What is claimed is:

1. A method of fabrication of a floating gate non-volatile memory device comprising the steps of :

    a) forming a bottom tunnel layer over a substrate; the bottom tunnel layer comprised of a lower oxide tunnel layer and an upper hafnium oxide tunnel layer;

    b) forming a charge storage layer over the bottom tunnel layer;

    c) forming a lower hafnium oxide storage layer over the charge storage layer;

    d) forming an upper cap oxide storage layer over the lower hafnium oxide storage layer; the lower hafnium oxide storage layer and the upper oxide storage layer form a top blocking layer;

    e) forming a gate electrode over the top blocking layer;

    f) patterning the bottom tunnel layer; the charge storage layer; and top blocking layer to form a gate structure;

    g) forming source/drain regions in the substrate adjacent to the gate structure.

2. The method of claim 1 which further comprises in a tunnel anneal step performed before the formation of the charge storage layer, annealing the bottom tunnel layer in a hydrogen containing atmosphere.

3. The method of claim 1 which further comprises in a tunnel anneal step performed before the formation of the charge storage layer, annealing the bottom tunnel layer in a hydrogen containing atmosphere at a temperature between about 300 and 450 degree C for a time between 5 and 60 seconds.

4. The method of claim 1 which further comprises in a storage anneal step performed after the formation of the charge storage layer and before the formation of the lower hafnium oxide storage layer, annealing charge storage layer over the bottom tunnel layer in a nitrogen containing atmosphere.

5. The method of claim 1 which further comprises in a storage anneal step performed after the formation of the charge storage layer and before the formation of the lower hafnium oxide storage layer, annealing charge storage layer over the bottom tunnel layer in a nitrogen containing atmosphere at a temperature between about 300 and 450 degree C for a time between 15 and 60 seconds.

6. The method of claim 1 which further comprises in a block anneal step performed after the formation of the lower hafnium oxide layer and before the formation of the upper cap oxide layer, annealing the top blocking layer in a nitrogen containing atmosphere.

7. The method of claim 1 which further comprises in a block anneal step performed after the formation of the lower hafnium oxide layer and before the formation of the upper cap oxide storage layer, annealing the top blocking layer in a nitrogen containing atmosphere at a temperature between about 300 and 450 degree C for a time between 15 and 60 seconds.

8. The method of claim 1 wherein the lower oxide tunnel layer is comprised essentially of silicon oxide and has a thickness between 0.2 and 1.0 nm.

9. The method of claim 1 wherein the upper hafnium oxide tunnel layer is comprised essentially of hafnium oxide and has a thickness between 1 and 6 nm.

10. The method of claim 1 wherein the charge storage layer is comprised of a tantalum oxide layer; the tantalum oxide layer is comprised essentially of tantalum oxide and has a thickness between 3 and 8 nm.

11. The method of claim 1 wherein the lower hafnium oxide storage layer is comprised essentially of hafnium oxide and has a thickness between 5 and 20 nm.

12. The method of claim 1 wherein the upper oxide storage layer is comprised essentially of silicon oxide and has a thickness between 0.2 and 1 nm.

13. A method of fabrication of a floating gate non-volatile memory device comprising the steps of :

a) forming a bottom tunnel layer over a substrate; the bottom tunnel layer comprised of a lower oxide tunnel layer and an upper hafnium oxide tunnel layer;

b) in a tunnel anneal step, annealing the bottom tunnel layer in a nitrogen containing atmosphere;

c) forming a charge storage layer over the bottom tunnel layer;

d) in a storage anneal step, annealing charge storage layer over the bottom tunnel layer in a nitrogen containing atmosphere;

e) forming a lower hafnium oxide layer over the charge storage layer;

f) in a block anneal step, annealing the lower hafnium oxide layer in a nitrogen containing atmosphere;

g) forming an upper cap oxide storage layer over the lower hafnium oxide storage layer; the lower hafnium oxide storage layer and an upper oxide storage layer form a top blocking layer;

h) forming a gate electrode over the top blocking layer;

i) patterning the bottom tunnel layer; the charge storage layer; and top blocking layer to form a gate structure;

j) forming source/drain regions in the substrate adjacent to the gate structure.

