Supplementary Material for *Conjugate Gradient Iterative Hard Thresholding: Observed Noise Stability for Compressed Sensing, by J.D. Blanchard, J. Tanner, K. Wei*

Jeffrey D. Blanchard, Jared Tanner, and Ke Wei

I. Outline of Supplementary Material

This document contains a representation of the full data generated for [1]. In [1] plots were selected to emphasize the most crucial information contained in the data. For completeness, this document includes all omitted plots. Figs. 1–6 present the 50% recovery phase transition curves for the compressed sensing problem to show the smooth decrease in the recovery region for all algorithms. Figures 7–24, labeled *Full data* in the list of figures, provide all data for each problem class tested: the 50% recovery phase transition curves for all algorithms, an algorithm selection map identifying the algorithm with minimum average recovery time among all algorithms tested, the minimum average recovery time, and a ratio of the average recovery time for each algorthm compared to the minimum average recovery time among all algorithms tested. For a more detailed view of the recovery performance for all values of ρ in the phase transition region, the full data also contains semi-log plots of the average computational times for successful recovery for the two values of δ which are closest to 0.1 and 0.3.

Consider $y = Ax + e$ where $x \in \mathbb{R}^n$ is k-sparse (i.e. the number of nonzeros entries in x is at most k, denoted $\|x\|_0 \leq k$), $A \in \mathbb{R}^{m\times n}$ and $e \in \mathbb{R}^m$ representing model misfit between representing y with k columns of A and/or additive noise. The compressed sensing recovery question asks one to identify the minimizer

$$\hat{x} = \arg\min_{z\in\mathbb{R}^n} \|y - Az\|_2 \quad \text{subject to} \quad \|z\|_0 \leq k. \tag{1}$$

The row-sparse approximation problem extends the compressed sensing problem to consider $Y = AX + E$ where $X \in \mathbb{R}^{n\times r}$ is k-row-sparse (i.e. the number of rows containing nonzero entries in X is at most k, denoted $\|X\|_{R0} \leq k$), $A \in \mathbb{R}^{m\times n}$ and $E \in \mathbb{R}^{m\times r}$ representing model misfit between representing Y with k columns of A and/or additive noise. The row-sparse approximation question asks one to identify the k-row-sparse minimizer

$$\hat{X} = \arg\min_{Z\in\mathbb{R}^{n\times r}} \|Y - AZ\|_F \quad \text{subject to} \quad \|Z\|_{R0} \leq k. \tag{2}$$

Question (1) is the special case of (2) with $r = 1$.

For the compressed sensing problem (1), the problem classes are defined in [1] and are denoted (Mat, B_ϵ). The measurement matrix $A \in \mathbb{R}^{m\times n}$ is drawn from the random matrix ensemble $Mat \in \{\mathcal{N}, \mathcal{S}_7, DCT\}$. $\mathcal{N}$ is the ensemble of dense, Gaussian matrices with entries drawn i.i.d. from $\mathcal{N}(0, m^{-1})$. $\mathcal{S}_7$ is the sparse ensemble with seven nonzero values per column drawn with equal probability from $\{-1/\sqrt{7}, 1/\sqrt{7}\}$ and with locations chosen uniformly. DCT is the ensemble of randomly subsampled discrete cosine transforms with m rows of the $n \times n$ DCT matrix chosen uniformly. The random vector $x \in \mathbb{R}^n$ is drawn from the sparse binary vector B with k locations chosen uniformly and nonzeros values of $\{-1, 1\}$ selected with equal probability. The random vector ensembles B_ϵ have the vector x drawn from B with the measurements defined by the model $y = Ax + e$ with $e \in \mathbb{R}^m$ a random misfit vector drawn uniformly from the sphere of radius $\epsilon\|Ax\|$.

The matrix ensembles for the row-sparse approximation problem (2) are identical to those from the compressed sensing problem. A problem class (Mat, B_ϵ) has the measurement matrix $A \in \mathbb{R}^{m\times n}$ drawn from a random matrix ensemble $Mat \in \{\mathcal{N}, \mathcal{S}_7, DCT\}$. The row-sparse matrix $X \in \mathbb{R}^{n\times r}$ drawn from the binary row-sparse matrix ensemble has its row-support chosen uniformly with nonzero values $\{-1, 1\}$ selected with equal probability. For the noise level ϵ, the measurements are defined by the model $Y = AX + E \in \mathbb{R}^{m\times r}$ with E a random misfit matrix; each column E_i of the misfit matrix E is drawn uniformly from the sphere of radius $\epsilon\|\tilde{Y}_i\|$ where $\{\tilde{Y}_i : i = 1, \ldots, r\}$ are the columns of $\tilde{Y} = AX$.

For both problems (1) and (2), the sparse ensemble B is equivalent to the ensemble B_ϵ with $\epsilon = 0$.

JDB is with the Department of Mathematics and Statistics, Grinnell College, Grinnell, IA 50112. JDB is supported by the National Science Foundation [DMS 11126152] and a 2013-2014 Harris Faculty Fellowship. (email: jeff@math.grinnell.edu)

JT and KW are with the Mathematics Institute, University of Oxford, Oxford, UK. JT is supported by an NVIDIA Professor Partnership. KW is supported by the China Scholarship Council. (email: {tanner,wei}@maths.ox.ac.uk)

LIST OF FIGURES

REFERENCES

[1] J. D. Blanchard, J. Tanner, and K. Wei, "Conjugate gradient iterative hard thresholding: Observed noise stability for compressed sensing," p. http://eprints.maths.ox.ac.uk/1806/, 2014.

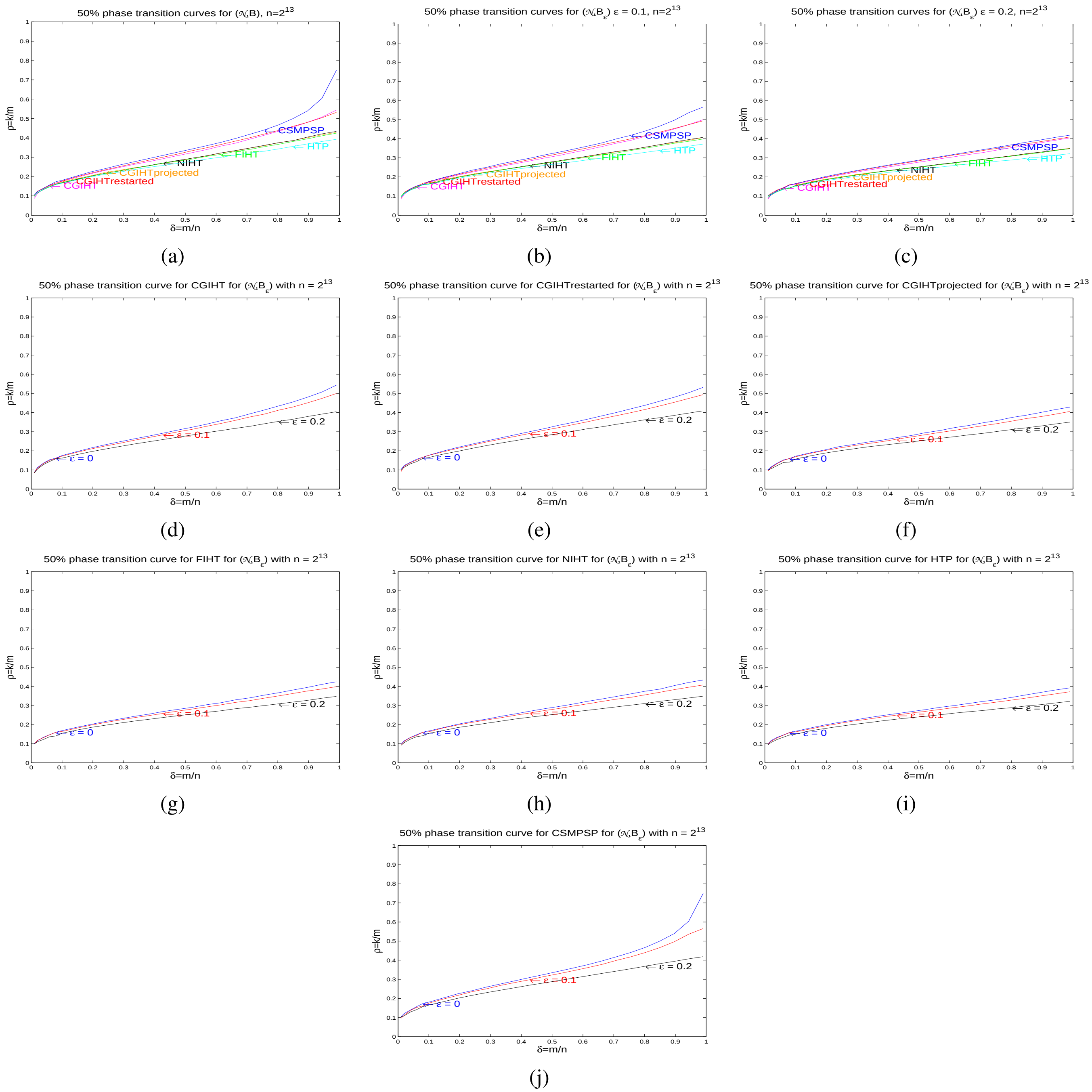


Fig. 1. 50% recovery probability logistic regression curves of all algorithms for problem classes $(\mathcal{N}, B_\epsilon)$ and $n = 2^{13}$ with (a) $\epsilon = 0$, (b) $\epsilon = 0.1$, and (c) $\epsilon = 0.2$. Stability of 50% recovery probability logistic regression curves for each algorithm for problem classes $(\mathcal{N}, B_\epsilon)$ and $n = 2^{13}$ with $\epsilon = 0, 0.1, 0.2$: (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP.

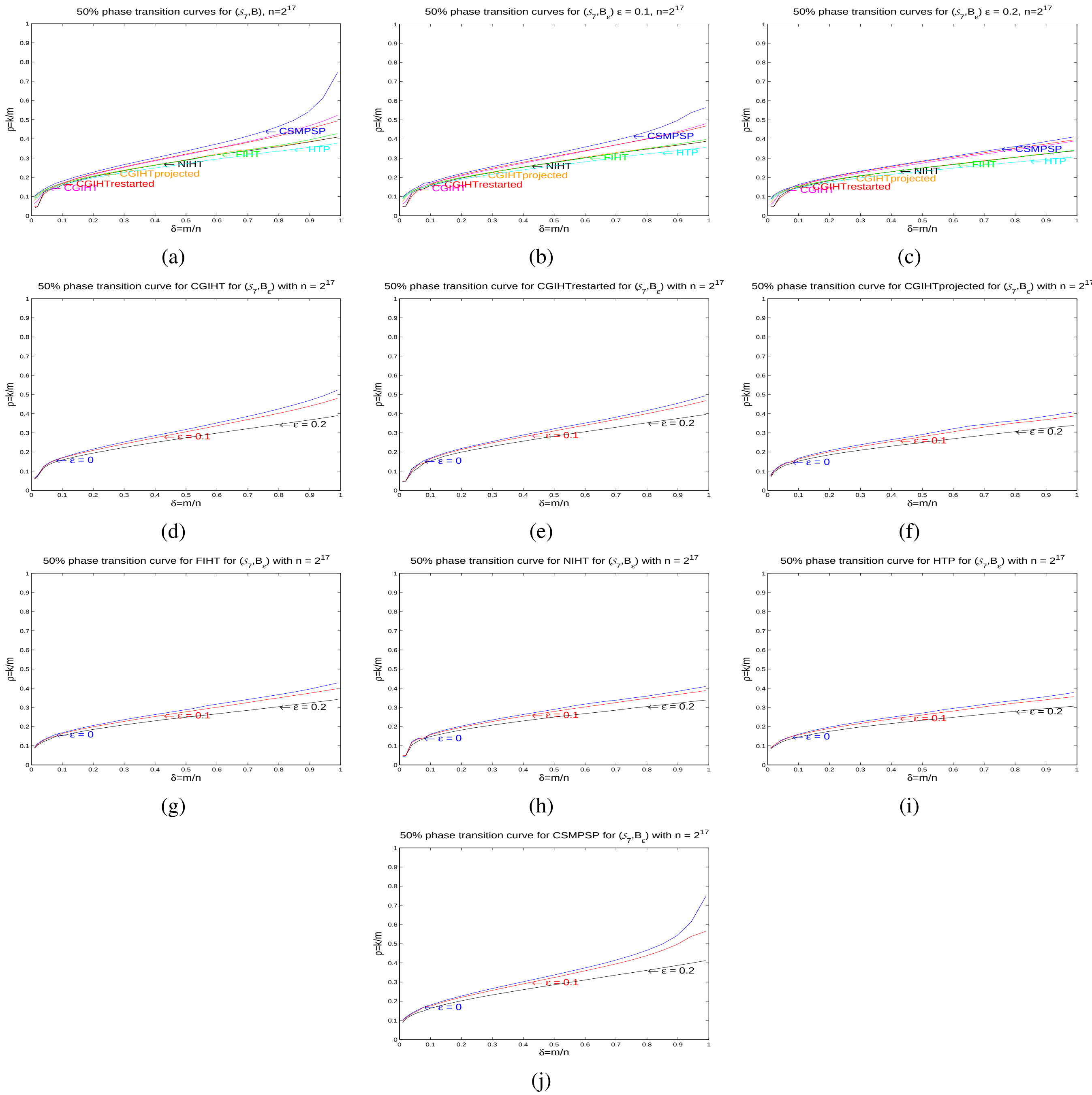


Fig. 2. 50% recovery probability logistic regression curves of all algorithms for problem classes $(\mathcal{S}_7, B_\epsilon)$ and $n = 2^{17}$ with (a) $\epsilon = 0$, (b) $\epsilon = 0.1$, and (c) $\epsilon = 0.2$. Stability of 50% recovery probability logistic regression curves for each algorithm for problem classes $(\mathcal{S}_7, B_\epsilon)$ and $n = 2^{17}$ with $\epsilon = 0, 0.1, 0.2$: (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP.

