

# Contents

|                                |     |
|--------------------------------|-----|
| Foreword to the second edition | vii |
| Preface                        | ix  |
| Introduction                   | 1   |

## **I Theory of computer arithmetic**

|  |     |
|--|-----|
| <b>1 First concepts</b>                                | 13  |
| 1.1 Ordered sets                                       | 13  |
| 1.2 Complete lattices and complete subnets             | 18  |
| 1.3 Screens and roundings                              | 24  |
| 1.4 Arithmetic operations and roundings                | 35  |
| <b>2 Ringoids and vectoids</b>                         | 43  |
| 2.1 Ringoids   | 43  |
| 2.2 Vectoids   | 54  |
| <b>3 Definition of computer arithmetic</b>             | 62  |
| 3.1 Introduction                                       | 62  |
| 3.2 Preliminaries                                      | 65  |
| 3.3 The traditional definition of computer arithmetic  | 69  |
| 3.4 Definition of computer arithmetic by semimorphisms | 70  |
| 3.5 A remark about roundings                           | 78  |
| 3.6 Uniqueness of the minus operator                   | 79  |
| 3.7 Rounding near zero                                 | 81  |
| <b>4 Interval arithmetic</b>                           | 87  |
| 4.1 Interval sets and arithmetic                       | 88  |
| 4.2 Interval arithmetic over a linearly ordered set    | 97  |
| 4.3 Interval matrices                                  | 101 |
| 4.4 Interval vectors                                   | 107 |
| 4.5 Interval arithmetic on a screen                    | 110 |

|      |   |     |
|------|---|-----|
| 4.6  | Interval matrices and interval vectors on a screen .....                | 118 |
| 4.7  | Complex interval arithmetic .....                                       | 126 |
| 4.8  | Complex interval matrices and interval vectors .....                    | 132 |
| 4.9  | Extended interval arithmetic .....                                      | 137 |
| 4.10 | Exception-free arithmetic for extended intervals .....                  | 141 |
| 4.11 | Extended interval arithmetic on the computer .....                      | 146 |
| 4.12 | Exception-free arithmetic for closed real intervals on the computer ... | 149 |
| 4.13 | Comparison relations and lattice operations .....                       | 152 |
| 4.14 | Algorithmic implementation of interval multiplication and division      | 153 |

## **II Implementation of arithmetic on computers**

|          |  |            |
|----------|--|------------|
| <b>5</b> | <b>Floating-point arithmetic</b>                                 | <b>157</b> |
| 5.1      | Definition and properties of the real numbers .....              | 157        |
| 5.2      | Floating-point numbers and roundings .....                       | 163        |
| 5.3      | Floating-point operations .....                                  | 172        |
| 5.4      | Subnormal floating-point numbers .....                           | 180        |
| 5.5      | On the IEEE floating-point arithmetic standard .....             | 181        |
| <b>6</b> | <b>Implementation of floating-point arithmetic on a computer</b> | <b>191</b> |
| 6.1      | A brief review of the realization of integer arithmetic .....    | 192        |
| 6.2      | Introductory remarks about the level 1 operations .....          | 201        |
| 6.3      | Addition and subtraction .....                                   | 206        |
| 6.4      | Normalization .....  | 210        |
| 6.5      | Multiplication .....   | 212        |
| 6.6      | Division .....   | 212        |
| 6.7      | Rounding .....   | 214        |
| 6.8      | A universal rounding unit .....                                  | 216        |
| 6.9      | Overflow and underflow treatment .....                           | 217        |
| 6.10     | Algorithms using the short accumulator .....                     | 220        |
| 6.11     | The level 2 operations .....                                     | 226        |
| <b>7</b> | <b>Hardware support for interval arithmetic</b>                  | <b>236</b> |
| 7.1      | Introduction .....   | 236        |

