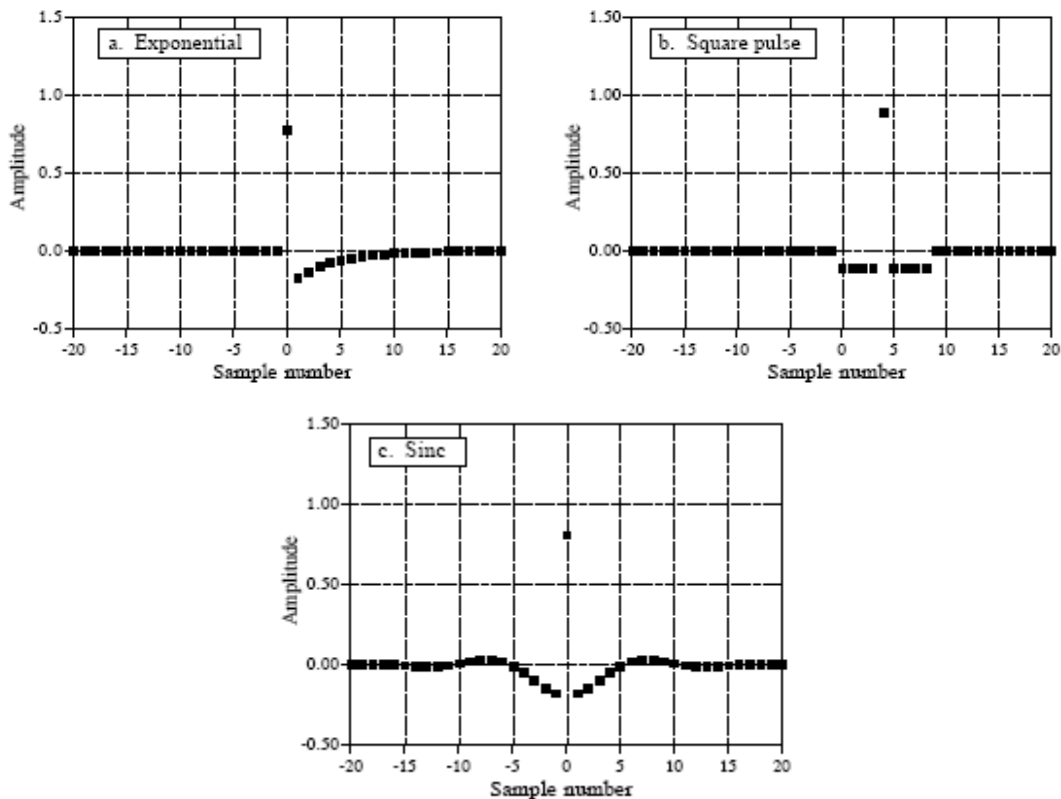


Fig. 4-5

Figura 4-5 prezintă trei nuclee de filtre trece-sus, derivate din nuclee corespunzătoare ale filtrelor trece-jos din figura 4-4. Aceasta este o strategie uzuală în proiectarea filtrelor: la început se concepe un filtru trece-jos și apoi se transformă acesta într-un filtru de care avem nevoie: trece-sus, trece-bandă, stop-bandă etc. Pentru a înțelege trecerea de la un filtru trece-jos, la un filtru trece-sus, să ne amintim că răspunsul la impuls a unei funcții delta permite trecerea întregului semnal, în timp ce răspunsul la impuls al unui filtru trece-jos permite trecerea numai a componentelor cu frecvență joasă. Prin superpoziție, nucleul unui filtru format dintr-o funcție delta minus nucleul unui filtru trece-jos, va permite trecerea întregului semnal mai puțin a componentelor cu frecvență joasă. Astfel se naște un filtru trece-sus. Așa cum se observă în figura 4-5 funcția delta este situată în centrul de simetrie sau în eșantionul zero dacă nucleul de filtru nu este simetric. Filtrul trece-sus prezintă amplificare zero la frecvență zero (DC), dacă suma tuturor punctelor din filtru este egală cu zero.

4.4 Semnale cauzale și noncauzale.

Să ne imaginăm un circuit analogic simplu. Dacă se aplică un impuls scurt la intrare, vom vedea un răspuns la ieșire. Acesta este un fel de cauză și efect pe care se bazează universul nostru. Orice efect trebuie să se întâmple după o cauză. Aceasta este caracteristica de bază a ceea ce numim noi timp. Acum să presupunem un sistem DSP care modifică semnalul de intrare într-un semnal de ieșire, ambele memorate în doi vectori într-un computer. Dacă acesta simulează un sistem din lumea reală, el trebuie să

urmeze același principiu al cauzalității. Spre exemplu, valoarea eșantionului opt al semnalului de intrare, poate afecta eșantionul opt sau mai mare în semnalul de ieșire. Sistemele care operează în acest mod sunt numite sisteme cauzale. Desigur, procesarea digitală nu necesită în mod obligatoriu o funcționare de acest fel. Din moment ce semnalele de intrare și cele de ieșire sunt vectori de numere, orice valoare a semnalului de intrare poate afecta orice valoare a semnalului de ieșire (figura 4-6).

4.5 Fază zero, fază liniară și fază neliniară.

Așa cum este prezentat în figura 4-7, un semnal are faza zero dacă acesta prezintă simetrie stânga - dreapta, în jurul eșantionului zero. Un semnal prezintă fază liniară dacă acesta are o simetrie stânga - dreapta, în jurul altui punct decât punctul zero. Asta înseamnă că orice semnal cu fază liniară poate fi modificat într-un semnal cu fază zero printr-o simplă deplasare la stânga sau la dreapta. În final, un semnal are o fază neliniară dacă acesta nu prezintă simetrie stânga - dreapta. A nu se face confuzie între liniaritatea sistemelor și liniaritatea fazei.

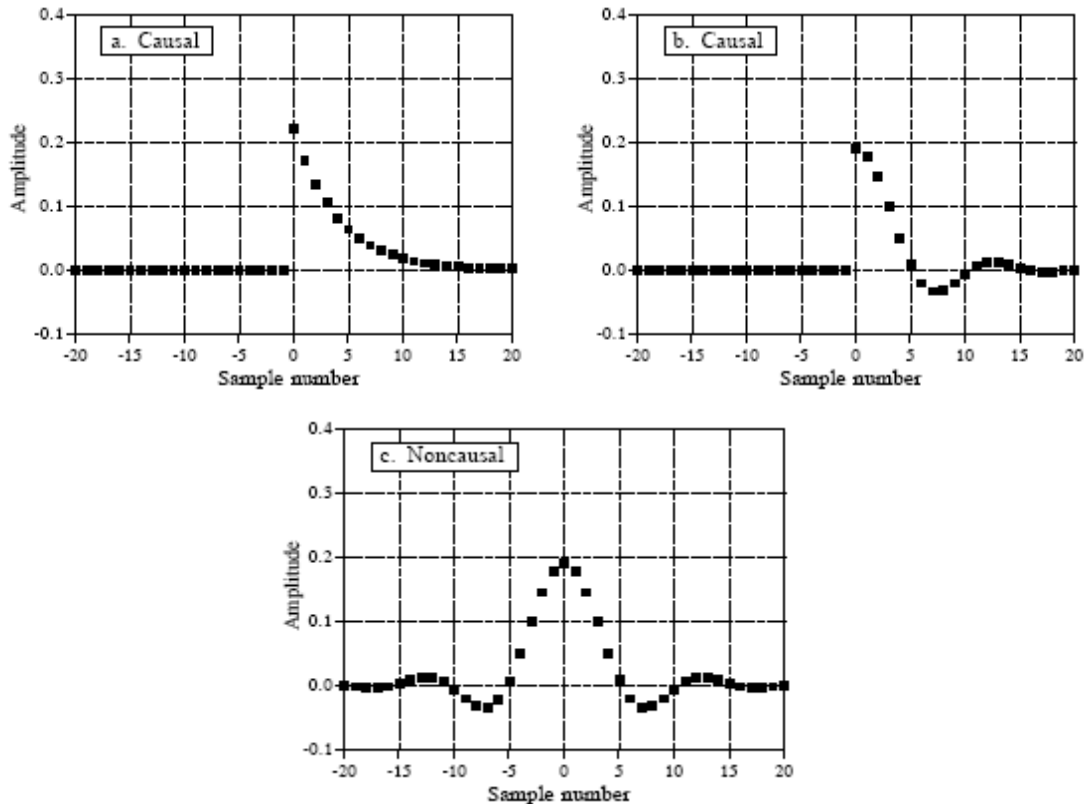


Fig. 4-6

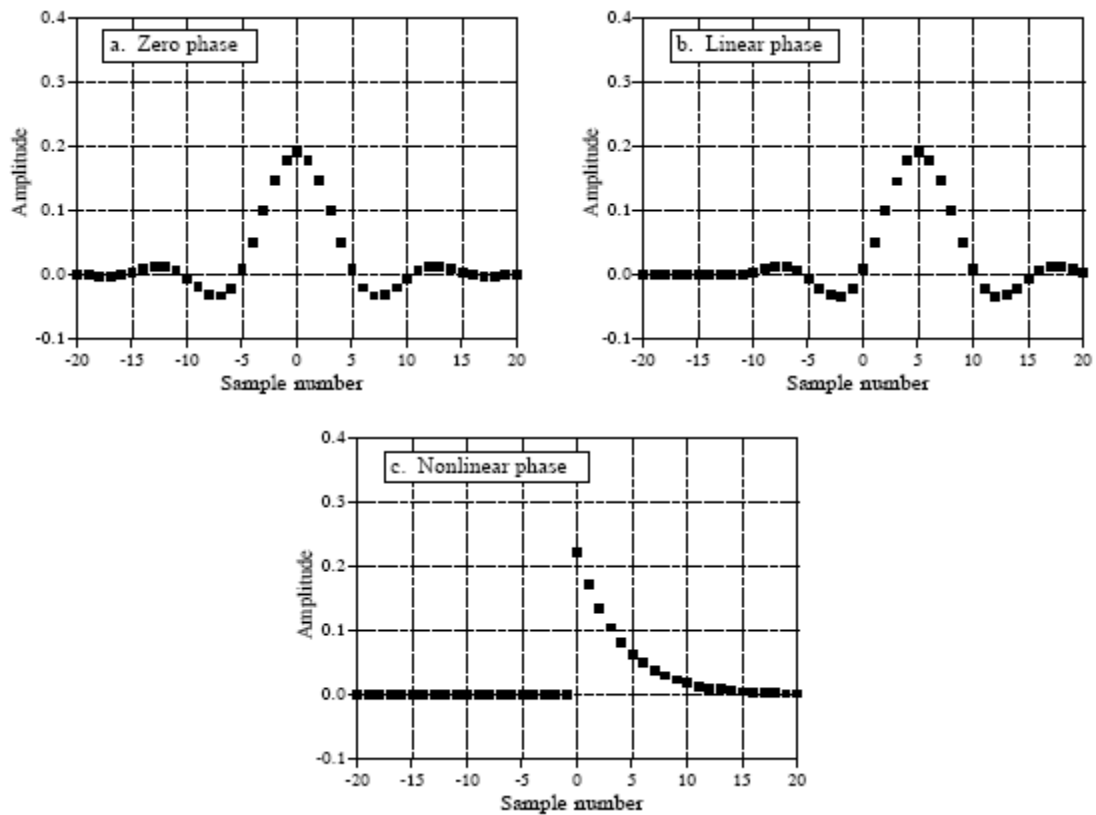


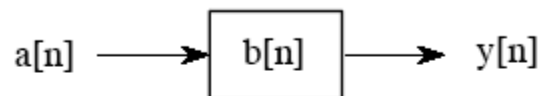
Fig. 4-7

4.6 Proprietăți matematice.

Proprietatea de comutativitate (figura 4-8):

$$a[n] * b[n] = b[n] * a[n] \quad \text{Ecuația 4-6}$$

IF



THEN

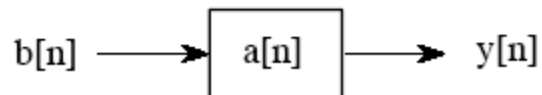
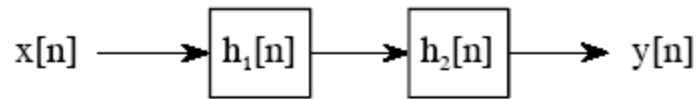


Fig. 4-8

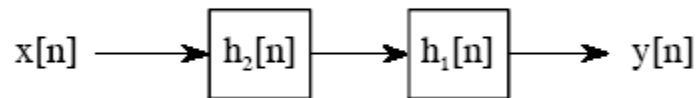
Proprietatea de asociativitate (figura 4-9):

$$(a[n] * b[n]) * c[n] = a[n] * (b[n] * c[n]) \quad \text{Ecuația 4-7}$$

IF



THEN



ALSO

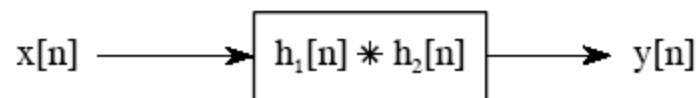
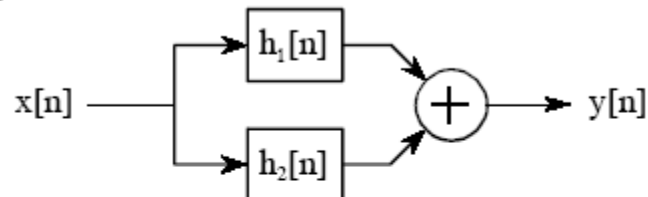


Fig. 4-9

Proprietatea de distributivitate (figura 4-10):

$$a[n] * b[n] + a[n] * c[n] = a[n] * (b[n] + c[n]) \quad \text{Ecuația 4-8}$$

IF



THEN

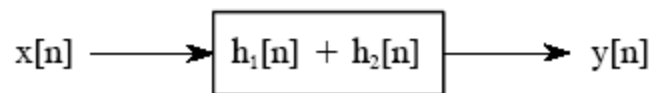


Fig. 4-10

4.7 Transferul dintre intrare și ieșire.

Așa cum este ilustrat în figura 4-11 să ne imaginăm un sistem liniar ce primește un semnal de intrare $x[n]$ și generează un semnal de ieșire $y[n]$. Acum să presupunem că semnalul de intrare este modificat (printr-un sistem liniar) rezultând un nou semnal de intrare, $x'[n]$. Va rezulta un nou semnal de ieșire $y'[n]$. Cum se modifică semnalul de

ieșire la o modificare a semnalului de intrare? Semnalul de ieșire se modifică exact în același mod ca și semnalul de intrare. Spre exemplu, dacă semnalul de intrare este amplificat cu doi, semnalul de ieșire va fi de asemenea amplificat cu doi. Dacă se scoate derivata semnalului de intrare, se va obține derivata semnalului de ieșire.

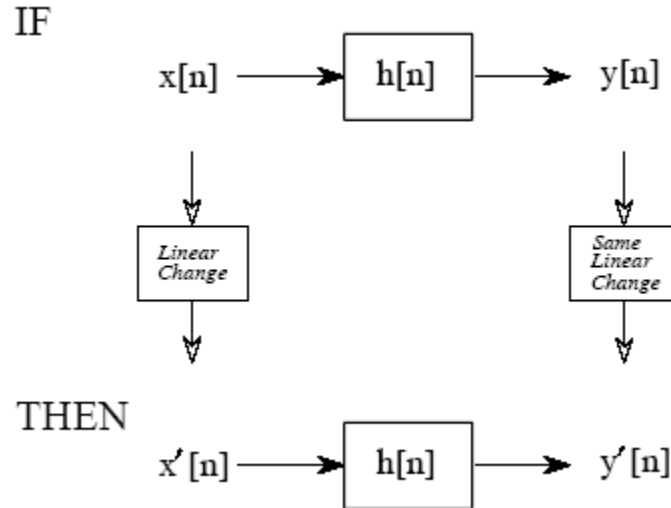


Fig. 4-11

4.8 Teorema limitei centrale.

Este un instrument important în teoria probabilității deoarece ea explică matematic de ce distribuția probabilității Gaussiene este observată așa de des în natură. Spre exemplu: amplitudinea zgomotului termic în circuitele electronice urmează o distribuție Gaussiană; secțiunea transversală a intensității unei raze laser este Gaussiană. Teorema Limitei Centrale spune că o distribuție Gaussiană rezultă atunci când variabila observată este suma a mai multor procese aleatorii. Chiar dacă procesele componente nu prezintă o distribuție Gaussiană, suma lor va prezenta.

Convoluția este o aplicație interesantă a Teoremei Limitei Centrale. Ca exemplu se ia semnalul din figura 4-12a. Dacă acesta se convoluționează cu el însuși de mai multe ori, se obține un Gaussian. Figura (b) prezintă rezultatul convoluției acestui semnal cu el însuși o singură dată. Figura (c) prezintă rezultatul obținut după trei operații de convoluție. Procedura converge foarte rapid la un Gaussian. Lățimea acestuia este egală cu lățimea impulsului original înmulțit cu rădăcina patrată a numărului de convoluție.

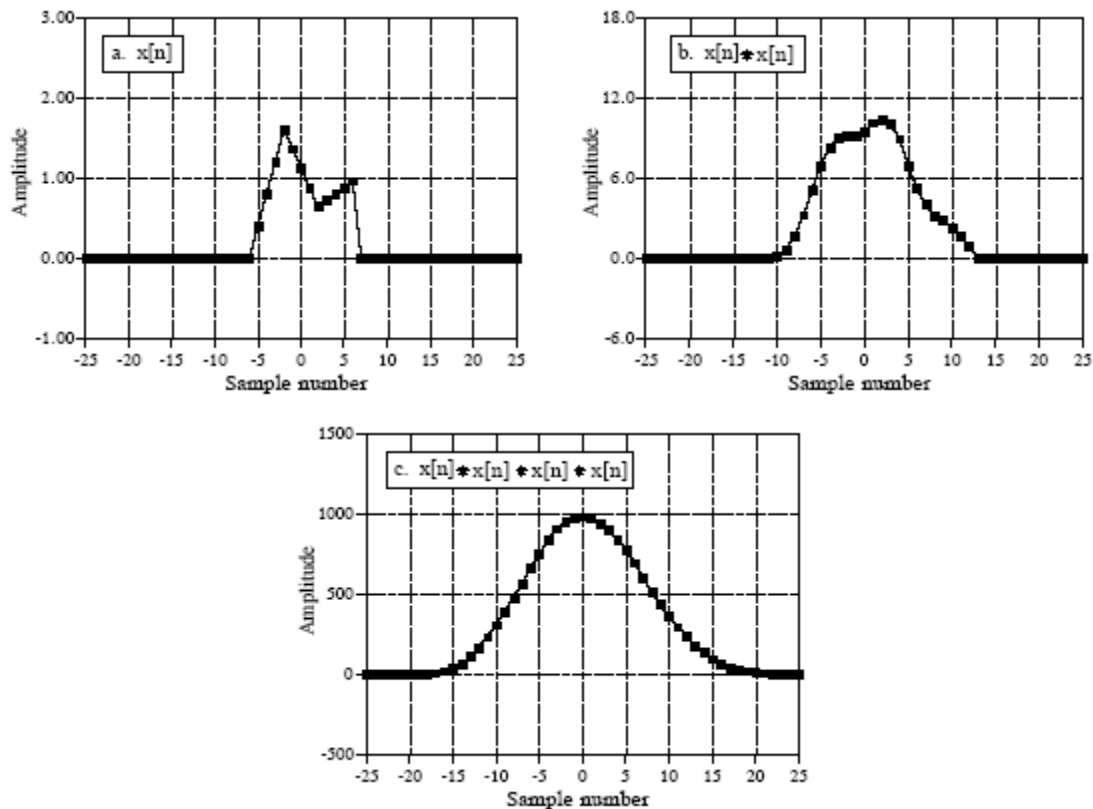


Fig. 4-12

4.9 Corelația

Conceptul de corelație poate fi cel mai bine prezentat cu un exemplu. Figura 4-13 prezintă elementele cheie a unui radar. O antenă proiectată special transmite un impuls radio într-o anumită direcție. Dacă unda propagată întâlnește un obiect, o cantitate mică de energie este reflectată înapoi la un receptor radio localizat în apropierea transmițătorului. Impulsul transmis are o anumită formă aleasă, cum ar fi triunghiulară ca în exemplu. Semnalul recepționat va consta din două părți: (1) o variantă scalată și deplasată a impulsului transmis și (2) zgomot, rezultând din interferența undelor radio, zgomot termic în componentele electronice etc. Din moment ce semnalul radio călătorește cu o anumită viteză (viteza luminii), deplasarea dintre impulsul transmis și cel recepționat este direct proporțională cu distanța până la obiectul detectat. Problema este: dându-se un semnal cu o anumită formă, care este cea mai bună metodă de a determina unde (sau dacă) acesta apare în alt semnal. Răspunsul este corelația.

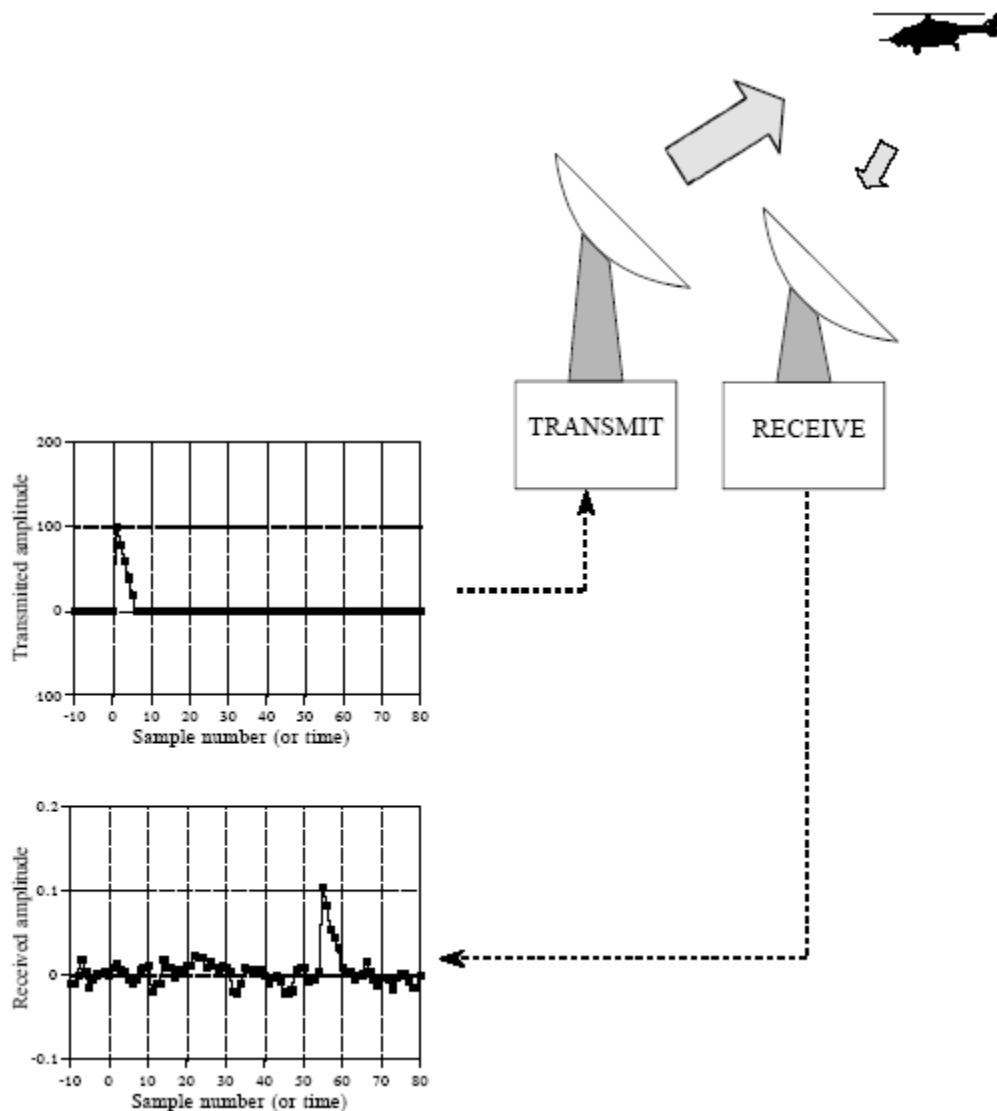


Fig. 4-13

Corelația este o operație matematică foarte asemănătoare cu convoluția. Figura 4-14, este o prezentare a mașinii de corelație. Semnalul care ne interesează, $t[n]$, numit și semnal țintă, este conținut în mașina de corelație. Fiecare eșantion din $y[n]$ este calculat prin deplasarea mașinii de corelație la dreapta sau la stânga. Amplitudinea fiecărui eșantion în semnalul de intercorelare $y[n]$ este o măsură a cât din semnalul țintă se găsește în semnalul recepționat. Asta înseamnă că va apărea un vârf în semnalul de intercorelare pentru fiecare semnal țintă care se regăsește în semnalul recepționat. Cu alte cuvinte, valoarea intercorelației este maximizată când semnalul țintă este aliniat cu un alt semnal țintă din semnalul recepționat.

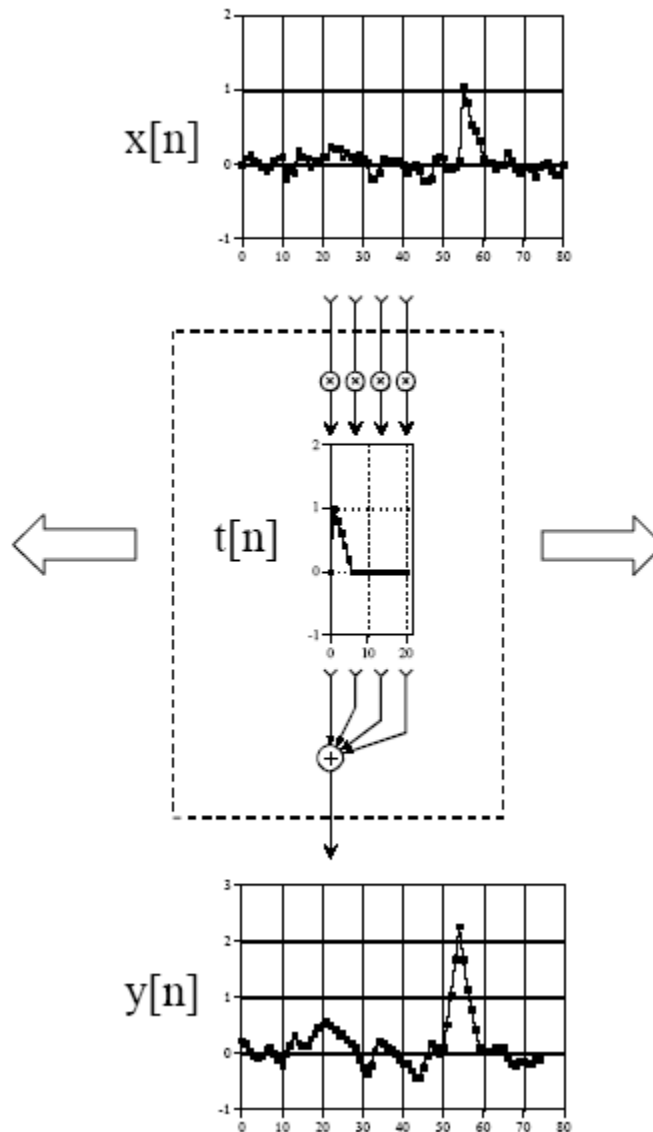


Fig. 4-14

Dacă avem zgomot în semnalul recepționat, el se va menține și în semnalul de intercorelare. Vârful generat în semnalul de intercorelare este simetric de la dreapta la stânga, chiar dacă semnalul țintă nu este simetric. În plus, lățimea acestui vârf este dublul lățimii semnalului țintă. Intercorelarea încearcă să detecteze semnalul țintă, nu să-l refacă.