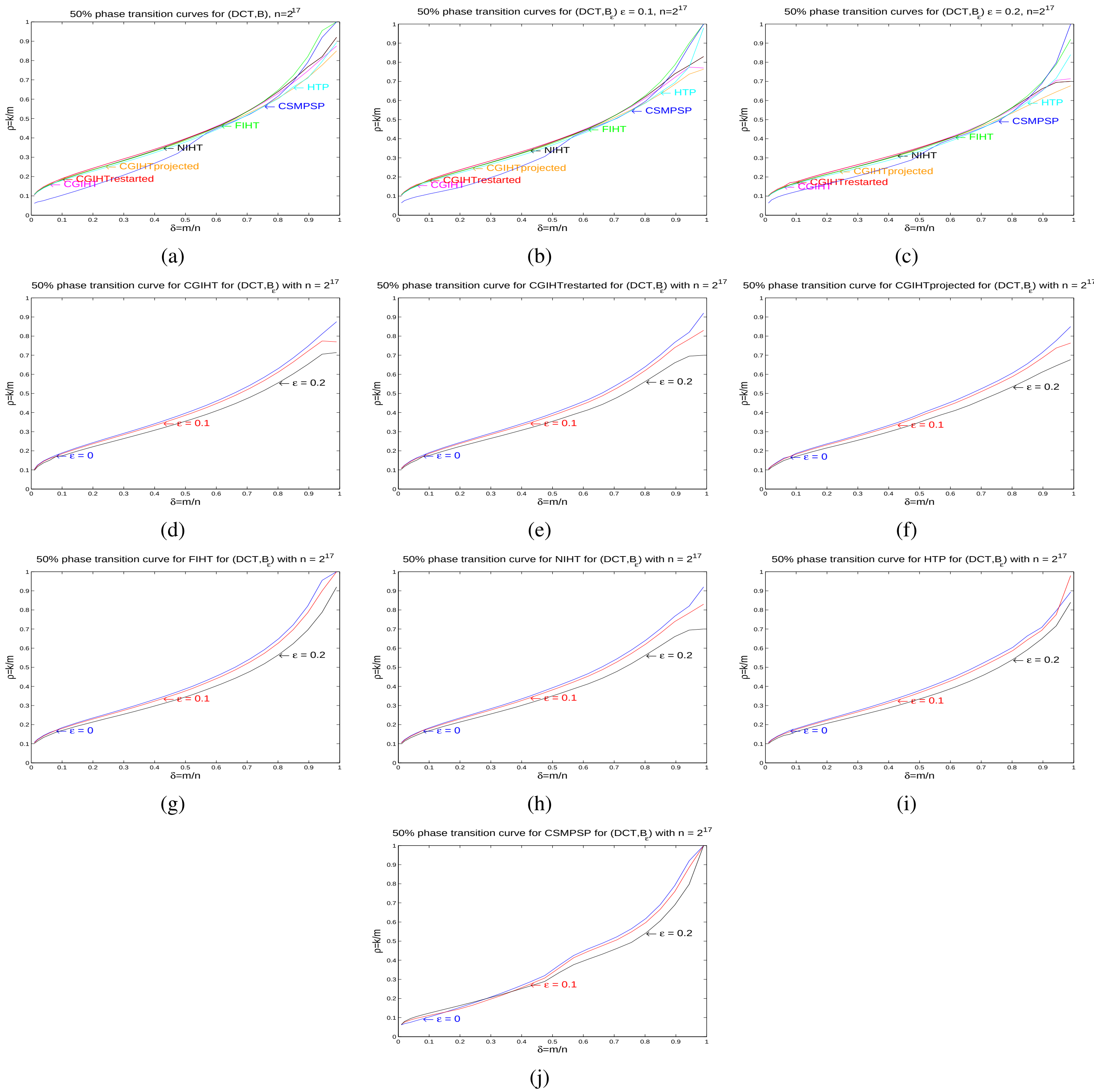


Fig. 3. 50% recovery probability logistic regression curves of all algorithms for problem classes (DCT, B_ϵ) and $n = 2^{17}$ with (a) $\epsilon = 0$, (b) $\epsilon = 0.1$, and (c) $\epsilon = 0.2$. Stability of 50% recovery probability logistic regression curves for each algorithm for problem classes (DCT, B_ϵ) and $n = 2^{17}$ with $\epsilon = 0, 0.1, 0.2$: (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP.

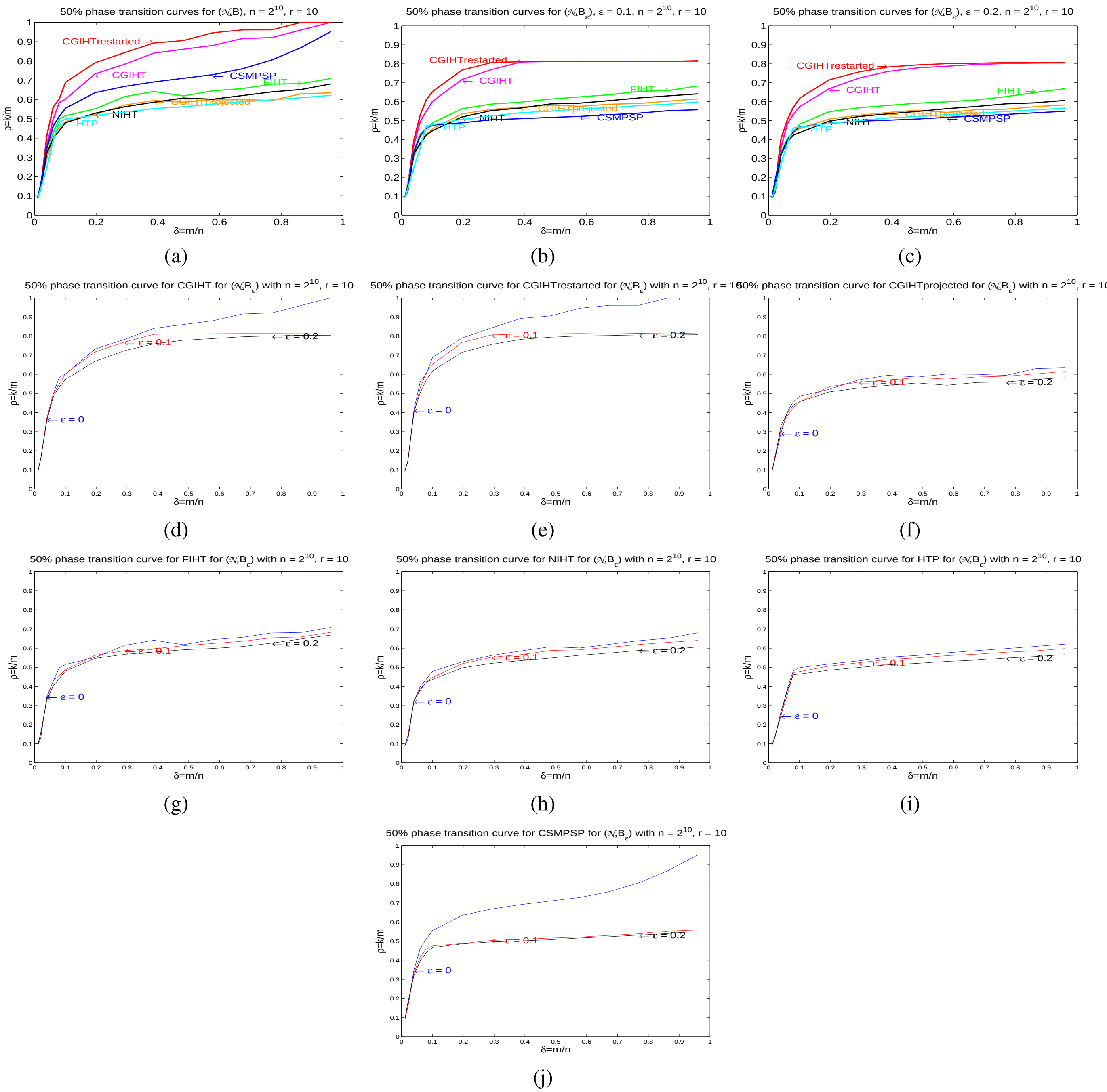


Fig. 4. 50% recovery probability logistic regression curves of all algorithms for problem classes $(\mathcal{N}, B_\epsilon)$ and $n = 2^{10}$, $r = 10$, with (a) $\epsilon = 0$, (b) $\epsilon = 0.1$, and (c) $\epsilon = 0.2$. Stability of 50% recovery probability logistic regression curves for each algorithm for problem classes $(\mathcal{N}, B_\epsilon)$ and $n = 2^{10}$, $r = 10$, with $\epsilon = 0, 0.1, 0.2$: (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP.

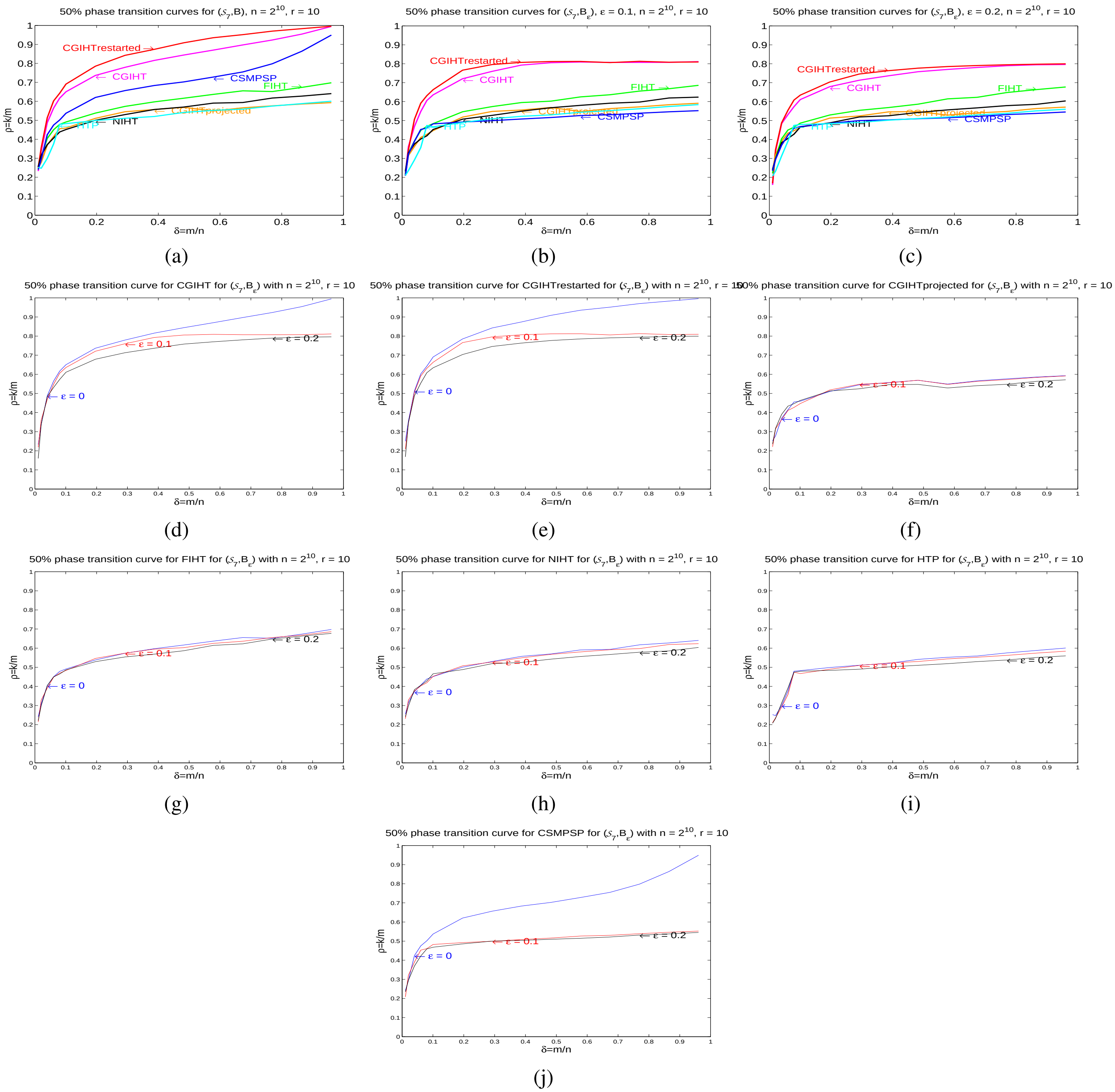
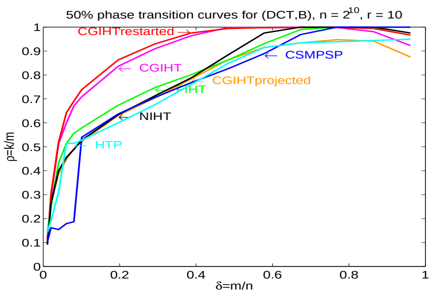
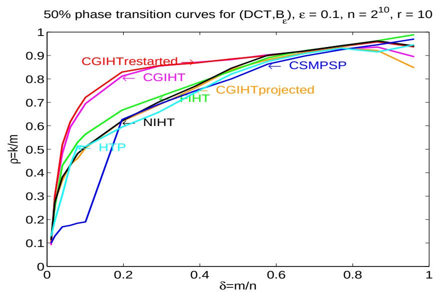


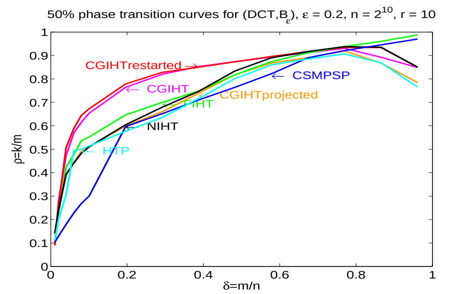
Fig. 5. 50% recovery probability logistic regression curves of all algorithms for problem classes $(\mathcal{S}_7, B_\epsilon)$ and $n = 2^{10}$, $r = 10$, with (a) $\epsilon = 0$, (b) $\epsilon = 0.1$, and (c) $\epsilon = 0.2$. Stability of 50% recovery probability logistic regression curves for each algorithm for problem classes $(\mathcal{S}_7, B_\epsilon)$ and $n = 2^{10}$, $r = 10$, with $\epsilon = 0, 0.1, 0.2$: (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP.



