**Table 1: Effectiveness of the proposed 3D-FNC.**

| Bench | Sparsity | AutoNCS [8] | | | | | 3D-FNC w/o. LM | | | | | 3D-FNC | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Area ($um^2$) | Wire ($um$) | #TSV | util | RT($s$) | Area ($um^2$) | Wire ($um$) | #TSV | util | RT($s$) | Area ($um^2$) | Wire ($um$) | #TSV | util | RT($s$) |
| n300 | 94.47% | 3925.17 | 93398.64 | 353 | 0.65 | 20.37 | 4058.70 | 96424.28 | 349 | 0.45 | 6.46 | 3814.78 | 94389.73 | 358 | 0.67 | 5.21 |
| n400 | 93.59% | 7318.64 | 269163.17 | 427 | 0.76 | 35.28 | 7381.36 | 276821.31 | 420 | 0.57 | 13.23 | 7153.36 | 272542.70 | 426 | 0.77 | 11.17 |
| n500 | 94.39% | 10521.18 | 407018.25 | 491 | 0.70 | 69.43 | 10994.21 | 429515.68 | 477 | 0.47 | 32.14 | 10502.32 | 411046.51 | 485 | 0.68 | 29.62 |
| avg. | 94.15% | 7255.00 | 256526.69 | 424 | 0.70 | 41.69 | 7478.09 | 267587.09 | 416 | 0.50 | 17.28 | 7156.82 | 259326.31 | 423 | 0.71 | 15.33 |
| ratio | – | 1.01 | 0.99 | 1.00 | 0.99 | 2.72 | 1.04 | 1.03 | 0.98 | 0.70 | 1.13 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

order to see the effect of the L-method, we compare the floorplanning results on the clustering results with and without the L-method. Since the size of crossbar, $s \times s$, is limited by the memristive technology, the maximum size of crossbars is set to $64 \times 64$ in this work [8]. Without the L-method, the number of clusters is determined by the way, in which iteratively increasing the number of clusters by one in hierarchical clustering until the size of the largest crossbar is below the size limitation. Then we run the 3D floorplanning on the clustering results of the case with and without L-method, respectively. The layer number is set to 2. The area of the neuron in $45nm$ technology is $2500um^2$ [20]. Besides each memristor and memristive crossbar use $4f^2$ and $s^2 \cdot 40f^2$ circuit area ($f$=feature size), respectively [11]. The experiment is tested on three Hopfield networks in [8] with the size of 300, 400, and 500. In Table 1, "3D-FNC w/o. LM" and "3D-FNC" list the average results. Column "util" gives the average utilization of all mapped crossbars. Columns "Area", "Wire" and "#TSV" represent chip area, total half-perimeter wirelength overhead, and the number of TSVs, respectively. We can see that in the case without L-method, the utilization of mapped crossbars are reduced by nearly 30%. In addition, area and wirelength are increased by around 4% and 3%, which demonstrates that the hierarchical clustering with L-method can result in highly hardware-efficient designs.

### 4.2 Comparison with Previous Work

We compare 3D-FNC with AutoNCS [8] on the hardware cost. In AutoNCS, the iterative spectral clustering is used for grouping the connections into the memristive crossbars. Since AutoNCS only generates 2D floorplanning output, to provide a fair comparison, we also implement a 3D floorplanning based on the iterative spectral clustering results of AutoNCS. The experiment is also tested on three Hopfield networks in [8] with the size of 300, 400, and 500. Table 1 lists the average statistic results. Column "RT" reports the total computational time in seconds. As shown in Table 1, compared with AutoNCS, 3D-FNC can achieve comparable area, wirelength cost and crossbar utilization, while the runtime of 3D-FNC is far less than AutoNCS. That is because in AutoNCS, the spectral clustering is executed in every iteration, thus the clustering method could be time consuming when the neural network is large.

## 5 CONCLUSION

In this paper, we have proposed an effective memristive crossbar mapping for neuromorphic computing systems on 3D IC, where both crossbar utilization and design cost are considered. Experimental results show that, compared with state-of-the-art, the proposed 3D-FNC can achieve highly hardware-efficient designs. Memristive crossbar gives hope for the anticipated efficient implementation of artificial neuromorphic networks, thus we expect to see a lot of researches to provide more efficient physical synthesis solutions.

## ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (NSFC) under grant No. 61732020 and The Research Grants Council of Hong Kong SAR (Project No. CUHK24209017). The authors would like to thank the Information Science Laboratory Center of USTC for hardware and software services, as well as Prof. Tsung-Yi Ho from National Tsing Hua University for providing test benches.

## REFERENCES

[1] Y. Chen, H. H. Li, C. Wu, C. Song, S. Li, C. Min, H.-P. Cheng, W. Wen, and X. Liu, "Neuromorphic computing's yesterday, today, and tomorrow–an evolutional view," *Integration, the VLSI Journal*, 2017.
[2] A. Ankit, A. Sengupta, and K. Roy, "Trannsformer: Neural network transformation for memristive crossbar based neuromorphic system design," *arXiv preprint arXiv:1708.07949*, 2017.
[3] F. Akopyan, J. Sawada, A. Cassidy, and et.al, "Truenorth: Design and tool flow of a 65 mw 1 million neuron programmable neurosynaptic chip," *IEEE TCAD*, vol. 34, no. 10, pp. 1537–1557, 2015.
[4] S. Acciarito, A. Cristini, G. Susi, and et.al, "Hardware design of lif with latency neuron model with memristive stdp synapses," *Integration, the VLSI Journal*, 2017.
[5] M. Hu, H. Li, Q. Wu, and G. S. Rose, "Hardware realization of bsb recall function using memristor crossbar arrays," in *Proc. DAC*, 2012, pp. 498–503.
[6] M. A. Ehsan, Z. Zhou, and Y. Yi, "Neuromorphic 3D integrated circuit: A hybrid, reliable and energy efficient approach for next generation computing," in *Proc. GLSVLSI*, 2017, pp. 221–226.
[7] H. An, M. A. Ehsan, Z. Zhou, F. Shen, and Y. Yi, "Monolithic 3D neuromorphic computing system with hybrid CMOS and memristor-based synapses and neurons," *Integration, the VLSI Journal*, 2017.
[8] W. Wen, C.-R. Wu, X. Hu, B. Liu, T.-Y. Ho, X. Li, and Y. Chen, "An eda framework for large scale hybrid neuromorphic computing systems," in *Proc. DAC*, 2015, pp. 1–6.
[9] C.-R. Wu, W. Wen, T.-Y. Ho, and Y. Chen, "Thermal optimization for memristor-based hybrid neuromorphic computing systems," in *Proc. ASPDAC*, 2016, pp. 274–279.
[10] J. Cui and Q. Qiu, "Towards memristor based accelerator for sparse matrix vector multiplication," in *Proc. ISCAS*, 2016, pp. 121–124.
[11] B. Liu, Y. Chen, B. Wysocki, and T. Huang, "The circuit realization of a neuromorphic computing system with memristor-based synapse design," in *Neural Information Processing*, 2012, pp. 357–365.
[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. NIPS*, 2012, pp. 1097–1105.
[13] Q. Xu and S. Chen, "Fast thermal analysis for fixed-outline 3D floorplanning," *Integration, the VLSI Journal*, vol. 59, pp. 157–167, 2017.
[14] P. H. Shiu, R. Ravichandran, S. Easwar, and S. K. Lim, "Multi-layer floorplanning for reliable system-on-package," in *Proc. ISCAS*, vol. 5, 2004, pp. V–69.
[15] S. Chen and T. Yoshimura, "Fixed-outline floorplanning: Enumerating block positions and a new objective function for calculating area costs," *IEEE TCAD*, vol. 27, no. 5, pp. 858–871, 2008.
[16] S. Chen, L. GE, M.-F. Chiang, and T. Yoshimura, "Lagrangian relaxation based interlayer signal via assignment for 3-D ICs," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 92, no. 4, pp. 1080–1087, 2009.
[17] R. C. Dubes and A. K. Jain, *Algorithms for clustering data.* New Jersey: Prentice Hall Englewood Cliffs, 1988.
[18] M. B. Alawieh, F. Wang, and X. Li, "Identifying wafer-level systematic failure patterns via unsupervised learning," *IEEE TCAD*, 2017.
[19] S. Salvador and P. Chan, "Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms," in *Proc. ICTAI*, 2004, pp. 576–584.
[20] J.-s. Seo, B. Brezzo, Y. Liu, and et.al, "A 45nm CMOS neuromorphic chip with a scalable architecture for learning in networks of spiking neurons," in *Proc. CICC*, 2011, pp. 1–4.