

Merged inner-product processor using the modified Booth algorithm

Une unité de calcul combinée de produit scalaire utilisant un algorithme de Booth modifié

Fayez Elguibaly*

This paper presents novel techniques for designing fast hardware for performing the multiply-accumulate (MAC) and inner-product (IP) operations. The hardware uses the modified Booth algorithm and incorporates an overflow detection and correction circuit that guarantees correct output even if every accumulation step results in an overflow. Four innovations are presented in this work: (a) A dependence graph (DG) is used to describe the merged MAC technique. (b) Extensive use of carry-save techniques is made in the Booth encoder, the Booth multiplier, and the multiplier final-adder stage. (c) An overflow detection and correction technique is proposed for extended arithmetic using a data path whose width is minimum, dictated only by the required precision, not the prevention of overflow. (d) A CMOS delay model is developed to describe multi-level gate delays taking into account input ramp and output loading. Numerical simulations were conducted using MATLAB, VHDL, Verilog, and C++ to verify the correctness of the proposed overflow technique and the functionality of the hardware. The delay model predicts that the new merged technique will produce hardware that is about three times as fast as other published MAC techniques.

Cet article présente une nouvelle approche de conception d'un dispositif matériel pour l'implantation des opérations de multiplication-accumulation (MA) et de produit scalaire (PS). Ce dispositif utilise un algorithme de Booth modifié et inclut la détection des dépassements et un circuit de correction qui assure une sortie correcte même lorsque chaque opération d'accumulation provoque un dépassement. Notre système présente quatre principales innovations: (a) un graphe de dépendance (GD) est utilisé pour décrire la technique de MA; (b) une utilisation intensive de techniques de retenue-sauvegarde dans l'encodeur de Booth, le multiplicateur de Booth, et l'étage d'accumulation final; (c) une technique de détection et de correction des dépassements est proposée pour l'arithmétique étendue et utilise un routage des données de largeur minimum dictée uniquement par le niveau de précision requis et non pas par la prévention des dépassements; (d) un modèle de délai CMOS est développé pour décrire les délais de portes multi-niveaux en tenant compte des entrées en rampes et les charges de sortie. Des simulations numériques ont été effectuées sur MATLAB, VHDL, Verilog, et en langage C++ pour vérifier l'exactitude de la technique proposée pour la détection des dépassements et pour assurer le fonctionnement du dispositif matériel. Le modèle à délai prédit que la nouvelle méthode permet de fabriquer une unité de calcul trois fois plus rapide que les autres techniques de MA actuellement proposées dans la littérature.

*The author is with the Department of Electrical and Computer Engineering, University of Victoria, Victoria, B.C. V8W 3P6.