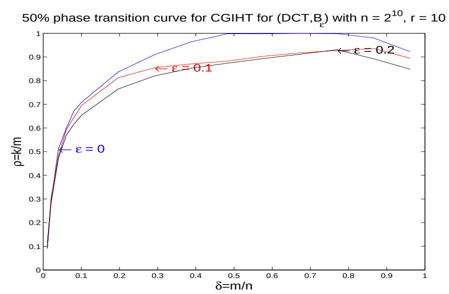
(a)



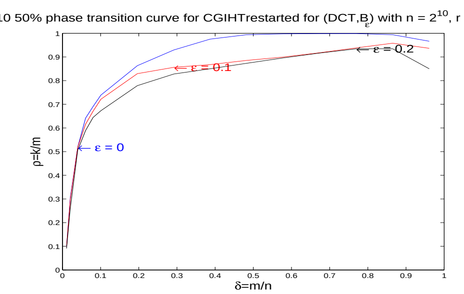
(b)



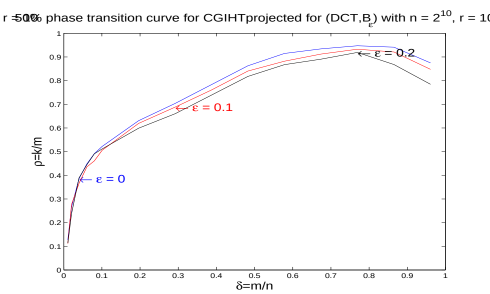
(c)



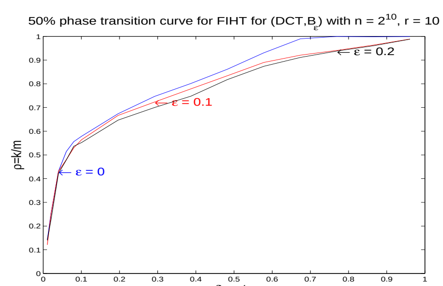
(d)



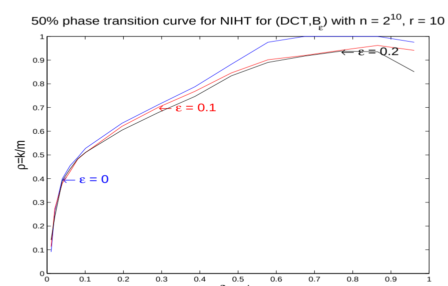
(e)



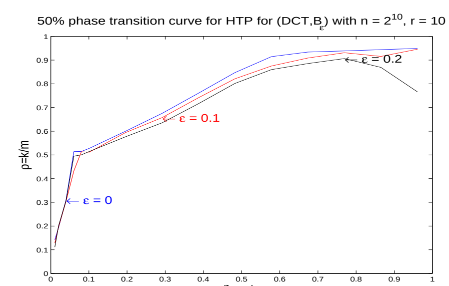
(f)



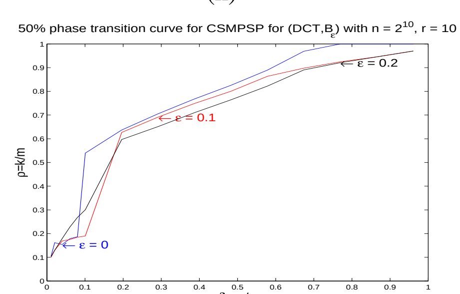
(g)



(h)



(i)



(j)

Fig. 6. 50% recovery probability logistic regression curves of all algorithms for problem classes (DCT, B_ϵ) and $n = 2^{10}$, $r = 10$, with (a) $\epsilon = 0$, (b) $\epsilon = 0.1$, and (c) $\epsilon = 0.2$. Stability of 50% recovery probability logistic regression curves for each algorithm for problem classes (DCT, B_ϵ) and $n = 2^{10}$, $r = 10$, with $\epsilon = 0, 0.1, 0.2$: (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP.

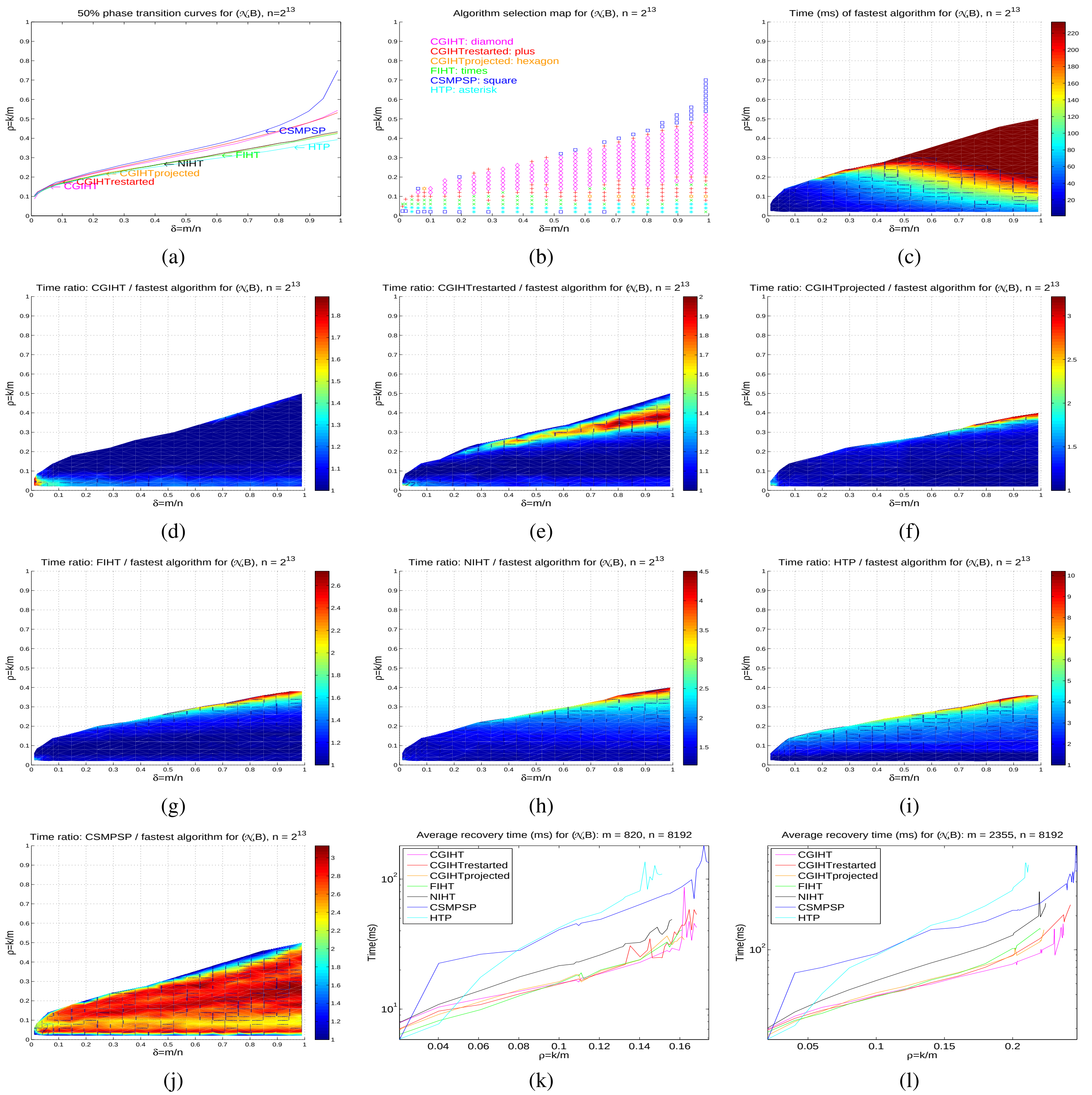


Fig. 7. Problem class $(\mathcal{N}, B)$ with $n = 2^{13}$. (a) 50% recovery probability logistic regression curves. (b) Algorithm selection map. (c) Time (ms) of least recovery time. Recovery time ratios: average recovery time over the least recovery time for (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP. Semi-log plot of average successful recovery time (ms) for all algorithms: (k) $\delta \approx 0.1$, (l) $\delta \approx 0.3$.

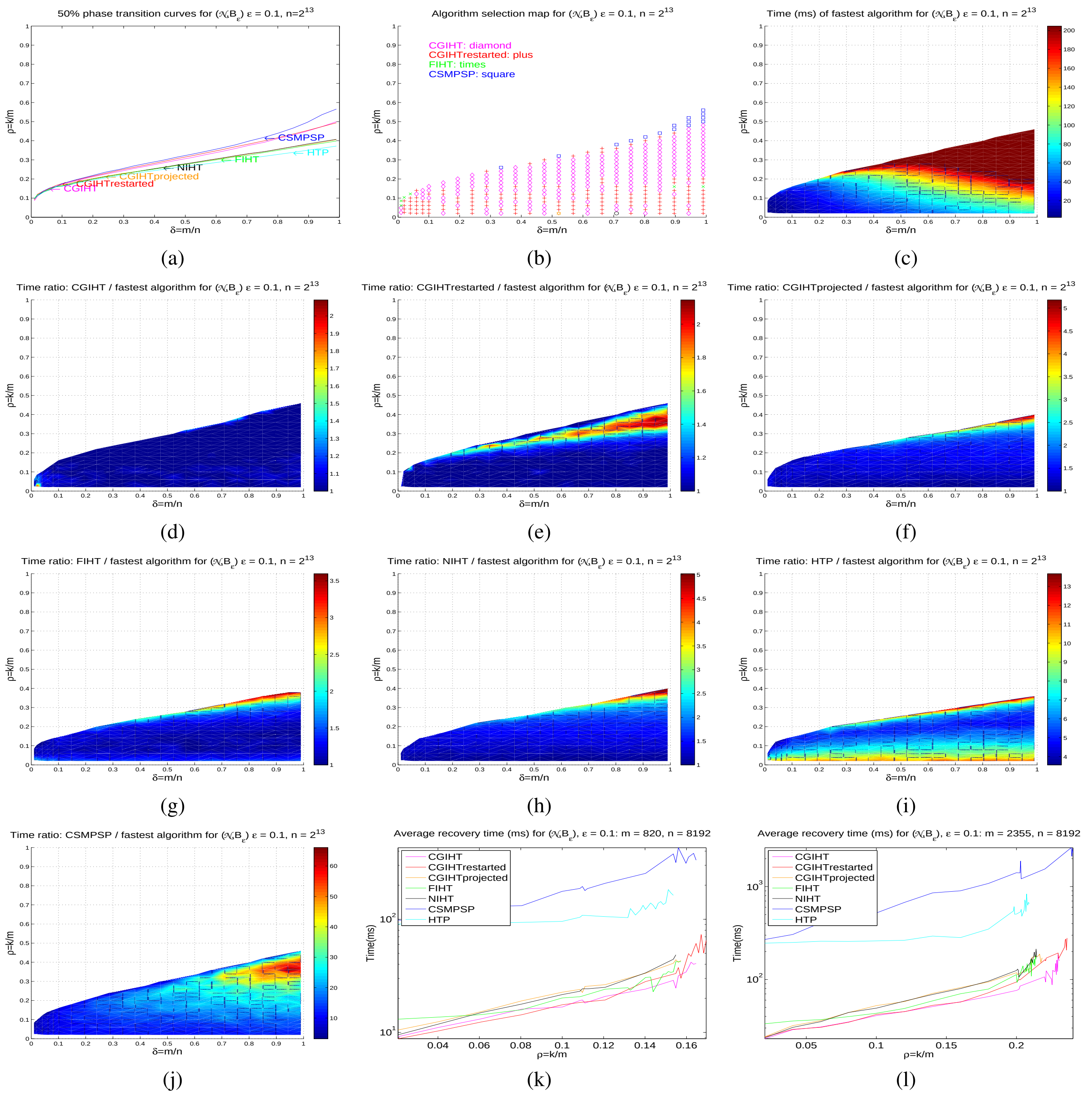


Fig. 8. Problem class $(\mathcal{N}, B_\epsilon)$ with $\epsilon = 0.1$ and $n = 2^{13}$. (a) 50% recovery probability logistic regression curves. (b) Algorithm selection map. (c) Time (ms) of least recovery time. Recovery time ratios: average recovery time over the least recovery time for (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP. Semi-log plot of average successful recovery time (ms) for all algorithms: (k) $\delta \approx 0.1$, (l) $\delta \approx 0.3$.