|          |   |            |
|----------|---|------------|
| 7.2      | Arithmetic interval operations                                | 237        |
| 7.2.1    | Algebraic operations  | 238        |
| 7.2.2    | Comments on the algebraic operations                          | 240        |
| 7.3      | Circuitry for the arithmetic interval operations              | 241        |
| 7.4      | Comparisons and lattice operations                            | 242        |
| 7.4.1    | Comments on comparisons and lattice operations                | 243        |
| 7.4.2    | Hardware support for comparisons and lattice operations       | 243        |
| 7.5      | Alternative circuitry for interval operations and comparisons | 244        |
| 7.5.1    | Hardware support for interval arithmetic on x86-processors    | 245        |
| 7.5.2    | Accurate evaluation of interval scalar products               | 247        |
| <b>8</b> | <b>Scalar products and complete arithmetic</b>                | <b>249</b> |
| 8.1      | Introduction and motivation                                   | 250        |
| 8.2      | Historical remarks  | 252        |
| 8.3      | The ubiquity of the scalar product in numerical analysis      | 257        |
| 8.4      | Implementation principles                                     | 260        |
| 8.4.1    | Long adder and long shift                                     | 262        |
| 8.4.2    | Short adder with local memory on the arithmetic unit          | 262        |
| 8.4.3    | Remarks   | 263        |
| 8.4.4    | Fast carry resolution   | 265        |
| 8.5      | Informal sketch for computing an exact dot product            | 267        |
| 8.6      | Scalar product computation units (SPUs)                       | 267        |
| 8.6.1    | SPU for computers with a 32 bit data bus                      | 269        |
| 8.6.2    | A coprocessor chip for the exact scalar product               | 272        |
| 8.6.3    | SPU for computers with a 64 bit data bus                      | 275        |
| 8.7      | Comments  | 278        |
| 8.7.1    | Rounding  | 278        |
| 8.7.2    | How much local memory should be provided on an SPU?           | 279        |
| 8.8      | The data format complete and complete arithmetic              | 281        |
| 8.8.1    | Low level instructions for complete arithmetic                | 282        |
| 8.8.2    | Complete arithmetic in high level programming languages       | 283        |
| 8.9      | Top speed scalar product units                                | 287        |
| 8.9.1    | SPU with long adder for 64 bit data word                      | 287        |
| 8.9.2    | SPU with long adder for 32 bit data word                      | 292        |
| 8.9.3    | An FPGA coprocessor for the exact scalar product              | 295        |

|       |   |     |
|-------|---|-----|
| 8.9.4 | SPU with short adder and complete register                  | 295 |
| 8.9.5 | Carry-free accumulation of products in redundant arithmetic | 301 |
| 8.10  | Hardware complete register window                           | 302 |

### **III Principles of verified computing**

|          |  |            |
|----------|--|------------|
| <b>9</b> | <b>Sample applications</b>   | <b>307</b> |
| 9.1      | Basic properties of interval mathematics                           | 309        |
| 9.1.1    | Interval arithmetic, a powerful calculus to deal with inequalities | 309        |
| 9.1.2    | Interval arithmetic as executable set operations                   | 310        |
| 9.1.3    | Enclosing the range of function values                             | 316        |
| 9.1.4    | Nonzero property of a function, global optimization                | 319        |
| 9.2      | Differentiation arithmetic, enclosures of derivatives              | 321        |
| 9.3      | The interval Newton method   | 329        |
| 9.4      | The extended interval Newton method                                | 332        |
| 9.5      | Verified solution of systems of linear equations                   | 333        |
| 9.6      | Accurate evaluation of arithmetic expressions                      | 340        |
| 9.6.1    | Complete expressions   | 341        |
| 9.6.2    | Accurate evaluation of polynomials                                 | 342        |
| 9.6.3    | Arithmetic expressions   | 346        |
| 9.7      | Multiple precision arithmetics                                     | 347        |
| 9.7.1    | Multiple precision floating-point arithmetic                       | 348        |
| 9.7.2    | Multiple precision interval arithmetic                             | 351        |
| 9.7.3    | Applications   | 356        |
| 9.7.4    | Adding an exponent part as a scaling factor to complete arithmetic | 358        |
| 9.8      | Remarks on Kaucher arithmetic                                      | 360        |
| 9.8.1    | The basic operations of Kaucher arithmetic                         | 364        |
| <b>A</b> | <b>Frequently used symbols</b>                                     | <b>367</b> |
| <b>B</b> | <b>On homomorphism</b>   | <b>369</b> |
|          | Bibliography   | 371        |
|          | List of figures  | 421        |
|          | List of tables   | 425        |
|          | Index  | 427        |