Corelația este metoda optimă de a detecta o formă de undă cunoscută într-un semnal zgomot. Mașinile de corelație și convoluție sunt identice, cu excepția unei mici diferențe. În mașina de convoluție semnalul din interior este inversat de la stânga la dreapta. În mașina de corelație această inversare nu are loc.

Din moment ce acest semnal inversat este singura diferență între cele două operații este posibil să reprezentăm corelația folosind aceeași matematică a convoluției. Spre exemplu,

când $a[n]$ și $b[n]$ sunt convoluționate pentru a produce $c[n]$, ecuația se scrie: $a[n]*b[n]=c[n]$. În comparație intercorelarea lui $a[n]$ și $b[n]$ poate fi scrisă: $a[n]*b[-n]=c[n]$.

Totuși cele două operații, chiar dacă sunt asemănătoare, reprezintă două proceduri DSP diferite. Convoluția reprezintă relația dintre semnalul de intrare, semnalul de ieșire și răspunsul la impuls al unui sistem. Corelația este un mod de a detecta o formă de undă cunoscută într-un zgomot de fond.

CAP 5. Transformata Fourier discretă

Analiza Fourier este o clasă de tehnici matematice ce se bazează pe descompunerea semnalelor în sinusoidă fiind folosită în prelucrarea semnalelor digitale.

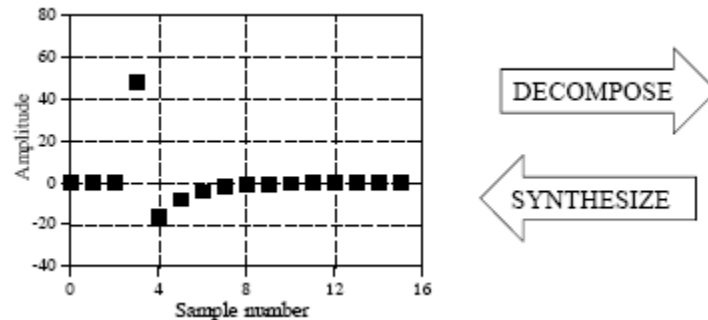
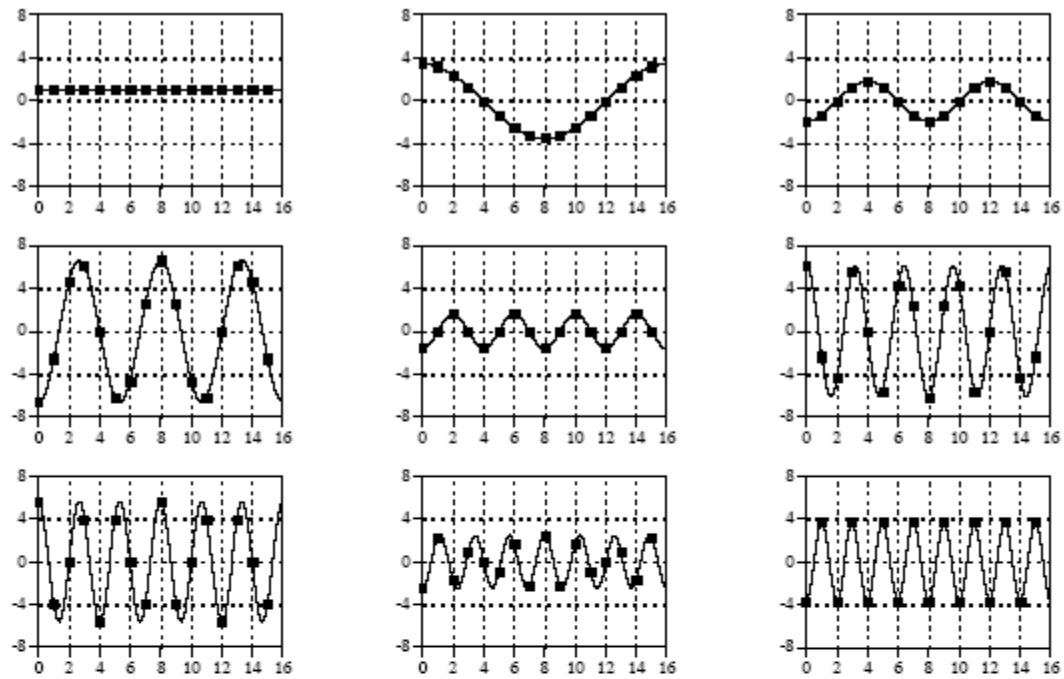


Fig. 5-1a

Figura 5-1 prezintă modul în care un semnal poate fi descompus în sinusoidă și cosinusoidă. În figura (a) avem un semnal de 16 puncte lungime, de la 0 la 15. Figura (b) prezintă descompunerea Fourier a acestui semnal, nouă sinusoidă și nouă cosinusoidă cu frecvențe și amplitudini diferite.

Scopul acestei descompuneri este ca în final să avem semnale cu care se poate lucra mai ușor decât cu semnalul original. Spre exemplu, descompunerea în impulsuri permite semnalelor să fie examinate un singur punct la un moment dat, ducând la convoluție. Componentele sinusoidale și cosinusoidale sunt mai simple decât semnalul original deoarece ele prezintă proprietatea pe care semnalul original nu o are: fidelitate sinusoidală. Sinusoidale sunt singurele forme de undă care au această proprietate.

Cosine Waves



Sine Waves

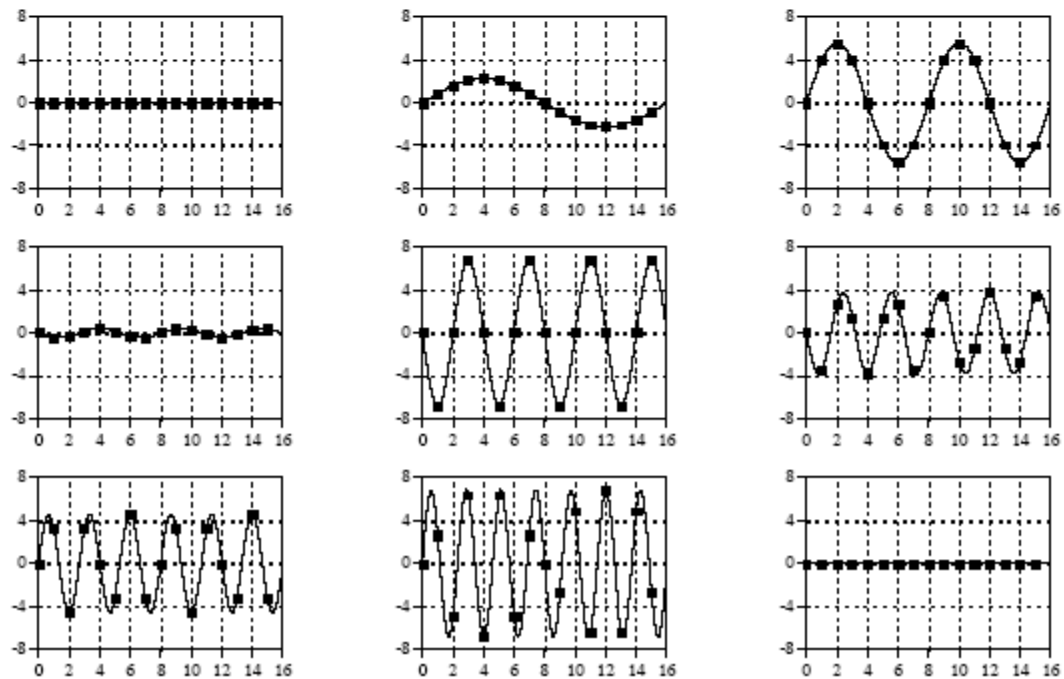


Fig. 5-1b

Transformata Fourier poate fi împărțită în patru categorii, rezultate din patru tipuri de bază ale semnalelor, ce pot fi întâlnite: continuu sau discret și periodic sau aperiodic. Din

combinația acestor două caracteristici se generează patru categorii descrise mai jos (figura 5-2).





Type of Transform	Example Signal
Fourier Transform <i>signals that are continuous and aperiodic</i>	
Fourier Series <i>signals that are continuous and periodic</i>	
Discrete Time Fourier Transform <i>signals that are discrete and aperiodic</i>	
Discrete Fourier Transform <i>signals that are discrete and periodic</i>	

Fig. 5-2

Aperiodic – Continuu

Aceste semnale se întind de la plus la minus infinit fără a se repeta într-o secvență continuă. Transformata Fourier pentru acest tip de semnal este numită simplu **Transformata Fourier**.

Periodic – Continuu

În acest tip sunt incluse sinusoidale, semnalele de ceas sau alte forme de undă care se repetă într-o secvență oarecare de la minus la plus infinit. Această variantă a transformatei Fourier este numită **Serie Fourier**.

Aperiodice – Discrete

Aceste semnale sunt definite ca puncte discrete între plus și minus infinit fără a se repeta în vreun mod periodic. Acest tip de transformată Fourier este numită **Transformată Fourier Discretă în Timp**.

Periodice – Discrete

Sunt semnale discrete care se repetă periodic de la minus la plus infinit. Această clasă a transformatei Fourier este numită **Transformată Fourier Discretă**.

Toate aceste patru clase de semnale se extind de la minus la plus infinit. Problema apare atunci când vrem să reținem aceste semnale într-un computer. Nu există nici o variantă a

transformatei Fourier ce folosește semnale cu lungime finită. Sinusoidele și cosinusoidele sunt definite ca fiind de la minus la plus infinit. Nu putem folosi un grup de semnale infinite pentru a sintetiza un semnal cu lungime finită. Modul de a ieși din această dilemă este de a face ca datele să arate ca un semnal infinit. Presupunem că semnalul nostru are un număr infinit de puncte egale cu zero la stânga și la dreapta sa. Semnalul rezultat va fi discret și aperiodic și se va aplica Transformata Fourier Discretă în Timp. Ca o alternativă, eșantioanele imaginare pot fi duplicate al celor actuale. În acest caz semnalul va fi discret și periodic, folosindu-se Transformata Fourier Discretă.

Problema care apare este că pentru a sintetiza un semnal aperiodic sunt necesare un număr infinit de sinusoid. Acest lucru face imposibilă folosirea Transformatei Fourier Directă în Timp, într-un algoritm. Rezultă că singurul algoritm care poate fi folosit în DSP este DFT. Cu alte cuvinte, calculatoarele digitale pot lucra cu informații discrete și finite ca lungime.

Să privim din nou exemplul DFT din figura 5-1. Semnalul de 16 puncte este descompus în 18 sinusoid, fiecare conținând 16 puncte. Semnalul din figura 5-1 (a) trebuie văzut ca o singură perioadă a unui semnal periodic și infinit. La fel și cele 18 sinusoid prezentate în (b) sunt segmente din 16 puncte ale unor semnale infinite.

5.1 DFT Reală.

Așa cum este prezentată în figura 5-3, transformata Fourier transformă un semnal de N puncte în două semnale de $N/2 + 1$ puncte. Vectorul de intrare conține semnalul care trebuie descompus, în timp ce vectorul de ieșire conține amplitudinile sinusoidelor și cosinusoidelor. Semnalul de intrare se află în domeniul timp, asta deoarece majoritatea tipurilor de semnal ce intră în DFT sunt compuse din eșantioane luate la intervale regulate de timp. La intrarea DFT - ului pot fi orice fel de tip de date eșantionate. Când apare termenul de “domeniu timp” în analiza Fourier, aceasta se referă la eșantioanele luate la diferite momente în timp. Termenul de “domeniu frecvență” este folosit pentru a descrie amplitudinea sinusoidelor și cosinusoidelor.

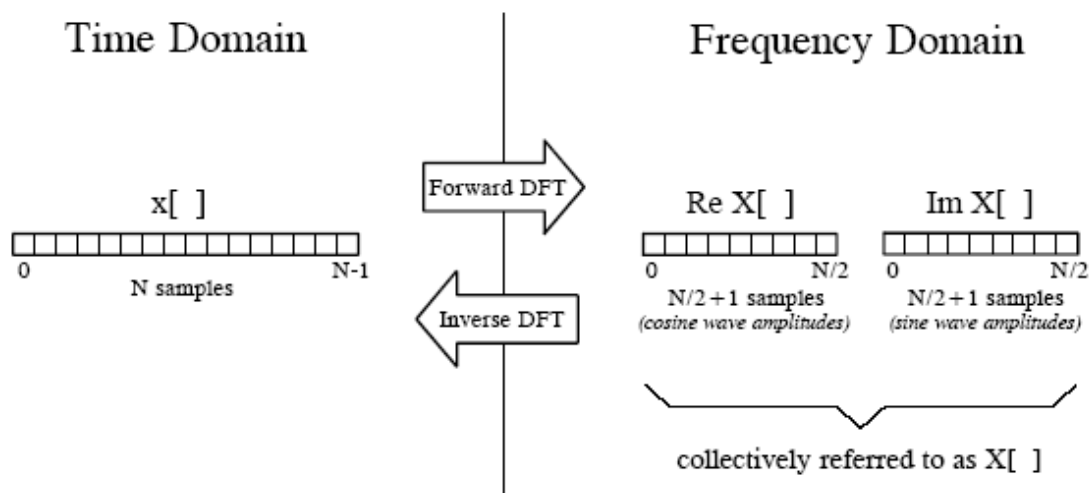


Fig. 5-3

Domeniul frecvență conține exact aceeași informație ca și domeniul timp, dar într-o formă diferită. Dacă cunoaștem un domeniu, putem calcula celălalt. Dându-se semnalul în domeniul timp, procesul de calculare a domeniului frecvență se numește descompunere, analiză, DFT direct sau simplu DFT. Dacă cunoaștem domeniul frecvență, calcularea domeniului timp se numește sinteză sau DFT inversă. Ambele, sinteza și analiza pot fi reprezentate în ecuații formând un algoritm de calcul.

De obicei, numărul de eșantioane în domeniul timp este reprezentat de variabila N . N poate fi orice număr întreg pozitiv și se alege ca fiind o putere a lui doi. Motivul pentru care se face asta este că cel mai eficient algoritm de calcul al DFT, Transformata Fourier Rapidă (FFT), operează de obicei cu N care este o putere a lui doi. Semnalele în domeniul timp sunt reprezentate cu caractere mici iar cele din domeniul frecvență cu caractere mari. Să presupunem un semnal de N elemente în domeniul timp $x[]$. Echivalentul în domeniul frecvență al acestui semnal este numit $X[]$, și este alcătuit din două părți de câte $N/2 + 1$ eșantioane. Ele sunt numite Partea Reală a lui $X[]$ scrisă $ReX[]$ și Partea Imaginară a lui $X[]$ scrisă $ImX[]$. În $ReX[]$ sunt amplitudinile cosinoidelor în timp ce în $ImX[]$ sunt amplitudinile sinusoidelor. Cele două semnale sunt de la $x[0]$ la $x[N-1]$, în domeniul timp și de la $ReX[0]$ la $ReX[N/2]$ și $ImX[0]$ la $ImX[N/2]$ în domeniul frecvență.

5.2 Variabilele independente ale domeniului frecvență.

Figura 5-4 prezintă un exemplu DFT cu $N = 128$. Semnalul din domeniul timp este de la $x[0]$ la $x[127]$. Semnalele din domeniul frecvență sunt conținute în doi vectori: $ReX[0]$ la $ReX[64]$ și $ImX[0]$ la $ImX[64]$.

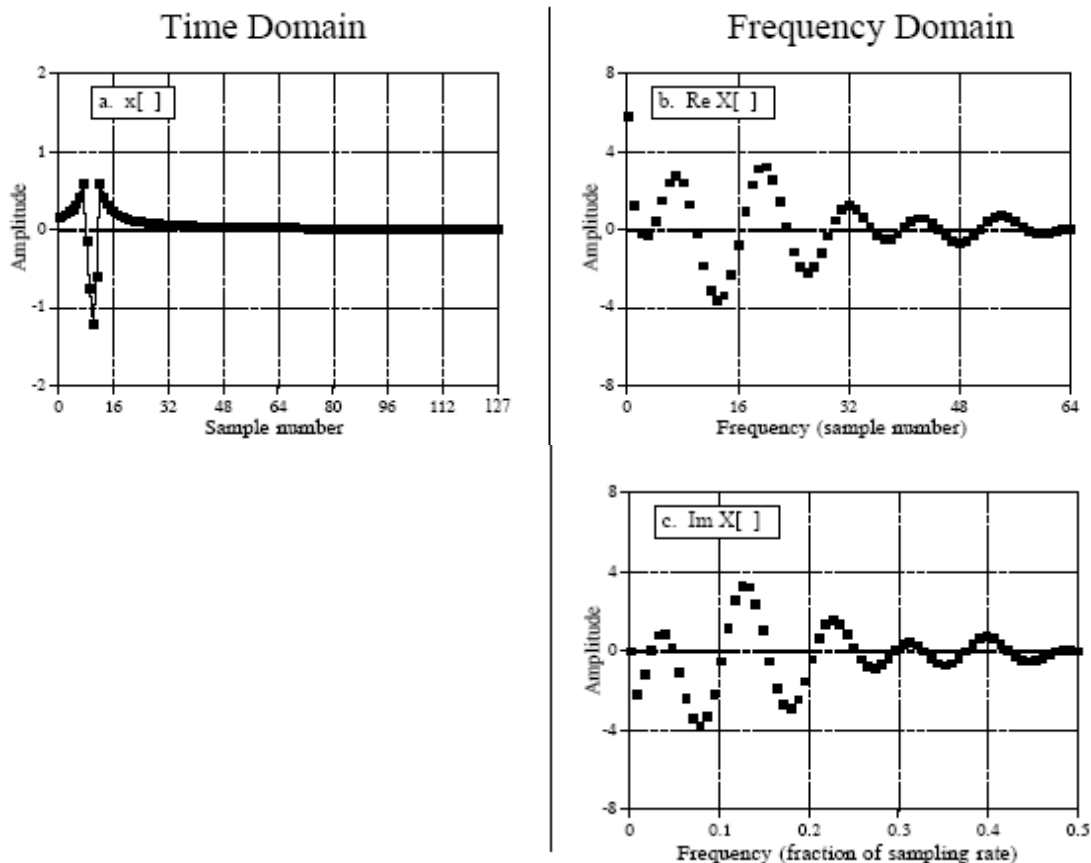


Fig. 5-4

Axa orizontală poate fi referită în patru moduri, toate utilizate în DSP. În prima metodă axa orizontală este notată de la 0 la 64, corespunzător cu 0 la $N/2$ eşantioane în vector. Când se foloseşte această notaţie, indexul în domeniul frecvenţă este un număr întreg, spre exemplu, $\text{Re}X[k]$ şi $\text{Im}X[k]$, unde k este de la 0 la $N/2$ cu pasul unu. Programatorii folosesc această metodă deoarece pot accesa fiecare locaţie a vectorului (fig. 5-4b).

În cea de-a doua metodă, folosită în (c), axa orizontală este notată ca fracţie a ratei de eşantionare. Asta înseamnă că valorile de pe axa orizontală sunt de la 0 la 0,5 din moment ce datele conţin frecvenţe cuprinse între 0 (DC) şi jumătate din rata de eşantionare. Indexul folosit în această notaţie este f , de la frecvenţă. Partea reală şi imaginară sunt scrise: $\text{Re}X[f]$ şi $\text{Im}X[f]$, unde f poate lua $N/2 + 1$ valori egal depărtate între 0 şi 0,5. Pentru a trece de la prima notaţie, k , la cea de-a doua, f , se împarte axa orizontală la N , $f = k/N$.

Al treilea mod este similar cu cel de-al doilea cu excepţia faptului că axa orizontală este înmulţită cu 2π . Indexul folosit în această notaţie este ω . Partea reală şi cea imaginară sunt scrise: $\text{Re}X[\omega]$ şi $\text{Im}X[\omega]$, unde ω ia $N/2 + 1$ valori egal depărtate între 0 şi π . Parametrul ω , este numit **frecvenţă naturală** şi are ca unitate de măsură **radianul**. Matematicienii folosesc această metodă deoarece fac ecuaţiile matematice mai scurte.

A patra metodă folosește notația directă a frecvențelor pe axa orizontală. Spre exemplu dacă sistemul face o eșantionare la 10kHz, domeniul de frecvență va fi între 0 și 5kHz.

5.3 Funcții de Bază ale DFT.

Sinusoidale și cosinusoidale folosite în DFT sunt numite funcții de bază DFT. Cu alte cuvinte, ieșirea DFT este un set de numere ce reprezintă amplitudini. Funcțiile de bază alcătuiesc un set de sinusoidale și cosinusoidale cu amplitudinea unitară. Dacă asignăm fiecare amplitudine (domeniul frecvență), la sinusoida sau cosinusoida corespunzătoare (funcțiile de bază), vom avea ca rezultat un set de sinusoidale și cosinusoidale scalate ce pot fi adunate pentru a forma semnalul din domeniul timp. Funcțiile de bază sunt generate din ecuațiile:

$$\begin{aligned} C_k[i] &= \cos(2\pi ki/N) \\ S_k[i] &= \sin(2\pi ki/N) \end{aligned} \quad \text{Ecuația 5-1}$$

unde: $c_k[i]$ este cosinusul pentru amplitudinea memorată în $\text{ReX}[k]$ și $s_k[i]$ este sinusul pentru amplitudinea memorată în $\text{ImX}[k]$. Spre exemplu, figura 5-5 prezintă câteva din cele 17 sinusoidale și cele 17 cosinusoidale folosite într-un DFT cu $N = 32$ de puncte. Din moment ce aceste sinusoidale sunt adunate pentru a forma semnalul de intrare, ele trebuie să aibă aceeași lungime ca și acesta. În acest caz, fiecare are 32 de puncte de la $i = 0$ la 31. Parametrul, k selectează frecvența fiecărei sinusoidale. În particular, $c_1[i]$ este cosinusoida ce face un ciclu complet în N puncte, $c_5[i]$ este cosinusoida ce face cinci cicli compleți în N puncte etc. Parametrul frecvență, k este egal cu numărul de cicli care apar în N puncte ale semnalului.

Figura (a) prezintă $c_0[i]$. Aceasta este o cosinusoidă cu o frecvență egală cu zero, care are o valoare constantă egală cu unu. Asta înseamnă că $\text{ReX}[0]$ reține valoarea medie a tuturor punctelor semnalului din domeniul timp. În electronică se spune că $\text{ReX}[0]$ reține valoarea de offset DC. Sinusoida pentru frecvența 0, $s_0[i]$ este prezentată în (b), toate punctele semnalului sunt egale cu zero. Acesta nu afectează semnalul din domeniul timp la sintetizare.

Figurile (c) și (d) prezintă $c_2[i]$ și $s_2[i]$, sinusoidale care completează doi cicli în N puncte. Acestea corespund cu $\text{ReX}[2]$ și $\text{ImX}[2]$. La fel, (e) și (f) prezintă $c_{10}[i]$, și $s_{10}[i]$, sinusoidale care completează 10 cicli în N puncte. Aceste sinusoidale corespund amplitudinilor reținute în $\text{ReX}[10]$ și $\text{ImX}[10]$. Problema este că figurile (e) și (f) nu arată ca niște sinusoidale. Dacă liniile continue nu ar fi prezente în aceste grafice am distinge cu greu forma acestor sinusoidale. Matematic ele sunt corecte.

Frecvențele maxime în funcțiile de bază sunt prezentate în (g) și (h). Acestea sunt $c_{N/2}[i]$ și $s_{N/2}[i]$ sau în exemplul nostru $c_{16}[i]$ și $s_{16}[i]$.

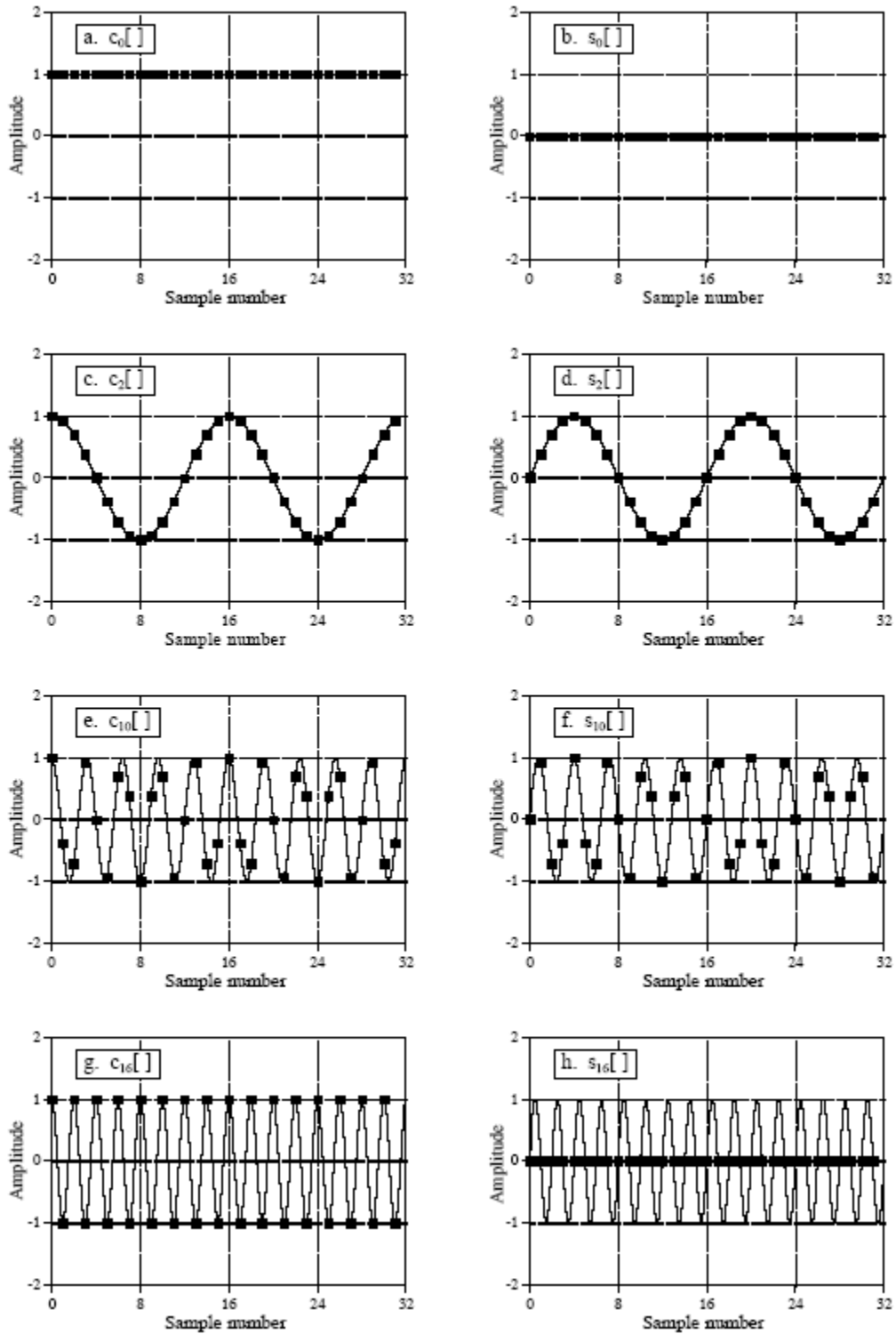


Fig. 5-5

Punctele de eșantionare în c_{16} sunt situate în vârfurile cosinusoidei între $+1$ și -1 . În contrast, sinusoida conține toate eșantioanele ca fiind zero. Acest lucru face ca valoarea

lui $\text{Im}X[N/2]$ să fie aceeași cu $\text{Im}X[0]$ și întotdeauna egală cu zero. Ele nu afectează sinteza domeniului timp.