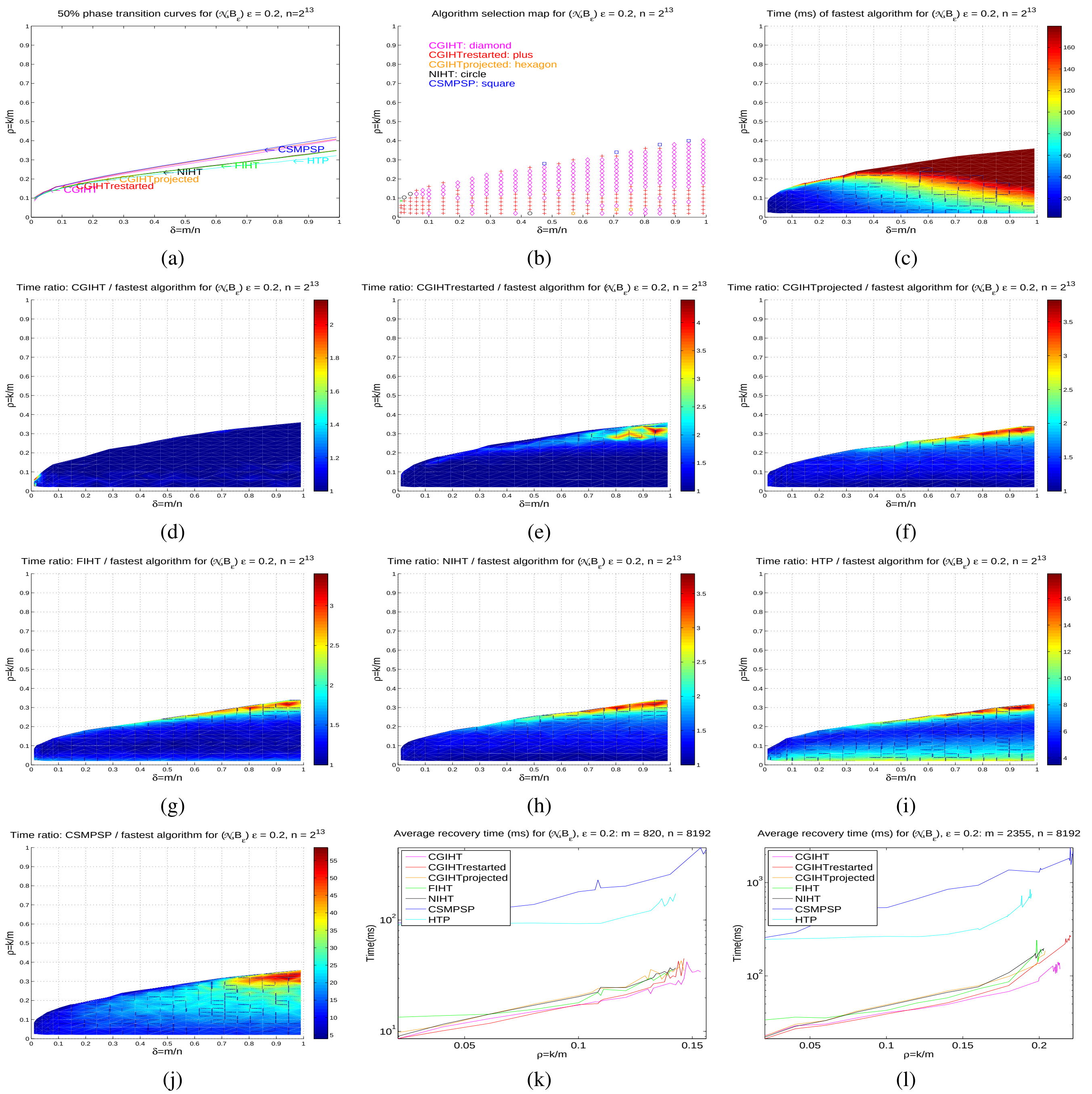


Fig. 9. Problem class $(\mathcal{N}, B_\epsilon)$ with $\epsilon = 0.2$ and $n = 2^{13}$. (a) 50% recovery probability logistic regression curves. (b) Algorithm selection map. (c) Time (ms) of least recovery time. Recovery time ratios: average recovery time over the least recovery time for (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP. Semi-log plot of average successful recovery time (ms) for all algorithms: (k) $\delta \approx 0.1$, (l) $\delta \approx 0.3$.

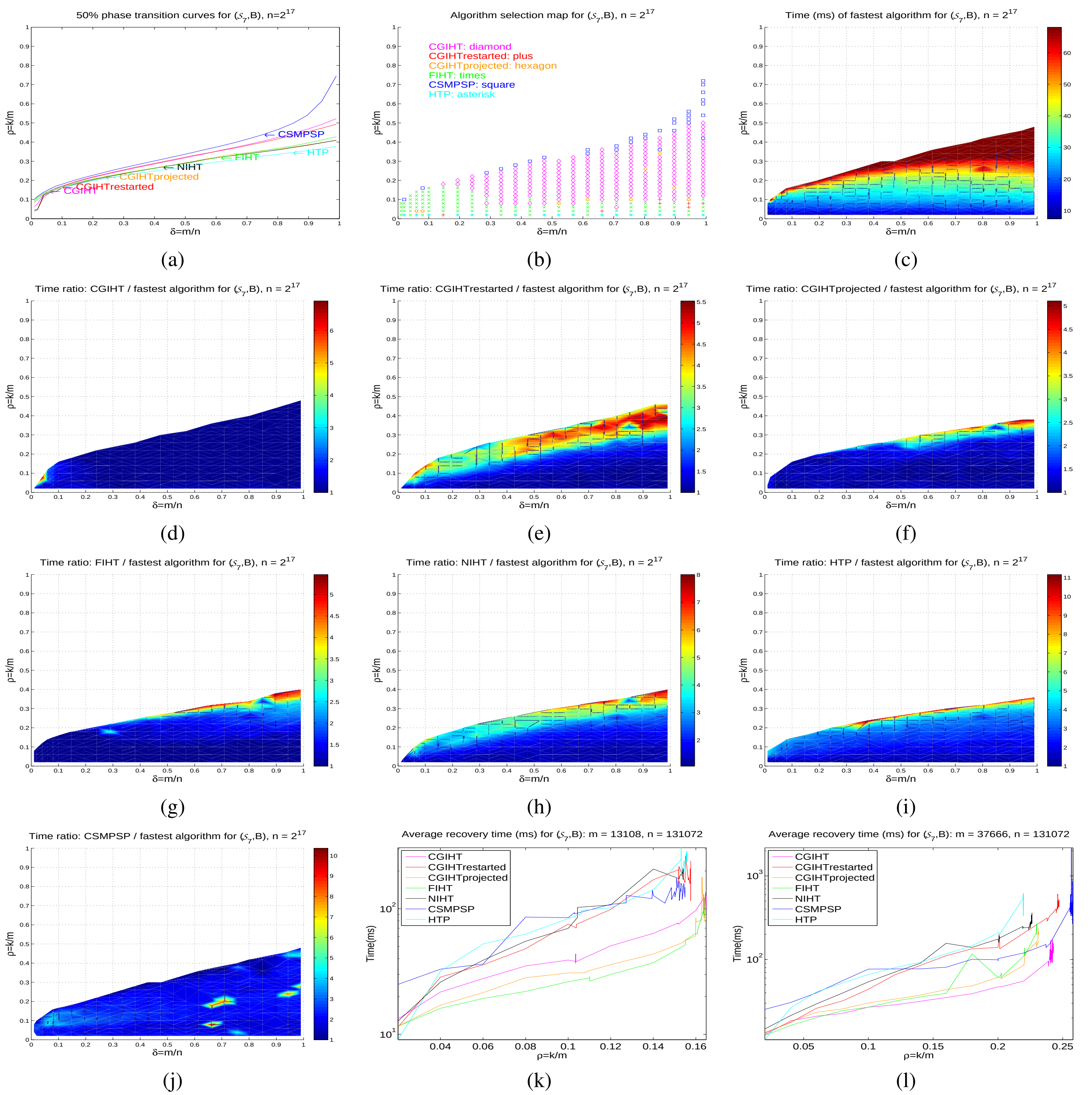


Fig. 10. Problem class $(\mathcal{S}_7, B)$ with $n = 2^{17}$. (a) 50% recovery probability logistic regression curves. (b) Algorithm selection map. (c) Time (ms) of least recovery time. Recovery time ratios: average recovery time over the least recovery time for (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP. Semi-log plot of average successful recovery time (ms) for all algorithms: (k) $\delta \approx 0.1$, (l) $\delta \approx 0.3$.

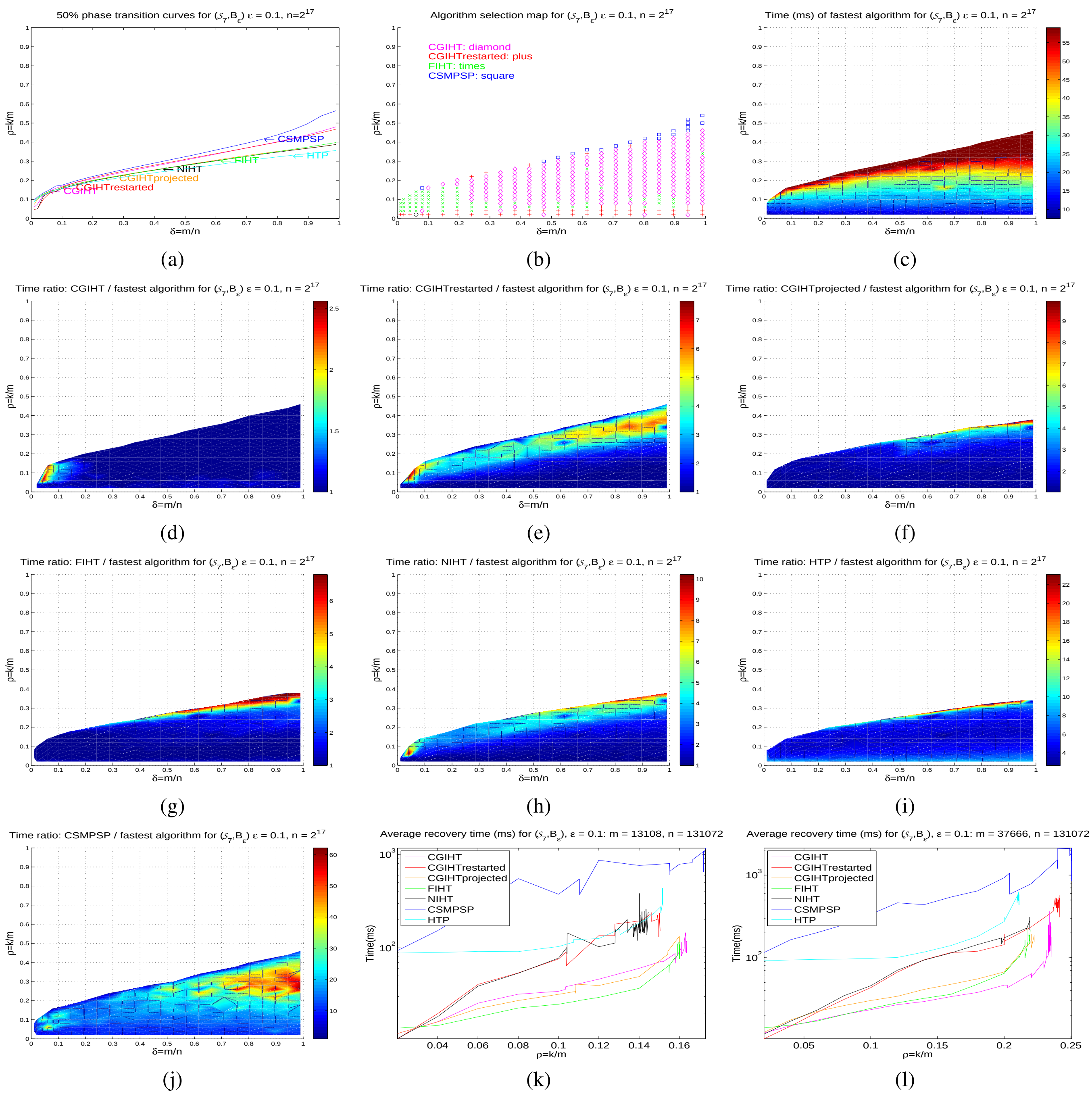


Fig. 11. Problem class $(\mathcal{S}_7, B_\epsilon)$ with $\epsilon = 0.1$ and $n = 2^{17}$. (a) 50% recovery probability logistic regression curves. (b) Algorithm selection map. (c) Time (ms) of least recovery time. Recovery time ratios: average recovery time over the least recovery time for (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP. Semi-log plot of average successful recovery time (ms) for all algorithms: (k) $\delta \approx 0.1$, (l) $\delta \approx 0.3$.

