CLAIMS

1. A matrix calculation apparatus, comprising a vector outer product processing engine and an accumulator, wherein

the vector outer product processing engine is configured to calculate vector outer products of N first column vectors and N first row vectors, to obtain N intermediate result matrices, wherein the first column vector comprises first element values and row coordinates of the first element values, the first row vector comprises second element values and column coordinates of the second element values, the intermediate result matrix comprises third element values and position coordinates of the third element values, the position coordinates comprise the row coordinates of the first element values and the column coordinates of the second element values, the N first column vectors are obtained by converting a first matrix in a compressed format, the N first row vectors are obtained by converting a second matrix in the compressed format, and N is an integer greater than or equal to 1; and

the accumulator is configured to:

write, in a generation sequence of the N intermediate result matrices, third element values in a first intermediate result matrix of the N intermediate result matrices into corresponding positions in a cache based on first position coordinates of the third element values in the first intermediate result matrix; and

read, based on second position coordinates of third element values in a second intermediate result matrix of the N intermediate result matrices, cached values that are at positions corresponding to the second position coordinates in the cache, and accumulate the third element values in the second intermediate result matrix and the cached values, to obtain a result matrix.

2. The apparatus according to claim 1, wherein the apparatus further comprises a matrix compression unit, wherein

the matrix compression unit is configured to compress the result matrix in the uncompressed format to obtain a result matrix in a compressed format.

3. The apparatus according to claim 1, wherein the accumulator is further specifically configured to:

sort third element values in the N intermediate result matrices based on position coordinates of the third element values; and

compare position coordinates in N intermediate result matrices obtained through sorting, add up third element values with same position coordinates, and delete position coordinates of a zero-element value, to obtain a result matrix in a compressed format.

4. The apparatus according to claim 1, wherein the apparatus further comprises a format conversion unit, wherein

the format conversion unit is configured to: obtain the first matrix and the second matrix, convert the first matrix into the N first column vectors, and convert the second matrix into the N first row vectors.

5. The apparatus according to claim 4, wherein the apparatus further comprises the format conversion unit, wherein

the format conversion unit is further configured to: obtain a fifth matrix and a sixth matrix, perform format conversion on the fifth matrix to obtain the first matrix, and perform format conversion on the sixth matrix to obtain the second matrix, wherein at least one of the fifth matrix and the sixth matrix is a matrix in an uncompressed format.

6. The apparatus according to any one of claims 1 to 5, wherein precision of element values comprised in the first column vector and the first row vector is first precision;

the vector outer product processing engine is further configured to calculate vector outer products of X second column vectors and X second row vectors to obtain X^2 fourth matrices, wherein the fourth matrix comprises fourth element values and position coordinates of the fourth element values, the position coordinates of the fourth element values comprise the row coordinates of the first element values and the column coordinates of the second element values, and precision of the fourth element values is the first precision, wherein the X second column vectors are obtained by splitting the first column vector, the X second row vectors are obtained by splitting the first row vector, precision of element values comprised in the second column vector and the second row vector is second precision, the first precision is higher than the second precision, and X is an integer greater than or equal to 2; and

the accumulator is further configured to accumulate, based on indexes of the position coordinates of the fourth element values, fourth element values with same position coordinates in the X^2 fourth matrices, to obtain the intermediate result matrix, wherein precision of the third element values in the intermediate result matrix is the first precision.