14. The method of claim 14 wherein the lower oxide tunnel layer is comprised essentially of silicon oxide and has a thickness between 0.2 and 1.0 nm.

15. The method of claim 14 wherein the upper hafnium oxide tunnel layer is comprised essentially of hafnium oxide and has a thickness between 1 and 6 nm.

16. The method of claim 14 wherein the tunnel anneal step comprises an anneal in a hydrogen containing atmosphere, at a temperature between about 300 and 450 degree C for a time between 5 and 60 seconds.

17. The method of claim 14 wherein the charge storage layer is comprised of a tantalum oxide layer; the tantalum oxide layer is comprised essentially of tantalum oxide and has a thickness between 3 and 8 nm.

18. The method of claim 14 wherein the storage anneal step comprises an anneal in a nitrogen or ammonia containing atmosphere, at a temperature between about 300 and 450 degree C for a time between 15 and 60 seconds.

19. The method of claim 14 wherein the lower hafnium oxide storage layer is comprised essentially of hafnium oxide and has a thickness between 5 and 20 nm.

20. The method of claim 14 wherein the block anneal step comprises an anneal in a nitrogen or ammonia containing atmosphere, at a temperature between about 300 and 450 degree C for a time between 15 and 60 seconds.

21. The method of claim 14 wherein the upper oxide layer is comprised essentially of silicon oxide and has a thickness between 0.2 and 1 nm.

22. A method of fabrication of a floating gate non-volatile memory device comprising the steps of :

a) forming a bottom tunnel layer over a substrate;

b) forming a charge storage layer over the bottom tunnel layer;

c) forming a top blocking layer over the charge storage layer;

d) forming a gate electrode over the top blocking layer;

e) patterning the bottom tunnel layer; the charge storage layer; and top blocking layer and the gate electrode to form a gate structure;

f) forming source/drain regions in the substrate adjacent to the gate structure.

23. The method of claim 22 which further includes in a tunnel anneal step performed after the formation of the bottom tunnel layer and before the formation of the charge storage layer, annealing the bottom tunnel layer in a hydrogen containing atmosphere.

24. The method of claim 22 which further includes in a tunnel anneal step performed after the formation of the bottom tunnel layer and before the formation of the charge storage layer , annealing the bottom tunnel layer in a hydrogen containing atmosphere; at a temperature between about 700 and 1000 degree C for a time between 15 and 60 seconds.

25. The method of claim 22 which further includes in a storage anneal step performed after the formation of the charge storage layer and before the formation of the top blocking layer, annealing charge storage layer in a nitrogen containing atmosphere.

26. The method of claim 22 which further includes in a storage anneal step performed after the formation of the charge storage layer and before the formation of the top blocking layer, annealing charge storage layer in a nitrogen containing atmosphere; at a temperature between about 300 and 450 degree C for a time between 15 and 60 seconds.

27. The method of claim 22 which further includes in a block plasma nitridation step performed after the formation of the top blocking layer, annealing the top blocking layer 240 in a nitrogen containing atmosphere to form a top silicon nitride layer on said top blocking layer.

28. The method of claim 22 which further includes in a block plasma nitridation step performed after the formation of the top blocking layer, annealing the top blocking layer in a nitrogen containing atmosphere at a temperature between about 300 and 500 degree C for a time between 20 and 60 seconds.

29. The method of claim 22 wherein the bottom tunnel layer is comprised essentially of silicon oxide and has a thickness between 1 and 4 nm.

30. The method of claim 22 wherein the charge storage layer 30 is comprised of a tantalum oxide layer 30; the tantalum oxide layer is comprised essentially of tantalum oxide and has a thickness between 3 and 8 nm.