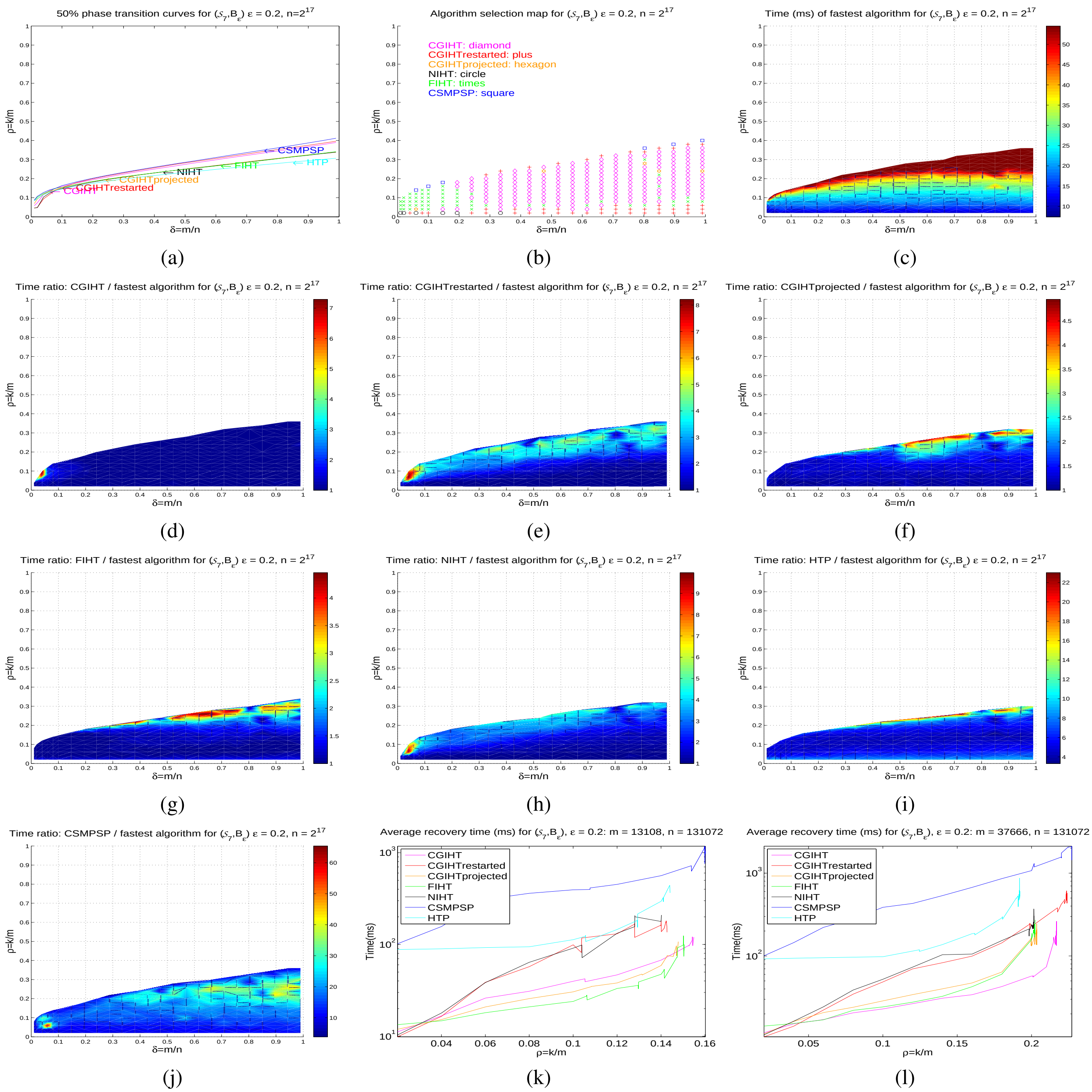


Fig. 12. Problem class $(\mathcal{S}_7, B_\epsilon)$ with $\epsilon = 0.2$ and $n = 2^{17}$. (a) 50% recovery probability logistic regression curves. (b) Algorithm selection map. (c) Time (ms) of least recovery time. Recovery time ratios: average recovery time over the least recovery time for (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP. Semi-log plot of average successful recovery time (ms) for all algorithms: (k) $\delta \approx 0.1$, (l) $\delta \approx 0.3$.

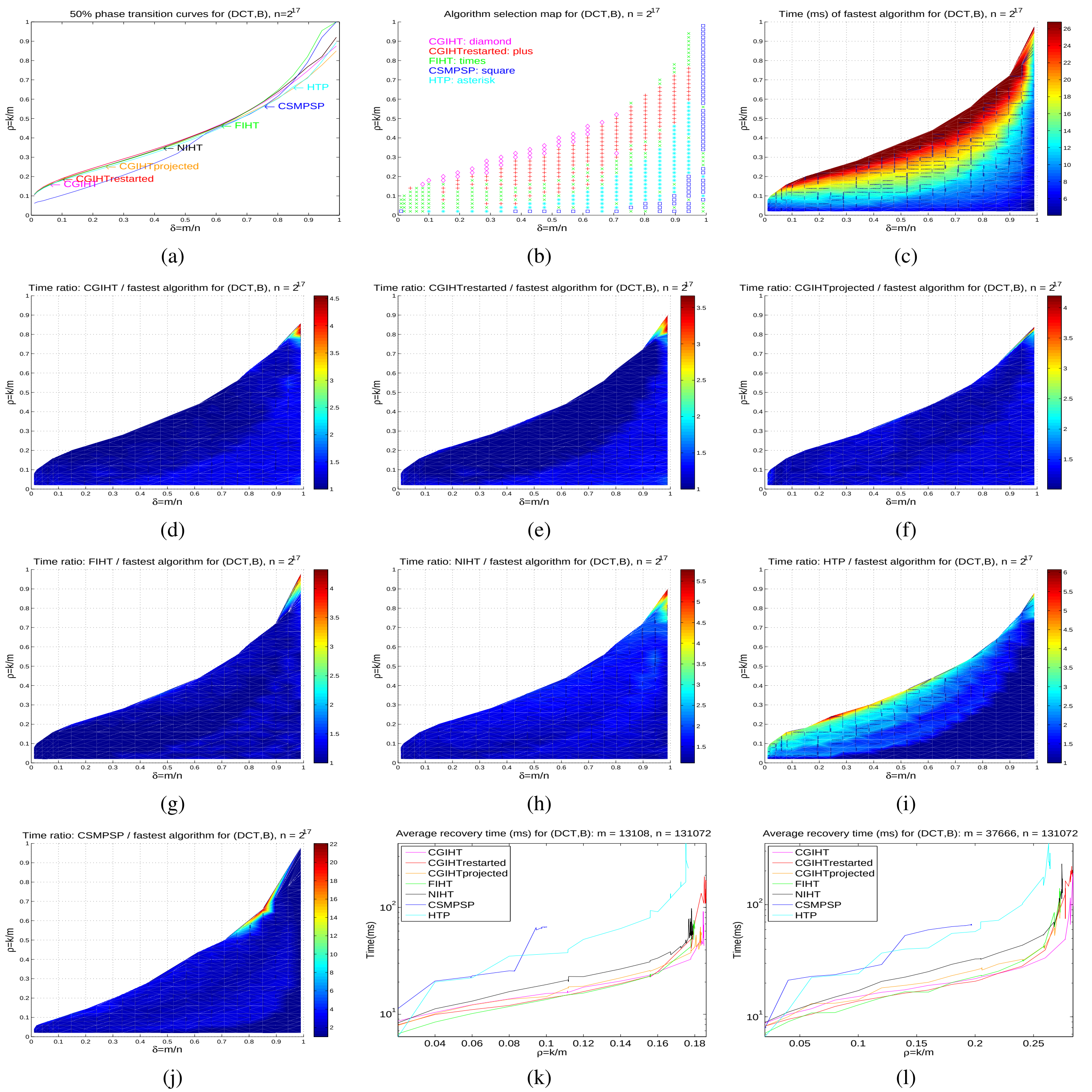


Fig. 13. Problem class (DCT, B) with $n = 2^{17}$. (a) 50% recovery probability logistic regression curves. (b) Algorithm selection map. (c) Time (ms) of least recovery time. Recovery time ratios: average recovery time over the least recovery time for (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP. Semi-log plot of average successful recovery time (ms) for all algorithms: (k) $\delta \approx 0.1$, (l) $\delta \approx 0.3$.

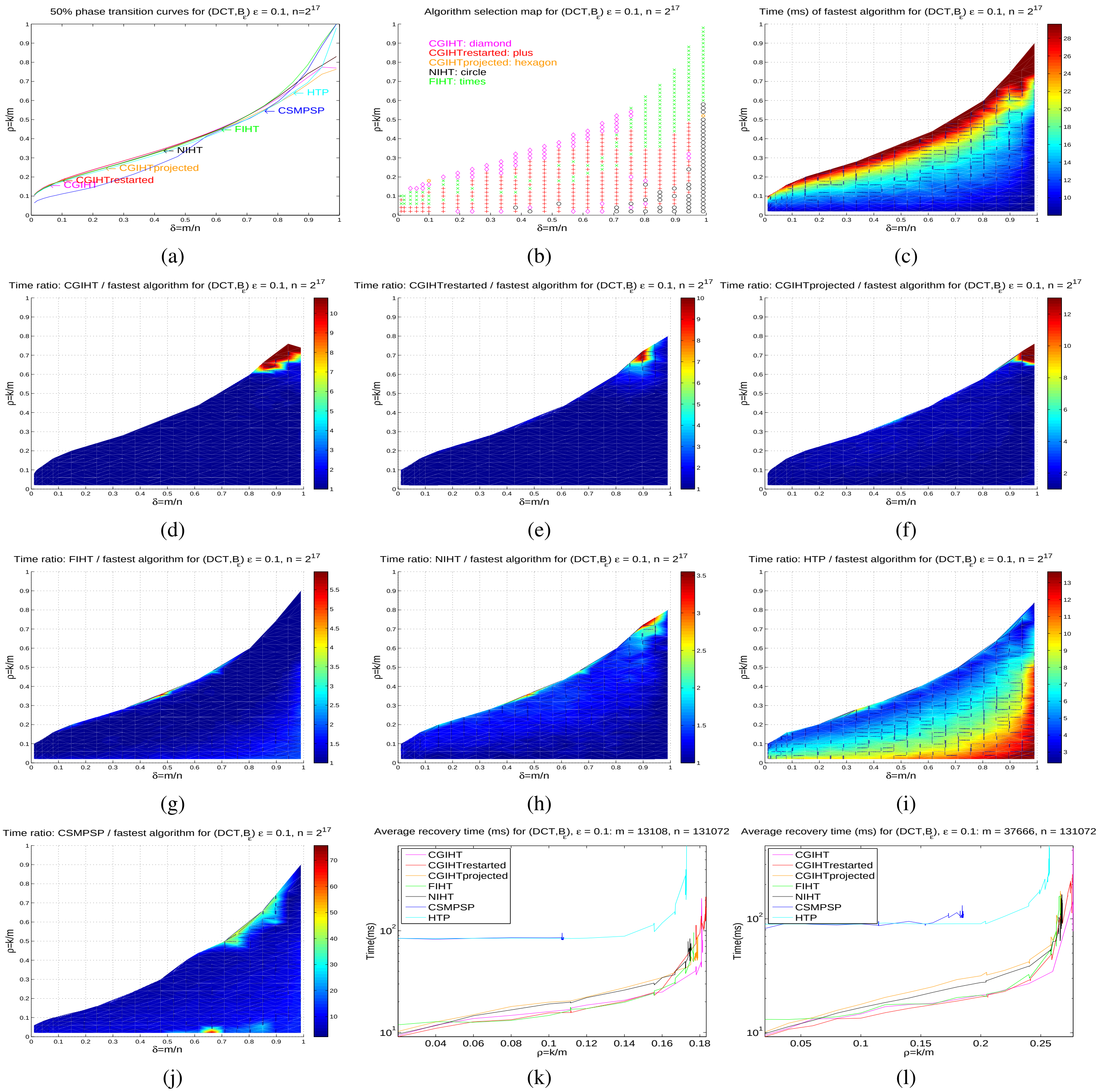


Fig. 14. Problem class (DCT, B_ϵ) with $\epsilon = 0.1$ and $n = 2^{17}$. (a) 50% recovery probability logistic regression curves. (b) Algorithm selection map. (c) Time (ms) of least recovery time. Recovery time ratios: average recovery time over the least recovery time for (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP. Semi-log plot of average successful recovery time (ms) for all algorithms: (k) $\delta \approx 0.1$, (l) $\delta \approx 0.3$.

