Supplementary Material

Appendix A
Bound in Remark 1

The bound follows directly from the application of a well-known Poincaré–Wirtinger inequality [32, Sec. 5.8.1], which yields:

$$\int_{\mathcal{A}_k} \|\mu^\star(\mathbf{x}) - \mu^\star_{\mathrm{d}}(\mathbf{x})\|\, d\mathbf{x} \leq \frac{\mathrm{diam}(\mathcal{A}_k)}{2} \int_{\mathcal{A}_k} |\nabla \mu^\star(\mathbf{x})| d\mathbf{x}. \quad (1)$$

Multiplying both sides by N, summing over $k = 1, \ldots, D$, and using the bound $\mathrm{diam}(\mathcal{A}_k) \leq \sqrt{\frac{2A}{D}}$ for square-sub-regions, we obtain

$$\sum_{k=1}^{D} \int_{\mathcal{A}_k} \|\lambda^\star(\mathbf{x}) - \lambda^\star_{\mathrm{d}}(\mathbf{x})\|\, d\mathbf{x} \leq N B_\star \sqrt{\frac{A}{2D}} \quad (2)$$

Substituting into (23) and proceeding as in the proof of Lemma 1, we obtain

$$E_{\mathrm{d}} \leq \frac{LNAB_\star}{D} \quad (3)$$

which is the required result.

Appendix B
Proof of Lemma 2

The proof follows along the lines of the symmetrization result in [20, Sec. 4.5]. Let $\{\mathbf{z}'_m\}_{m=1}^M$ be an independent copy of the training sample with the same distribution as $\{\mathbf{z}_m\}_{m=1}^M$. Defining

$$\hat{R}'(\boldsymbol{\lambda}_{\mathrm{d}}) := \frac{1}{M} \sum_{m=1}^{M} r(\boldsymbol{\lambda}_{\mathrm{d}}, \mathbf{z}'_m) \quad (4)$$

and using $\mathbb{E}'$ to denote the expectation with respect to $\{\mathbf{z}'_m\}_{m=1}^M$, we can write

$$\begin{aligned}\max_{\boldsymbol{\lambda}_{\mathrm{d}} \in \boldsymbol{\Lambda}} \left(R(\boldsymbol{\lambda}_{\mathrm{d}}) - \hat{R}(\boldsymbol{\lambda}_{\mathrm{d}})\right) &= \max_{\boldsymbol{\lambda}_{\mathrm{d}} \in \boldsymbol{\Lambda}} \left(\mathbb{E}'[\hat{R}'(\boldsymbol{\lambda}_{\mathrm{d}})] - \hat{R}(\boldsymbol{\lambda}_{\mathrm{d}})\right) \\ &= \max_{\boldsymbol{\lambda}_{\mathrm{d}} \in \boldsymbol{\Lambda}} \left(\mathbb{E}'[\hat{R}'(\boldsymbol{\lambda}_{\mathrm{d}}) - \hat{R}(\boldsymbol{\lambda}_{\mathrm{d}})]\right) \\ &\leq \mathbb{E}'[\max_{\boldsymbol{\lambda}_{\mathrm{d}} \in \boldsymbol{\Lambda}} \left(\hat{R}'(\boldsymbol{\lambda}_{\mathrm{d}}) - \hat{R}(\boldsymbol{\lambda}_{\mathrm{d}})\right)] \quad (5)\end{aligned}$$

where we have used the convexity of the max function. Next, we use the observation that a symmetric random variable X has the same distribution as εX, where ε is an independent Rademacher random variable. Applying it to $X = r(\boldsymbol{\lambda}, \mathbf{z}'_m) - r(\boldsymbol{\lambda}, \mathbf{z}_m)$, substituting into (5), and taking expectation with respect to $\{\mathbf{z}_m\}$, we obtain

$$\begin{aligned}&\mathbb{E}[\mathbb{E}'[\max_{\boldsymbol{\lambda}_{\mathrm{d}} \in \boldsymbol{\Lambda}} \left(\hat{R}'(\boldsymbol{\lambda}_{\mathrm{d}}) - \hat{R}(\boldsymbol{\lambda}_{\mathrm{d}})\right)]] \\ &= \mathbb{E}[\max_{\boldsymbol{\lambda}_{\mathrm{d}} \in \boldsymbol{\Lambda}} \frac{1}{M} \sum_{m=1}^{M} \varepsilon_m \left(r(\boldsymbol{\lambda}_{\mathrm{d}}, \mathbf{z}'_m) - r(\boldsymbol{\lambda}_{\mathrm{d}}, \mathbf{z}_m)\right)] \\ &\leq \mathbb{E}[\max_{\boldsymbol{\lambda}_{\mathrm{d}} \in \boldsymbol{\Lambda}} \frac{1}{M} \sum_{m=1}^{M} \varepsilon_m r(\boldsymbol{\lambda}_{\mathrm{d}}, \mathbf{z}'_m)] \\ &+ \mathbb{E}[\max_{\boldsymbol{\lambda}_{\mathrm{d}} \in \boldsymbol{\Lambda}} \frac{1}{M} \sum_{m=1}^{M} -\varepsilon_m r(\boldsymbol{\lambda}_{\mathrm{d}}, \mathbf{z}_m)]. \quad (6)\end{aligned}$$