Dacă N eșantioane intră în DFT și $N + 2$ eșantioane ies din acesta, de unde provine această extrainformație? Răspunsul este că două din semnalele de ieșire nu conțin informație ($\text{Im}X[0]$ și $\text{Im}X[N/2]$).

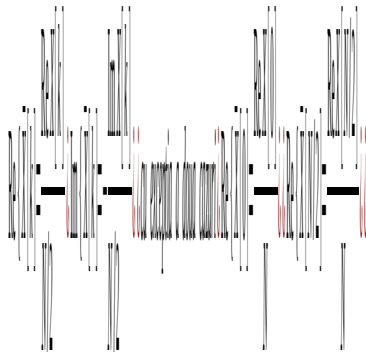
5.4 Sinteza. Calculul DFT Inversă

Ecuția de sinteză este:

$$x[i] = \sum_{k=0}^{N/2} \text{Re} \{ \bar{X}[k] \cos(2\pi ki/N) \} + \sum_{k=0}^{N/2} \text{Im} \{ \bar{X}[k] \sin(2\pi ki/N) \} \quad \text{Ecuția 5-2}$$

Cu alte cuvinte, orice semnal de N puncte, $x[i]$ poate fi creat prin adunarea a $N/2 + 1$ cosinusoide și $N/2 + 1$ sinusoide. Amplitudinile cosinusoidelor și sinusoidelor sunt reținute în vectorii $\text{Im} \{ \bar{X}[k] \}$ și $\text{Re} \{ \bar{X}[k] \}$. Ecuția de sinteză înmulțește aceste amplitudini cu funcțiile de bază pentru a crea un set de sinusoide și cosinusoide scalate. Adunând aceste sinusoide și cosinusoide scalate se generează semnalul în domeniul timp, $x[i]$.

În ecuația 5-2, vectorii sunt numiți $\text{Im} \{ \bar{X}[k] \}$ și $\text{Re} \{ \bar{X}[k] \}$ și nu $\text{Im}X[k]$ și $\text{Re}X[k]$. Asta deoarece amplitudinile folosite la sinteză (numite $\text{Im} \{ \bar{X}[k] \}$ și $\text{Re} \{ \bar{X}[k] \}$) sunt puțin diferite față de domeniul frecvență a unui semnal (notate prin $\text{Im}X[k]$ și $\text{Re}X[k]$). Ecuțiile de conversie sunt:



Ecuția 5-3.

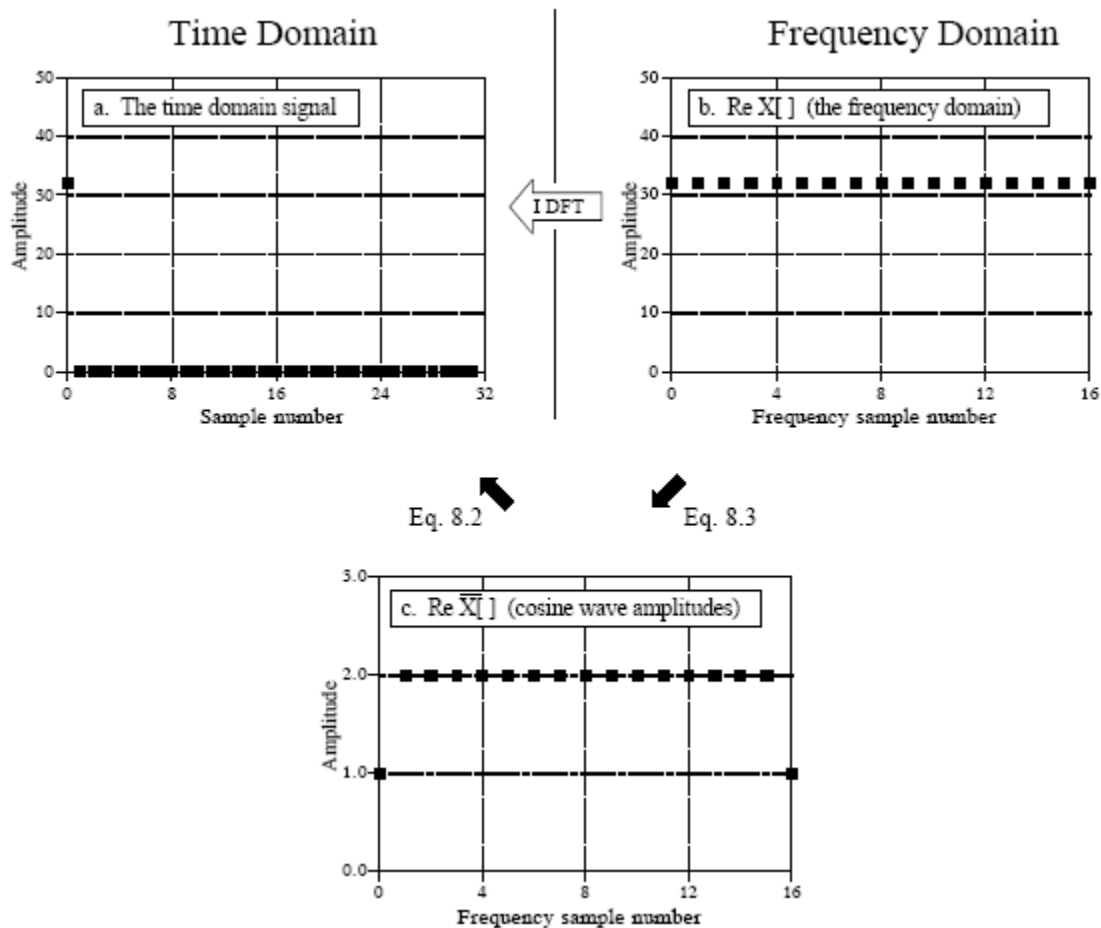


Fig. 5-6

Figura 5-6 prezintă operația DFT inversă și diferența dintre domeniul frecvență și amplitudinile folosite la sinteză. În figura 5-6 (a) este semnalul ce se dorește a fi sintetizat, un impuls cu amplitudinea de 32 pe eșantionul zero. Figura 5-6 (b) prezintă semnalul în domeniul frecvență. Partea reală a domeniului frecvență este o valoare constantă de 32. Partea imaginară (neprezentată în figură) este alcătuită din zerouri. Un impuls în domeniul timp corespunde unei valori constante în domeniul frecvență. Lucrul important este că 5-6 (b) este DFT-ul lui (a) și (a) este DFT-ul invers al lui (b).

Ecuția 5-3 este folosită pentru a face conversia semnalului din domeniul frecvență (b), în amplitudini ale cosinusoidelor (c). Toate cosinusoidelor au amplitudinea egală cu doi, cu excepția eșantioanelor 0 și 16 care au valoarea 1. Amplitudinea sinusoidelor nu este prezentată în acest exemplu deoarece ele au valoarea zero și nu contribuie în nici un fel la sinteză. Ecuția de sinteză, 5-2 este folosită pentru a converti amplitudinile cosinusoidelor (b), în semnalul din domeniul timp (a).

Diferența dintre domeniul frecvență și amplitudinile sinusoidelor apare deoarece domeniul frecvență este considerat ca o **densitate specială** (figura 5-7). Exemplul din această figură este partea reală a unui semnal de 32 de puncte în domeniul frecvență. Eșantioanele sunt de la 0 la 16, reprezentând 17 frecvențe egal depărtate între zero și

jumătate din frecvența de eșantionare. Densitatea spectrală descrie cât semnal (amplitudine) este prezent pe unitatea lățimii de bandă. Pentru a converti amplitudinile sinusoidale în densitate spectrală, împărțim fiecare amplitudine printr-o lățime de bandă. Cum determinăm lățimea de bandă a fiecărei frecvențe în domeniul frecvență?

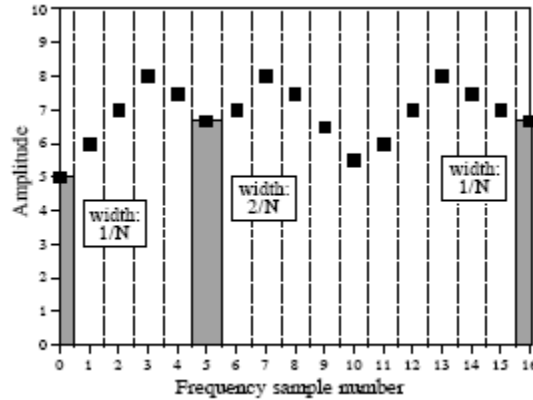


Fig. 5-7

Așa cum este prezentată în figură, lățimea de bandă poate fi definită prin trasarea de linii între eșantioane. Exprimată ca fracție, lățimea de bandă a fiecărui eșantion este $2/N$. Ca excepție sunt eșantioanele de la capete care au lățimea de bandă de $1/N$. De aici apare factorul de scalare $2/N$ între amplitudinile sinusoidale și domeniul frecvență, cât și factorul $1/N$ pentru primul și ultimul eșantion.

Partea imaginară apare negată doar pentru a face DFT-ul real corespunzător cu DFT-ul complex. Multe aplicații nu vor include acești factori de scalare sau această negare. Ele au fost studiate dintr-un motiv important: cel mai eficient mod de calcul al DFT-ului este algoritmul FFT. FFT-ul generează domeniul frecvență conform ecuațiilor 5-2 și 5-3.

5.5 Analiza. Calculul DFT-ului.

DFT-ul poate fi calculat în trei moduri complet diferite. Primul, prin rezolvarea unui sistem de ecuații. Această metodă este folositoare în înțelegerea DFT-ului, dar este prea ineficientă pentru a fi folosită în practică. A doua metodă folosește o procedură asemănătoare corelației. Ea se bazează pe detecția unei forme de undă cunoscută într-un alt semnal. A treia metodă, numită și Transformata Fourier Rapidă (FFT), este un algoritm ingenios ce descompune un DFT de N puncte în N DFT-uri fiecare de un singur punct. FFT-ul este de sute de ori mai rapid decât celelalte metode. Primele două metode vor fi prezentate acum în timp ce FFT-ul va fi discutat într-unul din capitolele următoare. Important este că toate cele trei metode generează ieșiri identice. În practică metoda a doua este folosită în cazurile în care DFT-ul are mai puțin de 32 de puncte, altfel se folosește FFT-ul.

5.5.1 DFT-ul reprezentat prin sisteme de ecuații

Dându-se N valori în domeniul timp se cere calcularea a N valori în domeniul frecvență (ignorând cele două valori ale domeniului frecvență care trebuie să fie zero).

Pentru a determina N necunoscute, trebuie să scriem N ecuații, liniare și independente. Pentru a face aceasta, luăm primul eșantion de la fiecare sinusoidă și le adunăm. Suma lor trebuie să fie egală cu primul eșantion a semnalului din domeniul timp, obținându-se prima ecuație. Pentru restul punctelor semnalului din domeniul timp se vor scrie ecuații asemănătoare, rezultând astfel N ecuații. Sistemul de ecuații se rezolvă prin metodele știute. Din păcate această metodă necesită un număr mare de calcule, motiv pentru care nu este folosit în DSP.

5.5.2 DFT-ul reprezentat prin corelație

Este metoda standard de calculare a DFT-ului. Să considerăm un exemplu pentru a reprezenta această metodă. Să calculăm DFT-ul unui semnal cu 64 de puncte. Asta înseamnă că trebuie să calculăm 33 de puncte pentru partea reală și 33 de puncte pentru partea imaginară a domeniului frecvență. În acest exemplu se va prezenta modul de calcul a unui singur eșantion, $ImX[3]$, amplitudinea sinusoidei ce execută trei cicli compleți între punctul 0 și punctul 63. Toate celelalte valori ale domeniului frecvență sunt calculate în mod similar.

Figura 5-8 prezintă folosirea corelației pentru a calcula $ImX[3]$. Figurile (a) și (b) prezintă două exemple de semnale în domeniul timp, numite: $x1[]$ și $x2[]$. Primul semnal $x1[]$ este alcătuit dintr-o sinusoidă care face trei cicli între punctele 0 și 63. În opoziție, $x2[]$ este alcătuit din mai multe sinusoidă și cosinusoidă, dar nici una nu execută trei cicli. Cele două semnale descriu ce ar trebui să facă algoritmul pentru a calcula pe $ImX[3]$. Când se întâlnește semnalul $x1[]$ algoritmul trebuie să genereze o valoare egală cu 32, amplitudinea sinusoidei prezent în semnal (modificat printr-un factor de scalare descris în ecuația 5-3). În schimb, când întâlnește celălalt semnal, $x2[]$ trebuie generată valoarea zero, indicându-se faptul că această sinusoidă nu este prezentă în semnalul dat. Pentru a detecta o formă de undă cunoscută într-un alt semnal, se înmulțesc cele două semnale și se adună toate punctele semnalului rezultat. Numărul obținut ca rezultat este o măsură a similarității celor două semnale. Figurile (c) și (d) prezintă semnalele care sunt căutate, o sinusoidă ce face trei cicli între punctul 0 și 63. Figura (e) reprezintă rezultatul înmulțirii dintre (a) și (c). La fel, (f) este rezultatul înmulțirii dintre (b) și (d). Suma punctelor din (e) este 32, în timp ce suma punctelor din (f) este 0, ceea ce trebuia de fapt demonstrat.

Celelalte eșantioane din domeniul frecvență sunt calculate în același mod. Această procedură este scrisă în ecuația de analiză, modul matematic de a calcula domeniul frecvență din domeniul timp.

$$\begin{aligned} \text{Re } X[k] &= \sum_{i=0}^{N-1} x[i] \cos(2\pi ki/N) \\ \text{Im } X[k] &= -\sum_{i=0}^{N-1} x[i] \sin(2\pi ki/N) \end{aligned} \quad \text{Ecuația 5-4}$$

Cu alte cuvinte, fiecare eșantion din domeniul frecvență este determinat prin înmulțirea semnalului din domeniul timp cu sinusoida sau cosinusoida căutată și adunarea punctelor rezultante.

Ecuția de analiză nu necesită o tratare specială a primului și ultimului punct, așa cum necesită ecuația de sinteză. Semnul minus din partea imaginară a ecuației 5-4 face ca DFT-real să corespundă cu DFT-complex și nu este întotdeauna inclus.

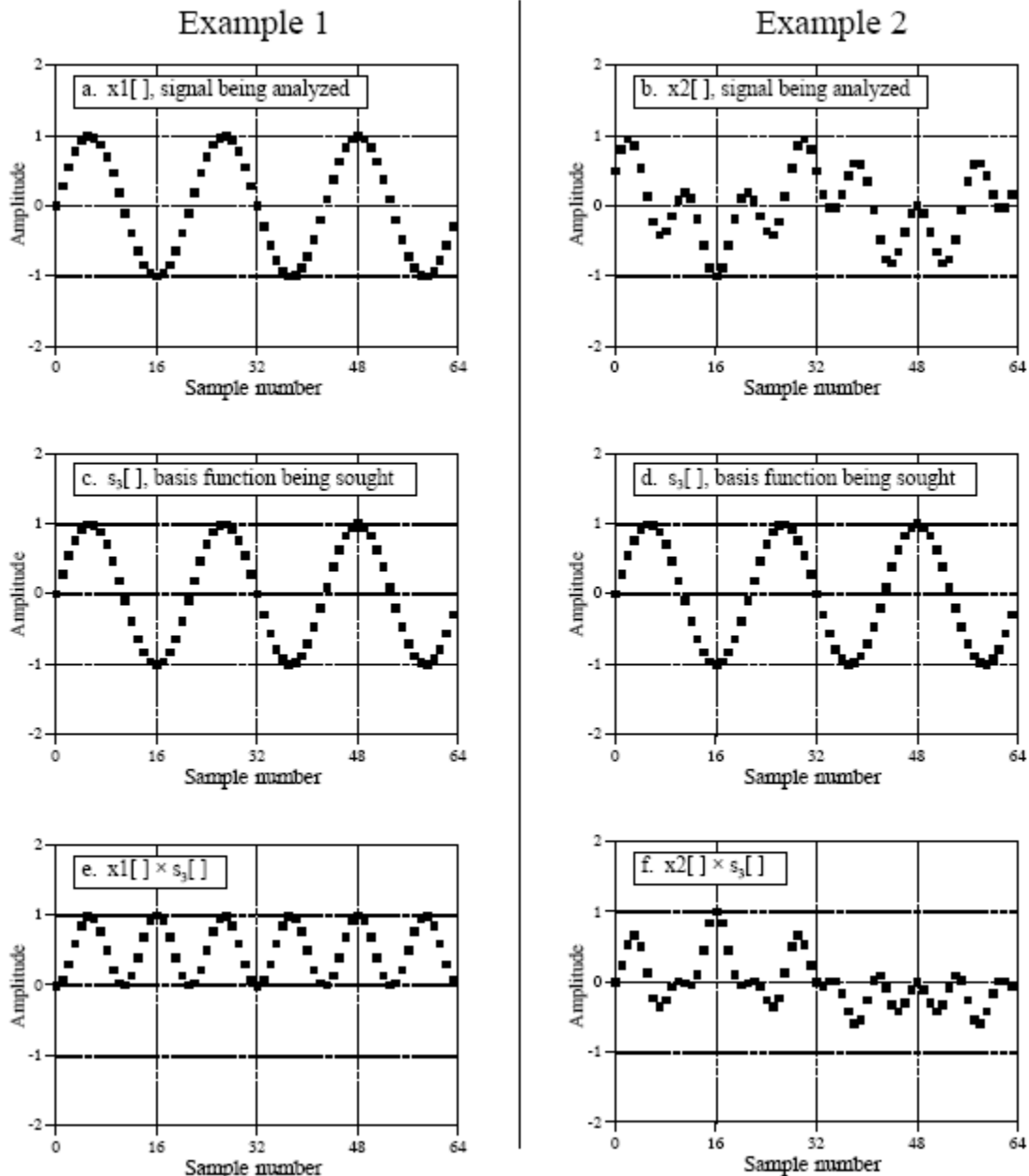


Fig. 5-8

Pentru ca acest algoritm de corelare să funcționeze, trebuie ca funcțiile de bază să aibă o proprietate: fiecare dintre ele trebuie să fie complet necorelabilă cu toate celelalte. Asta

înseamnă că dacă înmulțim oricare două funcții de bază, suma punctelor rezultante va fi egală cu zero. Funcțiile de bază care au această proprietate se numesc ortogonale.

5.6 Dualitatea

Ecuațiile de sinteză și analiză (5-2 și 5-4) sunt foarte asemănătoare. Pentru a trece dintr-un domeniu în altul, valorile cunoscute sunt înmulțite cu funcțiile de bază iar rezultatele produselor adunate. Faptul că DFT-ul și DFT-ul invers folosesc ecuații matematice asemănătoare este un lucru remarcabil, având în vedere modul total diferit de construcție a celor două proceduri.

De fapt, singura diferență semnificativă dintre cele două ecuații este că rezultatul în domeniul timp este un semnal de N puncte, în timp ce domeniul frecvență prezintă două semnale de $N/2 + 1$ puncte. DFT-ul complex definește domeniul timp și frecvență ca semnale complexe de câte N puncte. Acestea fac cele două domenii perfect simetrice.

Această simetrie dintre domeniul timp și domeniul frecvență este numită **dualitate** și prezintă numeroase proprietăți. Spre exemplu, un singur punct în domeniul frecvență corespunde unei sinusoide în domeniul timp. Prin dualitate, inversa este de asemenea adevărată, un singur punct în domeniul timp corespunde unei sinusoide în domeniul frecvență.

5.7 Notăția Polară

După cum se știe, domeniul frecvență este un grup de amplitudini de sinusoide și cosinusoide (cu mici modificări de scalare). Aceasta este numită notația **rectangulară**. Domeniul frecvență poate fi exprimat și în forma **polară**. În această notație, $ReX[j]$ și $ImX[j]$ sunt înlocuiți cu alți doi vectori numiți **Amplitudinea lui $X[j]$** scris în ecuații ca: **$MagX[j]$** și **Faza lui $X[j]$** , scris ca: **$PhaseX[j]$** . Amplitudinea și faza sunt perechi cu părțile reale și imaginare. Spre exemplu, $MagX[0]$ și $PhaseX[0]$ sunt calculate folosind numai $ReX[0]$ și $ImX[0]$. Pentru a înțelege această conversie să vedem ce se întâmplă atunci când adunăm o cosinusoidă cu o sinusoidă cu aceeași frecvență. Rezultatul este o cosinusoidă cu aceeași frecvență, dar cu o altă amplitudine și fază. În formă matematică cele două reprezentări arată astfel:

$$A \cos(x) + B \sin(x) = M \cos(x + \theta) \quad \text{Ecuația 5-5}$$

Un lucru important este acela că nu se pierde nici un fel de informație în acest proces. Cu alte cuvinte, informația conținută în A și B este de asemenea conținută în valorile M și θ . Figura 5-9 descrie modul de reprezentare a variabilelor A și B în sistemul de coordonate rectangulare, în timp ce M și θ sunt reprezentate în coordonate polare.

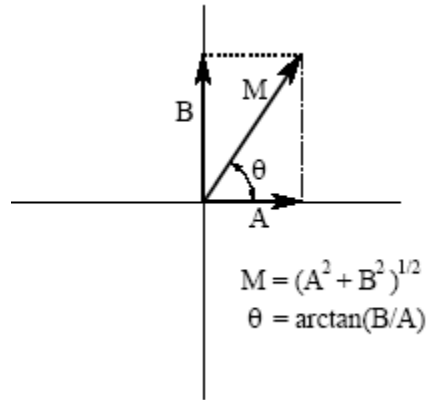


Fig. 5-9

În notația polară, $MagX[k]$ reține amplitudinea cosinusoidei, în timp ce $PhaseX[k]$ reține unghiul defazaj al cosinusoidei. Ecuațiile de conversie a domeniului frecvență din forma rectangulară în cea polară și invers sunt:

$$MagX[k] = (\text{Re } X[k]^2 + \text{Im } X[k]^2)^{1/2}$$

$$PhaseX[k] = \arctan\left(\frac{\text{Im } X[k]}{\text{Re } X[k]}\right)$$

Ecuația 5-6

$$\text{Re } X[k] = MagX[k] \cos(PhaseX[k])$$

$$\text{Im } X[k] = MagX[k] \sin(PhaseX[k])$$

Ecuația 5-7

Notația polară și cea rectangulară ne permite să vedem DFT-ul în două moduri diferite. În notația rectangulară, DFT-ul descompune un semnal de N puncte în $N/2 + 1$ cosinusoide și $N/2 + 1$ sinusoide, fiecare cu o amplitudine specifică. În notația polară DFT-ul descompune un semnal de N puncte în $N/2 + 1$ cosinusuri, fiecare cu amplitudini specifice și $N/2 + 1$ defazaje. De ce notația polară folosește cosinusoide în locul sinusoidelor? Sinusoidalele nu pot reprezenta componentele DC a unui semnal, o sinusoidă de frecvență zero este alcătuită din zerouri (vezi figura 5-5 a și b).

Chiar dacă reprezentarea polară și cea rectangulară conțin exact aceeași informație, există multe exemple în care ne e mai ușor să folosim una din notații decât cealaltă. Spre exemplu, figura 5-10 prezintă un semnal în domeniul frecvență descris în ambele forme. Ele reprezintă răspunsul în frecvență a unui filtru trece-jos.

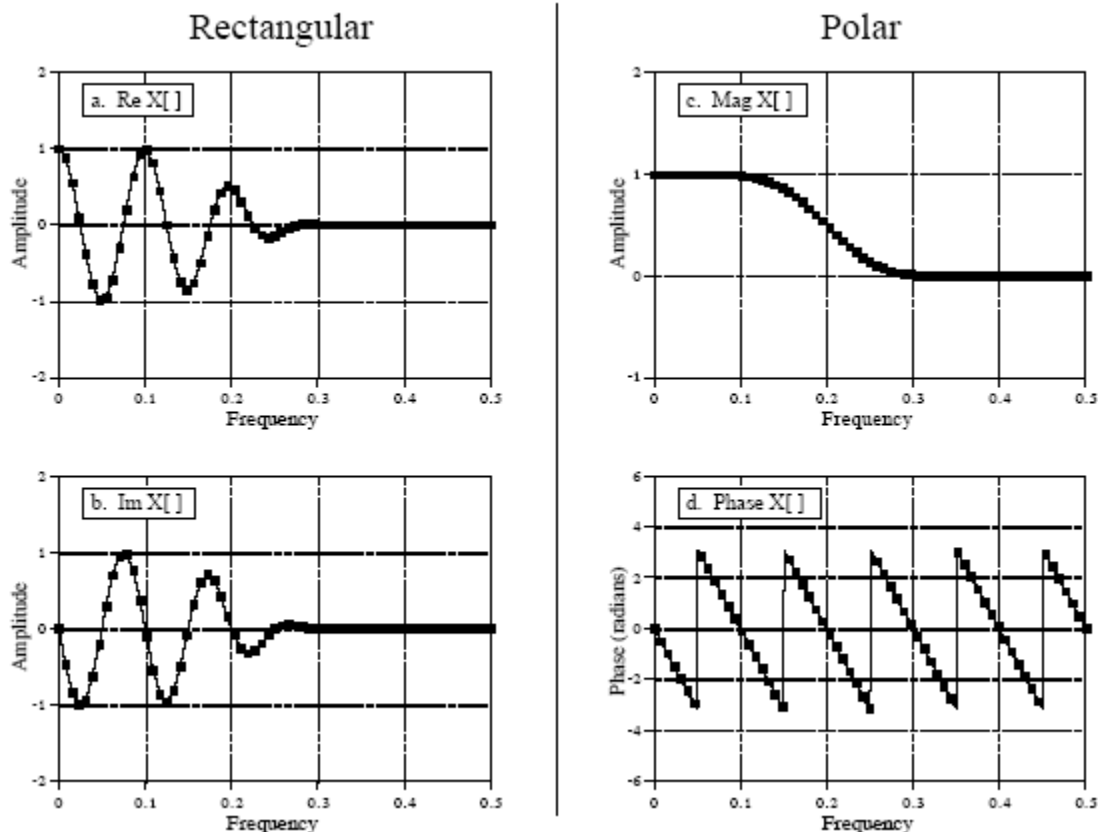


Fig. 5-10

Când folosim notația rectangulară și când folosim notația polară? Notația rectangulară este de obicei cea mai bună soluție pentru calcule, cum ar fi ecuații și programe software. În schimb, graficele sunt aproape întotdeauna reprezentate în forma polară. Este aproape imposibil pentru om să înțeleagă caracteristicile unui semnal în domeniul frecvență privind partea reală și cea imaginară a acestuia. Într-un program software, semnalele din domeniul frecvență sunt reținute în notația rectangulară. În momentul când ele trebuie privite și înțelese, trebuie făcută conversia rectangular-polar.