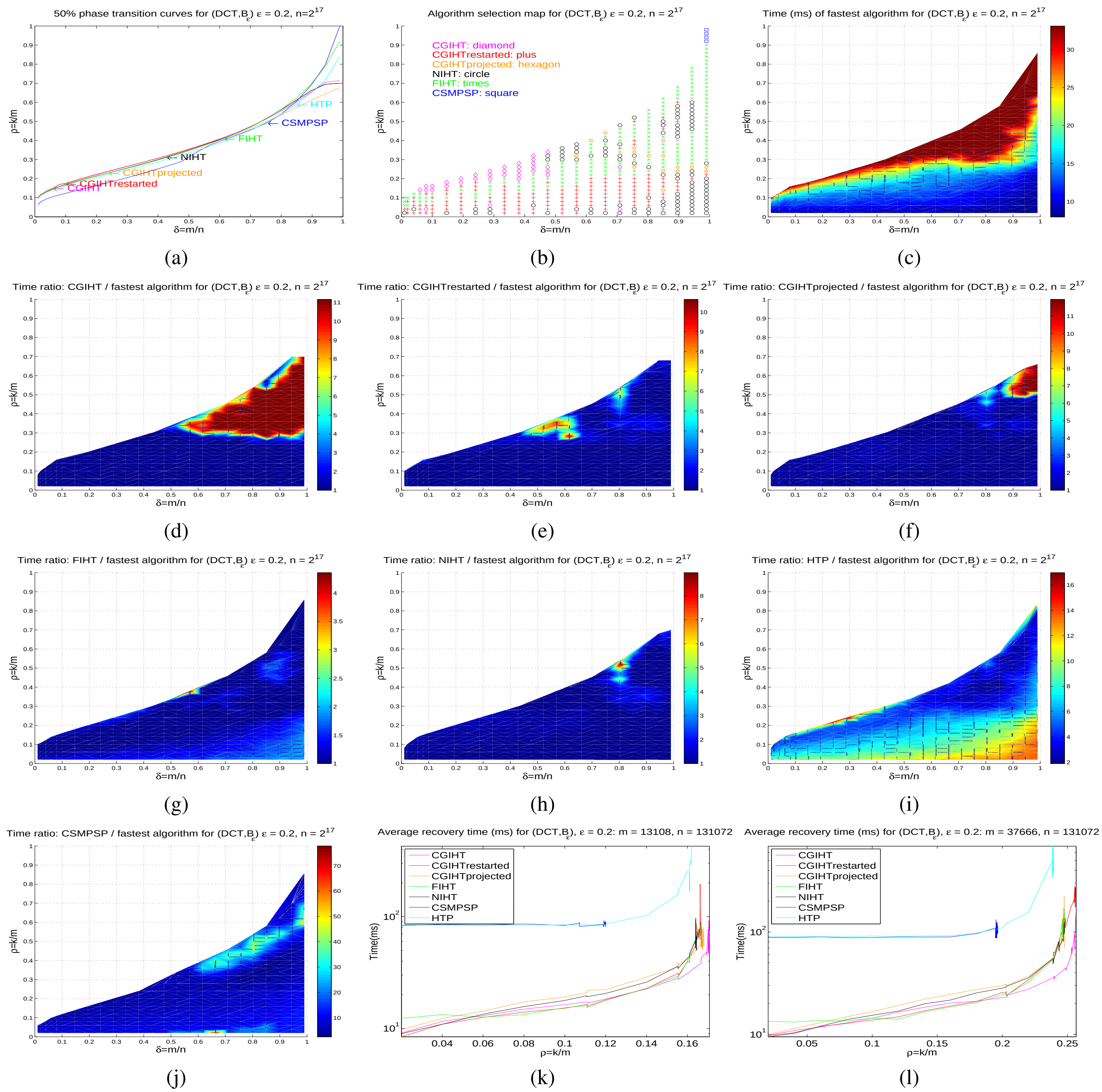


Fig. 15. Problem class (DCT, B_ϵ) with $\epsilon = 0.2$ and $n = 2^{17}$. (a) 50% recovery probability logistic regression curves. (b) Algorithm selection map. (c) Time (ms) of least recovery time. Recovery time ratios: average recovery time over the least recovery time for (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP. Semi-log plot of average successful recovery time (ms) for all algorithms: (k) $\delta \approx 0.1$, (l) $\delta \approx 0.3$.

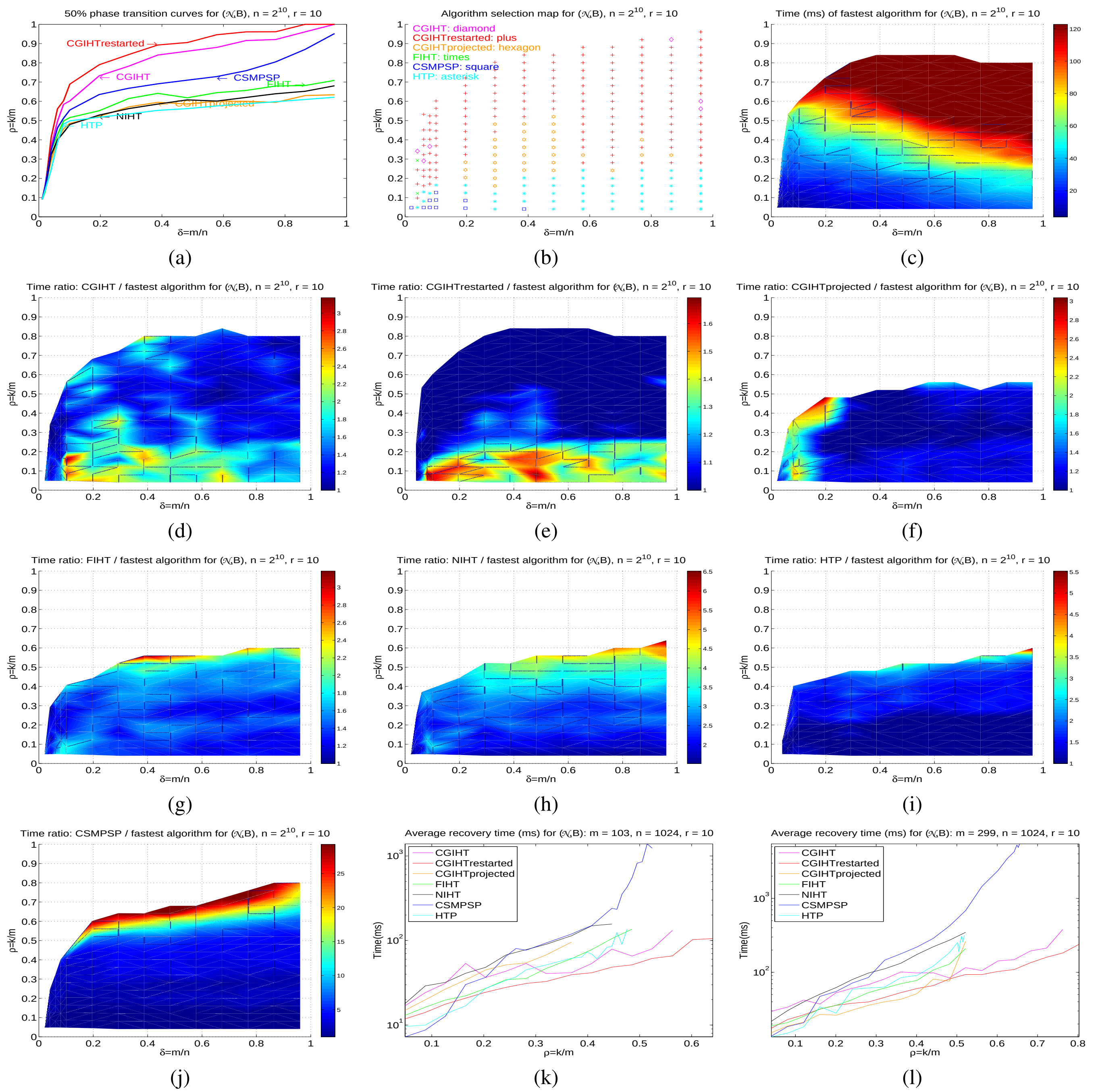


Fig. 16. Problem class $(\mathcal{N}, B)$ with $n = 2^{10}$ and $r = 10$. (a) 50% recovery probability logistic regression curves. (b) Algorithm selection map. (c) Time (ms) of least recovery time. Recovery time ratios: average recovery time over the least recovery time for (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP. Semi-log plot of average successful recovery time (ms) for all algorithms: (k) $\delta \approx 0.1$, (l) $\delta \approx 0.3$.

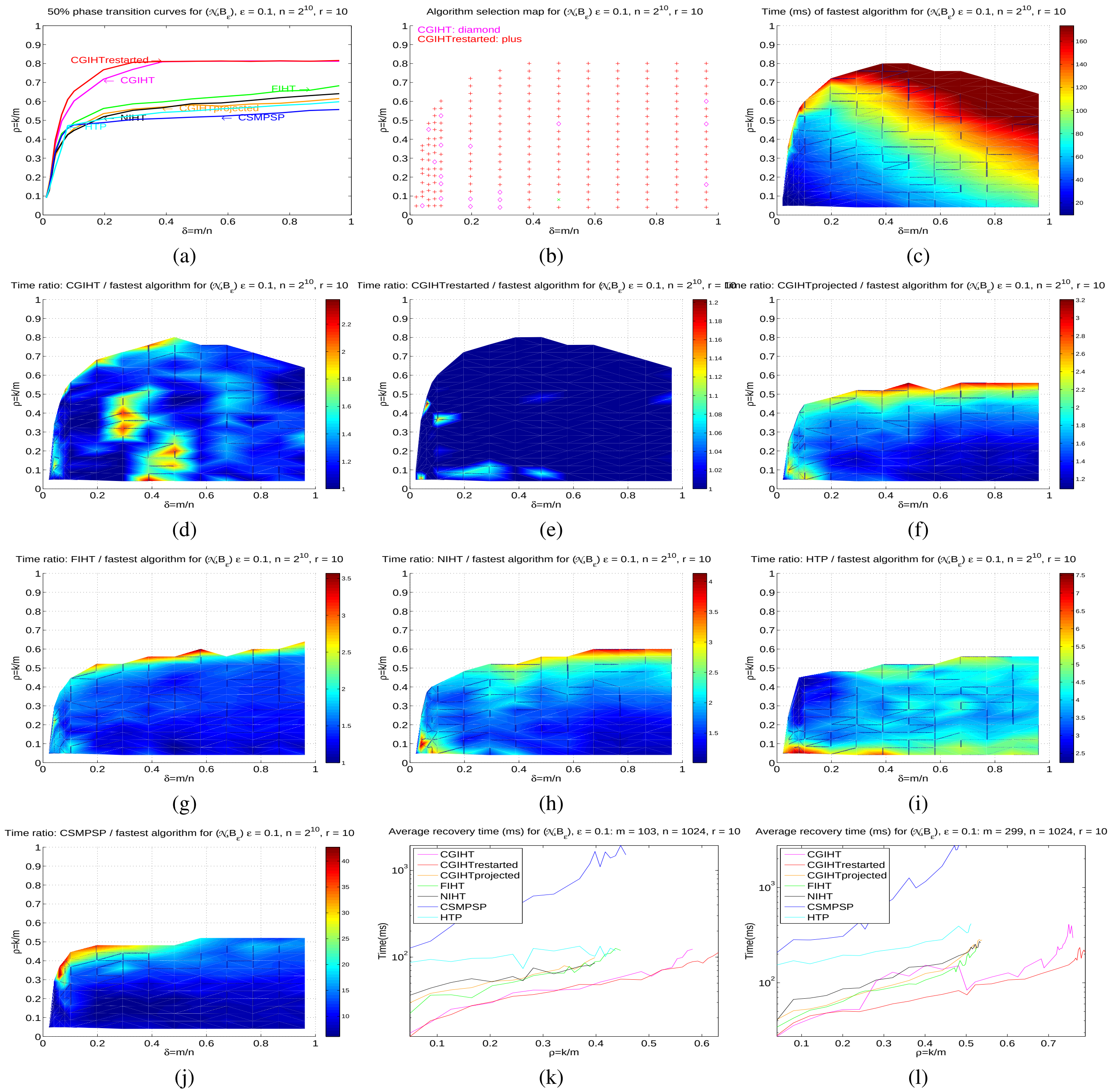


Fig. 17. Problem class $(\mathcal{N}, B_\epsilon)$ with $\epsilon = 0.1$ and $n = 2^{10}$ and $r = 10$. (a) 50% recovery probability logistic regression curves. (b) Algorithm selection map. (c) Time (ms) of least recovery time. Recovery time ratios: average recovery time over the least recovery time for (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP. Semi-log plot of average successful recovery time (ms) for all algorithms: (k) $\delta \approx 0.1$, (l) $\delta \approx 0.3$.

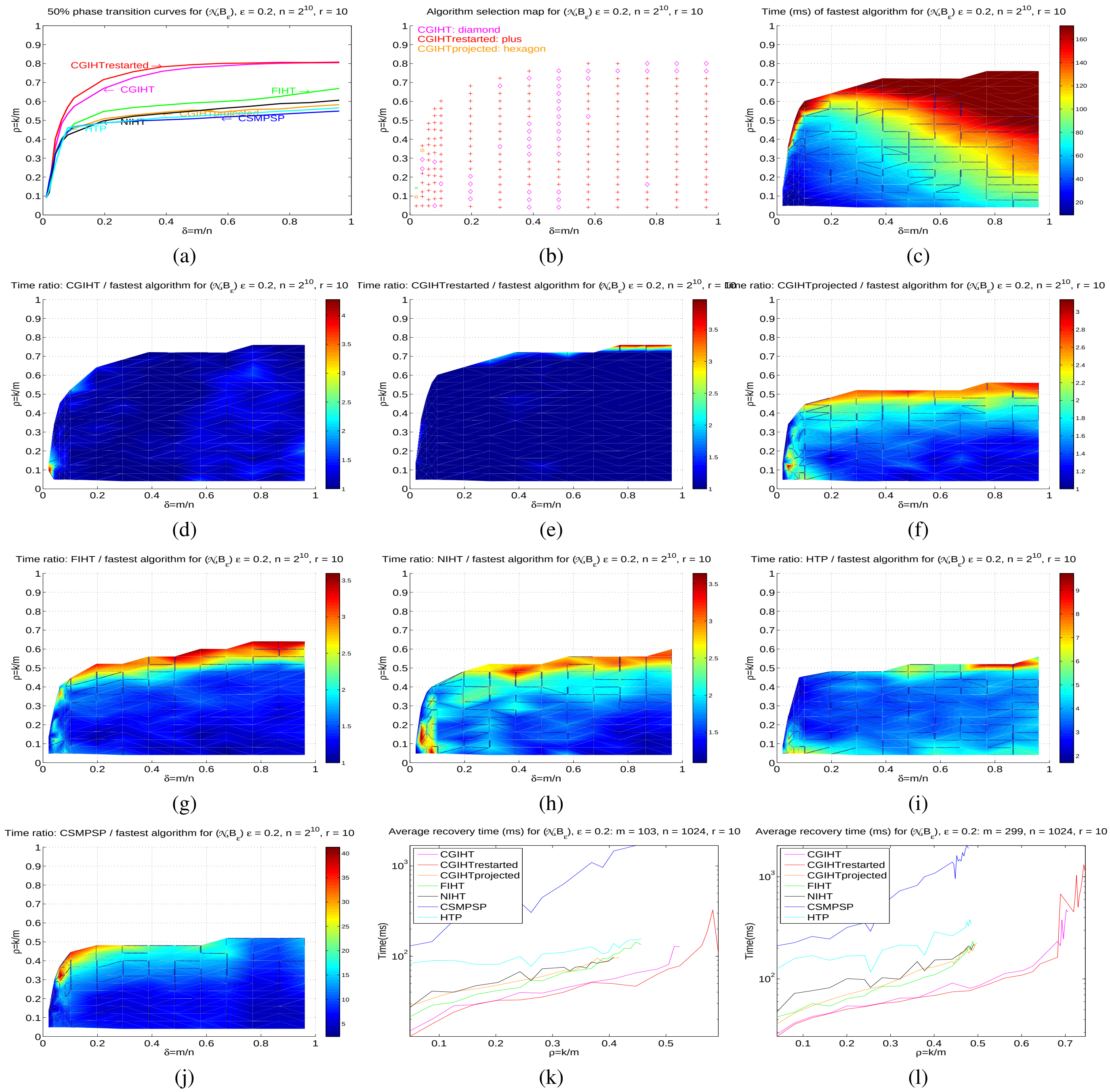


Fig. 18. Problem class $(\mathcal{N}, B_\epsilon)$ with $\epsilon = 0.2$ and $n = 2^{10}$ and $r = 10$. (a) 50% recovery probability logistic regression curves. (b) Algorithm selection map. (c) Time (ms) of least recovery time. Recovery time ratios: average recovery time over the least recovery time for (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP. Semi-log plot of average successful recovery time (ms) for all algorithms: (k) $\delta \approx 0.1$, (l) $\delta \approx 0.3$.