31. The method of claim 22 wherein the top blocking layer is comprised essentially of silicon oxide and has a thickness between 2 and 10 nm.

32. A method of fabrication of a floating gate non-volatile memory device comprising the steps of :

a) forming a bottom tunnel layer over a substrate;

b) in a tunnel anneal step, annealing the bottom tunnel layer in a hydrogen containing atmosphere;

c) forming a charge storage layer over the bottom tunnel layer;

d) in a storage anneal step, annealing charge storage layer in a nitrogen containing atmosphere;

e) forming a top blocking layer over the charge storage layer;

f) in a block plasma nitridation step, annealing the top blocking layer in a nitrogen containing atmosphere;

g) forming a gate electrode over the top blocking layer;

h) patterning the bottom tunnel layer; the charge storage layer; and top blocking layer and the gate electrode to form a gate structure;

i) forming source/drain regions in the substrate adjacent to the gate structure.

33. The method of claim 32 which further includes the tunnel anneal step comprises an anneal in a hydrogen containing atmosphere, at a temperature between about 700 and 1000 degree C for a time between 15 and 60 seconds.

34. The method of claim 32 which further includes the storage anneal step comprises an anneal in a nitrogen or ammonia containing atmosphere, at a temperature between about 300 and 450 degree C for a time between 15 and 60 seconds.

35. The method of claim 32 which further includes the block plasma nitridation step comprises an anneal in a nitrogen or ammonia containing atmosphere, at a temperature between about 300 and 500 degree C for a time between 20 and 60 seconds.

36. The method of claim 32 wherein the bottom tunnel layer is comprised essentially of silicon oxide and has a thickness between 1 and 4 nm.

37. The method of claim 32 wherein the charge storage layer is comprised of a tantalum oxide layer; the tantalum oxide layer is comprised essentially of tantalum oxide and has a thickness between 3 and 8 nm.

38. The method of claim 32 wherein the top blocking layer is comprised essentially of silicon oxide and has a thickness between 2 and 10 nm.

39. A floating gate non-volatile memory device comprising:

a) a bottom tunnel layer over a substrate; the bottom tunnel layer comprised of a lower oxide tunnel layer and a upper hafnium oxide tunnel layer;

b) a charge storage layer over the bottom tunnel layer;

c) a lower hafnium oxide layer over the charge storage layer;

d) an upper cap oxide layer over the lower hafnium oxide storage layer; the a lower hafnium oxide storage layer and an upper oxide storage layer form a top blocking layer;

e) a gate electrode over the top blocking layer;

f) source/drain regions in the substrate adjacent to the gate electrode.

40. The floating gate non-volatile memory device of claim 39 wherein the lower oxide tunnel layer is comprised essentially of silicon oxide and has a thickness between 0.2 and 1.0 nm.

41. The floating gate non-volatile memory device of claim 39 wherein the upper hafnium oxide tunnel layer is comprised essentially of hafnium oxide and has a thickness between 1 and 6 nm.

42. The floating gate non-volatile memory device of claim 39 wherein the charge storage layer is comprised of a tantalum oxide layer; the tantalum oxide layer is comprised essentially of tantalum oxide and has a thickness between 3 and 8 nm.

43. The floating gate non-volatile memory device of claim 39 wherein the lower hafnium oxide storage layer is comprised essentially of hafnium oxide and has a thickness between 5 and 20 nm.

44. The floating gate non-volatile memory device of claim 39 wherein the upper oxide storage layer is comprised essentially of silicon oxide and has a thickness between 0.2 and 1 nm.

45. A floating gate non-volatile memory device comprising:

a) forming a bottom tunnel layer over a substrate;

b) a charge storage layer over the bottom tunnel layer;

c) a top blocking layer over the charge storage layer;

d) a gate electrode over the top blocking layer;

e) source/drain regions in the substrate adjacent to the gate electrode.

46. The floating gate non-volatile memory device of claim 45 which further includes a top silicon nitride layer on the top blocking layer.