CAP 6. Aplicații ale DFT

Este un lucru normal ca informația să fie codată în sinusoidale care la rândul lor formează un semnal. Aceasta este adevărată pentru semnalele naturale cât și pentru cele generate de om. Spre exemplu, vorbirea este rezultatul vibrației corzilor vocale; stelele și planetele își modifică luminozitatea datorită rotirii acestora în jurul axei proprii cât și una în jurul alteia; o piatră aruncată în apă generează oscilații periodice ale apei etc. Forma semnalului în domeniul timp nu este importantă; informația se află în frecvență, faza și amplitudinea componentelor sinusoidale. DFT-ul este folosit pentru a extrage această informație.

Să presupunem că dorim să analizăm sunetele ce traversează oceanele. Pentru început, se plasează un microfon în apă iar semnalul electronic rezultat este amplificat la un nivel rezonabil (câțiva volți). Un filtru trece-jos este folosit pentru a scoate frecvențele mai mari de 80 de hertzi, astfel că semnalul poate fi digitizat cu 160 de eșantioane pe secundă.

Figura 6-1a prezintă 256 de eșantioane din semnalul nostru. Tot ceea ce se poate vedea este un zgomot ce prezintă puține informații pentru ochiul uman. Următorul pas este înmulțirea semnalului cu o fereastră Hamming, pentru a “curăța” semnalul de frecvențele aflate între frecvențele de bază, semnal prezentat în (b). Acesta are ca rezultat semnalul de 256 de puncte în care eșantioanele de la capete au amplitudinea redusă (c).

Trecând acest semnal printr-un DFT și convertindu-l la notația polară, rezultă un spectru de frecvențe de 129 de puncte în (d). Din nefericire acesta arată tot ca un zgomot. Asta deoarece nu există destulă informație în cele 256 de puncte. Folosind un DFT mai lung nu rezolvăm această problemă. Spre exemplu, dacă este folosit un DFT de 2048 de puncte, spectrul de frecvențe devine de 1025 de puncte lungime. Chiar dacă cele 2048 de puncte conțin mai multă informație, numărul mare de puncte din spectru diluează informația cu același factor. Un DFT mai lung oferă o mai bună rezoluție, dar același nivel al zgomotului.

Soluția este să folosim mai multe semnale originale, astfel încât ele să nu mărească numărul de puncte din spectrul de frecvențe. Asta se poate face prin “spargerea” semnalului de intrare în mai multe segmente de câte 256 de puncte. Fiecare din aceste segmente este înmulțit cu o fereastră Hamming, trecut apoi printr-un DFT și convertit în formă polară. Spectrele rezultate sunt mediate pentru a forma un singur spectru de 129 de puncte. Figura (e) prezintă un exemplu în care s-au mediat 100 de spectre de forma celui din figura (d). Îmbunătățirea este evidentă; zgomotul s-a redus la un nivel ce permite observarea unor caracteristici interesante ale semnalului. Numai magnitudinea domeniului frecvență este mediată în acest mod; faza este de obicei ignorată deoarece ea nu conține informații folositoare. Zgomotul se reduce de un număr de ori egal cu rădăcina pătrată din numărul de segmente. Dacă 100 de segmente sunt tipice, alte aplicații pot media milioane de segmente pentru a obține câteva caracteristici.

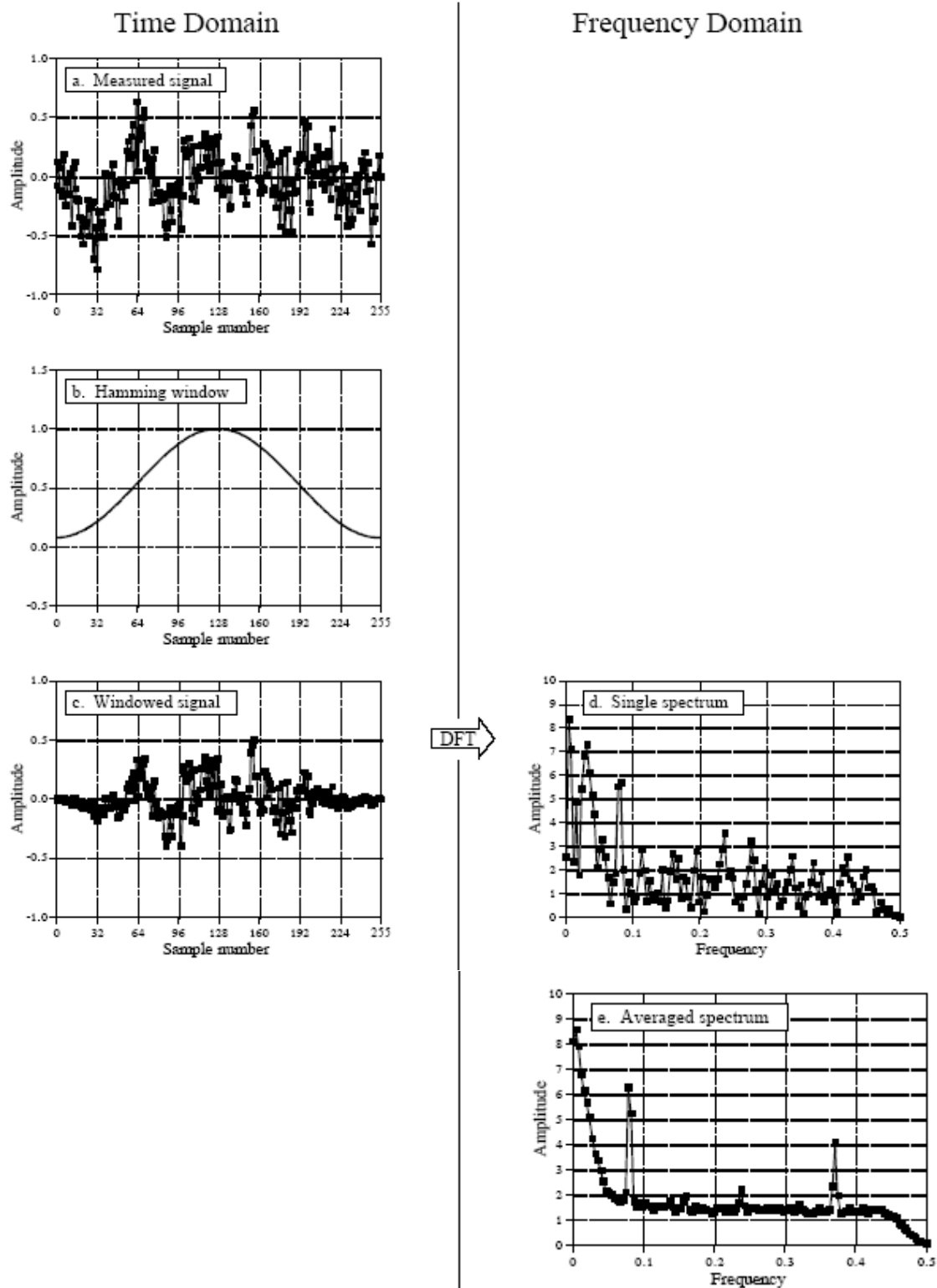


Fig. 6-1

Figura 6-2 prezintă un exemplu de spectru captat de la microfonul submarin. Între 10 și 70 de hertzi, semnalul constă dintr-o regiune relativ plată: Acest palier este numit zgomot

alb deoarece el conține o cantitate aproximativ egală din toate frecvențele, la fel ca și lumina albă. Spre exemplu, mișcarea electronilor în circuitele electronice generează zgomot alb. Sunetul generat de picăturile de ploaie ce cad pe o suprafață de apă este zgomot alb. Zgomotul alb prezentat în figura 6-2 poate fi din mai multe surse, incluzând circuitele analogice sau oceanul însuși.

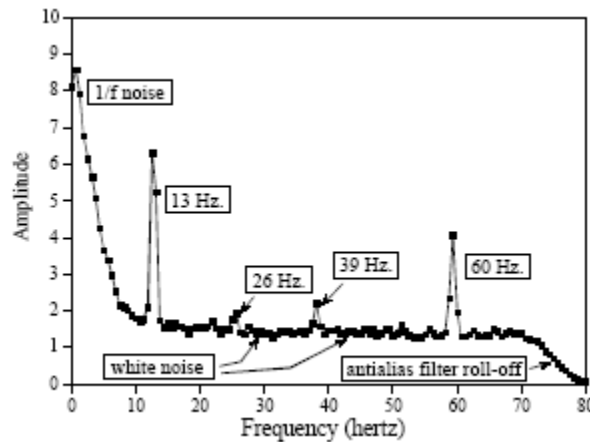


Fig. 6-2

Mai sus de 70 de hertzi, zgomotul alb descrește rapid în amplitudine. Această descreștere este rezultatul filtrului antialiasing. Un filtru ideal va permite trecerea frecvențelor mai mici de 80 de hertzi și va opri trecerea frecvențelor mai mari. În practică, o tăiere perfectă a acestor frecvențe este imposibilă, ea se va face gradual ca în figură.

Mai jos de 10 hertzi, zgomotul crește rapid generând așa numitul zgomot 1/f. Zgomotul 1/f este un mister. El a fost măsurat în diferite sisteme cum ar fi densitatea de trafic pe autostrăzi și zgomotul electric în tranzistoare. El, poate fi măsurat în toate sistemele dacă analizăm și frecvențele foarte joase. În ciuda largii răspândiri, o teorie generală și o înțelegere a zgomotului 1/f nu a fost prezentată.

Să explicăm vârfurile care apar în figura 6-2. Cel mai ușor de explicat este cel de la 60 de hertzi, ca rezultat al interferenței electromagnetice cu rețeaua electrică. De asemenea, ne așteptăm să observăm vârfuri mai mici pe multiplii ai acestei frecvențe (120, 180, 240 hertzi etc.), din moment ce semnalul din rețeaua electrică nu este o sinusoidă perfectă. Vom putea observa vârfuri de interferență între 25-40KHz datorate comutatoarelor electrice. De asemenea, vârfuri de frecvență pot apare și datorită stațiilor de radio și televiziune în banda megahertzilor.

Revenind la semnalul nostru, observăm un vârf mare la 13 hertzi, cu vârfuri mai mici la 26 și 39 hertzi. Acesta este un spectru de frecvențe a unei forme de undă nesinusoidală și periodice. Vârful de la 13 hertzi este numit frecvență fundamentală, în timp ce vârfurile de la 26 și 39 hertzi sunt numite a doua și a treia armonică. Celelalte vârfuri de la 52,65, 78... hertzi nu le observăm deoarece ele se pierd în zgomotul alb. Acest 13 hertzi poate fi generat, spre exemplu, de elicea cu trei pale a unui submarin, având o perioadă de rotație de 4,33 secunde. Acesta este un exemplu de sonar pasiv.

CAP 7. Proprietățile Transformatei Fourier

Domeniul timp sau domeniul frecvență sunt două moduri de reprezentare a semnalelor. Dacă un semnal este modificat într-un domeniu, el se va modifica de asemenea în celălalt domeniu, de obicei nu în același mod. Spre exemplu, convoluția semnalelor în domeniul timp are ca rezultat înmulțirea spectrelor acestora în domeniul frecvență. Alte operații matematice cum ar fi adunarea, scalarea și deplasarea, au de asemenea operații echivalente în celălalt domeniu. Aceste operații se numesc Proprietăți ale Transformatei Fourier și descriu modul cum o operație matematică într-un domeniu se reflectă în celălalt domeniu.

7.1 Liniaritatea Transformatei Fourier

Transformata Fourier este liniară, asta înseamnă că ea posedă proprietățile de omogenitate și aditivitate (figura 7-1). Figura (a) prezintă un semnal oarecare în domeniul timp cu spectrul în domeniul frecvență corespunzător (b). Numim aceste semnale $x[t]$ și respectiv $X[f]$. Omogenitatea înseamnă că o modificare în amplitudine într-un domeniu produce o modificare identică în amplitudine în celălalt domeniu.

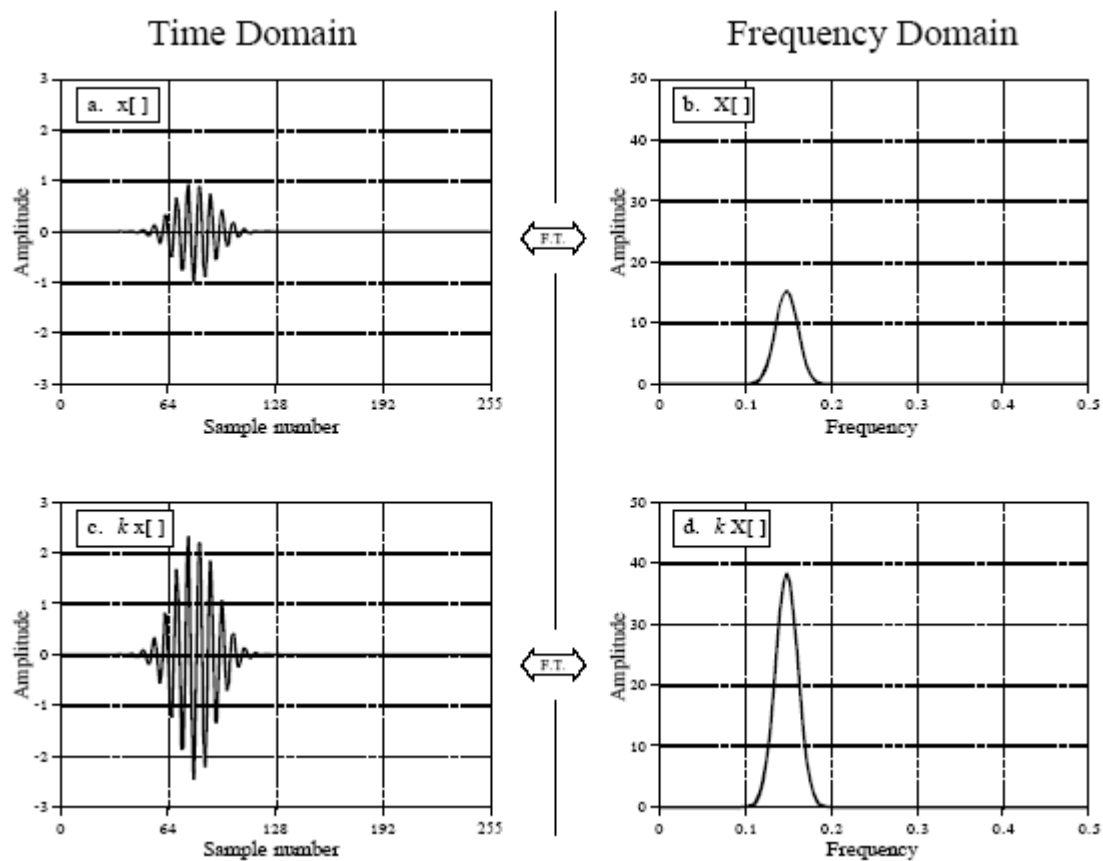


Fig. 7-1

Dacă $x[f]$ și $X[f]$ sunt perechi ale Transformatei Fourier, atunci $kx[f]$ și $kX[f]$ sunt de asemenea perechi ale Transformatei Fourier pentru oricare ar fi k . Dacă domeniul frecvență este reprezentat în formă rectangulară, $kX[f]$ înseamnă că partea reală și cea imaginară este înmulțită cu k . Dacă domeniul frecvență este reprezentat în formă polară, $kX[f]$ înseamnă că amplitudinea este înmulțită cu k , în timp ce faza rămâne neschimbată.

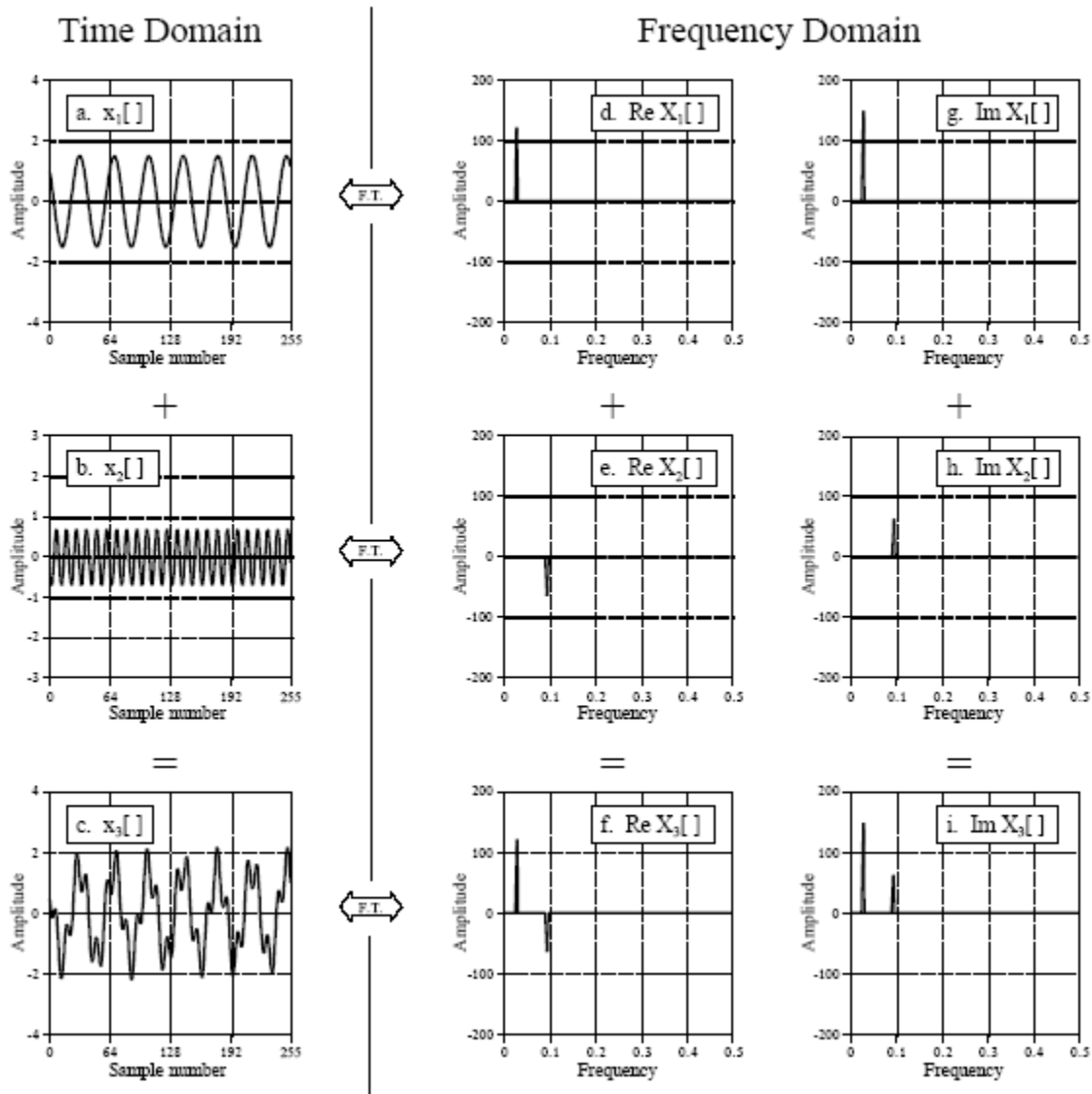


Fig. 7-2

Aditivitatea Transformatei Fourier înseamnă că sumarea într-un domeniu corespunde cu sumarea în celălalt domeniu. Un exemplu al acestuia este prezentat în figura 7-2. În această figură (a) și (b) sunt semnalele în domeniu timp numite $x_1[f]$ și $x_2[f]$. Adunând aceste semnale se obține un al treilea semnal numit $x_3[f]$ prezentat în (c). Fiecare din aceste trei semnale are un spectru de frecvențe constând din partea reală și partea imaginară: prezentat în figurile (d) la (i). Dacă cele două semnale în domeniul timp au fost adunate pentru a obține un al treilea semnal, atunci vom aduna cele două spectre corespunzătoare pentru a genera un al treilea spectru. Spectrele de

frecvențe reprezentate în forma rectangulară sunt adunate prin sumarea părților reale între ele și a părților imaginare între ele. Dacă $x_1[n] + x_2[n] = x_3[n]$, atunci: $ReX_1[f] + ReX_2[f] = ReX_3[f]$ și $ImX_1[f] + ImX_2[f] = ImX_3[f]$. Adunând toate cosinusoidale (partea reală) și toate sinusoidale (partea imaginară) nu apare nici o interacțiune între cele două.

Spectrul de frecvențe reprezentat în forma polară nu poate fi adunat direct; el trebuie convertit în forma rectangulară, adunat și apoi reconvertit înapoi în forma polară. Să ne imaginăm că adunăm două sinusoidale ce au aceeași frecvență, dar cu amplitudini diferite (A_1 și A_2) și faze (ϕ_1 și ϕ_2). Dacă cele două faze se întâmplă să fie egale, ($\phi_1 = \phi_2$), amplitudinile se vor aduna ($A_1 + A_2$) la adunarea sinusoidelor. Dacă cele două faze sunt opuse ($\phi_1 = -\phi_2$), amplitudinile se vor scădea ($A_1 - A_2$), la sumarea sinusoidelor.

În ciuda liniarității, Transformata Fourier nu este invariantă la deplasare. Cu alte cuvinte o deplasare în domeniul timp nu corespunde cu o deplasare în domeniul frecvență.

7.2 Caracteristicile fazei

Dacă $x[n] \leftrightarrow \text{Mag}X[f]$ și $\text{Phase}X[f]$, atunci o deplasare în domeniul timp este reprezentată prin: $x[n + s] \leftrightarrow \text{Mag}X[f]$ și $\text{Phase}X[f] + 2\pi sf$, (unde f este exprimat ca o fracție a ratei de eșantionare, cuprinsă între 0 și 0,5). Cu alte cuvinte, o deplasare cu s eșantioane în domeniul timp lasă amplitudinea neschimbată, dar adună un termen liniar la fază, $2\pi sf$.

Figura 7-3 prezintă modul în care este afectată faza atunci când semnalul din domeniul timp este deplasat la dreapta sau la stânga. Amplitudinea nu a fost inclusă în această figură deoarece ea nu prezintă interes; ea nu se modifică la o deplasare a domeniului timp. În figurile (a) la (d), forma de undă este deplasată gradual de la stânga la dreapta, vârf centrat pe eșantionul 128 până la vârf centrat pe eșantionul 0. Această secvență de grafice ia în considerare faptul că DFT-ul vede domeniul timp ca fiind circular; când o porțiune din semnal dispare la dreapta, el re apare la stânga.

În figura 7-3 forma de undă în domeniul timp este simetrică în jurul axei verticale. Semnalele cu acest tip de simetrie sunt numite cu faza liniară, deoarece faza spectrului lor de frecvență este o linie dreaptă. La fel, semnalele ce nu au această simetrie stânga-dreapta sunt numite cu fază neliniară, iar faza lor este orice altceva în afara unei linii drepte. Figurile (e) la (h) prezintă faza semnalelor din (a) până la (d).

Când forma de undă din domeniul timp este deplasată la dreapta, faza rămâne o linie dreaptă dar panta drepte scade. Când domeniul timp este deplasat la stânga, panta drepte crește. O deplasare în domeniul timp corespunde cu o modificare a pantei fazei. Figurile (b) și (f) prezintă unicul caz în care faza este zero. Aceasta apare când semnalul în domeniul timp este simetric în jurul punctului zero. La prima vedere,

această simetrie nu este evidentă în (b); el apare simetric în jurul eșantionului cu numărul 256.

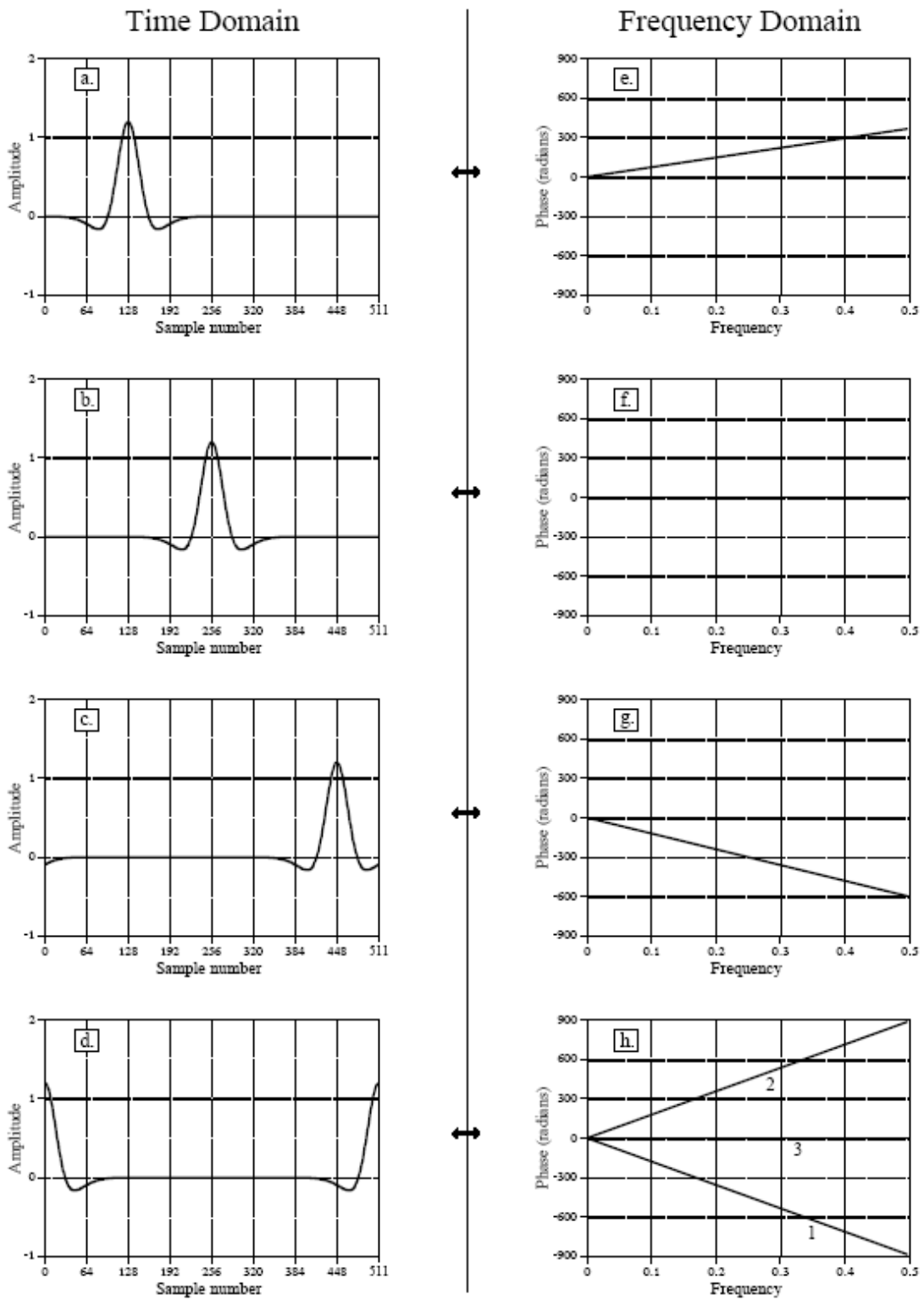


Fig. 7-3

Să ne amintim că DFT-ul vede domeniul timp ca fiind circular, cu eșantionul 0 legat de eșantionul $N - 1$. Orice semnal care este simetric în jurul eșantionului 0 va fi de asemenea simetric și în jurul lui $N/2$ și vice-versa. Când folosim membrii ai familiei Transformatei Fourier care nu văd domeniul timp ca fiind periodic (cum ar fi DTFT-ul), simetria trebuie să fie în jurul eșantionului zero pentru a genera o fază zero.

