Fast Fourier Transform

Key Papers in Computer Science
Seminar 2005

Dima Batenkov

Weizmann Institute of Science

dima.batenkov@gmail.com

Fast Fourier Transform - Overview

J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19:297–301, 1965

Fast Fourier Transform - Overview

J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19:297–301, 1965

- A fast algorithm for computing the Discrete Fourier Transform

Fast Fourier Transform - Overview

» Fast Fourier Transform - Overview

J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19:297–301, 1965

- A fast algorithm for computing the Discrete Fourier Transform
- (Re)discovered by Cooley & Tukey in 1965[1] and widely adopted thereafter

Fast Fourier Transform - Overview

J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19:297–301, 1965

- A fast algorithm for computing the Discrete Fourier Transform
- (Re)discovered by Cooley & Tukey in 1965[1] and widely adopted thereafter
- Has a long and fascinating history

Fourier Analysis

Fourier Series

- Expresses a (real) periodic function $x(t)$ as a sum of trigonometric series $(-L < t < L)$

$$x(t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n \cos\frac{\pi n}{L}t + b_n \sin\frac{\pi n}{L}t)$$

- Coefficients can be computed by

$$a_n = \frac{1}{L}\int_{-L}^{L} x(t)\cos\frac{\pi n}{L}t\,dt$$
$$b_n = \frac{1}{L}\int_{-L}^{L} x(t)\sin\frac{\pi n}{L}t\,dt$$

Fourier Series

- Generalized to complex-valued functions as

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{i\frac{\pi n}{L}t}$$

$$c_n = \frac{1}{2L}\int_{-L}^{L} x(t) e^{-i\frac{\pi n}{L}t}\, dt$$

Fourier Series

- Generalized to complex-valued functions as

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{i\frac{\pi n}{L}t}$$

$$c_n = \frac{1}{2L}\int_{-L}^{L} x(t) e^{-i\frac{\pi n}{L}t}\,dt$$

- Studied by D.Bernoulli and L.Euler
- Used by Fourier to solve the heat equation

Fourier Series

- Generalized to complex-valued functions as

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{i\frac{\pi n}{L}t}$$

$$c_n = \frac{1}{2L}\int_{-L}^{L} x(t) e^{-i\frac{\pi n}{L}t}\, dt$$

- Studied by D.Bernoulli and L.Euler
- Used by Fourier to solve the heat equation
- Converges for almost all “nice” functions (piecewise smooth, L^2 etc.)

Continuous Fourier Transform

- Generalization of Fourier Series for infinite domains

$$x(t) = \int_{-\infty}^{\infty} \mathcal{F}(f) e^{-2\pi i f t} \, df$$

$$\mathcal{F}(f) = \int_{-\infty}^{\infty} x(t) e^{2\pi i f t} \, dt$$

- Can represent continuous, aperiodic signals
- Continuous frequency spectrum

Discrete Fourier Transform

- If the signal $X(k)$ is periodic, band-limited and sampled at Nyquist frequency or higher, the DFT represents the CFT exactly [14]

$$A(r) = \sum_{k=0}^{N-1} X(k)\mathbf{W}_N^{rk}$$

where $\mathbf{W}_N = e^{-\frac{2\pi i}{N}}$ and $r = 0, 1, \ldots, N-1$

- The inverse transform:

$$X(j) = \frac{1}{N}\sum_{k=0}^{N-1} A(k)\mathbf{W}_N^{-jk}$$

Useful properties[6]

- Orthogonality of basis functions

$$\sum_{k=0}^{N-1} \mathbf{W}_N^{r'k} \mathbf{W}_N^{rk} = N\delta_N(r - r')$$

- Linearity
- (Cyclic) Convolution theorem

$$(x * y)_n = \sum_{k=0}^{N-1} x(k) y(n-k)$$
$$\Rightarrow \mathcal{F}\{(x * y)\} = \mathcal{F}\{X\}\mathcal{F}\{Y\}$$

- Symmetries: real $\leftrightarrow$ hermitian symmetric, imaginary $\leftrightarrow$ hermitian antisymmetric
- Shifting theorems

Applications

- Approximation by trigonometric polynomials
  - Data compression (MP3...)
- Spectral analysis of signals
- Frequency response of systems
- Partial Differential Equations
- Convolution via frequency domain
  - Cross-correlation
  - Multiplication of large integers
  - Symbolic multiplication of polynomials

Methods known by 1965

Available methods

- In many applications, large digitized datasets are becoming available, but cannot be processed due to prohibitive running time of DFT
- All (publicly known) methods for efficient computation utilize the symmetry of trigonometric functions, but still are $O(N^2)$
- Goertzel's method [7] is fastest known

Goertzel's algorithm[7]

- Given a particular $0 < j < N-1$, we want to evaluate

$$A(j) = \sum_{k=0}^{N-1} x(k) \cos\left(\frac{2\pi j}{N} k\right) \quad B(j) = \sum_{k=0}^{N-1} x(k) \sin\left(\frac{2\pi j}{N} k\right)$$

Goertzel's algorithm[7]

- Given a particular $0 < j < N-1$, we want to evaluate

$$A(j) = \sum_{k=0}^{N-1} x(k) \cos\left(\frac{2\pi j}{N} k\right) \quad B(j) = \sum_{k=0}^{N-1} x(k) \sin\left(\frac{2\pi j}{N} k\right)$$

- Recursively compute the sequence $\{u(s)\}_{s=0}^{N+1}$

$$u(s) = x(s) + 2\cos\left(\frac{2\pi j}{N}\right) u(s+1) - u(s+2)$$
$$u(N) = u(N+1) = 0$$

Goertzel's algorithm [7]

- Given a particular $0 < j < N-1$, we want to evaluate

$$A(j) = \sum_{k=0}^{N-1} x(k)\cos\left(\frac{2\pi j}{N}k\right) \quad B(j) = \sum_{k=0}^{N-1} x(k)\sin\left(\frac{2\pi j}{N}k\right)$$

- Recursively compute the sequence $\{u(s)\}_{s=0}^{N+1}$

$$u(s) = x(s) + 2\cos\left(\frac{2\pi j}{N}\right)u(s+1) - u(s+2)$$
$$u(N) = u(N+1) = 0$$

- Then

$$B(j) = u(1)\sin\left(\frac{2\pi j}{N}\right)$$
$$A(j) = x(0) + \cos\left(\frac{2\pi j}{N}\right)u(1) - u(2)$$

Goertzel's algorithm [7]

- Requires N multiplications and only one sine and cosine
- Roundoff errors grow rapidly [5]
- Excellent for computing a very small number of coefficients ($< \log N$)
- Can be implemented using only 2 intermediate storage locations and processing input signal as it arrives
- Used today for Dual-Tone Multi-Frequency detection (tone dialing)

Inspiration for the FFT

Factorial experiments

Purpose: investigate the effect of n factors on some measured quantity

- Each factor can have t distinct levels. For our discussion let $t = 2$, so there are 2 levels: "$+$" and "$-$".
- So we conduct t^n experiments (for every possible combination of the n factors) and calculate:
  - *Main effect* of a factor: what is the average change in the output as the factor goes from "$-$" to "$+$"?
  - *Interaction effects* between two or more factors: what is the difference in output between the case when both factors are present compared to when only one of them is present?

Factorial experiments

Example of $t = 2$, $n = 3$ experiment

- Say we have factors A, B, C which can be be "high" or "low". For example, they can represent levels of 3 different drugs given to patients
- Perform 2^3 measurements of some quantity, say blood pressure
- Investigate how each one of the drugs and their combination affect the blood pressure

Factorial experiments

Exp.	Combination	A	B	C	Blood pressure
1	(1)	$-$	$-$	$-$	x_1
2	a	$+$	$-$	$-$	x_2
3	b	$-$	$+$	$-$	x_3
4	ab	$+$	$+$	$-$	x_4
5	c	$-$	$-$	$+$	x_5
6	ac	$+$	$-$	$+$	x_6
7	bc	$-$	$+$	$+$	x_7
8	abc	$+$	$+$	$+$	x_8

Factorial experiments

- The main effect of A is $(ab - b) + (a - (1)) + (abc - bc) + (ac - c)$
- The interaction of A, B is $(ab - b) - (a - (1)) - ((abc - bc) - (ac - c))$
- etc.