47. The floating gate non-volatile memory device of claim 45 wherein the bottom tunnel layer is comprised essentially of silicon oxide and has a thickness between 1 and 4 nm.

48. The floating gate non-volatile memory device of claim 45 wherein the charge storage layer is comprised of a tantalum oxide layer; the tantalum oxide layer is comprised essentially of tantalum oxide and has a thickness between 3 and 8 nm.

49. The floating gate non-volatile memory device of claim 45 wherein the top blocking layer is comprised essentially of silicon oxide and has a thickness between 2 and 10 nm.

What is claimed is:

1. A method of fabrication of a non-volatile memory device comprising the steps of:

a) forming a bottom tunnel layer over a substrate; the bottom tunnel layer comprised of a lower oxide tunnel layer and an upper hafnium oxide tunnel layer;

b) forming a charge storage layer over the bottom tunnel layer; said charge storage layer is comprised of tantalum oxide;

b1) in a storage anneal step, annealing the charge storage layer in a nitrogen containing atmosphere;

c) forming a lower hafnium oxide blocking layer over the charge storage layer;

d) forming an upper cap oxide layer over the lower hafnium oxide blocking layer; the lower hafnium oxide blocking layer and the upper cap oxide layer form a top blocking layer;

e) forming a gate electrode over the top blocking layer;

f) patterning the bottom tunnel layer, the charge storage layer, the upper cap oxide layer, the lower hafnium oxide blocking layer, and the gate electrode to form a gate structure; and

g) forming source/drain regions in the substrate adjacent to the gate structure.

2. The method of claim 1 which further comprises in a tunnel anneal step performed before the formation of the charge storage layer, annealing the bottom tunnel layer in a hydrogen containing atmosphere.

3. The method of claim 1 which further comprises in a tunnel anneal step performed before the formation of the charge storage layer, annealing the bottom tunnel layer in a hydrogen containing atmosphere comprised of NH_3 or H_2 at a temperature between about 300 and 450 degrees C for a time between 5 and 60 seconds.

4. The method of claim 1 wherein the storage anneal step is performed at a temperature between about 300 and 450 degrees C for a time between 15 and 60 seconds.

5. The method of claim 1 which further comprises in a block anneal step performed after the formation of the lower hafnium oxide blocking layer and before the formation of the upper cap oxide layer, annealing the lower hafnium oxide blocking layer in a nitrogen containing atmosphere.

6. The method of claim 1 which further comprises in a block anneal step performed after the formation of the lower hafnium oxide blocking layer and before the formation of the upper cap oxide layer, annealing the lower hafnium oxide blocking layer in a nitrogen containing atmosphere at a temperature between about 300 and 450 degrees C for a time between 15 and 60 seconds.

7. The method of claim 1 wherein the lower oxide tunnel layer consists essentially of silicon oxide and has a thickness between 0.2 and 1.0 nm;

the upper hafnium oxide tunnel layer consists essentially of hafnium oxide and has a thickness between 1 and 6 nm;

the charge storage layer consists essentially of tantalum oxide and has a thickness between 3 and 8 nm;

the lower hafnium oxide blocking layer consists essentially of hafnium oxide and has a thickness between 5 and 20 nm; and

the upper cap oxide layer consists essentially of silicon oxide and has a thickness between 0.2 and 1 nm.

8. The method of claim 1 wherein the upper hafnium oxide tunnel layer is comprised essentially of hafnium oxide and has a thickness between 1 and 6 nm.

9. The method of claim 1 wherein the charge storage layer consists essentially of tantalum oxide and has a thickness between 3 and 8 nm.

10. The method of claim 1 wherein the lower hafnium oxide blocking layer consists essentially of hafnium oxide and has a thickness between 5 and 20 nm.

11. The method of claim 1 wherein

the lower oxide tunnel layer consists essentially of silicon oxide and has a thickness between 0.2 and 1.0 nm: and

the upper cap oxide layer consists essentially of silicon oxide and has a thickness between 0.2 and 1 nm.