Figurile (d) și (h) necesită o discuție mai amănunțită. Prima dată, să ne imaginăm că (d) s-a format prin deplasarea semnalului din (c) puțin mai la dreapta. Asta înseamnă că faza în (h) va avea o pantă puțin mai negativă decât cea din (g). Această fază este prezentată prin linia 1. Apoi, să ne imaginăm că (d) a fost formată prin deplasarea lui (a) puțin mai la stânga. În acest caz, faza trebuie să aibă o pantă puțin mai mare decât (e), așa cum este ilustrat prin linia 2. În final, observăm că (d) este simetrică în jurul lui $N/2$ și de aceea ar trebui să aibă fază zero ilustrată prin linia 3. Care din aceste trei faze sunt corecte? Toate trei sunt corecte, depinde de ambiguitățile care apar la defazajele cu π și 2π . Spre exemplu, fiecare eșantion din linia 2 diferă de eșantionul corespunzător din linia 1 printr-un multiplu de 2π , făcându-le egale. Diferența dintre linia 3 și linia 1 și 2 este π .

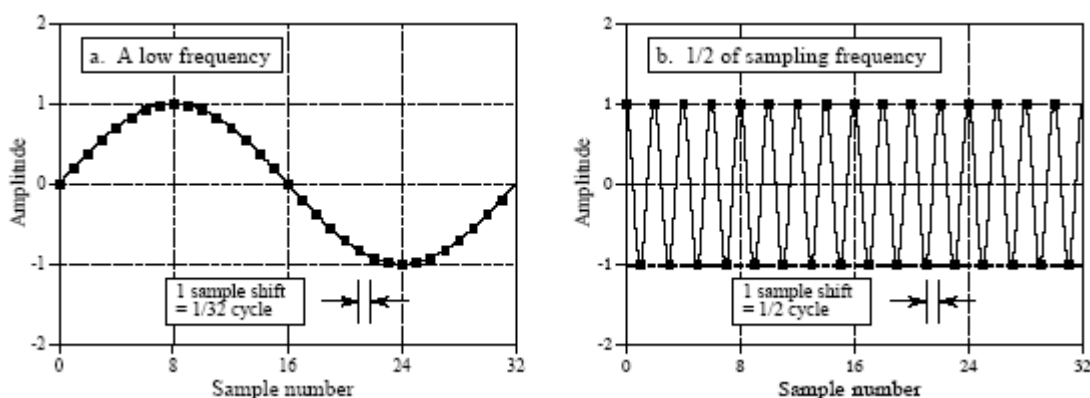


Fig. 7-4

Pentru a explica acest comportament al fazei, să ne imaginăm că semnalul se deplasează cu un eșantion la dreapta. Asta înseamnă că toate sinusoidalele care compun acest semnal trebuie să se deplaseze cu un eșantion la dreapta. Figura 7-4 prezintă două sinusoidale ce pot aparține acestui semnal. În (a), sinusoida are o frecvență foarte joasă iar o deplasare cu un singur eșantion reprezintă doar o mică fracție dintr-o perioadă. În (b), sinusoida are frecvența egală cu jumătate din frecvența de eșantionare, frecvența maximă ce poate exista în semnal. O deplasare cu un eșantion este egală cu o jumătate de ciclu sau cu π radiani. Asta înseamnă că atunci când o deplasare este exprimată într-o modificare a fazei, ea devine proporțională cu frecvența sinusoidalei care este deplasată.

Să considerăm un semnal ce este simetric în jurul lui zero, având fază zero. Figura 7-5(a) prezintă modul cum faza acestui semnal se modifică atunci când acesta este deplasat la stânga sau la dreapta. La cea mai mare frecvență (jumătate din rata de eșantionare) faza crește cu π pentru fiecare eșantion deplasat la stânga și descrește cu π pentru

fiecare eșantion deplasat la dreapta. La frecvența zero, nu există o deplasare a fazei, iar toate frecvențele cuprinse între acestea urmează o linie dreaptă.

Toate exemplele folosite au o fază liniară. Figura 7-5(b) arată că o fază neliniară reacționează la deplasare în același fel. În acest exemplu faza neliniară este o linie dreaptă cu două impulsuri rectangulare. Atunci când domeniul timp este deplasat, aceste caracteristici neliniare se suprapun peste panta modificată.

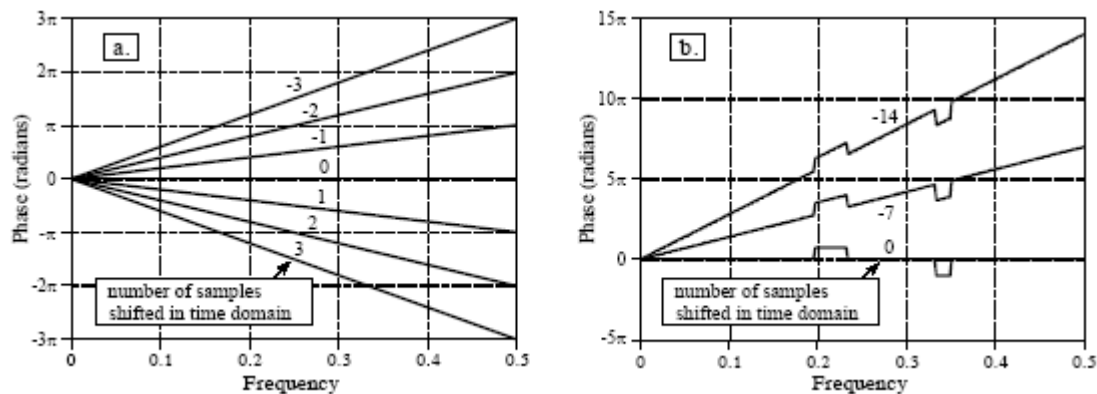


Fig. 7-5

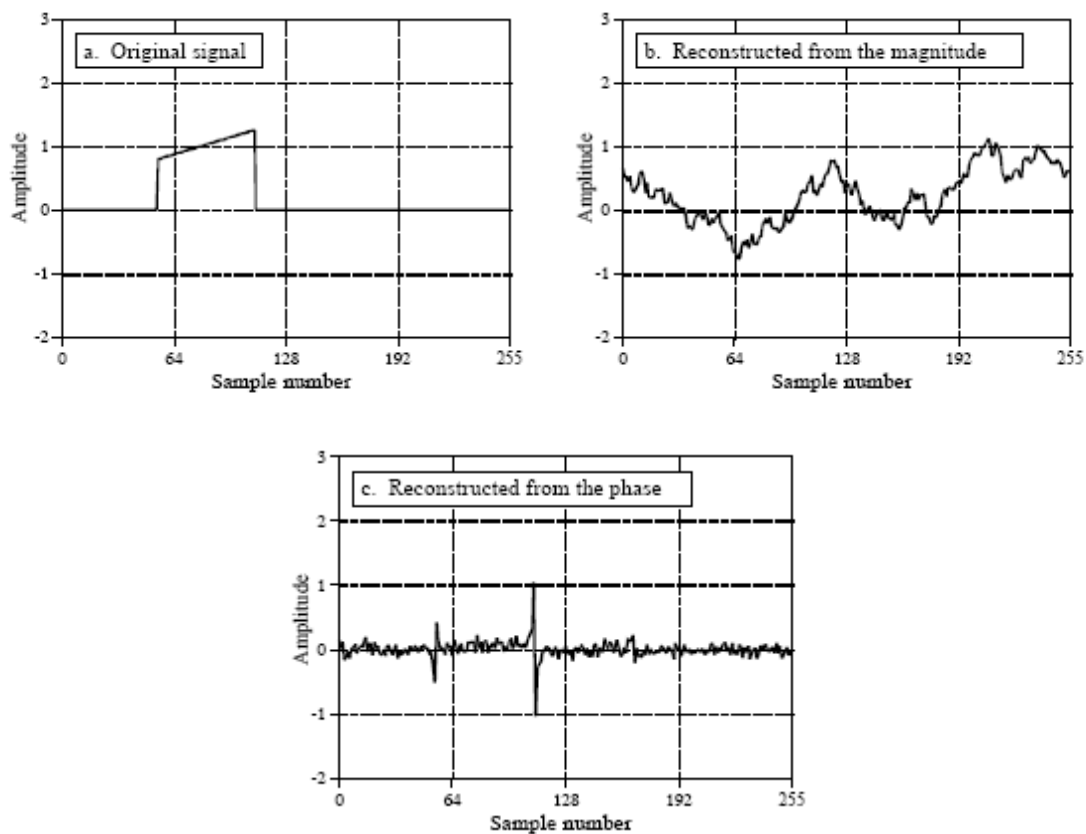


Fig. 7-6

Figura 7-6 este o demonstrație interesantă asupra informației conținută în fază și amplitudine. În (a) avem un semnal cu două caracteristici distincte: un front crescător la eșantionul cu numărul 55 și un front descrescător în eșantionul cu numărul 110. Fronturile sunt foarte importante când informația este codată în forma semnalului. Ele indică faptul că ceva s-a întâmplat. Pentru a începe demonstrația, se aplică un DFT semnalului din (a), spectrul frecvență fiind convertit în forma polară. Semnalul prezentat în (b) a fost format prin înlocuirea fazei cu numere cuprinse între $-\pi$ și π , folosindu-se apoi un DFT Invers pentru a reconstrui semnalul din domeniul timp. Cu alte cuvinte, (b) este bazată numai pe informația conținută în amplitudine. Într-un mod similar, (c) este formată prin înlocuirea amplitudinii cu un șir de numere aleatoare înainte de folosirea DFT-ului Invers. Acest lucru face reconstrucția lui (c) bazată numai pe informația conținută în fază.

Rezultatul arată că fronturile sunt poziționate la locul lor în (c), dar total absente în (b). Asta deoarece un front este format atunci când mai multe sinusoidale "cresc" în același punct. Majoritatea informației despre frecvența unui semnal în domeniul timp este conținută în fază și mai puțin în amplitudine, în contrast cu semnalele care au informația codată în domeniul frecvență, cum ar fi semnalele audio. Amplitudinea joacă un rol mai important în aceste semnale în raport cu faza.

De ce simetria stânga-dreapta corespunde unei faze zero (sau liniară)? Figura 7-7 ne dă răspunsul. Un semnal poate fi descompus în jumătatea stângă și jumătatea dreaptă așa cum este prezentat în (a), (b) și (c). Eșantionul din centrul de simetrie (zero în acest caz) împarte semnalul în două jumătăți în oglindă. Amplitudinea celor două semnale va fi identică, așa cum putem observa în (e) și (f) în timp ce faza va fi opusă ca semn (h) și (i). Rezultă două lucruri importante din aceasta:

- fiecare semnal care este simetric între partea stângă și partea dreaptă prezintă o rază liniară deoarece faza neliniară a jumătății din stânga anulează faza neliniară a jumătății din dreapta;
- să ne imaginăm o rotire a semnalului din (b) rezultând semnalul (c). Această rotire stânga-dreapta în domeniul timp nu modifică amplitudinea dar schimbă semnul fiecărui punct al fazei.

La fel, schimbarea semnului fazei are ca rezultat o rotire stânga-dreapta a semnalului în domeniul timp. Dacă semnalele sunt continue, rotirea este în jurul lui 0 . Dacă semnalele sunt discrete, rotirea este în jurul lui 0 și a lui $N/2$ simultan.

Schimbarea semnului fazei este o operație destul de uzuală așa că i s-a dat un nume și un simbol. Numele este complex conjugată iar simbolul este reprezentat printr-o stelută în partea dreapta-sus a variabilei. Spre exemplu, dacă $X[f]$ constă din $MagX[f]$ și $PhaseX[f]$, atunci $X^*[f]$ este numită complex conjugată și este compusă din $MagX[f]$ și $-PhaseX[f]$. În notația rectangulară, complex conjugată este determinată prin lăsarea părții reale nemodificată și schimbarea semnului părții imaginare. Dacă $X[f]$ este alcătuit din $ReX[f]$ și $ImX[f]$, atunci $X^*[f]$ este alcătuită din $ReX[f]$ și $-ImX[f]$. Există mai multe exemple legate de folosirea complex conjugatei în DSP. Dacă $x[n]$ are o Transformată Fourier $X[f]$, atunci $x[-n]$ are o Transformată Fourier $X^*[f]$. Un alt exemplu, să ne amintim că, corelația poate fi executată ca o convoluție, doar printr-o simplă rotire a unuia din semnale. În formă matematică, $a[n]*b[n]$ este convoluție, în timp ce $a[n]*b^*[-n]$ este corelație. În domeniul frecvență aceste operații

corespund cu $A[f] \times B[f]$ și respectiv $A[f] \times B^*[f]$. Un ultim exemplu, să considerăm un semnal arbitrar, $x[n]$ și spectrul său $X[f]$. Spectrul de frecvențe poate fi modificat la fază zero prin înmulțirea lui cu complex conjugatul său $X[f] \times X^*[f]$. Oricare ar fi faza lui $X[f]$, ea va fi anulată prin adunarea opusului (când spectrele se înmulțesc, fazele lor se adună). În domeniul timp asta înseamnă că $x[n] \times x^*[-n]$ (un semnal convoluționat cu simetricul său) va avea o simetrie stânga-dreapta în jurul lui zero, oricare ar fi $x[n]$.

7.3 Natura periodică a DFT-ului

Diferit de celelalte trei Transformate Fourier, DFT-ul vede domeniul timp și domeniul frecvență ca fiind periodice. Acest lucru poate fi un inconvenient din moment ce majoritatea semnalelor folosite în DSP sunt neperiodice.

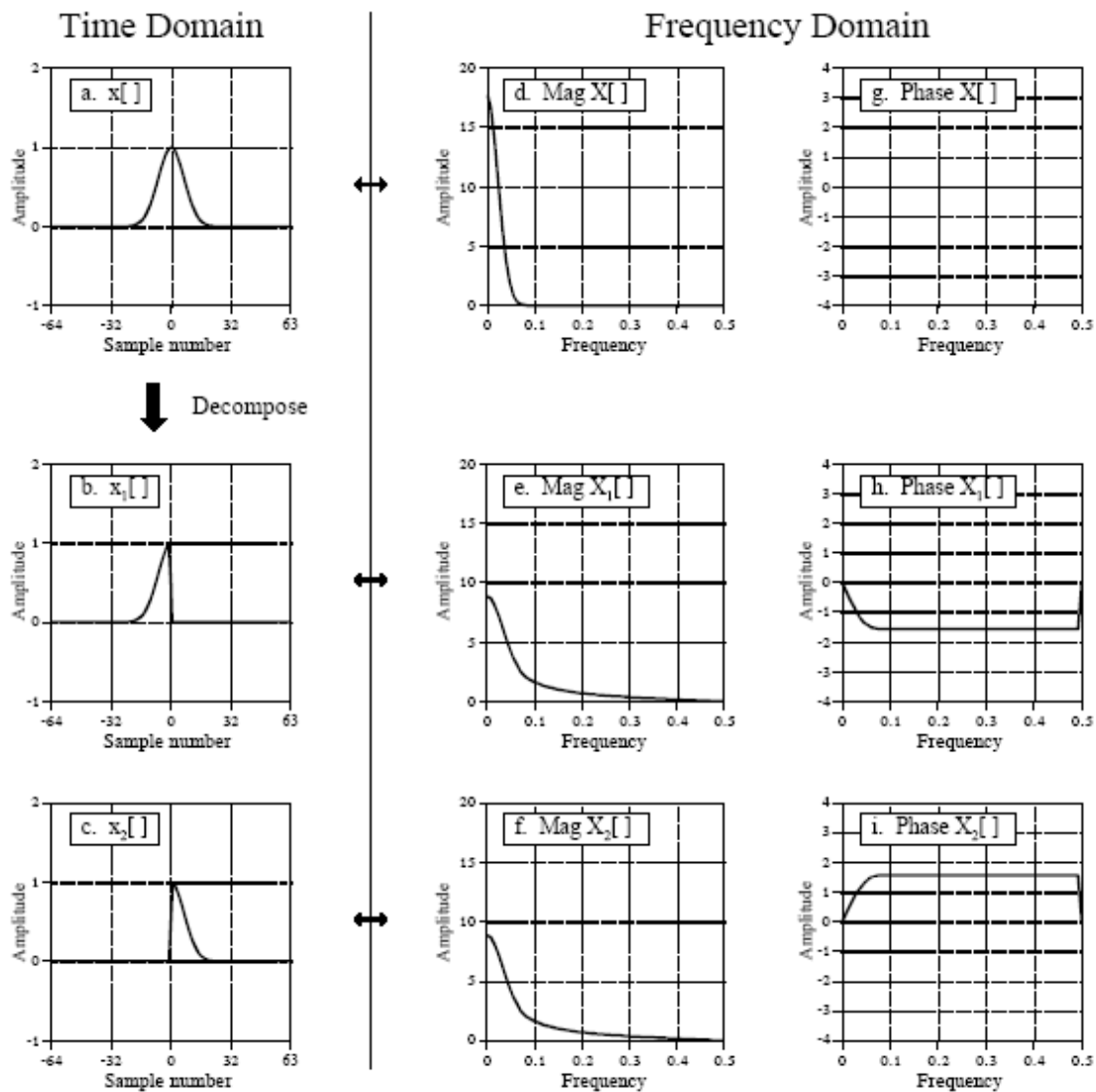


Fig. 7-7

Figura 7-8 prezintă două interpretări diferite ale semnalului în domeniul timp. Pentru început să privim semnalul de sus, un semnal de N puncte în domeniul timp. Acesta reprezintă modul în care un semnal digital este recepționat în experimentele științifice și în aplicațiile ingineresti. Spre exemplu, aceste 128 de eșantioane au fost captate prin eșantionarea unui parametru la intervale regulate de timp. Eșantionul 0 este discret și separat de eșantionul 127 deoarece ele sunt reținute la momente diferite. Nu există nici un motiv să credem că eșantioanele din stânga semnalului au vreo legătură cu cele din dreapta.

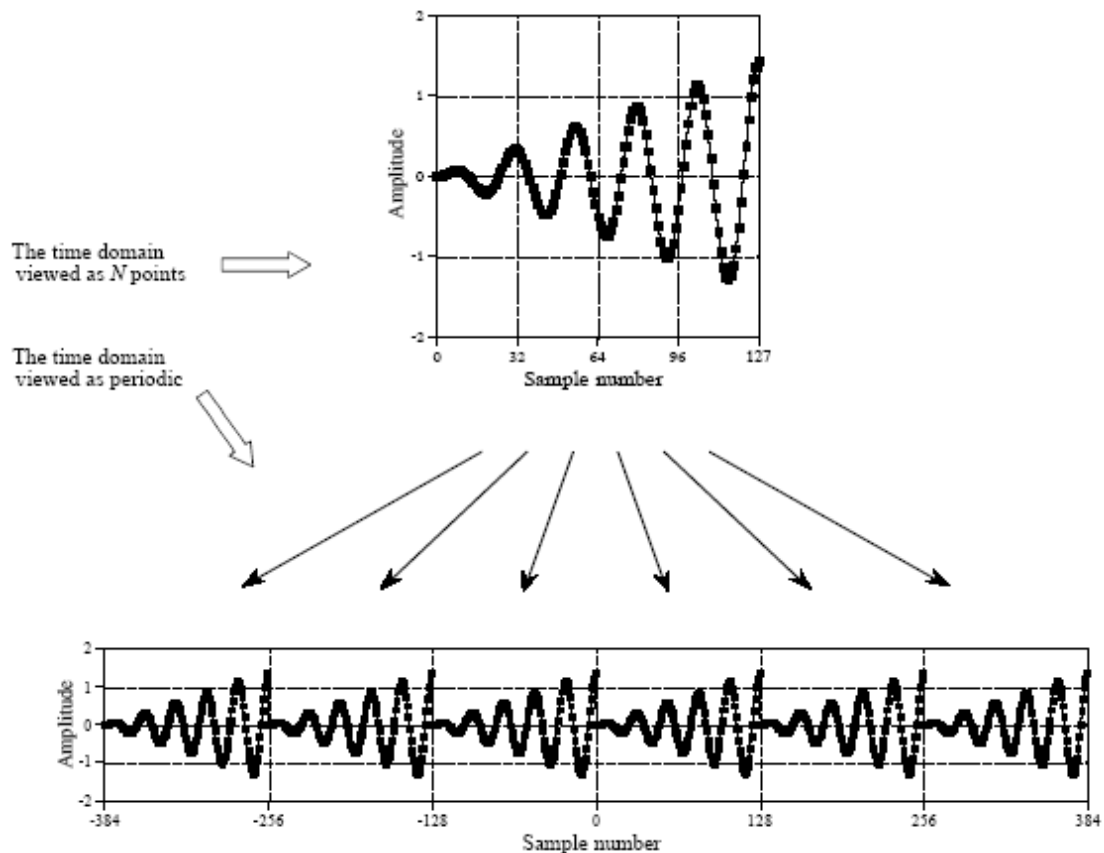


Fig. 7-8

Din nefericire, DFT-ul nu vede lucrurile în acest fel. Așa cum este prezentat în josul figurii, DFT-ul vede acele 128 de puncte, ca pe o perioadă a unui semnal periodic de lungime infinită. Asta înseamnă că partea dreaptă a semnalului recepționat este legată de partea stângă a perioadei identice. În această figură eșantionul 0 urmează după eșantionul 127. Aceasta justifică folosirea vectorilor circulari.

Cea mai serioasă consecință a periodicității domeniul timp este aliasing-ul în domeniul timp. Pentru a demonstra aceasta, să presupunem că avem un semnal în domeniul timp pe care-l trecem printr-un DFT pentru a determina spectrul său. Putem apoi, imediat, să refacem semnalul original, trecând spectrul printr-un DFT Invers, dar întreaga procedură nu ar fi prea interesantă. În locul acesteia, putem modifica spectrul de frecvențe după un anumit algoritm și apoi să-l trecem printr-un DFT Invers. Spre

exemplu putem șterge anumite frecvențe, le putem modifica faza sau amplitudinea, le putem deplasa etc. Din nefericire aceste modificări în domeniul frecvență pot genera un semnal în domeniul timp ce este prea lung pentru a încăpea într-o singură perioadă. Aceasta face ca semnalul rezultat să se suprapună peste perioada adiacentă. Când domeniul timp este văzut ca fiind circular, porțiunea din semnal care depășește partea dreaptă re apare în partea stângă a semnalului și vice-versa. Asta înseamnă ca o porțiune din semnalul în domeniul timp să ocupe o altă zonă în acest domeniu. Dacă se întâmplă ca în acea zonă să existe deja un alt semnal atunci cele două se vor aduna, rezultând o pierdere de informație.

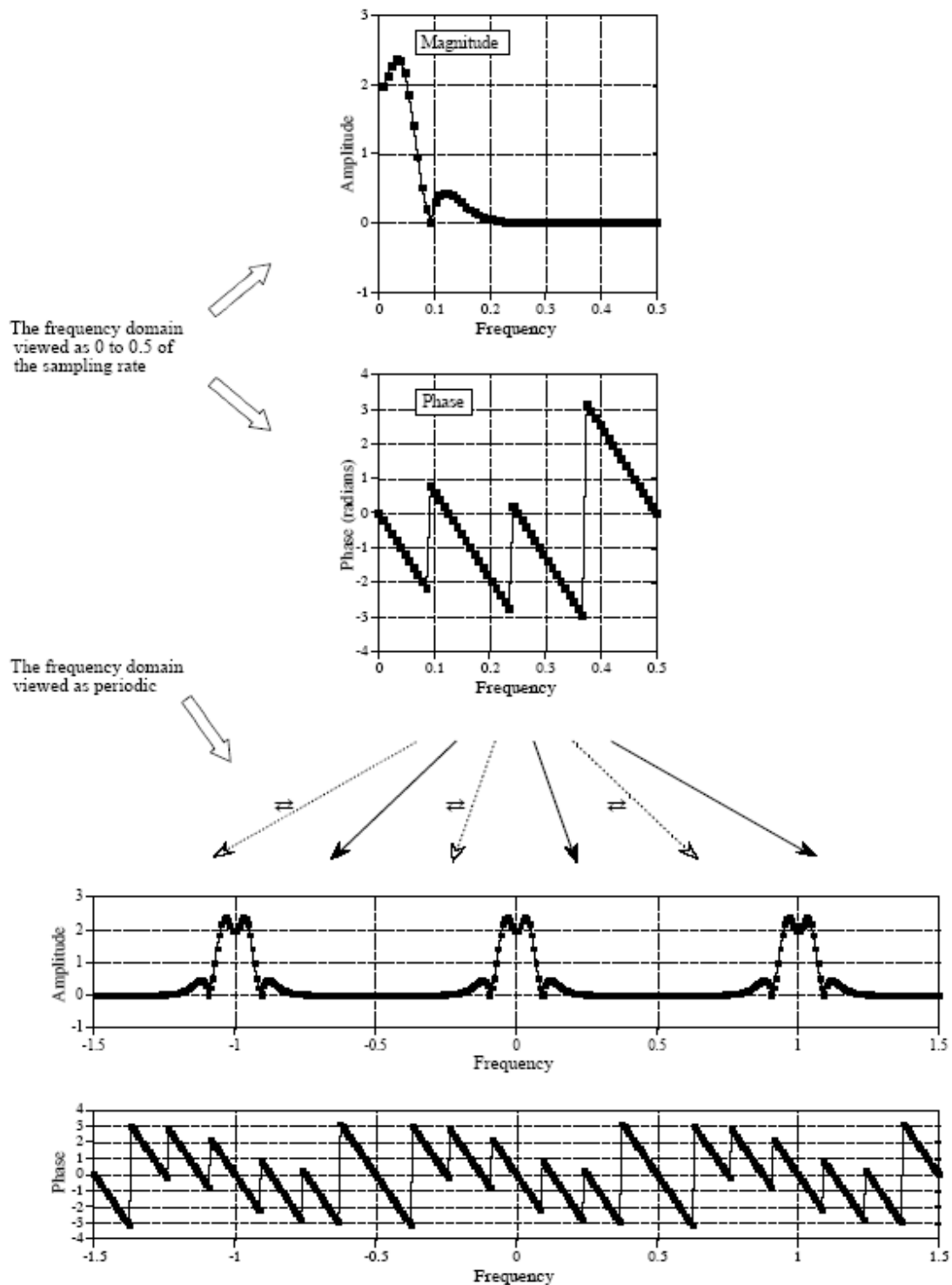


Fig. 7-9

Periodicitatea în domeniul frecvență se comportă aproximativ la fel, dar este mult mai complicată. Figura 7-9 prezintă un exemplu. Figurile de sus descriu amplitudinea și faza spectrului frecvență alcătuite din $N/2 + 1$ puncte cuprinse între 0 și 0,5 din frecvența de eșantionare. Acesta este cel mai simplu mod de a vedea spectrul frecvență.

Figura 7-9 descrie modul cum DFT-ul vede această frecvență ca fiind periodică. Caracteristica principală este aceea că spectrul de frecvențe cuprins între 0 și 0,5 are o imagine în oglindă între 0 și -0,5. Această imagine în oglindă (de frecvențe negative) este puțin diferită între magnitudine și fază. În magnitudine semnalul este rotit de la stânga la dreapta. În fază, semnalul este rotit de la stânga la dreapta și i se modifică și semnul. Aceste două tipuri de simetrii au și denumiri: magnitudinea este un semnal par (ea are simetrie pară), în timp ce faza este un semnal impar (ea are simetrie impară). Dacă frecvența este convertită în partea reală și cea imaginară, partea reală va fi întotdeauna pară, în timp ce partea imaginară va fi impară.