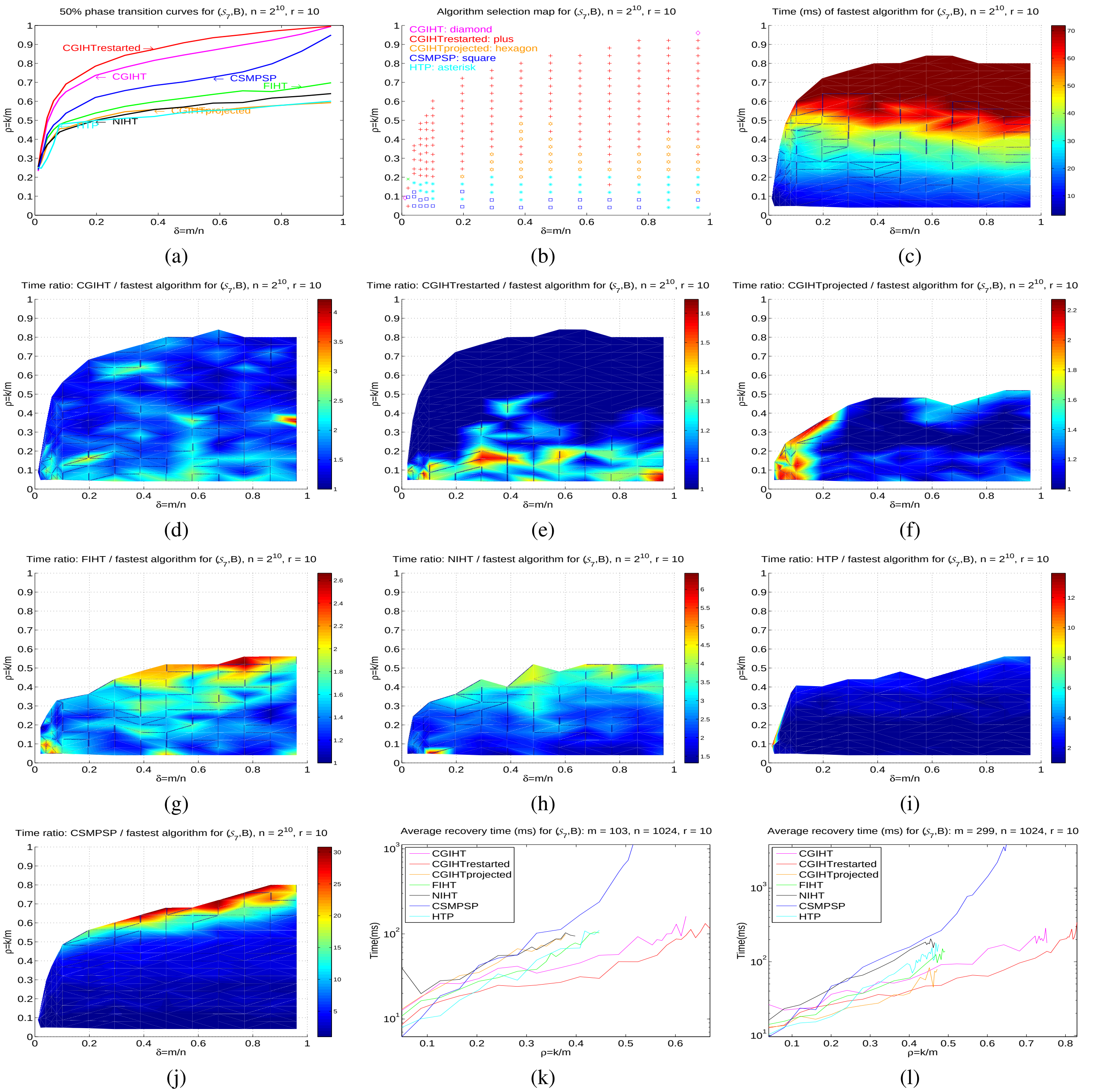


Fig. 19. Problem class $(\mathcal{S}_7, B)$ with $n = 2^{10}$ and $r = 10$. (a) 50% recovery probability logistic regression curves. (b) Algorithm selection map. (c) Time (ms) of least recovery time. Recovery time ratios: average recovery time over the least recovery time for (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP. Semi-log plot of average successful recovery time (ms) for all algorithms: (k) $\delta \approx 0.1$, (l) $\delta \approx 0.3$.

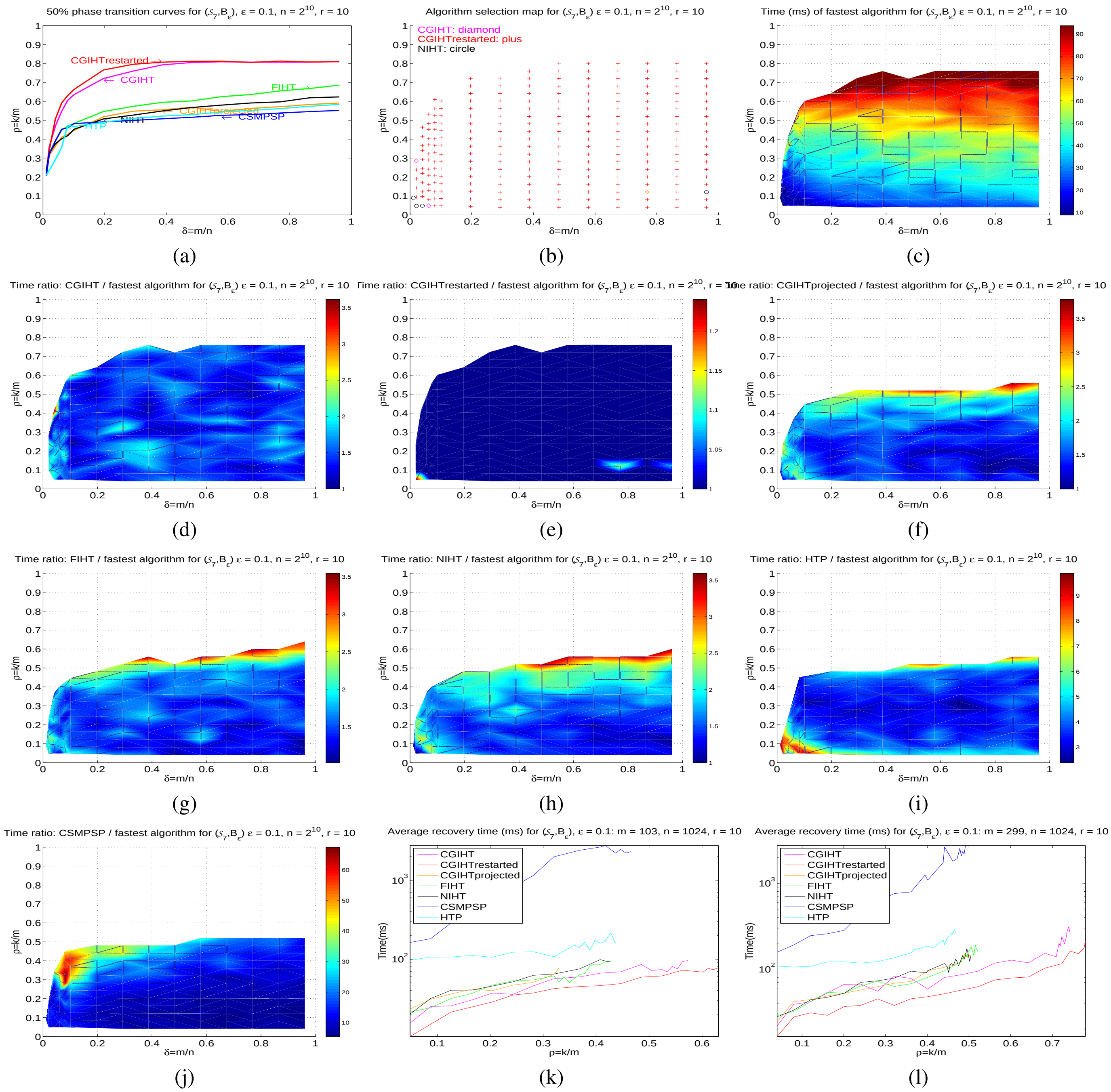


Fig. 20. Problem class $(\mathcal{S}_7, B_\epsilon)$ with $\epsilon = 0.1$ and $n = 2^{10}$ and $r = 10$. (a) 50% recovery probability logistic regression curves. (b) Algorithm selection map. (c) Time (ms) of least recovery time. Recovery time ratios: average recovery time over the least recovery time for (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP. Semi-log plot of average successful recovery time (ms) for all algorithms: (k) $\delta \approx 0.1$, (l) $\delta \approx 0.3$.

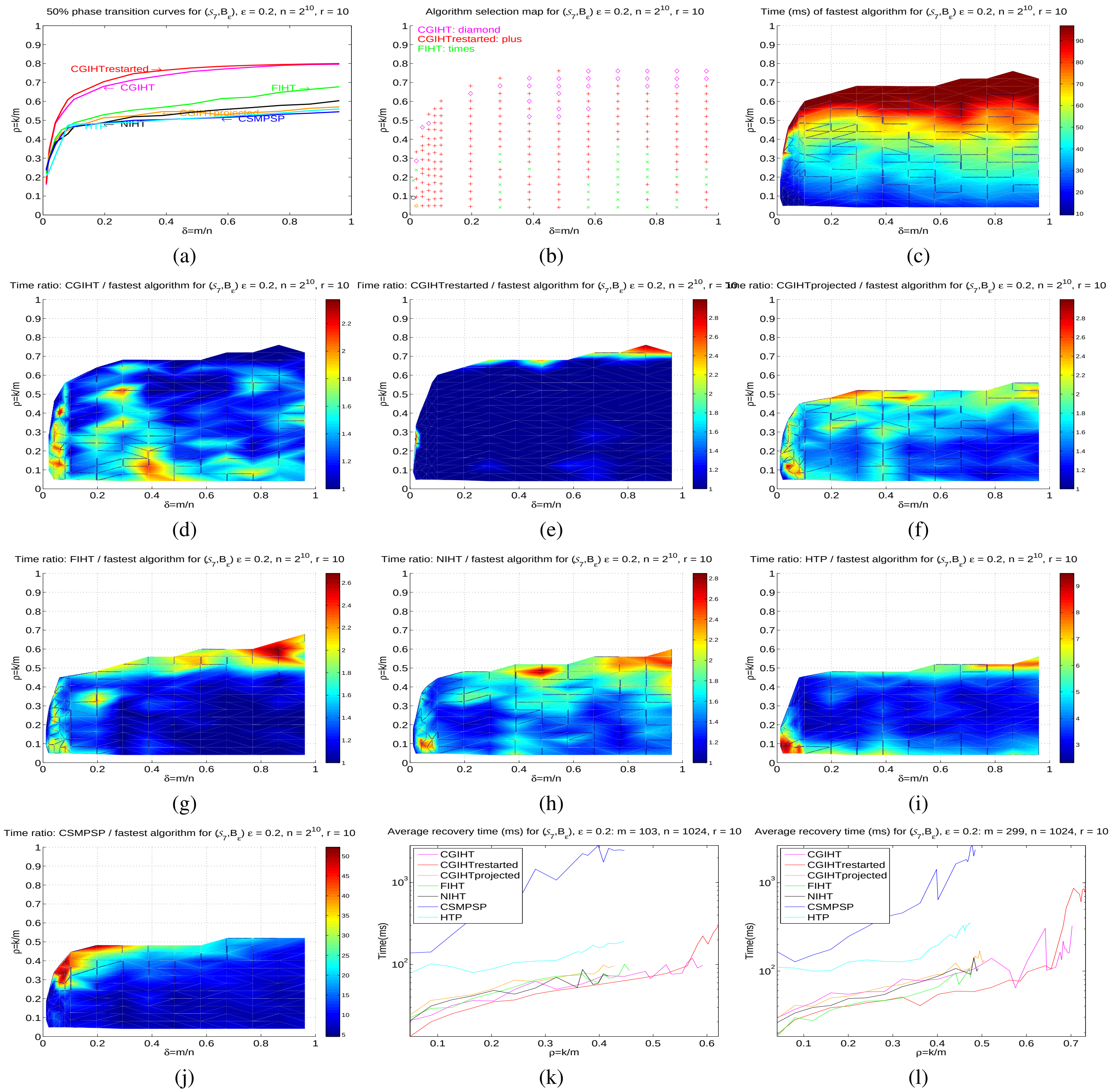


Fig. 21. Problem class $(\mathcal{S}_7, B_\epsilon)$ with $\epsilon = 0.2$ and $n = 2^{10}$ and $r = 10$. (a) 50% recovery probability logistic regression curves. (b) Algorithm selection map. (c) Time (ms) of least recovery time. Recovery time ratios: average recovery time over the least recovery time for (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP. Semi-log plot of average successful recovery time (ms) for all algorithms: (k) $\delta \approx 0.1$, (l) $\delta \approx 0.3$.