12. A method of fabrication of a non-volatile memory device comprising the steps of:

a) forming a bottom tunnel layer over and in contact with a substrate; the bottom tunnel layer comprised of a lower oxide tunnel layer and an upper hafnium oxide tunnel layer; the lower oxide tunnel layer is below and in contact with the upper hafnium oxide tunnel layer;

b) in a tunnel anneal step, annealing the bottom tunnel layer in a hydrogen containing atmosphere;

c) forming a charge storage layer over and in contact with the bottom tunnel layer; said charge storage layer consists essentially of tantalum oxide;

d) in a storage anneal step, annealing the charge storage layer in a nitrogen containing atmosphere;

e) forming a lower hafnium oxide blocking layer over and in contact with the charge storage layer;

f) in a block anneal step, annealing the lower hafnium oxide blocking layer in a nitrogen containing atmosphere;

g) forming an upper cap oxide layer over and in contact with the lower hafnium oxide blocking layer; the lower hafnium oxide blocking layer and the upper cap oxide layer form a top blocking layer;

h) forming a gate electrode over and in contact with the top blocking layer;

i) patterning the bottom tunnel layer, the charge storage layer, and the top blocking layer and the gate electrode to form a gate structure; and

j) forming source/drain regions in the substrate adjacent to the gate structure.

13. The method of claim 12 wherein the lower oxide tunnel layer consists essentially of silicon oxide and has a thickness between 0.2 and 1.0 nm;

the upper hafnium oxide tunnel layer consists essentially of hafnium oxide and has a thickness between 1 and 6 nm;

the charge storage layer consists essentially of tantalum oxide and has a thickness between 3 and 8 nm;

the lower hafnium oxide blocking layer consists essentially of hafnium oxide and has a thickness between 5 and 20 nm; and

the upper cap oxide layer consists essentially of silicon oxide and has a thickness between 0.2 and 1 nm.

14. The method of claim 12 wherein the upper hafnium oxide tunnel layer is comprised essentially of hafnium oxide and has a thickness between 1 and 6 nm.

15. The method of claim 12 wherein the tunnel anneal step comprises an anneal in a hydrogen containing atmosphere, at a temperature between about 300 and 450 degrees C for a time between 5 and 60 seconds.

16. The method of claim 12 wherein the charge storage layer consists essentially of tantalum oxide and has a thickness between 3 and 8 nm.

17. The method of claim 12 wherein the storage anneal step comprises an anneal in a nitrogen or ammonia containing atmosphere, at a temperature between about 300 and 450 degrees C for a time between 15 and 60 seconds.

18. The method of claim 12 wherein the lower hafnium oxide blocking layer consists essentially of hafnium oxide and has a thickness between 5 and 20 nm.

19. The method of claim 12 wherein the block anneal step comprises an anneal in a nitrogen or ammonia containing atmosphere, at a temperature between about 300 and 450 degrees C for a time between 15 and 60 seconds.

20. The method of claim 12 wherein the upper cap oxide layer consists essentially of silicon oxide and has a thickness between 0.2 and 1 nm.

21. The method of claim 12 wherein the tunnel anneal step comprises an anneal in a hydrogen containing atmosphere, at a temperature between about 300 and 450 degrees C for a time between 5 and 60 seconds;

the storage anneal step comprises an anneal in a nitrogen or ammonia containing atmosphere, at a temperature between about 300 and 450 degrees C for a time between 15 and 60 seconds; and

the block anneal step comprises an anneal in a nitrogen or ammonia containing atmosphere, at a temperature between about 300 and 450 degrees C for a time between 15 and 60 seconds.

22. The method of claim 12 wherein the tunnel anneal step comprises an anneal in a hydrogen containing atmosphere, at a temperature between about 300 and 450 degrees C;

the storage anneal step comprises an anneal in a nitrogen or ammonia containing atmosphere, at a temperature between about 300 and 450 degrees C; and

the block anneal step comprises an anneal in a nitrogen or ammonia containing atmosphere, at a temperature between about 300 and 450 degrees C.