Luând în considerare aceste frecvențe negative, DFT-ul vede domeniul frecvență ca fiind periodic, cu o perioadă de 1 din frecvența de eșantionare, de la -0,5 la 0,5 sau de la 0 la 1. Asta face lungimea perioadei domeniului frecvență egal cu N , la fel ca și în domeniul timp.

Periodicitatea domeniului frecvență îl face pe acesta susceptibil de aliasing în domeniul frecvență, la fel ca și aliasing-ul în domeniul timp. Să ne imaginăm un semnal în domeniul timp ce prezintă un spectru de frecvențe. Dacă semnalul în domeniul timp este modificat, și spectrul său se va modifica. Dacă spectrul de frecvențe deja modificat este prea mare pentru spațiul unei perioade, el va fi plasat în perioadele următoare. Acest aliasing cauzează două probleme: frecvențele nu sunt acolo unde ar trebui să fie și adunarea frecvențelor suprapuse din diferite perioade, duce la distrugerea informației.

Aliasing-ul în domeniul frecvență este mult mai dificil de înțeles ca cel în domeniul timp. Să considerăm o singură frecvență ce este forțată să se deplaseze de la 0,01 la 0,49 în domeniul frecvență. Frecvența negativă corespunzătoare se va deplasa de la -0,01 la -0,49. Când frecvența pozitivă trece peste bariera de 0,5, frecvența negativă trece peste bariera de -0,5. Din moment ce domeniul frecvență este periodic, aceleași lucruri pot apărea și în alte perioade cum ar fi între 0,5 și 1,5. Să vedem cum ar arăta o deplasare a frecvenței dacă am vedea numai banda cuprinsă între 0 și 0,5. Dacă frecvența ce se deplasează ar dispărea în partea dreaptă, ea va reapare tot în partea dreaptă dar deplasându-se în direcție opusă.

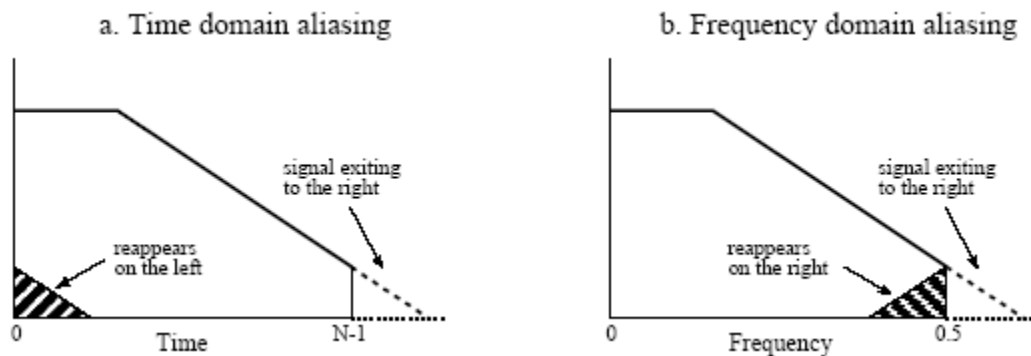


Fig. 7-10

Figura 7-10 prezintă apariția aliasing-ului în domeniile timp și frecvență pe lungimea unei singure perioade. Așa cum este prezentat în (a), dacă un capăt al semnalului în domeniul timp este prea lung pentru a se încadra pe lungimea unei singure perioade, partea care iese în afara perioadei va apărea în cealaltă parte. În (b) este prezentat domeniul frecvență. Dacă un semnal depășește o perioadă, capătul ce iese din lungimea

unei perioade reapare în aceeași parte dar inversat de la dreapta la stânga. În ambele cazuri aliasing-ul duce la distrugerea informației.

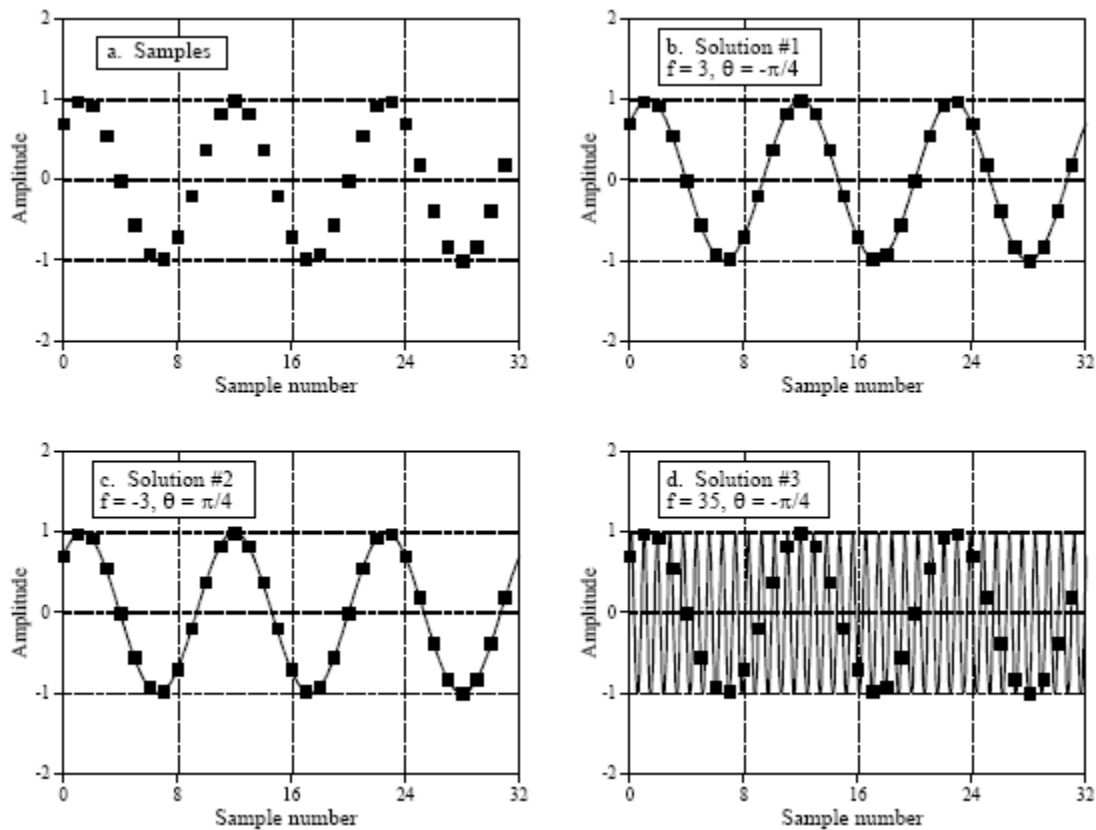


Fig. 7-11

Figura 7-11 ne prezintă ce înseamnă de fapt frecvența negativă. Figura (a) este un semnal discret alcătuit din 32 de eșantioane. Să presupunem că trebuie să găsim spectrul de frecvență corespunzător celor 32 de puncte. Pentru a ne face munca mai ușoară să considerăm că aceste puncte reprezintă o cosinusoidă. Cu alte cuvinte, trebuie să găsim frecvența și deplasarea de fază (f și θ) cum ar fi $x[n] = \cos(2\pi nf/N + \theta)$, ce se potrivește cu eșantioanele date. Se observă că o soluție ar fi cea din figura (b), care este $f = 3$ și $\theta = -\pi/4$.

Dacă ne oprim cu analiza aici vom obține doar o treime din soluția problemei. Asta deoarece mai sunt încă două soluții pe care nu le-am luat în calcul. Așa cum este arătat în (c), a doua soluție este $f = -3$ și $\theta = \pi/4$. Chiar dacă ideea de frecvență negativă este mai grea de înțeles ea nu schimbă faptul că acestea sunt soluții valide reieșite din calcul matematic. Fiecare sinusoidă cu frecvență pozitivă poate fi exprimată și ca o sinusoidă cu frecvență negativă. Asta se aplică atât semnalelor continue cât și celor discrete.

Cea de-a treia soluție nu este singulară, ci reprezintă o familie infinită de soluții. Așa cum este prezentat în (d), sinusoida cu $f = 35$ și $\theta = -\pi/4$ trece prin toate punctele discrete deci este o soluție corectă, chiar dacă ea are mai multe oscilații între eșantioane. La fel și pentru $f = \pm 29$, $f = \pm 61$ și $f = \pm 67$. Fiecare din aceste trei soluții, corespunde unei

secțiuni diferite ale spectrului de frecvență. Pentru semnalele discrete, prima soluție corespunde frecvențelor cuprinse între 0 și 0,5 din rata de eșantionare. A doua soluție aparține frecvențelor cuprinse între 0 și $-0,5$. În final, a treia soluție aparține unui număr infinit de frecvențe simetrice, mai jos de $-0,5$ și mai sus de 0,5. Dacă semnalul pe care îl analizăm este continuu, prima soluție aparține frecvențelor de la 0 la $+\infty$, în timp ce a doua soluție aparține frecvențelor de la 0 la $-\infty$. Al treilea grup de soluții nu există pentru semnalele continue.

7.4 Compresia și Expandarea

Așa cum este arătat în figura 7-12 o compresie a unui semnal într-un domeniu rezultă într-o expandare a semnalului în celălalt domeniu și viceversa. Pentru semnalele continue, dacă $X(f)$ este Transformata Fourier a lui $x(t)$, atunci $1/k \times X(f/k)$ este Transformata Fourier a lui $x(kt)$, unde k este parametrul ce controlează expansiunea sau contracția semnalului. Dacă un eveniment se petrece rapid (este compresat în timp) el trebuie să fie compus din frecvențe înalte. Dacă un eveniment se petrece în mod lent (este expandat în timp), el trebuie să fie alcătuit din frecvențe joase. Aceasta este valabilă și pentru cele două extreme care apar. Dacă un semnal în domeniul timp este compresat la infinit, devenind un impuls, spectrul corespondent este expandat la infinit, devenind o valoare constantă. La fel, dacă domeniul timp este expandat până când el devine o valoare constantă, domeniul frecvență devine un impuls.

Semnalele discrete au o comportare asemănătoare, prezentând în plus câteva probleme. Prima este aliasing-ul. Să ne imaginăm că impulsul din (a) este compresat de mai multe ori decât este prezentat. Spectrul de frecvențe este expandat cu un factor egal cu cel de compresie iar “cocoase” ale semnalului (b) sunt trecute dincolo de 0,5.

Aliasing-ul strică relația dintre extensie și compresie. Acest tip de aliasing poate avea loc și în domeniul timp. Să ne imaginăm că spectrul de frecvențe din (f) este compresat mult mai tare, rezultând o extindere a semnalului din domeniul timp (e) în perioadele învecinate.

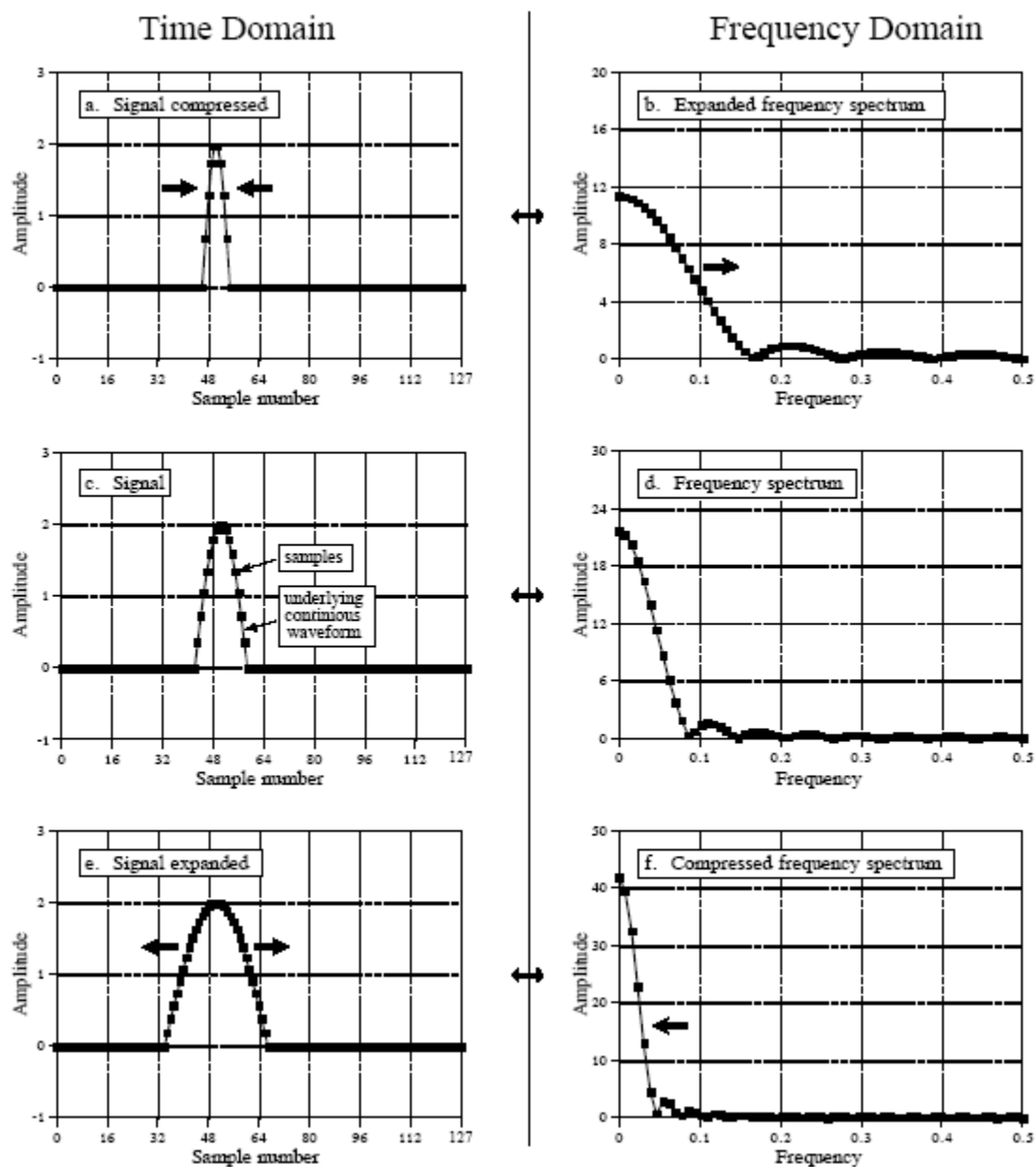


Fig. 7-12

O a doua problemă este să definim exact ce înseamnă să comprimăm sau să expandăm un semnal. Așa cum este prezentat în figura 7-12a, un semnal discret este comprimat prin comprimarea liniei continue ce leagă eșantioanele și apoi reeșantionarea noii linii continue pentru a găsi noul semnal discret. Același proces și pentru expandarea semnalelor discrete, prezentat în (e). Când un semnal discret este comprimat, evenimentele din semnal (cum ar fi lățimea impulsului) au loc într-un număr mai mic de eșantioane. La fel, evenimentele într-un semnal expandat se petrec într-un număr mai mare de eșantioane.

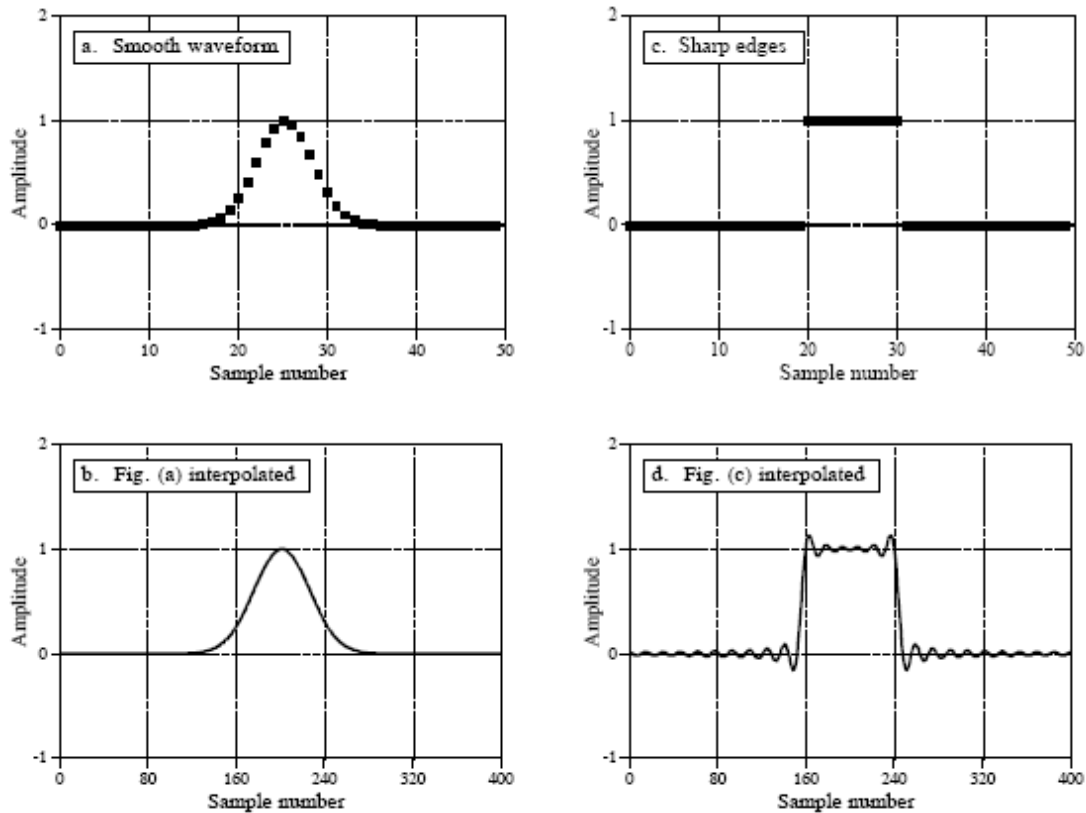


Fig. 7-13

Aceeași procedură poate fi executată prin păstrarea liniilor continue dar cu o eșantionare la o rată diferită. Spre exemplu, să privim figura 7-13(a), un Gaussian discret alcătuit din 50 de eșantioane. În (b), aceeași curbă continuă este reprezentată prin 400 de eșantioane. Schimbarea dintre (a) și (b) poate fi văzută în două feluri: (1) Rata de eșantionare a fost menținută constantă, dar linia continuă a fost expandată de opt ori și (2) Linia continuă a fost menținută constantă dar frecvența de eșantionare a fost mărită de opt ori. Aceste metode de modificare a ratei de eșantionare în acest mod este numită tehnică multiratăă. Dacă se mărește numărul de eșantioane atunci vorbim de interpolare. Dacă se micșorează numărul de eșantioane atunci este vorba de decimare.

Când un semnal prezintă informația codată în domeniul frecvență, ignorăm domeniul timp și ne concentrăm asupra spectrului de frecvență. Așa cum știm, o rezoluție mare a spectrului de frecvențe (mai multe eșantioane între 0 și 0,5) pot fi obținute prin “umplerea” semnalului în domeniul timp cu zerouri înainte de a-l trece printr-un DFT.

Dualitatea ne permite să facem aceasta și în sens invers. Dacă dorim o eșantionare mai “fină” în domeniul timp (interpolare) “umplem” spectrul de frecvențe cu zerouri înainte de a-l trece printr-un DFT invers. Să presupunem că vrem să interpolăm un semnal de 50 de puncte într-unul de 400. Asta se va face în felul următor: (1) Se iau cele 50 de eșantioane la care se adună zerouri pentru a-l face de 64 eșantioane lungime. (2) Se folosește un DFT de 64 de puncte pentru determinarea spectrului, ce constă din 33 puncte în partea reală și 33 de puncte în partea imaginară. (3) Se adaugă în partea dreaptă a spectrului (în partea reală și cea imaginară) 224 de zerouri, pentru a face spectrul de

frecvență de 257 lungime. (4) Se folosește un DFT Invers de 512 puncte pentru a transforma data înapoi în domeniul timp. Va rezulta un semnal cu 512 eșantioane având o rezoluție mult mai mare decât cel cu 64 de eșantioane. Primele 400 de eșantioane din acest semnal reprezintă o versiune interpolată a semnalului original de 50 de eșantioane.

Caracteristica de bază a acestei tehnici este aceea că semnalul interpolat este alcătuit din aceleași frecvențe ca și semnalul original. Uneori ea nu poate garanta o bună comportare în domeniul timp. Spre exemplu, figurile 7-13(a) și (b) prezintă un semnal de 50 de eșantioane interpolat într-un semnal de 400 de eșantioane prin această metodă. Interpolarea este o linie ce unește punctele semnalului original. În schimb (c) și (d) prezintă un alt exemplu în care domeniul timp este determinat incorect. Oscilațiile prezentate în (d) apar la fronturi abrupte sau alte discontinuități din semnal. Ele se numesc efect Gibbs.

7.5 Înmulțirea Semnalelor (Modularea în Amplitudine).

O proprietate importantă a Transformatei Fourier este aceea că convoluția într-un domeniu corespunde cu înmulțirea în celălalt domeniu. Semnalele în domeniul timp pot fi convoluționate prin înmulțirea spectrelor lor. Modularea în amplitudine este un exemplu a situației inverse, înmulțirea în domeniul timp corespunde convoluției în domeniul frecvență. În plus, modularea în amplitudine prezintă un excelent exemplu despre cum frecvențele negative apar în problemele de știință și inginerie.

Semnalele audio sunt bune pentru combinația pe distanțe lungi; când vorbim, cineva care se află în aceeași încăpere ne va auzi. Pe de altă parte, frecvențele radio sunt bune pentru comunicații pe distanțe lungi. Spre exemplu, dacă o sinusoidă cu amplitudinea de 100 volți și frecvența 1MHz este trecută printr-o antenă, unda radio rezultantă poate fi detectată în camera vecină, țara vecină sau chiar planeta vecină. Modularea este procesul de combinare a două semnale pentru a forma un al treilea având caracteristicile ambelor semnale. Asta implică întotdeauna procese neliniare cum ar fi înmulțirea; nu putem aduna pur și simplu cele două semnale. În comunicațiile radio semnalele radio rezultate prin modulare se pot propaga pe distanțe lungi și pot duce cu ele semnale audio sau orice altă informație.

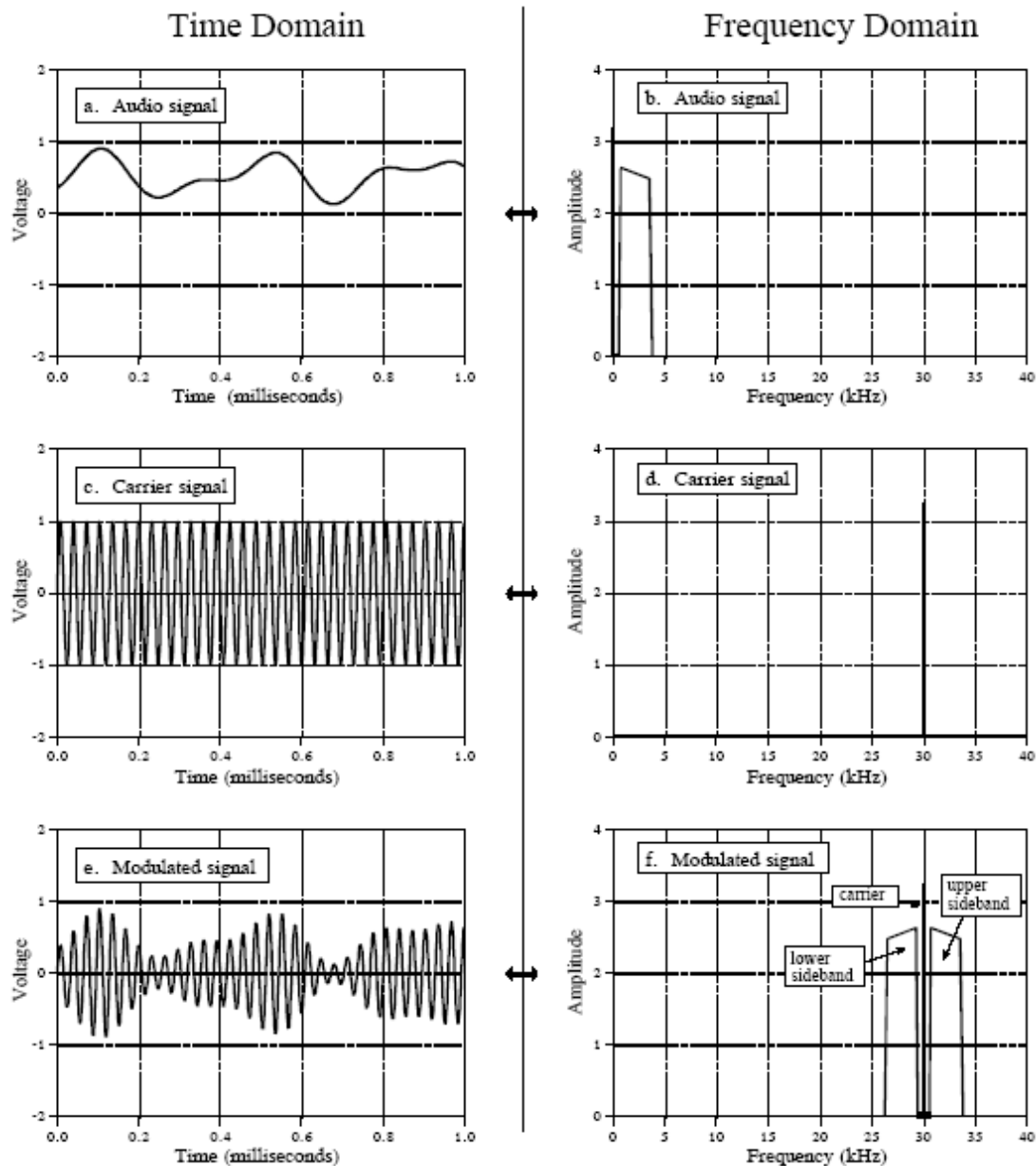


Fig. 7-14

Comunicația radio este o disciplină foarte dezvoltată fiind concepute multe scheme de modulare. Una din cele mai simple este numită modularea în amplitudine. Figura 7-14 prezintă un exemplu al felului cum modularea în amplitudine apare în domeniul timp și frecvență. Se vor folosi semnale continue în acest exemplu, datorat faptului că modularea este o aplicație folosită în circuitele analogice. Oricum, întreaga procedură poate fi transpusă și în forma discretă.

Figura (a) prezintă un semnal audio ce prezintă un offset astfel încât semnalul prezintă numai valori pozitive. Figura (b) descrie spectrul de frecvențe al semnalului, compus din frecvențe de la 300Hz la 3kHz, gama vocală, plus componenta DC. Celelalte frecvențe au fost scoase prin filtrare analogică. Figura (c) și (d) prezintă unda purtătoare, o sinusoidă

pură cu frecvența mult mai mare decât cea a semnalului audio. În domeniul timp, modularea în amplitudine constă din înmulțirea semnalului audio cu purtătoarea. Rezultatul, prezentat în (e) o formă de undă oscilatorie cu amplitudinea instantanee proporțională cu semnalul audio original. În jargon, anvelopa undei purtătoare este egală cu semnalul modulator. Acest semnal poate fi trecut printr-o antenă, convertit într-o undă radio și apoi detectat de o antenă de recepție. Rezultatul este un semnal identic cu (e) fiind generat de partea electronică a receptorului radio. Un circuit detector sau demodulator este apoi folosit pentru a converti semnalul din (e) în cel din (a).