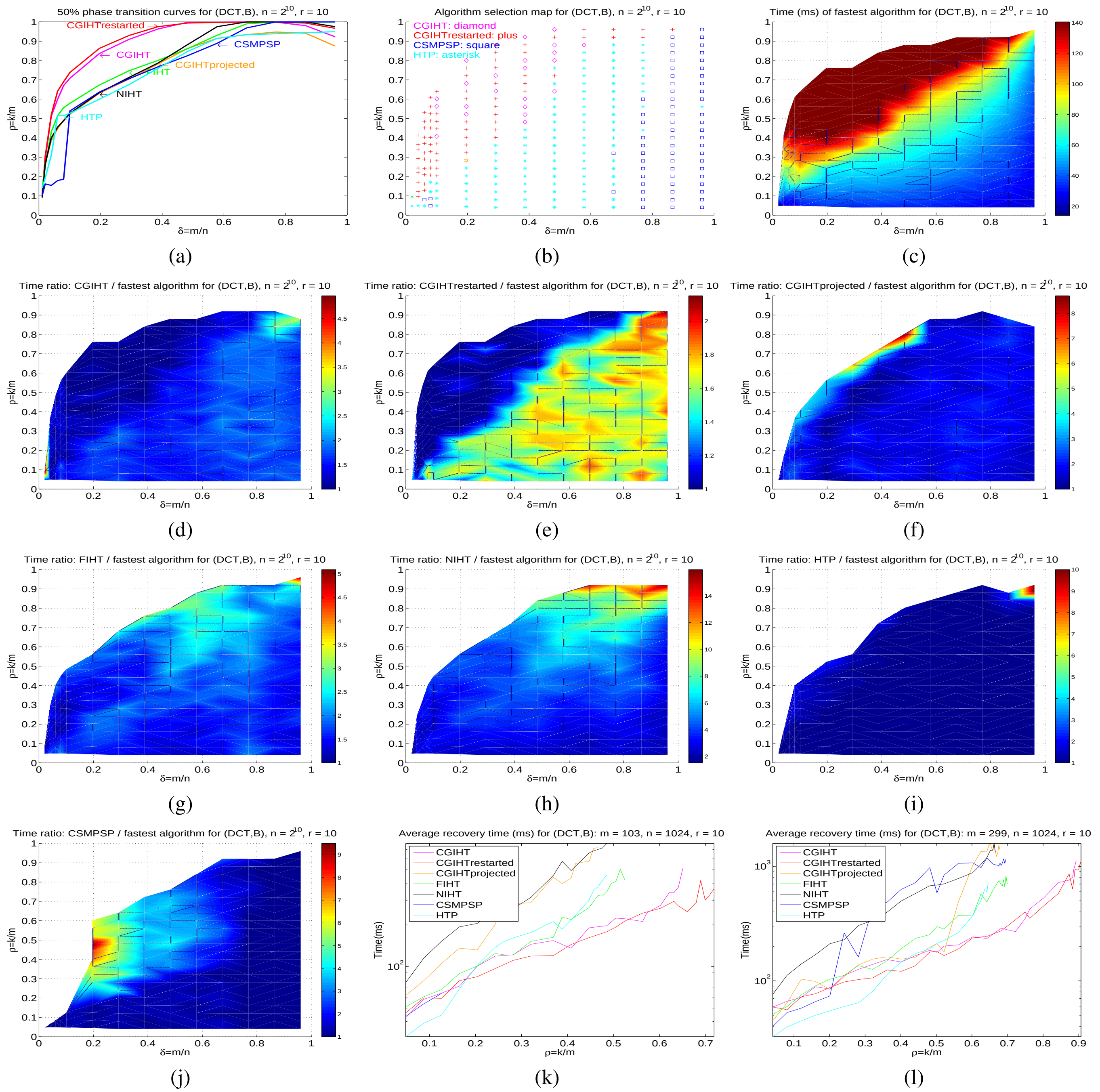


Fig. 22. Problem class (DCT, B) with $n = 2^{10}$ and $r = 10$. (a) 50% recovery probability logistic regression curves. (b) Algorithm selection map. (c) Time (ms) of least recovery time. Recovery time ratios: average recovery time over the least recovery time for (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP. Semi-log plot of average successful recovery time (ms) for all algorithms: (k) $\delta \approx 0.1$, (l) $\delta \approx 0.3$.

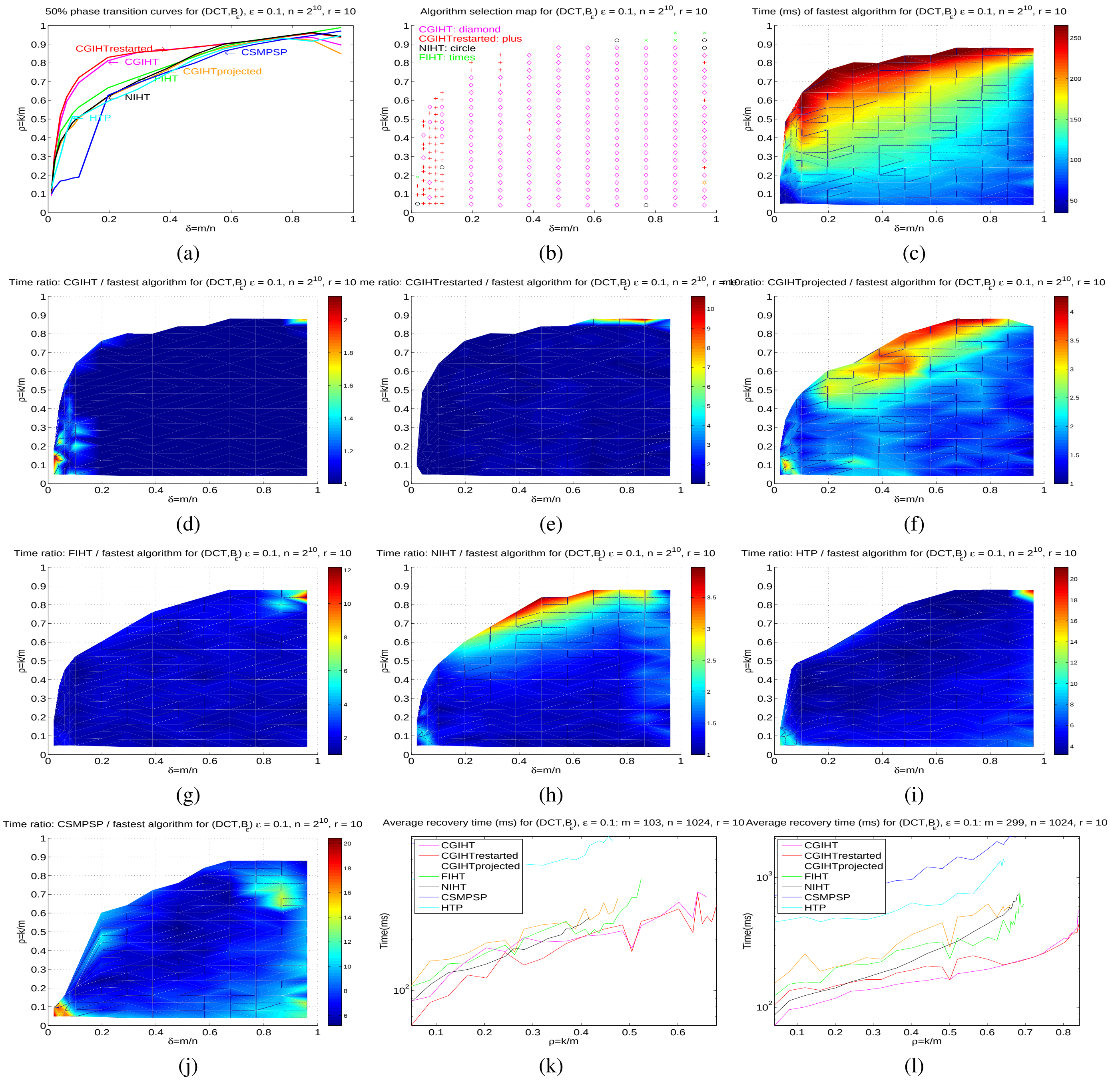


Fig. 23. Problem class (DCT, B_ϵ) with $\epsilon = 0.1$ and $n = 2^{10}$ and $r = 10$. (a) 50% recovery probability logistic regression curves. (b) Algorithm selection map. (c) Time (ms) of least recovery time. Recovery time ratios: average recovery time over the least recovery time for (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP. Semi-log plot of average successful recovery time (ms) for all algorithms: (k) $\delta \approx 0.1$, (l) $\delta \approx 0.3$.

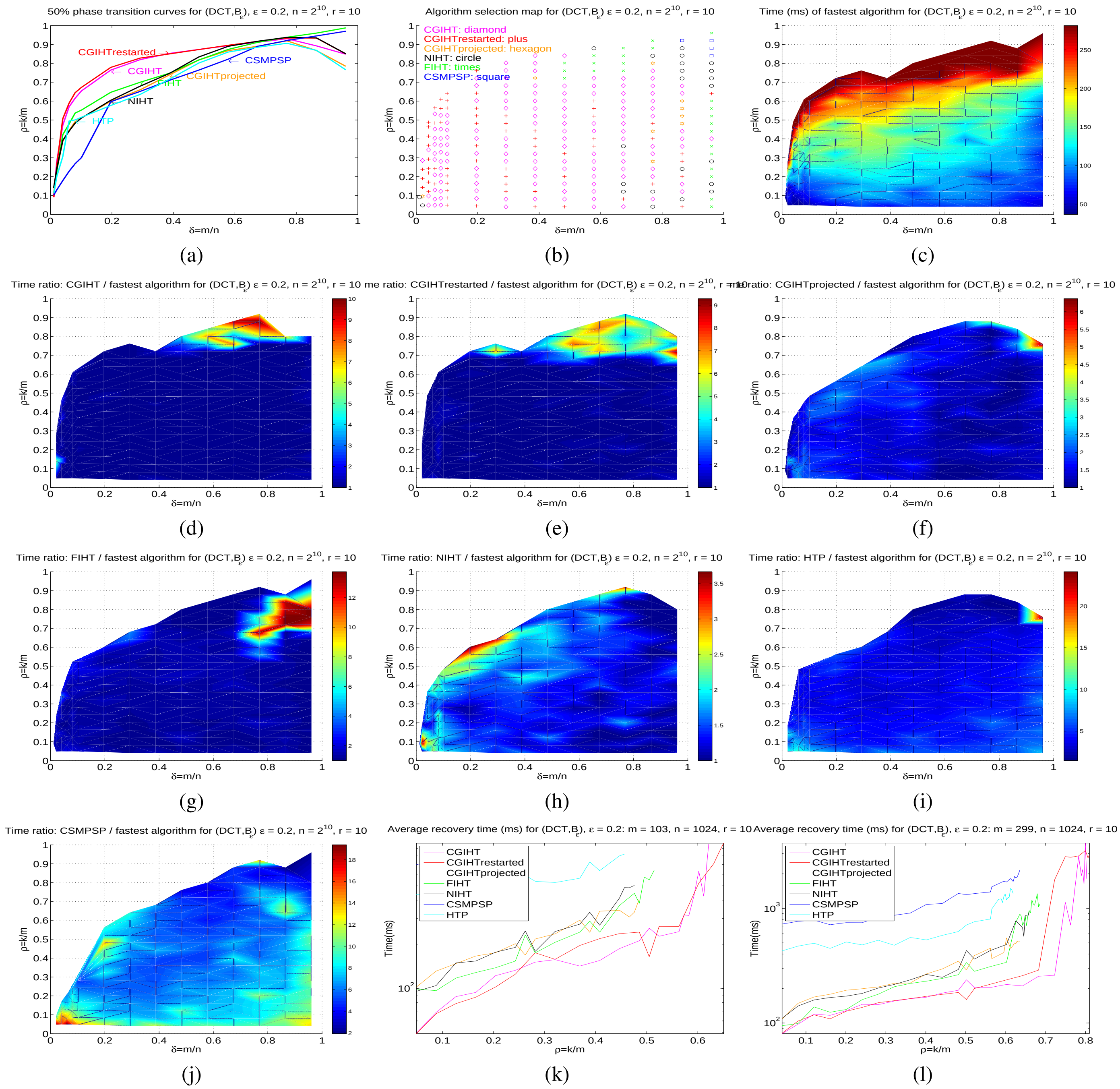


Fig. 24. Problem class (DCT, B_ϵ) with $\epsilon = 0.2$ and $n = 2^{10}$ and $r = 10$. (a) 50% recovery probability logistic regression curves. (b) Algorithm selection map. (c) Time (ms) of least recovery time. Recovery time ratios: average recovery time over the least recovery time for (d) CGIHT, (e) CGIHT restarted, (f) CGIHT projected, (g) FIHT, (h) NIHT, (i) HTP, (j) CSMPSP. Semi-log plot of average successful recovery time (ms) for all algorithms: (k) $\delta \approx 0.1$, (l) $\delta \approx 0.3$.