23. The method of claim 12 wherein the tunnel anneal step comprises an anneal in a hydrogen containing atmosphere, at a temperature between about 300 and 450 degrees C for a time between 5 and 60 seconds;

the storage anneal step comprises an anneal in a nitrogen or ammonia containing atmosphere, at a temperature between about 300 and 450 degrees C for a time between 15 and 60 seconds;

the block anneal step comprises an anneal in a nitrogen or ammonia containing atmosphere, at a temperature between about 300 and 450 degrees C for a time between 15 and 60 seconds;

the lower oxide tunnel layer consists essentially of silicon oxide and has a thickness between 0.2 and 1.0 nm;

the upper hafnium oxide tunnel layer is comprised essentially of hafnium oxide and has a thickness between 1 and 6 nm;

the charge storage layer consists essentially of tantalum oxide and has a thickness between 3 and 8 nm;

the lower hafnium oxide blocking layer consists essentially of hafnium oxide and has a thickness between 5 and 20 nm; and

the upper cap oxide layer consists essentially of silicon oxide and has a thickness between 0.2 and 1 nm.

24. The method of claim 1 wherein the charge storage layer consists essentially of tantalum oxide.

25. The method of claim 1 wherein

the lower oxide tunnel layer is on and in contact with the substrate;

the upper hafnium oxide tunnel layer is on and in contact with the lower oxide tunnel layer;

the charge storage layer is on and in contact with upper hafnium oxide tunnel layer; said charge storage layer consists essentially of tantalum oxide;

the lower hafnium oxide blocking layer is on and in contact with the charge storage layer; and

the upper cap oxide layer is on in contact with the lower hafnium oxide blocking layer.

26. The method of claim 1 wherein the lower oxide tunnel layer consists essentially of silicon oxide;

the upper hafnium oxide tunnel layer consists essentially of hafnium oxide; the charge storage layer consists essentially of tantalum oxide;

the lower hafnium oxide blocking layer consists essentially of hafnium oxide; and

the upper cap oxide layer consists essentially of silicon oxide.

27. The method of claim 1 wherein

the lower oxide tunnel layer is on and in contact with the substrate;

the upper hafnium oxide tunnel layer is on and in contact with the lower oxide tunnel layer;

the charge storage layer is on and in contact with the upper hafnium oxide tunnel layer; said charge storage layer consists essentially of tantalum oxide;

the lower hafnium oxide blocking layer is on and in contact with the charge storage layer;

the upper cap oxide layer is on and in contact with the lower hafnium oxide blocking layer;

the lower oxide tunnel layer consists essentially of silicon oxide;

the upper hafnium oxide tunnel layer consists essentially of hafnium oxide; the charge storage layer consists essentially of tantalum oxide;

the lower hafnium oxide blocking layer consists essentially of hafnium oxide; and

the upper cap oxide layer consists essentially of silicon oxide.

28. A method of fabrication of a non-volatile memory device comprising the steps of:

a) forming a bottom tunnel layer over a substrate; the bottom tunnel layer comprised of a lower oxide tunnel layer and an upper hafnium oxide tunnel layer;

b) forming a charge storage layer over the bottom tunnel layer; said charge storage layer is comprised of tantalum oxide;

c) forming a lower hafnium oxide blocking layer over the charge storage layer;

c1) in a block anneal step, annealing the lower hafnium oxide blocking layer in a nitrogen containing atmosphere;

d) forming an upper cap oxide layer over the lower hafnium oxide blocking layer; the lower hafnium oxide blocking layer and the upper cap oxide layer form a top blocking layer;

e) forming a gate electrode over the top blocking layer;

f) patterning the bottom tunnel layer, the charge storage layer, the upper cap oxide layer, the lower hafnium oxide blocking layer, and the gate electrode to form a gate structure; and

g) forming source/drain regions in the substrate adjacent to the gate structure.