Factorial experiments

- The main effect of A is
$(ab - b) + (a - (1)) + (abc - bc) + (ac - c)$
- The interaction of A, B is
$(ab - b) - (a - (1)) - ((abc - bc) - (ac - c))$
- etc.

If the main effects and interactions are arranged in vector $\mathbf{y}$, then we can write $\mathbf{y} = A\mathbf{x}$ where

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix}$$

Factorial experiments

- F.Yates devised[15] an iterative n-step algorithm which simplifies the calculation. Equivalent to decomposing $A = B^n$, where (for our case of $n = 3$):

$$B = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 \end{pmatrix}$$

so that much less operations are required

Good's paper [9]

I. J. Good. The interaction algorithm and practical Fourier analysis. *Journal Roy. Stat. Soc.*, 20:361–372, 1958

- Essentially develops a "prime-factor" FFT
- The idea is inspired by Yates' efficient algorithm, but Good proves general theorems which can be applied to the task of efficiently multiplying a N-vector by certain $N \times N$ matrices
- The work remains unnoticed until Cooley-Tukey publication
- Good meets Tukey in 1956 and tells him briefly about his method
- It is revealed later [4] that the FFT had not been discovered much earlier by a coincidence

Good's paper[9]

Definition 1 *Let $M^{(i)}$ be $t \times t$ matrices, $i = 1 \ldots n$. Then the direct product*

$$M^{(1)} \times M^{(2)} \times \cdots \times M^{(n)}$$

is a $t^n \times t^n$ matrix C whose elements are

$$C_{r,s} = \prod_{i=1}^{n} M^{(i)}_{r_i,s_i}$$

where $r = (r_1, \ldots, r_n)$ and $s = (s_1, \ldots s_n)$ are the base-t representation of the index

Good's paper [9]

Definition 1 *Let $M^{(i)}$ be $t \times t$ matrices, $i = 1 \ldots n$. Then the direct product*

$$M^{(1)} \times M^{(2)} \times \cdots \times M^{(n)}$$

is a $t^n \times t^n$ matrix C whose elements are

$$C_{r,s} = \prod_{i=1}^{n} M^{(i)}_{r_i,s_i}$$

where $r = (r_1, \ldots, r_n)$ and $s = (s_1, \ldots s_n)$ are the base-t representation of the index

Theorem 1 *For every M, $M^{[n]} = A^n$ where*

$$A_{r,s} = \{M_{r_1,s_n}\delta^{s_1}_{r_2}\delta^{s_2}_{r_3}\ldots\delta^{s_{n-1}}_{r_n}\}$$

Interpretation: A has at most t^{n+1} nonzero elements. So if, for a given B we can find M such that $M^{[n]} = B$, then we can compute $B\mathbf{x} = A^n\mathbf{x}$ in $O(nt^n)$ instead of $O(t^{2n})$.

Good's paper[9]

Good's observations

- For 2^n factorial experiment:

$$M = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

Good's paper[9]

Good's observations

- n-dimensional DFT, size N in each dimension

$$A(s_1, \ldots, s_n) = \sum_{(r_1,\ldots,r_n)}^{(N-1,\ldots,N-1)} x(\mathbf{r}) \mathbf{W}_N^{r_1 s_1} \times \cdots \times \mathbf{W}_N^{r_n s_n}$$

can be represented as

$$A = \Omega^{[n]}\mathbf{x}, \, \Omega = \{\mathbf{W}_N^{rs}\} \quad r, s = 0, \ldots, N-1$$

By the theorem, $A = \Psi^n \mathbf{x}$ where Ψ is sparse

Good's paper[9]

Good's observations

- n-dimensional DFT, size N in each dimension

$$A(s_1,\ldots,s_n) = \sum_{(r_1,\ldots,r_n)}^{(N-1,\ldots,N-1)} x(\mathbf{r})\mathbf{W}_N^{r_1 s_1} \times \cdots \times \mathbf{W}_N^{r_n s_n}$$

  can be represented as

$$A = \Omega^{[n]}\mathbf{x},\ \Omega = \{\mathbf{W}_N^{rs}\} \quad r,s = 0,\ldots,N-1$$

  By the theorem, $A = \Psi^n \mathbf{x}$ where Ψ is sparse

- n-dimensional DFT, N_i in each dimension

$$A = (\Omega^{(1)} \times \cdots \times \Omega^{(n)})\mathbf{x} = \Phi^{(1)}\Phi^{(2)}\ldots\Phi^{(n)}\mathbf{x}$$

$$\Omega^{(i)} = \{\mathbf{W}_{N_i}^{rs}\}$$

$$\Phi^{(i)} = I_{N_1} \times \cdots \times \Omega^{(i)} \times \cdots \times I_{N_n}$$

Good's paper[9]

If $N = N_1 N_2$ and $(N_1, N_2) = 1$ then N-point 1-D DFT can be represented as 2-D DFT

$$r = N_2 r_1 + N_1 r_2$$

(*Chinese Remainder Thm.*) $s = N_2 s_1 (N_2^{-1}) \mod N_1 + N_1 s_2 (N_1^{-1}) \mod N_2$

$$\Rightarrow s \equiv s_1 \mod N_1, \quad s \equiv s_2 \mod N_2$$

These are index mappings $s \leftrightarrow (s_1, s_2)$ and $r \leftrightarrow (r_1, r_2)$

$$A(s) = A(s_1, s_2) = \sum_{r=0}^{N} x(r) \mathbf{W}_N^{rs} = \sum_{r_1=0}^{N_1-1} \sum_{r_2=0}^{N_2-1} x(r_1, r_2) \mathbf{W}_{N_1 N_2}^{rs}$$

(*substitute* r, s *from above*) $= \sum_{r_1=0}^{N_1-1} \sum_{r_2=0}^{N_2-1} x(r_1, r_2) \mathbf{W}_{N_1}^{r_1 s_1} \mathbf{W}_{N_2}^{r_2 s_2}$

This is exactly 2-D DFT seen above!

Good's paper[9]

So, as above, we can write

$$\begin{aligned} A &= \Omega \mathbf{x} \\ &= P(\Omega^{(1)} \times \Omega^{(2)}) Q^{-1} \mathbf{x} \\ &= P \Phi^{(1)} \Phi^{(2)} Q^{-1} \mathbf{x} \end{aligned}$$

where P and Q are permutation matrices. This results in $N(N_1 + N_2)$ multiplications instead of N^2.

