

1128-65-127

**Yves Nievergelt\*** (ynievergelt@ewu.edu), Eastern Washington University, Department of Mathematics, 216 Kingston Hall, Cheney, WA 99004. *Binary Floating-Point Subtraction of a Floating-Point Square Accurate to the Antepenultimate Digit by Deflation Without Fused Multiply-Subtract or Fused Multiply-Add.*

Differences of the form  $r^2 - s$  occur, for instance, in Newton's Method to compute  $\sqrt{s}$ , and in the calculation of the discriminant of a monic quadratic polynomial  $x^2 + 2rx + s$ . To compute  $r^2 - s$  accurately to the antepenultimate digit on computing systems lacking fused multiply-add and fused multiply-subtract, an algorithm is presented here that produces floating-point numbers  $\hat{r}$  and  $\hat{s}$  with smaller magnitudes and more trailing zeroes such that  $\hat{r}^2 - \hat{s} = r^2 - s$ . The algorithm may be iterated or its first result  $(\hat{r}, \hat{s})$  delivered to W. Kahan's DISC algorithm to compute  $\text{DISC}(1, \hat{r}, \hat{s}) = \hat{r}^2 - \hat{s} \cdot 1$  ([www.dtic.mil/dtic/tr/fulltext/u2/a206859.pdf](http://www.dtic.mil/dtic/tr/fulltext/u2/a206859.pdf)). While DISC bases the size of each reduction on  $r$ , the algorithm presented here uses  $s$ . (Received February 21, 2017)