În timp ce semnalele din domeniul timp sunt înmulțite, spectrul de frecvențe corespunzător este convoluționat. Asta înseamnă că (f) este determinat prin convoluția lui (b) cu (d). În timp ce spectrul purtătoarei este o funcție delta deplasată, spectrul semnalului modulat este egal cu spectrul audio deplasat cu frecvența purtătoarei. Rezultatul, spectrul modulat, este compus din trei componente: purtătoarea, banda superioară și banda inferioară.

Acestea corespund celor trei părți ale semnalului audio original: componenta DC, frecvențele pozitive cuprinse între 0,3 și 3kHz și respectiv frecvențele negative cuprinse între -0,3 și -3kHz. Deși frecvențele negative în semnalul audio original sunt ceva abstract frecvențele rezultate în banda inferioară sunt foarte reale.

Inginerii în comunicații trăiesc din analiza acestui tip de domeniu frecvență. Spre exemplu, să considerăm spectrul de frecvență pentru transmisia folosită în televiziune. Un semnal standard TV are un spectru de frecvențe cuprins între DC și 6MHz. Prin folosirea acestor tehnici de deplasare în frecvență, 82 de canale cu lățimea de 6MHz sunt așezate una lângă alta. Spre exemplu, canalul 3 este de la 60 la 66MHz, canalul 4 de la 66 la 72MHz, canalul 83 de la 884 la 890 MHz etc. Receptorul TV mută canalul dorit înapoi în banda DC - 6MHz pentru a-l afișa pe ecran. Această schemă este numită, multiplexare în domeniul frecvență.

7.6 Transformata Fourier Discretă în Timp.

Transformata Fourier Discretă în Timp (DTFT) este membră a familiei transformatei Fourier care operează cu semnale discrete și aperiodice. Cel mai ușor mod de a înțelege DTFT-ul este de a găsi legătura lui cu DFT-ul. Pentru început să ne imaginăm că am cules un semnal de N eșantioane și dorim să-i determinăm spectrul său. Prin folosirea DFT-ului semnalul poate fi descompus în $N/2 + 1$ sinusoidale și cosinusoidale cu frecvențe egal distanțate între zero și jumătate din rata de eșantioane. Așa cum am văzut, “umplerea” semnalului în domeniul timp cu zerouri, face perioada domeniului timp mai lungă cât și distanța între eșantioanele domeniului frecvență mai mică. Dacă N tinde către infinit, domeniul timp devine aperiodic și domeniul frecvență devine un semnal continuu. Acesta este DTFT-ul, Transformata Fourier care leagă un semnal discret și aperiodic cu unul continuu și periodic în spectrul frecvență.

$$\begin{aligned} \operatorname{Re} X(\omega) &= \sum_{n=-\infty}^{+\infty} x[n] \cos(\omega n) \\ \operatorname{Im} X(\omega) &= - \sum_{n=-\infty}^{+\infty} x[n] \sin(\omega n) \end{aligned}$$

Ecuatia 7-1

$$x[n] = \frac{1}{\pi} \int_0^{\pi} \operatorname{Re} X(\omega) \cos(\omega n) - \operatorname{Im} X(\omega) \sin(\omega n) d\omega$$

Ecuatia 7-2

Matematica DTFT-ului poate fi înțeleasă prin studierea ecuațiilor de sinteză și analiză ale DFT-ului (Ec. 5-2, 5-3 și 5-4) și făcând N tinzând la infinit. Există câteva detalii subtile în aceste relații. Primul, semnalul în domeniul timp, $x[n]$ este încă discret și de aceea este reprezentat prin paranteze pătrate. În schimb, semnalele în domeniul frecvență, $\operatorname{Re} X(m)$ și $\operatorname{Im} X(m)$, sunt conținute, fiind scrise cu paranteze rotunde. Din moment ce domeniul frecvență este continuu, ecuația de sinteză trebuie scrisă ca o integrală în locul sumei.

Așa cum am discutat, frecvența este reprezentată în domeniul frecvență al DFT-ului de una din cele trei variabile: k , un întreg ce are valoarea între 0 și $N/2$; f , o fracție din rata de eșantioane, de la 0 la 0,5; sau ω o fracție din rata de eșantionare exprimată ca o frecvență naturală, de la 0 la π . Spectrul DTFT-ului este continuu așa că poate fi folosit fie f fie ω . Mai uzual este ω , deoarece face ecuațiile mai scurte prin eliminarea factorului 2π . Să ne amintim că atunci când este folosit ω , spectrul de frecvențe este cuprins între 0 și π , ce corespunde lui DC la jumătate din rata de eșantionare.

Când calculăm DFT-ul invers, eșantioanele 0 și $N/2$ trebuie împărțite prin doi (Ec. 5-3) înainte de a trece la ecuația de sinteză (Ec. 5-2). Aceasta nu este necesară în DTFT. Așa cum ne amintim, aceasta acțiune în DFT are legătură cu spectrul de frecvență ce a fost definit ca o densitate spectrală, amplitudine pe unitatea lățimii de bandă. Când spectrul devine continuu, tratamentul special al capetelor dispare. Factorul de normalizare care trebuie inclus în DFT, $2/N$ (Ec. 5-3) devine $1/\pi$ în DTFT (Ec. 7-2). Unii autori plasează acești termeni în fața ecuației de sinteză, în timp ce alții în plasează în fața ecuației de analiză. Să presupunem că începem cu un semnal în domeniul timp. După ce îl trecem printr-o transformată Fourier și apoi printr-o transformată Fourier inversă, dorim să ajungem la semnalul pe care l-am avut la început. Asta înseamnă că termenul $1/\pi$ (sau $2/N$) trebuie să apară undeva în toată această procedură, în transformata directă sau cea inversă. Unii autori chiar împart acest termen între cele două ecuații prin plasarea lui $1/\sqrt{\pi}$ la începutul ambelor ecuații.

Din moment de DTFT-ul implică sumări și integrări infinite, ea nu poate fi calculată cu ajutorul calculatoarelor digitale. Este folosită în problemele teoretice înlocuind DFT-ul. Spre exemplu, să presupunem că dorim să găsim răspunsul în frecvență al unui sistem din răspunsul său la impuls. Dacă răspunsul său la impuls arată ca un șir de numere, obținut dintr-o măsurătoare experimentală, sau o simulare pe calculator, un program DFT ne poate rezolva problema. El determină spectrul de frecvențe ca fiind un alt vector de numere, egal distanțate între 0 și 0,5 din rata de eșantionare.

În alt caz, răspunsul la impuls poate fi cunoscut ca fiind o ecuație cum ar fi o funcție sinc sau o sinusoidă a cărei amplitudine scade exponențial. Aici este folosit DTFT-ul pentru calcularea matematică a domeniului frecvență ca fiind o altă ecuație, specificând întreaga curbă continuă între 0 și 0,5. DFT-ul poate fi folosit și el în acest calcul, el determinând o ecuație pentru eșantioanele răspunsului la frecvență, și nu pentru întreaga curbă.

CAP 8. Transformata Fourier Rapidă

Există mai multe moduri de a calcula Transformata Fourier Discretă (DFT), cum ar fi rezolvarea unui sistem de ecuații liniare, sau prin aplicarea corelației descrise într-un capitol anterior. Transformata Fourier Rapidă FFT este o altă metodă pentru a calcula DFT-ul. Ea produce aceleași rezultate ca și celelalte metode dar este mult mai eficientă, adesea reducând timpul de calcul de sute de ori. Dacă FFT-ul nu ar fi fost descoperit multe tehnici în DSP nu ar fi fost practice. Chiar dacă FFT-ul conține numai câteva zeci de linii de cod, este unul din cei mai complicați algoritmi în DSP.

8.1 Descrierea algoritmului FFT

FFT-ul este un algoritm complicat iar detaliile sunt de obicei lăsate pe seama acelor specializați în asemenea lucruri. Acest capitol descrie modul de operare al FFT-ului. Va trebui să ne amintim câte ceva despre matematica numerelor complexe.

În notația complexă, fiecare din domeniile timp și frecvență conțin un semnal alcătuit din N puncte complexe. Fiecare din aceste numere complexe este alcătuit la rândul său din două numere, partea reală și partea imaginară. Spre exemplu, atunci când ne referim la eșantionul complex $X[42]$, vorbim de fapt de combinația dintre $ReX[42]$ și $ImX[42]$. Cu alte cuvinte, fiecare variabilă complexă reține două numere. Când două variabile complexe sunt înmulțite, trebuie combinate patru componente individuale, pentru a forma cele două componente ale produsului. În continuare, fiecare termen ca: semnal, punct, eșantion sau valoare va face referire la combinația dintre partea reală și partea imaginară.

FFT-ul operează prin descompunerea unui semnal de N puncte în domeniul timp în N semnale în domeniul timp, fiecare compus dintr-un singur punct. Al doilea pas este acela N spectre de frecvență corespunzătoare celor N semnale în domeniu timp. În final, cele N spectre sunt sintetizate într-un singur spectru de frecvență.

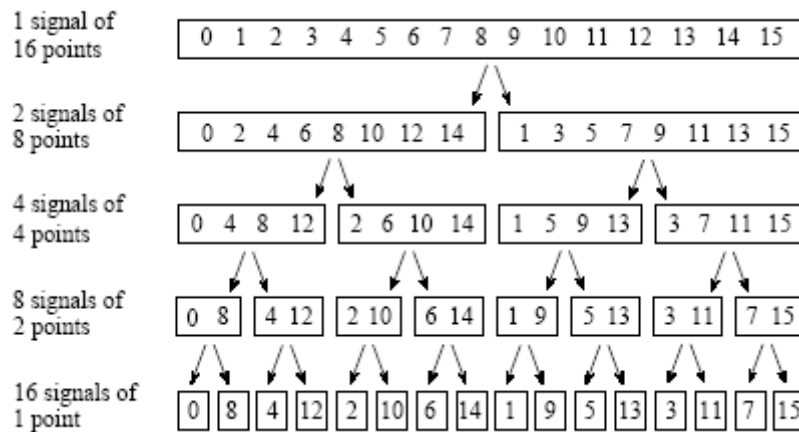


Fig. 8-1

Figura 8-1 prezintă un exemplu de descompunere în domeniu timp folosită în FFT. În acest exemplu, un semnal de 16 puncte este descompus pe parcursul a patru pași. Primul pas, împarte semnalul de 16 puncte în două semnale, fiecare alcătuit din câte 8 puncte. În al doilea pas se face o descompunere în patru semnale de câte 4 puncte. Acest algoritm continuă până când se obțin N semnale de câte un punct fiecare. Este folosită o descompunere întrețesută de fiecare dată când semnalul este descompus în două. Asta înseamnă că semnalul este separat în eșantioanele aflate pe locații pare și cele de pe locații impare, (Fig. 8-1). Este de nevoie de $\log_2 N$ pași pentru a rezolva acest algoritm.

Sample numbers in normal order		Sample numbers after bit reversal	
Decimal	Binary	Decimal	Binary
0	0000	0	0000
1	0001	8	1000
2	0010	4	0100
3	0011	12	1100
4	0100	2	0010
5	0101	10	1010
6	0110	6	0100
7	0111	14	1110
8	1000	1	0001
9	1001	9	1001
10	1010	5	0101
11	1011	13	1101
12	1100	3	0011
13	1101	11	1011
14	1110	7	0111
15	1111	15	1111

Fig. 8-2

Acum, că am văzut structura acestei descompuneri, vom vedea că ea poate fi mult simplificată. Descompunerea nu este nimic mai mult decât o reordonare a eșantioanelor în semnal. Figura 8-2 prezintă algoritmul folosit în această reordonare. În partea stângă, numărul eșantioanelor semnalului original este listat în ordinea crescătoare, cu

echivalenții lor binari. În partea dreaptă sunt listate numărul eșantioanelor rearanjate împreună cu echivalenții lor binari. Se observă că numerele binare sunt inversele celorlalte. Spre exemplu, eșantionul 3 (0011) este schimbat cu eșantionul 12 (1100). La fel, eșantionul cu numărul 14 (1110) este schimbat cu eșantionul 7 (0111) și așa mai departe. Descompunerea FFT în domeniul timp este generată de un algoritm de sortare prin inversarea biților. Asta implică modificarea ordinii celor N eșantioane ale domeniului timp prin această incrementare binară asociată cu o rotire a biților de la dreapta la stânga așa cum este prezentat în coloana cea mai din dreapta a figurii 8-2.

Următorul pas în algoritmul FFT este acela de a găsi spectrul de frecvențe pentru fiecare din semnalele de câte un singur punct în domeniul timp. Nimic nu poate fi mai ușor; spectrul de frecvențe al unui semnal de un singur punct este egal cu el însuși. Asta înseamnă că nu este nimic de făcut la acest pas. Totuși nu trebuie să uităm că, acum, semnalele de un punct sunt spectre de frecvență și nu semnale în domeniul timp.

Ultimul pas în FFT este acela de a combina cele N spectre de frecvență în ordinea inversă în care a avut loc descompunerea în domeniul timp. Acesta este punctul în care algoritmul creează probleme. Din păcate, metoda inversării pe bit nu este aplicabilă, trebuind să mergem pas cu pas până la refacerea întregului semnal. În primul pas, 16 spectre de frecvență (fiecare de câte un punct) sunt sintetizate în 8 spectre de frecvență (fiecare de câte două puncte). În al doilea pas, cele 8 spectre de frecvență sunt sintetizate în 4 spectre de frecvență (de câte 4 puncte fiecare) și așa mai departe. Ultimul pas va genera în ieșirea FFT-ului, un spectru de frecvență de 16 puncte.

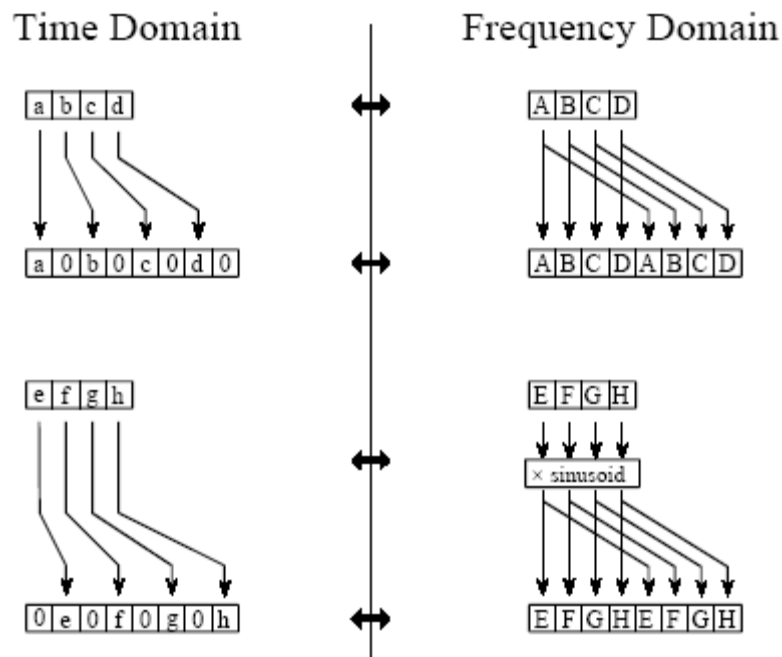


Fig. 8-3

Figura 8-3 prezintă modul în care două spectre de frecvențe, fiecare compuse din câte 4 puncte, sunt combinate într-un singur spectru de frecvențe de 8 puncte. Această sinteză

trebuie să refacă descompunerea întreșută făcută în domeniul timp. Cu alte cuvinte, operarea în domeniul frecvență trebuie să corespundă procedurii de combinare prin întreșere a două semnale de câte 4 puncte în domeniul timp. Considerăm două semnale în domeniul timp, $abcd$ și $fegh$. Un semnal de 8 puncte în domeniul timp poate fi format în doi pași: se introduce câte un 0 între cele patru valori ale fiecărui semnal obținându-se 2 semnale de câte 8 puncte și apoi se adună semnalele între ele. Asta înseamnă că $abcd$ devine $a0b0c0d0$ și $fegh$ devine $0e0f0g0h$. Adunând aceste două semnale de câte 8 puncte rezultă $aebfcgdh$. Așa cum este prezentat în figura 8-3, diluarea domeniul timp cu zerouri corespunde unei duplicări a spectrului de frecvențe. De aceea, spectrele de frecvențe sunt combinate în FFT prin duplicarea lor și apoi adunarea acestor spectre duplicate.

Pentru a realiza potrivirea acestor semnale la adunare ele sunt diluate cu zerouri într-un mod puțin diferit. Într-un semnal punctele impare sunt zero în timp ce în celălalt semnal punctele pare sunt zero. Cu alte cuvinte unul din semnalele din domeniul timp ($0e0f0g0h$ în figura 8-3) este deplasat la dreapta cu un eșantion. Această deplasare în domeniul timp corespunde cu înmulțirea spectrului cu o sinusoidă. Pentru a demonstra asta, să ne amintim că deplasarea în domeniul timp este echivalentă cu o convoluție a semnalului cu o funcție delta deplasată aceasta înmulțește spectrul semnalului cu spectrul unei funcții delta deplasată. Spectrul unei funcții delta deplasată este o sinusoidă (vezi fig. 8-4).

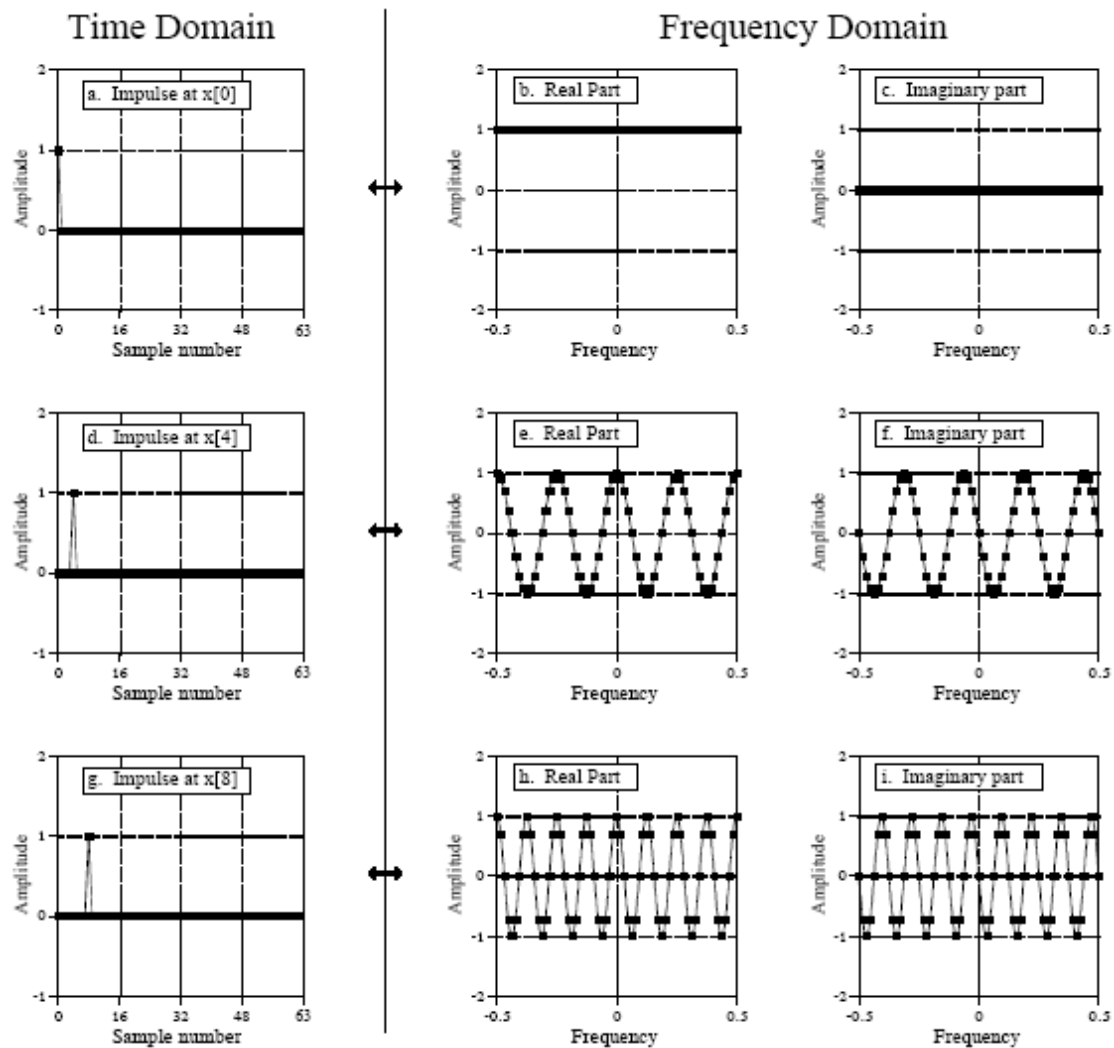


Fig. 8-4

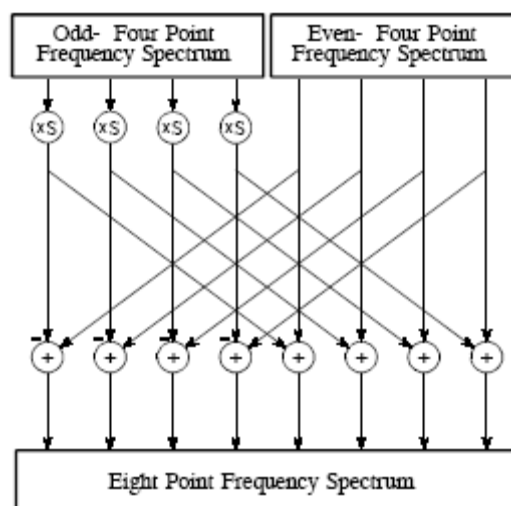


Fig. 8-5

Figura 8-5 prezintă diagrama pentru combinarea a două spectre de câte 4 puncte într-un singur spectru de 8 puncte. Figura 8-6 prezintă schema de bază a acestei operații, schema din figura 8-5 fiind obținută prin repetarea de patru ori a acesteia.

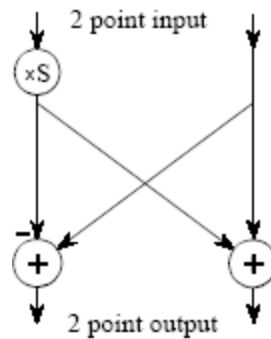


Fig. 8-6

Această diagramă simplificată este numită fluture (butterfly) datorită asemănării schemei cu două aripi deschise. Fluturele este elementul principal de calcul al FFT-ului, transformând două puncte complexe în alte două puncte complexe.

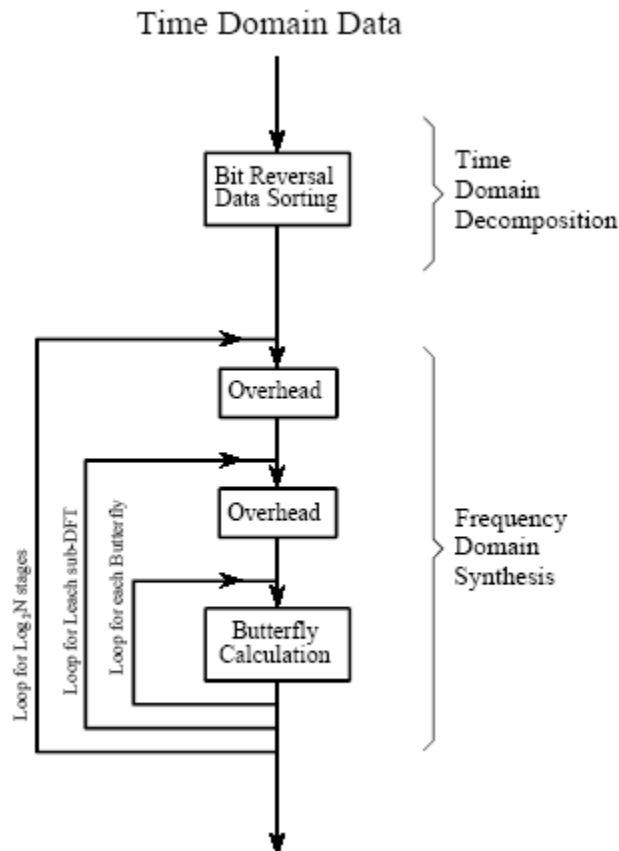


Fig. 8-7

Figura 8-7 prezintă structura întregului FFT. Descompunerea în domeniul timp este realizată cu un algoritm de sortare prin inversarea biților. Transformarea datelor descompuse din domeniul timp în domeniul frecvență nu implică nimic și de aceea nu apare în figură.

Sinteza domeniului frecvență necesită trei bucle. Buclea externă are $\log_2 N$ pași (fiecare nivel din figura 8-1 începând de jos și urcând spre nivelul superior). Buclea din mijloc trece prin fiecare spectru individual ce este procesat în momentul respectiv (prin fiecare căsuță a fiecărui nivel din figura 8-1). Buclea internă folosește fluturile pentru calculul punctelor în fiecare spectru de frecvență (trecând prin fiecare eșantion prin interiorul unei căsuțe). Secvențele “overhead” din figura 8-7 calculează indecșii de început și de sfârșit ale buclelor cât și sinusoidale folosite în fluturi.

8.2 Comparații de viteză și precizie.

Când DFT-ul este calculat prin corelație, programul folosește două bucle imbricate, fiecare de câte N puncte. Asta înseamnă că numărul total de operații este proporțional cu $N \times N$. De aceea timpul necesar pentru a executa un program este dat de relația:

$$TimpulDeExecuție = k_{DFT} \times N^2$$

unde N este numărul de puncte al DFT-ului și k_{DFT} este o constantă de proporționalitate. Dacă valorile sinusurilor și cosinusurilor sunt calculate în interiorul buclelor, k_{DFT} este egal cu aproximativ 25 de microsecunde pe un Pentium la 100MHz dacă aceste valori sunt precalculate și reținute într-un tabel k_{DFT} devine aproximativ 7 secunde. Spre exemplu, un DFT de 1024 de puncte va necesita aproximativ 25 de secunde sau aproape 25 de milisecunde pentru un punct, ceea ce este foarte mult.

Folosind aceeași strategie putem determina timpul de execuție pentru FFT. Timpul necesar pentru rutina de inversare a biților este neglijabil. În fiecare din cei $\log_2 N$ pași există $N/2$ fluturi de calculat:

$$TimpulDeExecuție = k_{FFT} \times N \times \log_2 N$$

Valoarea lui k_{FFT} este aproximativ 10 microsecunde pentru un sistem Pentium la 100MHz. Un FFT de 1024 de puncte necesită aproximativ 70 de milisecunde pentru a fi executat sau 70 microsecunde pentru calcularea unui punct. Aceasta este cu mai mult de 300 de ori mai rapidă decât DFT-ul calculat prin corelație.