Good’s results - summary

- n-dimensionanal, size N_i in each dimension

$$\left(\sum_n N_i\right)\left(\prod_n N_i\right) \text{ instead of } \left(\prod_n N_i\right)^2$$

- 1-D, size $N = \prod_n N_i$ where N_i’s are mutually prime

$$\left(\sum_n N_i\right) N \text{ instead of } N^2$$

- Still thinks that other “tricks” such as using the symmetries and sines and cosines of known angles might be more useful

The article

James W.Cooley (1926-)

- 1949 - B.A. in Mathematics (Manhattan College, NY)
- 1951 - M.A. in Mathematics (Columbia University)
- 1961 - Ph.D in Applied Mathematics (Columbia University)

- 1962 - Joins IBM Watson Research Center
- The FFT is considered his most significant contribution to mathematics
- Member of IEEE Digital Signal Processing Committee
- Awarded IEEE fellowship on his works on FFT
- Contributed much to establishing the terminology of DSP

John Wilder Tukey (1915-2000)

- 1936 - Bachelor in Chemistry (Brown University)
- 1937 - Master in Chemistry (Brown University)
- 1939 - PhD in Mathematics (Princeton) for “Denumerability in Topology”

- During WWII joins Fire Control Research Office and contributes to war effort (mainly invlolving statistics)
- From 1945 also works with Shannon and Hamming in AT&T Bell Labs
- Besides co-authoring the FFT in 1965, has several major contributions to statistics

The story - meeting

- Richard Garwin (Columbia University IBM Watson Research center) meets John Tukey (professor of mathematics in Princeton) at Kennedy's Scientific Advisory Committee in 1963

The story - meeting

- Richard Garwin (Columbia University IBM Watson Research center) meets John Tukey (professor of mathematics in Princeton) at Kennedy's Scientific Advisory Committee in 1963
  - Tukey shows how a Fourier series of length N when $N = ab$ is composite can be expressed as a-point series of b-point subseries, thereby reducing the number of computations from N^2 to $N(a+b)$

The story - meeting

- Richard Garwin (Columbia University IBM Watson Research center) meets John Tukey (professor of mathematics in Princeton) at Kennedy's Scientific Advisory Committee in 1963
  - Tukey shows how a Fourier series of length N when $N = ab$ is composite can be expressed as a-point series of b-point subseries, thereby reducing the number of computations from N^2 to $N(a+b)$
  - Notes that if N is a power of two, the algorithm can be repeated, giving $N \log_2 N$ operations

The story - meeting

- Richard Garwin (Columbia University IBM Watson Research center) meets John Tukey (professor of mathematics in Princeton) at Kennedy's Scientific Advisory Committee in 1963
  - Tukey shows how a Fourier series of length N when $N = ab$ is composite can be expressed as a-point series of b-point subseries, thereby reducing the number of computations from N^2 to $N(a+b)$
  - Notes that if N is a power of two, the algorithm can be repeated, giving $N \log_2 N$ operations
  - Garwin, being aware of applications in various fields that would greatly benefit from an efficient algorithm for DFT, realizes the importance of this and from now on, "pushes" the FFT at every opportunity

The story - publication

- Garwin goes to Jim Cooley (IBM Watson Research Center) with the notes from conversation with Tukey and asks to implement the algorithm

The story - publication

- Garwin goes to Jim Cooley (IBM Watson Research Center) with the notes from conversation with Tukey and asks to implement the algorithm
- Says that he needs the FFT for computing the 3-D Fourier transform of spin orientations of He^3. It is revealed later that what Garwin really wanted was to be able to detect nuclear exposions from seismic data

The story - publication

- Garwin goes to Jim Cooley (IBM Watson Research Center) with the notes from conversation with Tukey and asks to implement the algorithm
- Says that he needs the FFT for computing the 3-D Fourier transform of spin orientations of He^3. It is revealed later that what Garwin really wanted was to be able to detect nuclear exposions from seismic data
- Cooley finally writes a 3-D in-place FFT and sends it to Garwin

The story - publication

- Garwin goes to Jim Cooley (IBM Watson Research Center) with the notes from conversation with Tukey and asks to implement the algorithm
- Says that he needs the FFT for computing the 3-D Fourier transform of spin orientations of He^3. It is revealed later that what Garwin really wanted was to be able to detect nuclear exposions from seismic data
- Cooley finally writes a 3-D in-place FFT and sends it to Garwin
- The FFT is exposed to a number of mathematicians

The story - publication

- Garwin goes to Jim Cooley (IBM Watson Research Center) with the notes from conversation with Tukey and asks to implement the algorithm
- Says that he needs the FFT for computing the 3-D Fourier transform of spin orientations of He^3. It is revealed later that what Garwin really wanted was to be able to detect nuclear exposions from seismic data
- Cooley finally writes a 3-D in-place FFT and sends it to Garwin
- The FFT is exposed to a number of mathematicians
- It is decided that the algorithm should be put in the public domain to prevent patenting

The story - publication

- Garwin goes to Jim Cooley (IBM Watson Research Center) with the notes from conversation with Tukey and asks to implement the algorithm
- Says that he needs the FFT for computing the 3-D Fourier transform of spin orientations of He^3. It is revealed later that what Garwin really wanted was to be able to detect nuclear exposions from seismic data
- Cooley finally writes a 3-D in-place FFT and sends it to Garwin
- The FFT is exposed to a number of mathematicians
- It is decided that the algorithm should be put in the public domain to prevent patenting
- Cooley writes the draft, Tukey adds some references to Good's article and the paper is submitted to Mathematics of Computation in August 1964

The story - publication

- Garwin goes to Jim Cooley (IBM Watson Research Center) with the notes from conversation with Tukey and asks to implement the algorithm
- Says that he needs the FFT for computing the 3-D Fourier transform of spin orientations of He^3. It is revealed later that what Garwin really wanted was to be able to detect nuclear exposions from seismic data
- Cooley finally writes a 3-D in-place FFT and sends it to Garwin
- The FFT is exposed to a number of mathematicians
- It is decided that the algorithm should be put in the public domain to prevent patenting
- Cooley writes the draft, Tukey adds some references to Good's article and the paper is submitted to Mathematics of Computation in August 1964
- Published in April 1965

What if...

- Good reveals in 1993[4] that Garwin visited him in 1957, but:
- “Garwin told me with great enthusiasm about his experimental work... Unfortunately, not wanting to change the subject..., I didn’t mention my FFT work at that dinner, although I had intended to do so. He wrote me on February 9, 1976 and said: “Had we talked about it in 1957, I could have stolen [publicized] it from you instead of from John Tukey...”... That would have completely changed the history of the FFT”

Cooley-Tukey FFT

$A(j) = \sum_{k=0}^{N-1} X(k)\mathbf{W}_N^{jk} \quad N = N_1 N_2$

Cooley-Tukey FFT

$A(j) = \sum_{k=0}^{N-1} X(k) \mathbf{W}_N^{jk} \quad N = N_1 N_2$

- Convert the 1-D indices into 2-D

$$j = j_0 + j_1 N_1 \qquad j_0 = 0, \ldots, N_1 - 1 \qquad j_1 = 0, \ldots, N_2 - 1$$

$$k = k_0 + k_1 N_2 \qquad k_0 = 0, \ldots, N_2 - 1 \qquad k_1 = 0, \ldots, N_1 - 1$$

Cooley-Tukey FFT

$A(j) = \sum_{k=0}^{N-1} X(k)\mathbf{W}_N^{jk} \quad N = N_1 N_2$

- Convert the 1-D indices into 2-D

$$j = j_0 + j_1 N_1 \qquad j_0 = 0, \ldots, N_1 - 1 \qquad j_1 = 0, \ldots, N_2 - 1$$
$$k = k_0 + k_1 N_2 \qquad k_0 = 0, \ldots, N_2 - 1 \qquad k_1 = 0, \ldots, N_1 - 1$$

- "Decimate" the original sequence into N_2 N_1-subsequences:

Cooley-Tukey FFT

$A(j) = \sum_{k=0}^{N-1} X(k) \mathbf{W}_N^{jk} \quad N = N_1 N_2$

- Convert the 1-D indices into 2-D

$$j = j_0 + j_1 N_1 \qquad j_0 = 0, \ldots, N_1 - 1 \qquad j_1 = 0, \ldots, N_2 - 1$$
$$k = k_0 + k_1 N_2 \qquad k_0 = 0, \ldots, N_2 - 1 \qquad k_1 = 0, \ldots, N_1 - 1$$

- "Decimate" the original sequence into N_2 N_1-subsequences:

$$A(j_1, j_0) = \sum_{k=0}^{N-1} X(k) \mathbf{W}_N^{jk}$$

Cooley-Tukey FFT

$A(j) = \sum_{k=0}^{N-1} X(k) \mathbf{W}_N^{jk} \quad N = N_1 N_2$

- Convert the 1-D indices into 2-D

$$j = j_0 + j_1 N_1 \qquad j_0 = 0, \ldots, N_1 - 1 \qquad j_1 = 0, \ldots, N_2 - 1$$
$$k = k_0 + k_1 N_2 \qquad k_0 = 0, \ldots, N_2 - 1 \qquad k_1 = 0, \ldots, N_1 - 1$$