7. A matrix calculation method, wherein the method is applied to a matrix calculation

apparatus, and the method comprises:

obtaining a first calculation instruction, wherein the first calculation instruction comprises N first column vectors and N first row vectors;

calculating vector outer products of the N first column vectors and the N first row vectors, to obtain N intermediate result matrices, wherein the first column vector comprises first element values and row coordinates of the first element values, the first row vector comprises second element values and column coordinates of the second element values, the intermediate result matrix comprises third element values and position coordinates of the third element values, the position coordinates comprise the row coordinates of the first element values and the column coordinates of the second element values, the N first column vectors are obtained by converting a first matrix in a compressed format, the N first row vectors are obtained by converting a second matrix in the compressed format, and N is an integer greater than or equal to 1; and

accumulating, based on indexes of the position coordinates of the third element values, third element values with same position coordinates in the N intermediate result matrices, to obtain a result matrix, wherein the accumulating comprises:

writing, in a generation sequence of the N intermediate result matrices, third element values in a first intermediate result matrix of the N intermediate result matrices into corresponding positions in a cache based on first position coordinates of the third element values in the first intermediate result matrix; and

reading, based on second position coordinates of third element values in a second intermediate result matrix of the N intermediate result matrices, cached values that are at positions corresponding to the second position coordinates in the cache, and accumulating the third element values in the second intermediate result matrix and the cached values.

8. The method according to claim 7, wherein the method further comprises:

compressing the result matrix in the uncompressed format to obtain a result matrix in a compressed format.

9. The method according to claim 7, wherein the accumulating, based on indexes of the position coordinates of the third element values, third element values with same position coordinates in the N intermediate result matrices, to obtain a result matrix comprises:

sorting third element values in the N intermediate result matrices based on position coordinates of the third element values; and

comparing position coordinates in N intermediate result matrices obtained through sorting, adding up third element values with same position coordinates, and deleting position coordinates of a zero-element value, to obtain a result matrix in a compressed format.

10. The method according to claim 7, wherein before the obtaining a first calculation instruction, the method further comprises:

obtaining a second calculation instruction, wherein the second calculation instruction comprises the first matrix and the second matrix; and

converting the first matrix into the N first column vectors, and converting the second matrix into the N first row vectors.

11. The method according to claim 10, wherein before the obtaining a second calculation instruction, the method further comprises:

obtaining a third calculation instruction, wherein the third calculation instruction comprises a fifth matrix and a sixth matrix, and at least one of the fifth matrix and the sixth matrix is a matrix in an uncompressed format; and

performing format conversion on the fifth matrix to obtain the first matrix in the compressed format, and performing format conversion on the sixth matrix to obtain the second matrix.

12. The method according to any one of claims 7 to 11, wherein precision of element values comprised in the first column vector and the first row vector is first precision;

the calculating vector outer products of the N first column vectors and the N first row vectors, to obtain N intermediate result matrices comprises:

calculating vector outer products of X second column vectors and X second row vectors to obtain X^2 fourth matrices, wherein the fourth matrix comprises fourth element values and position coordinates of the fourth element values, the position coordinates of the fourth element values comprise the row coordinates of the first element values and the column coordinates of the second element values, and precision of the fourth element values is the first precision, wherein the X second column vectors are obtained by splitting the first column vector, the X second row vectors are obtained by splitting the first row vector, precision of element values comprised in the second column vector and the second row vector is second precision, the first precision is higher than the second precision, and X is an integer greater than or equal to 2; and

the method further comprises:

accumulating, based on indexes of the position coordinates of the fourth element values, fourth element values with same position coordinates in the X_2 fourth matrices, to obtain the intermediate result matrix, wherein precision of the third element values in the intermediate result matrix is the first precision.

13. A matrix calculation circuit, wherein the matrix calculation circuit is configured to perform the method according to any one of claims 7 to 12.

14. A matrix calculation system, wherein the system comprises a processor and a matrix calculation apparatus, the processor is configured to send a calculation instruction to the matrix calculation apparatus, and the matrix calculation apparatus is configured to perform the method according to any one of claims 7 to 12.

15. A chip, wherein the chip comprises a processor, a matrix calculation apparatus is integrated into the processor, and the matrix calculation apparatus is configured to perform the matrix calculation method according to any one of claims 7 to 12.

16. A matrix calculation device, wherein the device comprises the matrix calculation system according to claim 14 or the chip according to claim 15.