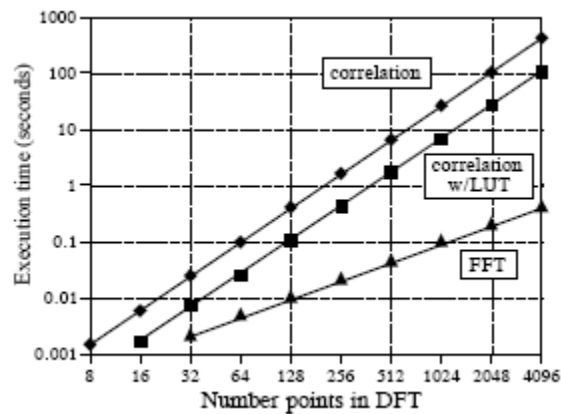


Fig. 8-8

Se observă că $N \times \log_2 N$ crește mult mai încet decât N^2 odată cu creșterea lui N . Spre exemplu un FFT de 32 de puncte este de aproximativ de 10 ori mai rapid decât metoda corelației. Un FFT de 4096 de puncte este de 1000 de ori mai rapid. Pentru valori mici ale lui N (să zicem între 32 și 128), FFT este important. Pentru valori mai mari ale lui N (mai sus de 1024 de puncte), FFT-ul devine critic. Figura 8-8 compară timpul de execuție a celor doi algoritmi.

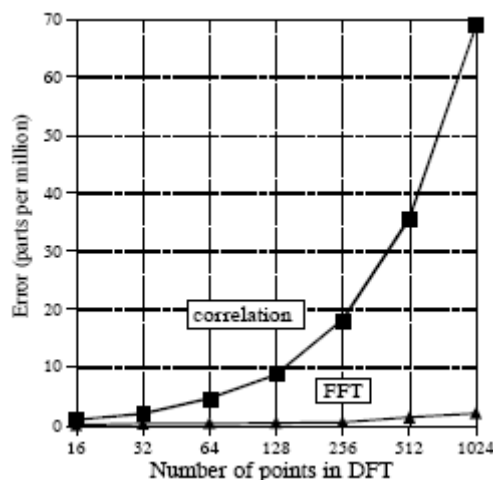


Fig. 8-9

FFT-ul mai prezintă și alte avantaje. FFT-ul este calculat mult mai precis deoarece are mai puține calcule matematice și mai puține erori rezultate din rotunjiri. Aceasta poate fi demonstrată prin execuția unui FFT asupra unui semnal arbitrar și apoi refacerea acestuia cu ajutorul unui FFT invers. Zgomotul rezultat din calcule matematice poate fi determinat prin calculul diferenței dintre cele două semnale. Pentru comparații, aceeași procedură poate fi repetată folosind un DFT calculat prin corelație și apoi un DFT invers corespunzător. Vom compara zgomotul generat de FFT cu cel generat de DFT. Vezi figura 8-9.

CAP 9. Filtre dedicate

Majoritatea filtrelor au unul din cele patru răspunsuri în frecvență standard: trece-jos, trece-sus, trece-bandă, stop-bandă. Acest capitol prezintă o metodă generală de proiectare a filtrelor digitale cu un răspuns în frecvență arbitrar, dedicat anumitor aplicații. În acest domeniu DSP excelează, rezolvând probleme ce nu fac față circuitelor analogice. Vom discuta despre două utilizări importante ale filtrelor dedicate: *deconvoluția*, o metodă de refacere a semnalelor ce au suferit convoluții nedorite, și *filtre optimale*, problema separării semnalelor cu suprapunerea spectrului de frecvențe.

9.1 Răspunsul în Frecvență Arbitrar

Să presupunem că dorim să producem un filtru care să genereze răspunsul în frecvență de forma 9-1(a). Se observă că acesta este foarte neregulat și aproape imposibil de obținut cu filtre analogice. Acest răspuns în frecvență, ideal, este definit ca un șir de numere și nu ca niște ecuații. În acest exemplu, sunt 513 eșantioane repartizate între 0 și 0,5 din frecvența de eșantionare. Pentru o reprezentare mai bună a răspunsului în frecvență se folosesc mai multe puncte, în timp ce dacă se folosesc mai puține puncte se reduce timpul de calcul. De obicei se folosesc 513 puncte pentru majoritatea aplicațiilor.

Pe lângă vectorul amplitudine, prezentat în figura (a), mai există și un vector fază nereprezentat în figură, acesta fiind în întregime zero. La fel răspunsul în frecvență poate fi reprezentat prin specificarea părții reale și imaginare, în locul amplitudinii și fazei.

Următorul pas este acela de a aplica un DFT invers asupra răspunsului în frecvență pentru a-l trece în domeniul timp. Cea mai rapidă metodă este aceea de a converti domeniul frecvență în forma rectangulară și apoi aplicat FFT-ul invers. Acesta are ca rezultat un semnal compus din 1024 de eșantioane de la 0 la 1023, prezentat în (b). Acesta este răspunsul la impuls corespunzător răspunsului în frecvență dorit. Deocamdată el nu poate fi folosit. Trebuie deplasat, trunchiat și ferestruit. În acest exemplu s-a ales $M = 40$, adică 41 de puncte, de la 0 la 40.

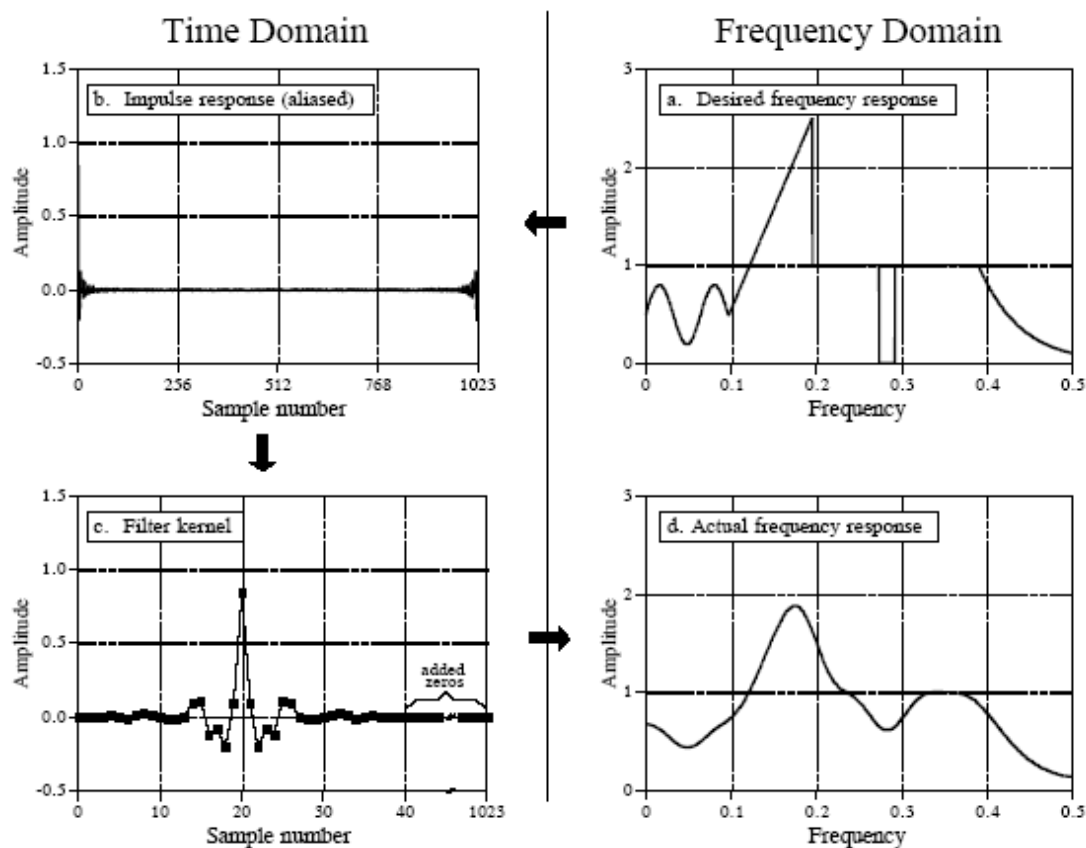


Fig. 9-1

Ultimul pas este acela de a testa nucleul de filtru obținut. Prin folosirea FFT-ului se obține răspunsul în frecvență prezentat în (d). Pentru a obține o rezoluție mai bună se "umple" nucleul de filtru cu zerouri înainte de aplicarea FFT-ului. Spre exemplu, folosind 1024 de puncte (41 din nucleul de filtru, plus 983 zerouri), rezultă 513 eșantioane între 0 și 0,5.

Așa cum se poate vedea în figura 9-2, lungimea nucleului determină o mai bună potrivire a răspunsului în frecvență obținut cu cel dorit. Teoretic, orice răspuns în frecvență poate fi obținut dacă se folosește un nucleu de filtru destul de lung.

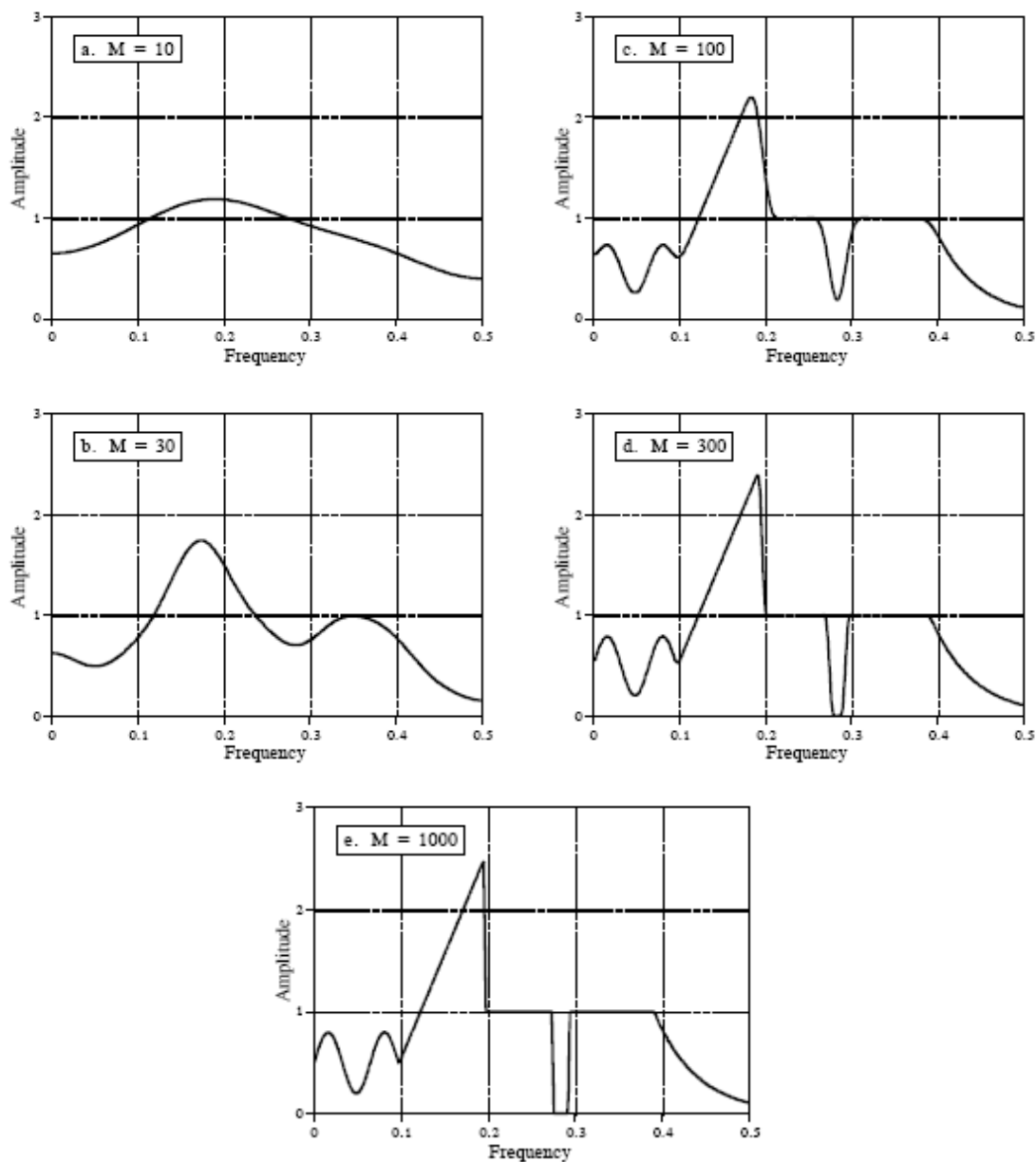


Fig. 9-2

Răspunsul la impuls din figura 9-1b nu poate fi folosit direct ca și nucleu de filtru. Dacă (a) este transformata Fourier a lui (b), convoluționând un semnal de intrare cu (b), nu vom produce răspunsul în frecvență dorit. El este supus fenomenului de aliasing. Pentru a elimina acest neajuns și a putea forma nucleul de filtru (c), acest impuls la răspuns este trunchiat la $M+1$ eșantioane, deplasate cu $M/2$ eșantioane la dreapta, și înmulțite cu o fereastră Hamming sau Blackman. În acest exemplu $M = 40$. Nucleul de filtru este testat prin umplerea cu zerouri a lui (c) până la 1024 de puncte, obținându-se astfel (d).

9.2 Deconvoluția

Convoluțiile nedorite sunt probleme inerente la transferul analog al informațiilor. Spre exemplu: imaginile neclare, ecourile apărute pe liniile telefonice pe distanțe lungi, limitările de bandă apărute la senzorii analogici etc. Deconvoluția este procesul de filtrare a semnalului pentru a elimina convoluțiile nedorite. Scopul deconvoluției este acela de a recrea semnalul așa cum era înainte de a avea loc convoluția. Acest lucru necesită a se ști caracteristicile convoluției (răspunsul la impuls sau răspunsul în frecvență), ceea ce îl face diferit de deconvoluția oarbă, unde caracteristicile convoluției parazite nu erau știute. Deconvoluția oarbă este o problemă mult mai dificilă, și de obicei nu are o soluție generală, particulară aplicației.

Deconvoluția este aproape imposibilă de înțeles în domeniul timp, dar mai simplă în domeniul frecvență. Fiecare sinusoidă ce compune semnalul original poate fi modificată în amplitudine și/sau fază de către o convoluție nedorită. Pentru a extrage semnalul original, filtrul pentru deconvoluție trebuie să refacă aceste amplitudini sau faze.

Pentru a ilustra deconvoluția vom folosi un detector de raze gama (figura 9-3). Acest dispozitiv este compus din două părți, un scintilator și un detector de lumină. Un scintilator este un material transparent special, cum ar fi iodura de sodiu sau germanați de bismut. Aceștia modifică energia radiațiilor gama în mici flash-uri de lumină. Lumina este convertită în semnal electric de către un senzor, cum ar fi o fotodiodă. Din moment ce fiecare rază gama este un impuls, ieșirea din detector este de fapt răspunsul la impuls a sistemului.

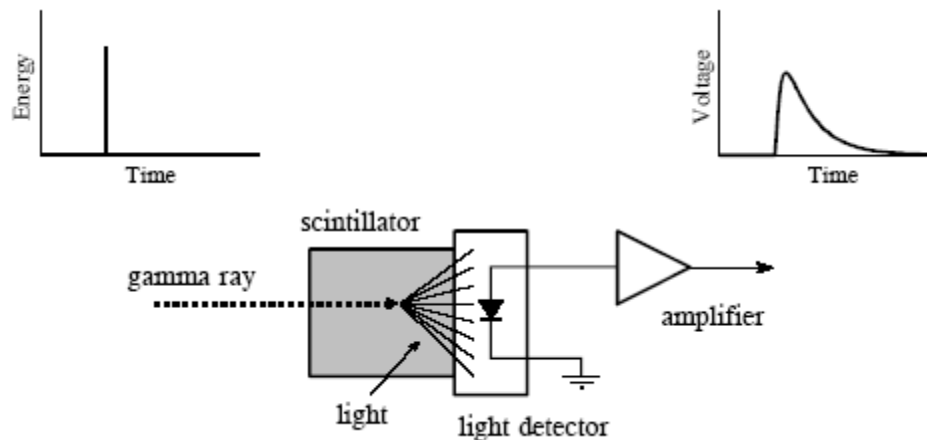


Fig. 9-3

Figura 9-4a prezintă un exemplu de impulsuri generate de detector, ca răspuns la razele gama ce sosesc aleator. Informația ce trebuie extrasă din acest semnal de ieșire este amplitudinea fiecărui impuls, ce este proporțională cu energia razei gama ce a generat-o. Astfel se pot determina informații medicale despre un anumit pacient, se poate determina vârsta galaxiilor îndepărtate, se pot detecta bombe în camera de bagaje a avioanelor etc.

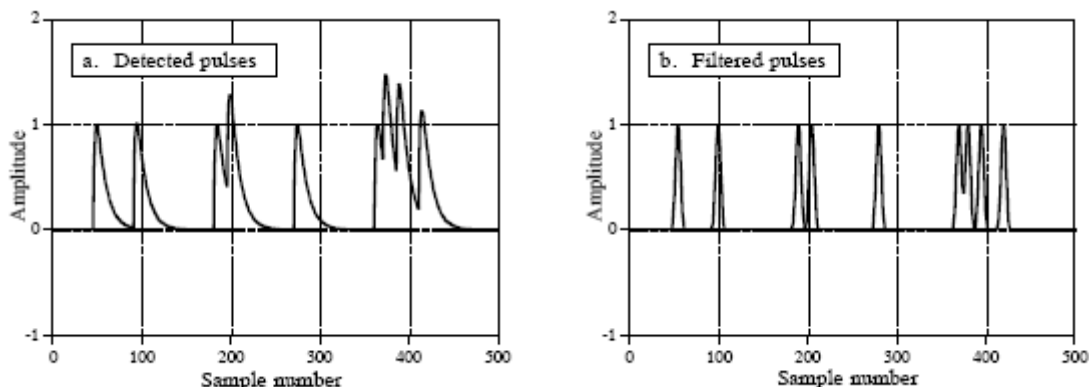


Fig. 9-4

După cum se poate observa în (a), două sau mai multe impulsuri se pot suprapune, mărind astfel amplitudinea semnalului măsurat. Acest lucru nu este prea indicat. Pentru a rezolva această problemă se aplică deconvoluția asupra semnalului de ieșire. Acest lucru face impulsurile mai "ascuțite" și deci nu se vor mai suprapune.

Chiar dacă semnalul este codat în domeniul timp, majoritatea analizei se va face în domeniul unde problema este mai ușor de înțeles. Figura 9-5a este semnalul produs de detector (acest lucru se știe). Figura (c) este semnalul pe care dorim să-l obținem (de asemenea, acest lucru se știe). Acest impuls a fost ales arbitrar, având forma unei ferestre Blackman, cu o lungime aproximativ egală cu o treime din lungimea impulsului produs de detector (a). Scopul este acela de a găsi un nucleu de filtru care convoluționat cu semnalul original (a), să producă semnalul din (c). Ca și ecuație: $a * e = c$, se cunosc a și c și trebuie determinat e .

Deconvoluția nu este chiar operația inversă convoluției (cel puțin în domeniul timp).

Din fericire, această problemă se simplifică în domeniul frecvență. Convoluția într-un domeniu, corespunde cu înmulțirea în celălalt domeniu. În figura 9-5, dacă $b \times f = d$, fiind date b și d , trebuie determinat f . Răspunsul în frecvență a filtrului (f), este spectrul frecvență a impulsului căutat (d), împărțit la spectrul frecvență al impulsului detectat (b).

Desigur, ideal ar fi ca impulsurile obținute (c), să fie cât mai "subțiri". Totuși apare o problemă. Dacă am face aceste impulsuri foarte subțiri, răspunsul în frecvență (f) a acestora va trebui să amplifice semnalele cu frecvență mare. Problema este că și zgomotul se situează undeva la frecvențe mari. Deci trebuie făcut un compromis între lățimea semnalului util și amplificarea de zgomot.

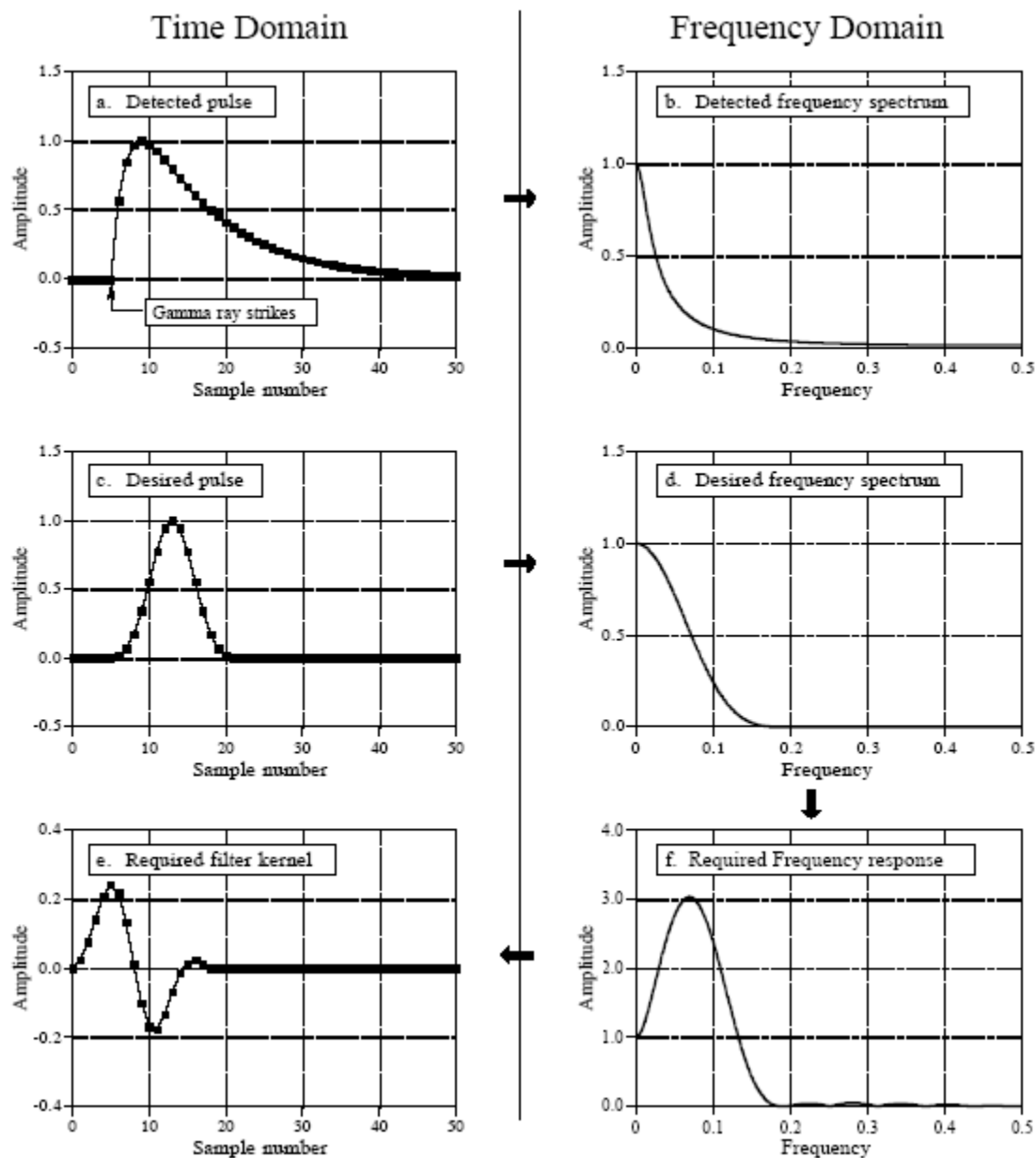


Fig. 9-5

Deconvoluția este de asemenea analiza semnalelor codate în domeniul frecvență. Un exemplu clasic este restaurarea înregistrărilor vechi (opera cântăreților Enrico Caruso 1873-1921). Figura 9-6 prezintă o descriere generală. Problema apare la înregistrarea semnalului. Datorită echipamentelor primitive, apare o rezonanță a sunetului în conul ce-l ghida pe acesta înspre microfon. Spectrul de frecvență a semnalului audio original este prezentat în figura (a). Figura (b) arată răspunsul în frecvență a echipamentului de înregistrare. Se observă câteva peak-uri datorită rezonanțelor. Spectrul semnalului înregistrat este prezentat în (c) și este egal cu spectrul original (a), înmulțit cu răspunsul în frecvență din (b). Scopul deconvoluției este acela de a contracara convoluția nedorită. Cu alte cuvinte, răspunsul în frecvență a filtrului de deconvoluție (d), trebuie să fie inversul lui (b). Asta înseamnă că fiecare vârf din (b) trebuie anulat de un vârf în jos în (d). Dacă acest filtru este proiectat corespunzător, semnalul rezultat va avea spectrul (e) identic cu cel original. Problema este că răspunsul în frecvență a echipamentului de înregistrare (b) nu poate fi știut. Cu alte cuvinte aceasta este o problemă de deconvoluție "oarbă". Se dă (c)-ul și trebuie determinat (d)-ul.

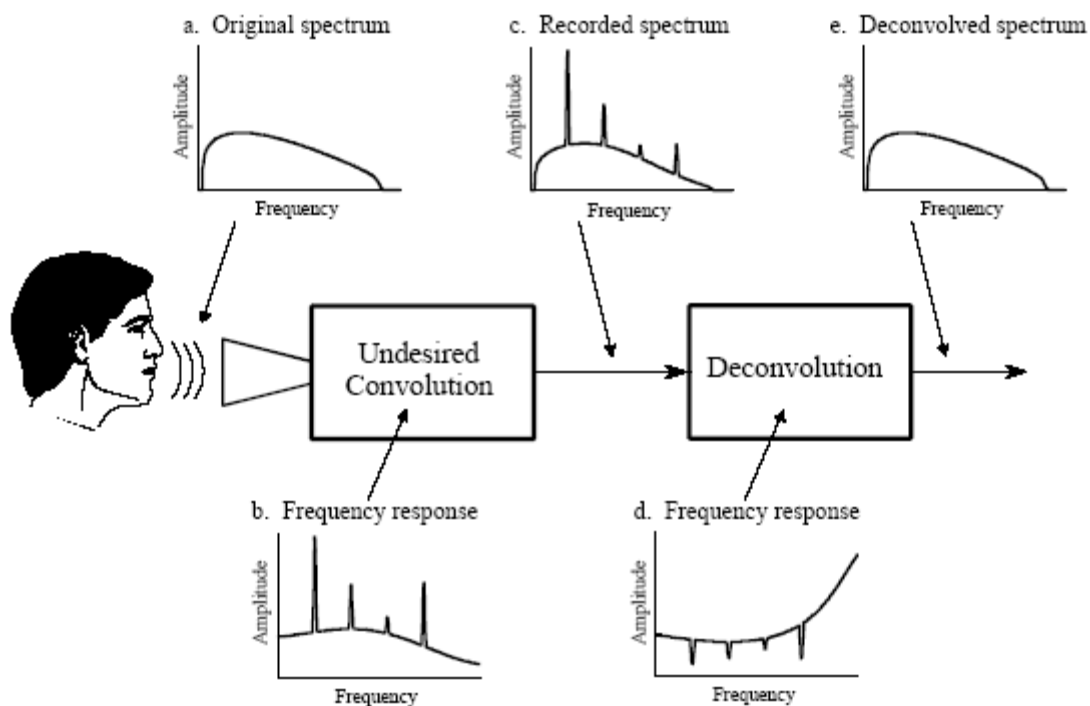


Fig. 9-6

Problemele de deconvoluție oarbă sunt legate de estimări sau presupuneri asupra unor parametri necunoscuți. Legat de acest exemplu, spectrul mediu a muzicii originale este determinat dintr-un spectru mediu a aceleași piese muzicale realizate cu instrumentele actuale. Simplist, răspunsul în frecvență necunoscut este luat ca o medie spectrală a vechii înregistrări, împărțită la media spectrală a noii înregistrări. Această metodă este mult mai complicată și se numește procesare *homomorfică*.