- "Decimate" the original sequence into N_2 N_1-subsequences:

$$A(j_1, j_0) = \sum_{k=0}^{N-1} X(k) \mathbf{W}_N^{jk}$$
$$= \sum_{k_0=0}^{N_2-1} \sum_{k_1=0}^{N_1-1} X(k_1, k_0) \mathbf{W}_{N_1 N_2}^{(j_0+j_1 N_1)(k_0+k_1 N_2)}$$

Cooley-Tukey FFT

$A(j) = \sum_{k=0}^{N-1} X(k)\mathbf{W}_N^{jk} \quad N = N_1 N_2$

- Convert the 1-D indices into 2-D

$$j = j_0 + j_1 N_1 \qquad j_0 = 0, \ldots, N_1 - 1 \qquad j_1 = 0, \ldots, N_2 - 1$$
$$k = k_0 + k_1 N_2 \qquad k_0 = 0, \ldots, N_2 - 1 \qquad k_1 = 0, \ldots, N_1 - 1$$

- "Decimate" the original sequence into N_2 N_1-subsequences:

$$\begin{aligned} A(j_1, j_0) &= \sum_{k=0}^{N-1} X(k)\mathbf{W}_N^{jk} \\ &= \sum_{k_0=0}^{N_2-1} \sum_{k_1=0}^{N_1-1} X(k_1, k_0)\mathbf{W}_{N_1 N_2}^{(j_0+j_1N_1)(k_0+k_1N_2)} \\ &= \sum_{k_0=0}^{N_2-1} \mathbf{W}_N^{j_0k_0}\mathbf{W}_{N_2}^{j_1k_0} \sum_{k_1=0}^{N_1-1} X(k_1, k_0)\mathbf{W}_{N_1}^{j_0k_1} \end{aligned}$$

Cooley-Tukey FFT

$A(j) = \sum_{k=0}^{N-1} X(k) \mathbf{W}_N^{jk} \quad N = N_1 N_2$

- Convert the 1-D indices into 2-D

$$
\begin{array}{lll}
j = j_0 + j_1 N_1 & j_0 = 0, \ldots, N_1 - 1 & j_1 = 0, \ldots, N_2 - 1 \\
k = k_0 + k_1 N_2 & k_0 = 0, \ldots, N_2 - 1 & k_1 = 0, \ldots, N_1 - 1
\end{array}
$$

- "Decimate" the original sequence into N_2 N_1-subsequences:

$$
\begin{aligned}
A(j_1, j_0) &= \sum_{k=0}^{N-1} X(k) \mathbf{W}_N^{jk} \\
&= \sum_{k_0=0}^{N_2-1} \sum_{k_1=0}^{N_1-1} X(k_1, k_0) \mathbf{W}_{N_1 N_2}^{(j_0 + j_1 N_1)(k_0 + k_1 N_2)} \\
&= \sum_{k_0=0}^{N_2-1} \mathbf{W}_N^{j_0 k_0} \mathbf{W}_{N_2}^{j_1 k_0} \sum_{k_1=0}^{N_1-1} X(k_1, k_0) \mathbf{W}_{N_1}^{j_0 k_1} \quad = \sum_{k_0=0}^{N_2-1} \mathbf{W}_N^{j_0 k_0} X_{j_0}(k_0) \mathbf{W}_{N_2}^{j_1 k_0}
\end{aligned}
$$

Cooley-Tukey FFT

- The sequences X_{j_0} have N elements and require NN_1 operations to compute
- Given these, the array A requires NN_2 operations to compute
- Overall $N(N_1 + N_2)$
- In general, if $N = \prod_{i=0}^{n} N_i$ then $N \left(\sum_{i=0}^{n} N_i\right)$ operations overall.

Cooley-Tukey FFT

If $N_1 = N_2 = \cdots = N_n = 2$ (i.e. $N = 2^n$):

$$j = j_{n-1}2^{n-1} + \cdots + j_0, \quad k = k_{n-1}2^{n-1} + \cdots + k_0$$

$$A(j_{n-1}, \ldots, j_0) = \sum_{k_0=0}^{1} \cdots \sum_{k_{n-1}=0}^{1} X(k_{n-1}, \ldots, k_0)\mathbf{W}_N^{jk}$$

$$= \sum_{k_0} \mathbf{W}_N^{jk_0} \sum_{k_1} \mathbf{W}_N^{jk_1 \times 2} \cdots \sum_{k_{n-1}} \mathbf{W}_N^{jk_{n-1} \times 2^{n-1}} X(k_{n-1}, \ldots, k_0)$$

$$\Rightarrow X_{j_0}(k_{n-2}, \ldots, k_0) = \sum_{k_{n-1}} X(k_{n-1}, \ldots, k_0)\mathbf{W}_2^{j_0 k_{n-1}}$$

$$X_{j_1,j_0}(k_{n-3}, \ldots, k_0) = \sum_{k_{n-2}} \mathbf{W}_4^{j_0 k_{n-2}} X_{j_0}(k_{n-2}, \ldots, k_0)\mathbf{W}_2^{j_1 k_{n-2}}$$

$$X_{j_{m-1},\ldots,j_0}(k_{n-m-1}, \ldots, k_0) = \sum_{k_{n-m}} \mathbf{W}_{2^m}^{(j_0 + \cdots + j_{m-2} \times 2^{m-2})k_{n-m}} X_{j_{m-2},\ldots,j_0}(k_{n-m}, \ldots, k_0) \times$$

$$\times \mathbf{W}_2^{j_{m-1} k_{n-m}}$$

Cooley-Tukey FFT

Finally, $A(j_{n-1}, \ldots, j_0) = X_{j_{n-1},\ldots,j_0}$

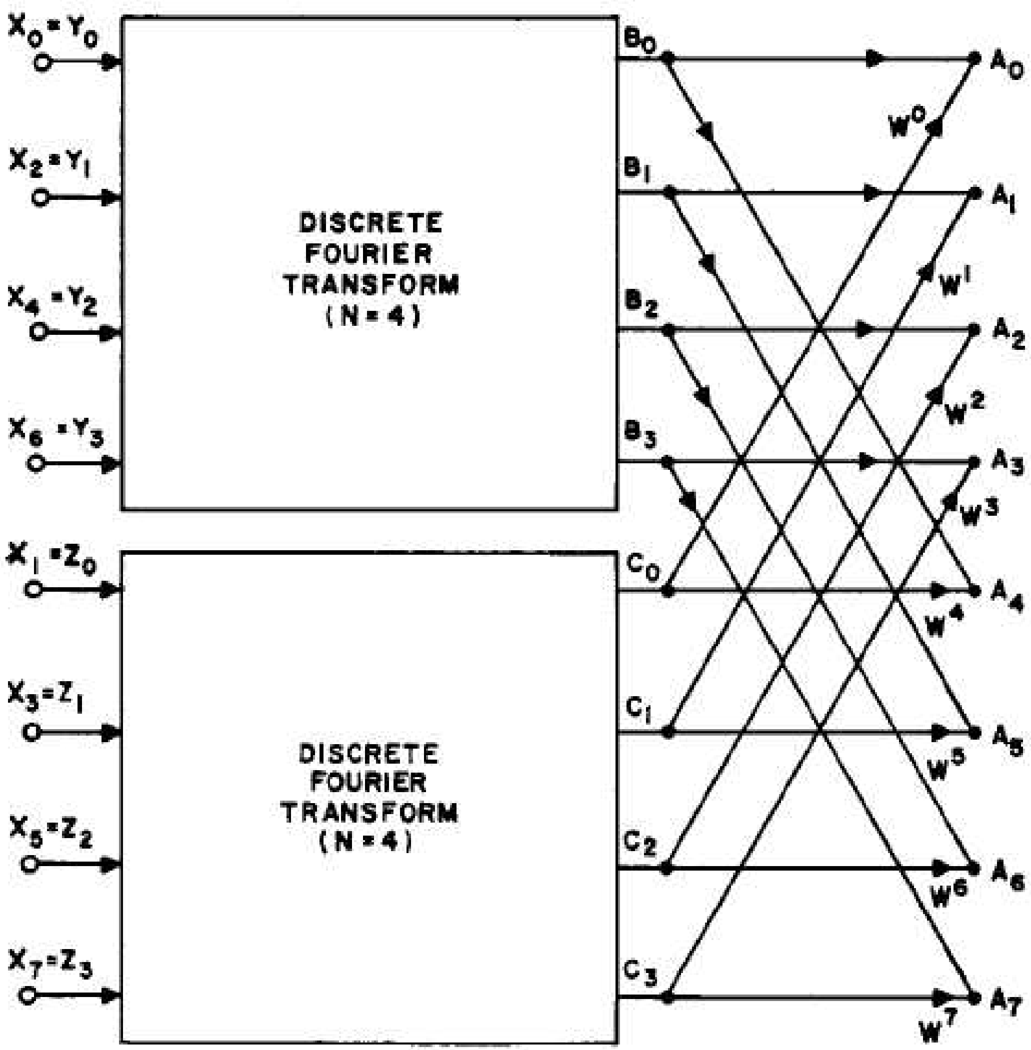
- Two important properties:
  - In-place: only two terms are involved in the calculation of every intermediate pair $X_{j_{m-1},\ldots,j_0}(\ldots)\ j_{m-1} = 0, 1$:
    - $X_{j_{m-2},\ldots,j_0}(0, \ldots)$
    - $X_{j_{m-2},\ldots,j_0}(1, \ldots)$

