12.4: Direct Fast Convolution and Rectangular Transforms

A relatively new approach uses index mapping directly to convert a one dimensional convolution into a multidimensional convolution. This can be done by either a type-1 or type-2 map. The short convolutions along each dimension are then done by Winograd's optimal algorithms. Unlike for the case of the DFT, there is no savings of arithmetic from the index mapping alone. All the savings comes from efficient short algorithms. In the case of index mapping with convolution, the multiplications must be nested together in the center of the algorithm in the same way as for the WFTA. There is no equivalent to the PFA structure for convolution. The multidimensional convolution can not be calculated by row and column convolutions as the DFT was by row and column DFTs.

It would first seem that applying the index mapping and optimal short algorithms directly to convolution would be more efficient than using DFTs and converting them to convolution to be calculated by the same optimal algorithms. In practical algorithms, however, the DFT method seems to be more efficient.

A method that is attractive for special purpose hardware uses distributed arithmetic. This approach uses a table look up of precomputed partial products to produce a system that does convolution without requiring multiplications.

Another method that requires special hardware uses number theoretic transforms to calculate convolution. These transforms are defined over finite fields or rings with arithmetic performed modulo special numbers. These transforms have rather limited flexibility, but when they can be used, they are very efficient.

Contributor

- ContribEEBurrus