    So the values can be overwritten and no more additional storage be used
  - Bit-reversal
    - It is convenient to store $X_{j_{m-1},\ldots,j_0}(k_{n-m-1}, \ldots, k_0)$ at index $j_0 2^{n-1} + \cdots + j_{m-1} 2^{n-m} + k_{n-m-1} 2^{m-n-1} + \cdots + k_0$
    - The result must be reversed
- Cooley & Tukey didn't realize that the algorithm in fact builds large DFT's from smaller ones with some multiplications along the way.
- The terms $\mathbf{W}_{2^m}^{(j_0+\cdots+j_{m-2}\times 2^{m-2})k_{n-m}}$ are later called[6] "twiddle factors".

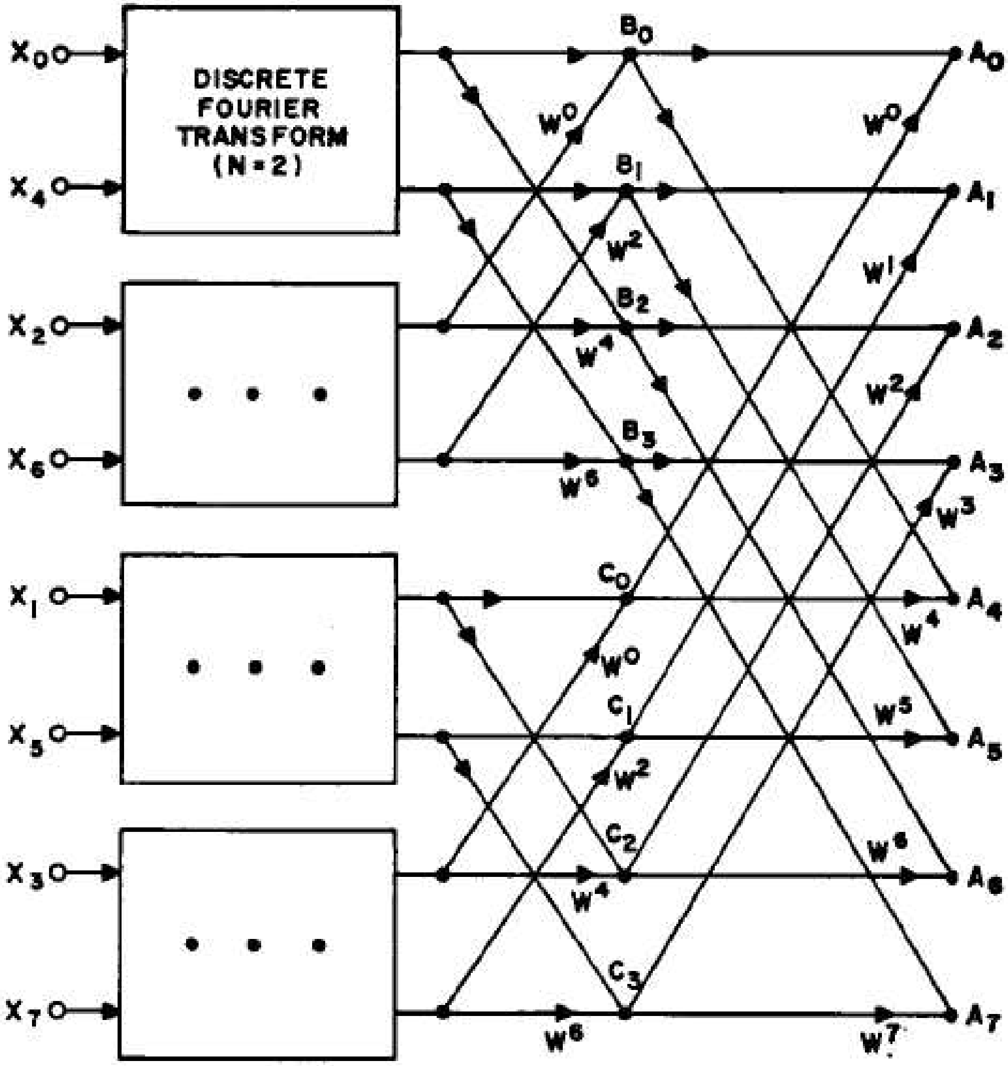
FFT properties

- Roundoff error significantly reduced compared to defining formula[6]
- A lower bound of $\frac{1}{2} n \log_2 n$ operations over $\mathbb{C}$ for linear algorithms is proved[12], so FFT is in a sense optimal
- Two "canonical" FFTs
  - Decimation-in-Time: the Cooley & Tukey version. Equivalent to taking $N_1 = N/2, \quad N_2 = 2$. Then $X_{j_0}(0)$ are the DFT coefficients of even-numbered samples and $X_{j_0}(1)$ - those of odd-numbered. The final coefficients are simply linear combination of the two.
  - Decimation-in-Frequency (Tukey-Sande[6]): $N_1 = 2, \quad N_2 = N/2$. Then $Y(*) = X_0(*)$ is the sum of even-numbered and odd-numbered samples, and $Z(*) = X_1(*)$ is the difference. The even Fourier coefficients are the DFT of Y, and the odd-numbered - "weighted" DFT of Z.

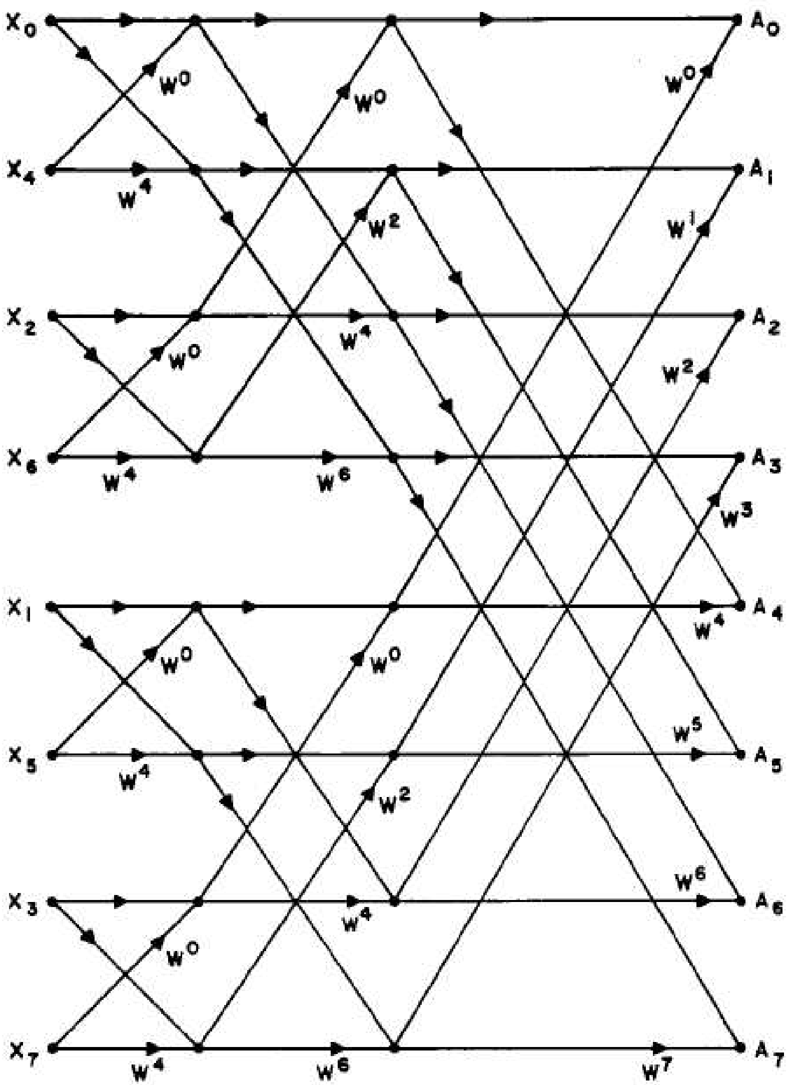
DIT signal flow



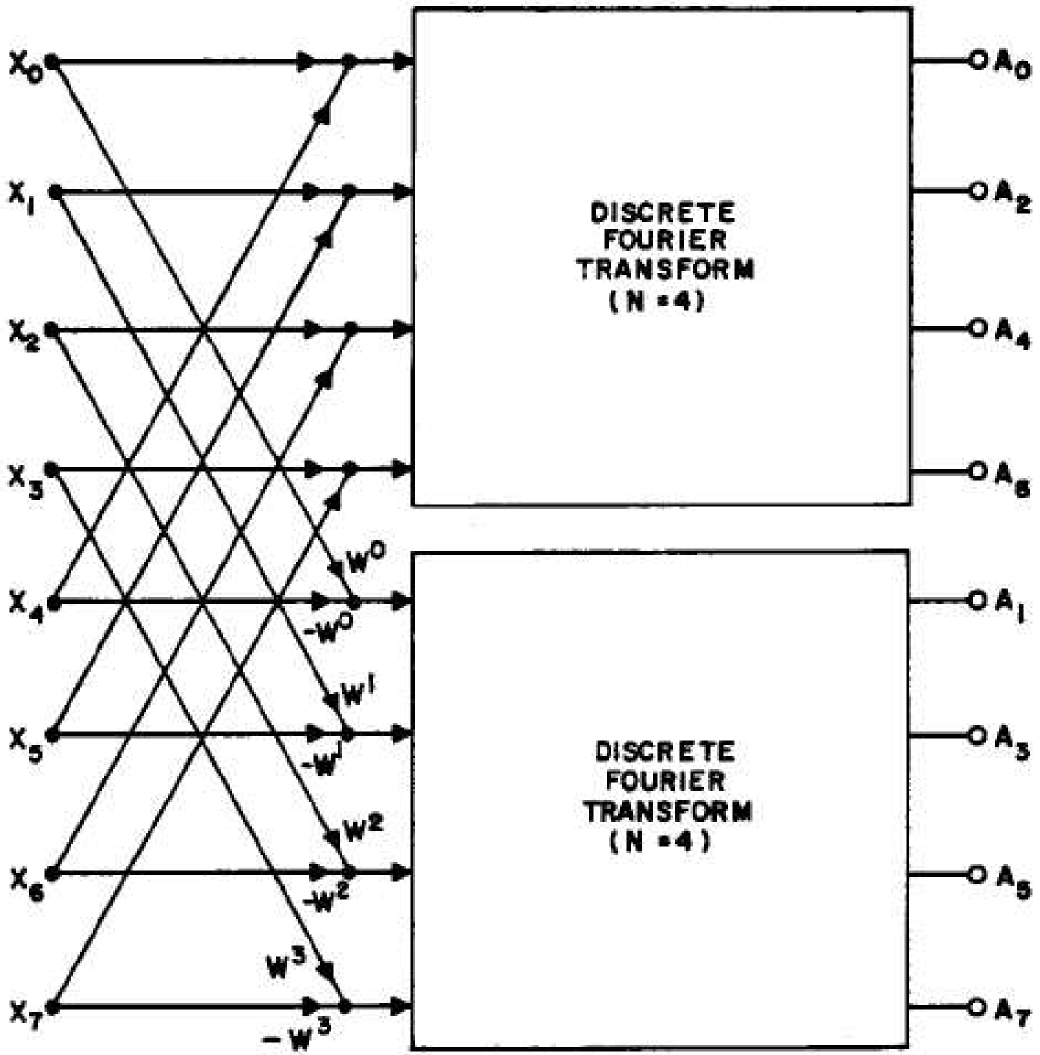
DIT signal flow



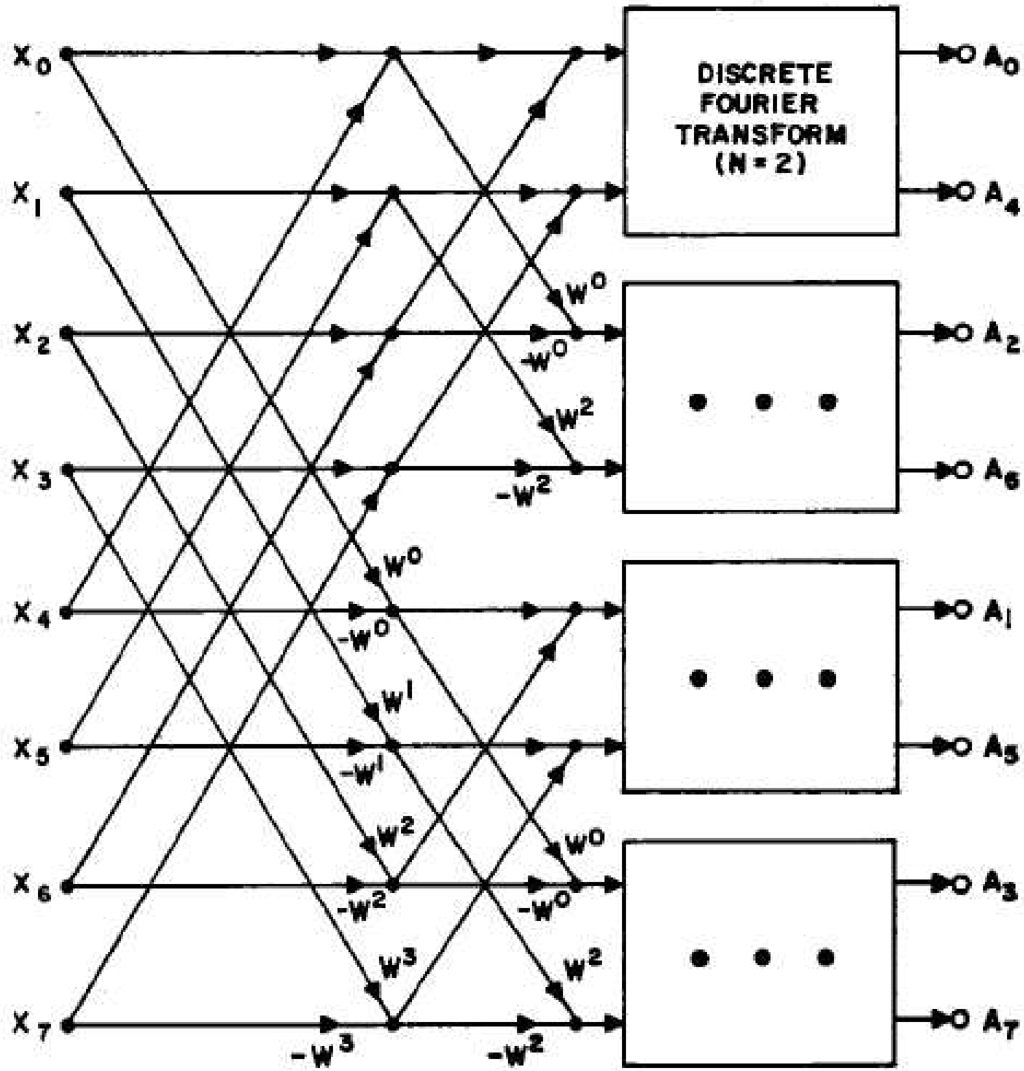
DIT signal flow



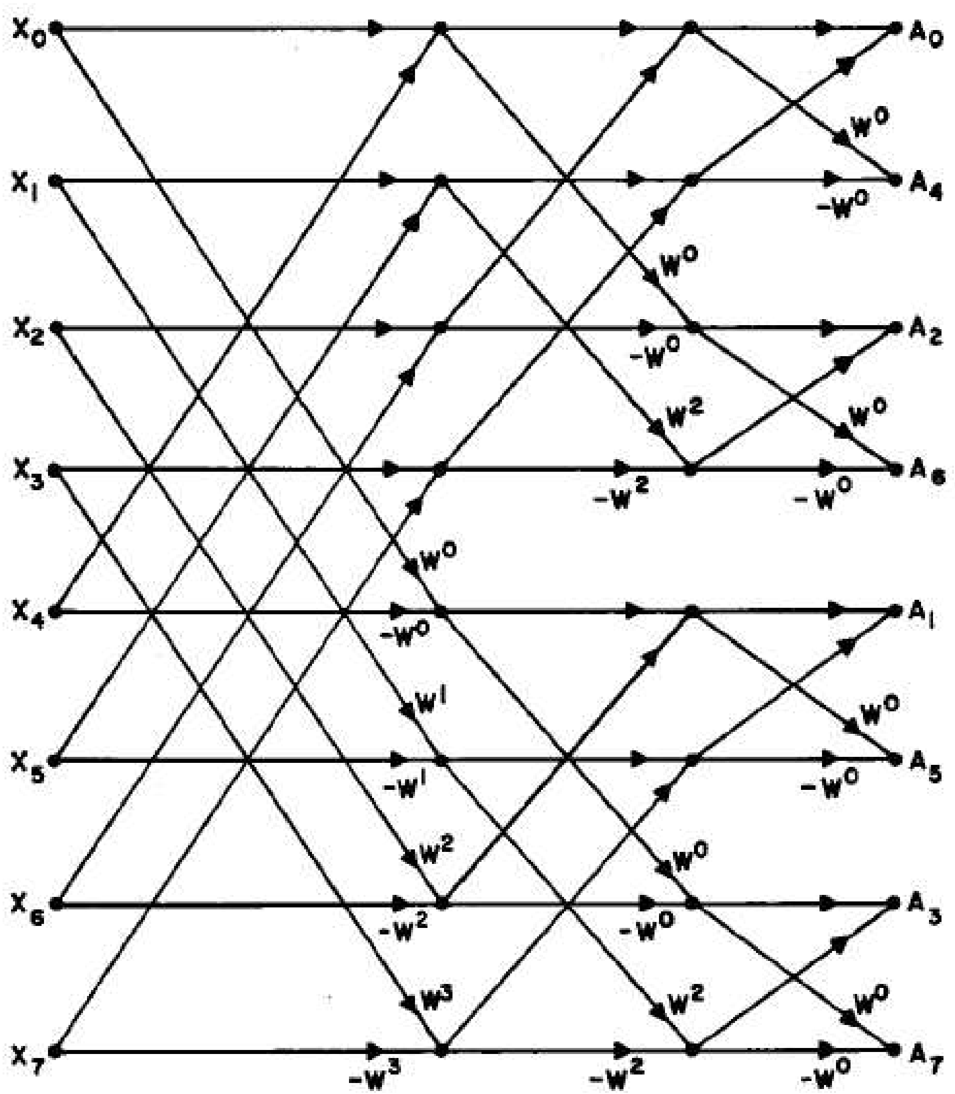
DIF signal flow



DIF signal flow



DIF signal flow



Newly discovered history of FFT

What happened after the publication

- Rudnick's note[13] mentions Danielson-Lanczos paper from 1942[3]. "Although ... less elegant and is phrased wholly in terms of real quantities, it yields the same results as the binary form of the Cooley-Tukey algorithm with a comparable number of arithmetical operations".
- "Historical Notes on FFT"[2]
  - Also mentions Danielson-Lanczos
  - Thomas and Prime Factor algorithm (Good) - quite distinct from Cooley-Tukey FFT
- A later investigation[10] revealed that Gauss essentially discovered the Cooley-Tukey FFT in 1805(!)

Danielson-Lanczos

- G. C. Danielson and C. Lanczos. Some improvements in practical Fourier analysis and their application to X-ray scattering from liquids. *J. Franklin Institute*, 233:365–380 and 435–452, 1942

Danielson-Lanczos

- G. C. Danielson and C. Lanczos. Some improvements in practical Fourier analysis and their application to X-ray scattering from liquids. *J. Franklin Institute*, 233:365–380 and 435–452, 1942
- "... the available standard forms become impractical for a large number of coefficients. We shall show that, by a certain transformation process, it is possible to double the number of ordinates with only slightly more than double the labor"

Danielson-Lanczos

- G. C. Danielson and C. Lanczos. Some improvements in practical Fourier analysis and their application to X-ray scattering from liquids. *J. Franklin Institute*, 233:365–380 and 435–452, 1942
- "... the available standard forms become impractical for a large number of coefficients. We shall show that, by a certain transformation process, it is possible to double the number of ordinates with only slightly more than double the labor"
- Concerned with improving efficiency of hand calculation

Danielson-Lanczos

- G. C. Danielson and C. Lanczos. Some improvements in practical Fourier analysis and their application to X-ray scattering from liquids. *J. Franklin Institute*, 233:365–380 and 435–452, 1942
- "... the available standard forms become impractical for a large number of coefficients. We shall show that, by a certain transformation process, it is possible to double the number of ordinates with only slightly more than double the labor"
- Concerned with improving efficiency of hand calculation
- "We shall now describe a method which eliminates the necessity of complicated schemes by reducing ... analysis for $4n$ coefficients to two analyses for $2n$ coefficients"

Danielson-Lanczos

- G. C. Danielson and C. Lanczos. Some improvements in practical Fourier analysis and their application to X-ray scattering from liquids. *J. Franklin Institute*, 233:365–380 and 435–452, 1942
- "... the available standard forms become impractical for a large number of coefficients. We shall show that, by a certain transformation process, it is possible to double the number of ordinates with only slightly more than double the labor"
- Concerned with improving efficiency of hand calculation
- "We shall now describe a method which eliminates the necessity of complicated schemes by reducing ... analysis for $4n$ coefficients to two analyses for $2n$ coefficients"
- "If desired, this reduction process can be applied twice or three times" (???!!)
- "Adoping these improvements the approximate times... are: 10 minutes for 8 coefficients, 25 min. for 16, 60 min. for 32 and 140 min. for 64" ($\approx 0.37N \log_2 N$)

Danielson-Lanczos

- Take the DIT Cooley-Tukey with $N_1 = N/2, \quad N_2 = 2$. Recall

$$A(j_0 + \frac{N}{2} j_1) = \sum_{k_1=0}^{N/2-1} X(2k_1) \mathbf{W}_{N/2}^{j_0 k_1} + \mathbf{W}_N^{j_0 + \frac{N}{2} j_1} \sum_{k_1=0}^{N/2-1} X(2k_1 + 1) \mathbf{W}_{N/2}^{j_0 k_1}$$

Danielson-Lanczos

- Take the DIT Cooley-Tukey with $N_1 = N/2, \quad N_2 = 2$. Recall

$$A(j_0 + \frac{N}{2} j_1) = \sum_{k_1=0}^{N/2-1} X(2k_1) \mathbf{W}_{N/2}^{j_0 k_1} + \mathbf{W}_N^{j_0 + \frac{N}{2} j_1} \sum_{k_1=0}^{N/2-1} X(2k_1+1) \mathbf{W}_{N/2}^{j_0 k_1}$$

- Danielson-Lanczos showed *completely analogous* result, under the framework of real trigonometric series:
  - "...the contribution of the even ordinates to a Fourier analysis of $4n$ coefficients is exactly the same as the contribution of all the ordinates to a Fourier analysis of $2n$ coefficients
  - "..contribution of the odd ordinates is also reducible to the same scheme... by means of a transformation process introducing a phase difference"
  - "...we see that, apart from the weight factors $2\cos(\pi/4n)k$ and $2\sin(\pi/4n)k$, the calculation is identical to the cosine analysis of half the number of ordinates"

Danielson-Lanczos

- Take the DIT Cooley-Tukey with $N_1 = N/2, \quad N_2 = 2$. Recall

$$A(j_0 + \frac{N}{2} j_1) = \sum_{k_1=0}^{N/2-1} X(2k_1) \mathbf{W}_{N/2}^{j_0 k_1} + \mathbf{W}_N^{j_0 + \frac{N}{2} j_1} \sum_{k_1=0}^{N/2-1} X(2k_1+1) \mathbf{W}_{N/2}^{j_0 k_1}$$

- Danielson-Lanczos showed *completely analogous* result, under the framework of real trigonometric series:
  - "...the contribution of the even ordinates to a Fourier analysis of $4n$ coefficients is exactly the same as the contribution of all the ordinates to a Fourier analysis of $2n$ coefficients
  - "..contribution of the odd ordinates is also reducible to the same scheme... by means of a transformation process introducing a phase difference"
  - "...we see that, apart from the weight factors $2\cos(\pi/4n)k$ and $2\sin(\pi/4n)k$, the calculation is identical to the cosine analysis of half the number of ordinates"
  - These "weight factors" are exactly the "twiddle factors" $\mathbf{W}_N^j$ above

Gauss

- Herman H. Goldstine writes in a footnote of “A History of Numerical Analysis from the 16th through the 19th Century” (1977)[8]:

  “This fascinating work of Gauss was neglected and rediscovered by Cooley and Tukey in an important paper in 1965”

- This goes largely unnoticed until the research by Heideman, Johnson and Burrus[10] in 1985
  - Essentially develops an DIF FFT for two factors and real sequences. Gives examples for $N = 12, 36$.
  - Declares that it can be generalized to more than 2 factors, although no examples are given
  - Uses the algorithm for solving the problem of determining orbit parameters of asteroids
  - The treatise was published posthumously (1866). By then other numerical methods were preferred to DFT, so nobody found this interesting enough...

Impact

Impact

- Certainly a "classic"
- Two special issues of IEEE trans. on Audio and Electroacoustics devoted entirely to FFT[11]
- Arden House Workshop on FFT[11]
  - Brought together people of very diverse specialities
  - "someday radio tuners will operate with digital processing units. I have heard this suggested with tongue in cheek, but one can speculate." - J.W.Cooley
- Many people "discovered" the DFT via the FFT

Further developments

- Real-FFT (DCT...)
- Parallel FFT
- FFT for prime N
- ...

Concluding thoughts

- It is obvious that prompt recognition and publication of significant achievments is an important goal

Concluding thoughts

- It is obvious that prompt recognition and publication of significant achievments is an important goal
- However, the publication in itself may not be enough

Concluding thoughts

- It is obvious that prompt recognition and publication of significant achievments is an important goal
- However, the publication in itself may not be enough
- Communication between mathematicians, numerical analysts and workers in a very wide range of applications can be very fruitful

Concluding thoughts

- It is obvious that prompt recognition and publication of significant achievments is an important goal
- However, the publication in itself may not be enough
- Communication between mathematicians, numerical analysts and workers in a very wide range of applications can be very fruitful
- Always seek new analytical methods and not rely solely on increase in processing speed

Concluding thoughts

- It is obvious that prompt recognition and publication of significant achievments is an important goal
- However, the publication in itself may not be enough
- Communication between mathematicians, numerical analysts and workers in a very wide range of applications can be very fruitful
- Always seek new analytical methods and not rely solely on increase in processing speed
- Careful attention to a review of old literature may offer some rewards

Concluding thoughts

- It is obvious that prompt recognition and publication of significant achievments is an important goal
- However, the publication in itself may not be enough
- Communication between mathematicians, numerical analysts and workers in a very wide range of applications can be very fruitful
- Always seek new analytical methods and not rely solely on increase in processing speed
- Careful attention to a review of old literature may offer some rewards
- Do not publish papers in Journal of Franklin Institute

Concluding thoughts

- It is obvious that prompt recognition and publication of significant achievments is an important goal
- However, the publication in itself may not be enough
- Communication between mathematicians, numerical analysts and workers in a very wide range of applications can be very fruitful
- Always seek new analytical methods and not rely solely on increase in processing speed
- Careful attention to a review of old literature may offer some rewards
- Do not publish papers in Journal of Franklin Institute
- Do not publish papers in neo-classic Latin

References

1. J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19:297–301, 1965.

2. J. W. Cooley, P. A. Lewis, and P. D. Welch. History of the fast Fourier transform. In *Proc. IEEE*, volume 55, pages 1675–1677, October 1967.

3. G. C. Danielson and C. Lanczos. Some improvements in practical Fourier analysis and their application to X-ray scattering from liquids. *J. Franklin Institute*, 233:365–380 and 435–452, 1942.

4. D.L.Banks. A conversation with I. J. Good. *Statistical Science*, 11(1):1–19, 1996.

5. W. M. Gentleman. An error analysis of Goertzel's (Watt's) method for computing Fourier coefficients. *Comput. J.*, 12:160–165, 1969.

6. W. M. Gentleman and G. Sande. Fast Fourier transforms—for fun and profit. In *Fall Joint Computer Conference*, volume 29 of *AFIPS Conference Proceedings*, pages 563–578. Spartan Books, Washington, D.C., 1966.

7. G. Goertzel. An algorithm for the evaluation of finite trigonometric series. *The American Mathematical Monthly*, 65(1):34–35, January 1958.

8. Herman H. Goldstine. *A History of Numerical Analysis from the 16th through the 19th Century*. Springer-Verlag, New York, 1977. ISBN 0-387-90277-5.

9. I. J. Good. The interaction algorithm and practical Fourier analysis. *Journal Roy. Stat. Soc.*, 20: 361–372, 1958.

10. M. T. Heideman, D. H. Johnson, and C. S. Burrus. Gauss and the history of the Fast Fourier Transform. *Archive for History of Exact Sciences*, 34:265–267, 1985.

11. IEEE. Special issue on fast Fourier transform. *IEEE Trans. on Audio and Electroacoustics*, AU-17: 65–186, 1969.

12. Morgenstern. Note on a lower bound of the linear complexity of the fast Fourier transform. *JACM: Journal of the ACM*, 20, 1973.

13. Philip Rudnick. Note on the calculation of Fourier series (in Technical Notes and Short Papers). *j-MATH-COMPUT*, 20(95):429–430, July 1966. ISSN 0025-5718.

14. C. E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37:10–21, January 1949.

15. F. Yates. *The design and analysis of factorial experiments*, volume 35 of *Impr. Bur. Soil Sci. Tech. Comm.* 1937.