where the expectations on the right are with respect to $\{\mathbf{z}_m, \mathbf{z}'_m, \varepsilon_m\}$. Again, since $\varepsilon_m r(\boldsymbol{\lambda}_{\mathrm{d}}, \mathbf{z}'_m)$ has the same distribution as $-\varepsilon_m r(\boldsymbol{\lambda}_{\mathrm{d}}, \mathbf{z}_m)$, the last two terms in (6) are the same. Hence, substituting into (5), we obtain

$$\begin{aligned}&\mathbb{E}[\max_{\boldsymbol{\lambda}_{\mathrm{d}} \in \boldsymbol{\Lambda}} \left(R(\boldsymbol{\lambda}_{\mathrm{d}}) - \hat{R}(\boldsymbol{\lambda}_{\mathrm{d}})\right)] \\ &\leq 2\mathbb{E}[\max_{\boldsymbol{\lambda}_{\mathrm{d}} \in \boldsymbol{\Lambda}} \frac{1}{M} \sum_{m=1}^{M} \varepsilon_m r(\boldsymbol{\lambda}_{\mathrm{d}}, \mathbf{z}'_m)] = 2R_M \quad (7)\end{aligned}$$

which is the required result. The proof for the second inequality also follows similarly.

Appendix C
Bound in (48)

Consider i.i.d. random variables $\{\mathsf{X}_k\}_{k=1}^K$ in $[0, B_r]$ with cumulative distribution $F(x)$ and mean $\mu := \mathbb{E}\mathsf{X}_k$. For any $0 \leq t \leq \mu$, applying Markov's inequality to the nonnegative random variable $B_r - \mathsf{X}_k$ gives [20, (1.5)]

$$\begin{aligned}F(\mu - t) &= \mathbb{P}\left[\mathsf{X}_k \leq \mu - t\right] = \mathbb{P}\left[B_r - \mathsf{X}_k \geq B_r - \mu + t\right] \\ &\leq \tfrac{B_r - \mu}{B_r - \mu + t} \leq 1 - \tfrac{t}{B_r}. \quad (8)\end{aligned}$$

Since $\mathsf{X}_{\max} := \max_k \mathsf{X}_k$ has distribution $F(t)^K$, the tail-integral formula yields:

$$\mathbb{E}[(\mu - \mathsf{X}_{\max})_+] = \int_0^\mu \mathbb{P}\left[\mathsf{X}_{\max} < \mu - t\right] dt \quad (9)$$

$$= \int_0^\mu F(\mu - t)^K\, dt \leq \int_0^{B_r} \left(1 - \frac{t}{B_r}\right)^K dt \leq \tfrac{B_r}{K}. \quad (10)$$

Appendix D
Lipschitz Continuity of coverage probabilities in (60) and (65)

In this section we show that the coverage probabilities in (60) and (65) are $L_\ell(e^{-1} + 4^{-1})$-Lipschitz and $L_\ell e^{-1}$-Lipschitz respectively, with respect to the W_1 metric, hence verifying Assumption 1. The approach is similar for both the coverage probability expressions. The functional mean-value theorem gives [33]

$$\begin{aligned}r(\lambda, \mathbf{z}) - r(\lambda', \mathbf{z}) &= \int_0^1 \int_{\mathcal{A}} \varphi_t(\mathbf{y})(\lambda(\mathbf{y}) - \lambda'(\mathbf{y})) d\mathbf{y} dt \\ &= \int_{\mathcal{A}} \varphi(\mathbf{y})(\lambda(\mathbf{y}) - \lambda'(\mathbf{y})) d\mathbf{y} \quad (11)\end{aligned}$$

where $\varphi_t(\mathbf{y}) := \frac{dr(\lambda_t, \mathbf{z})}{d\lambda}(\mathbf{y})$, $\lambda_t = (1-t)\lambda' + t\lambda$, and $\varphi(\mathbf{y}) := \int_0^1 \varphi_t(\mathbf{y}) dt$. If φ is L_φ-Lipschitz, we can write $\frac{1}{L_\varphi}\varphi \in \mathcal{L}_1$ and use (11) to obtain

$$|r(\lambda, \mathbf{z}) - r(\lambda', \mathbf{z})| \leq L_\varphi W_1(\lambda, \lambda') \quad (12)$$

It remains to determine L_φ for the two cases.

A. Lipschitz continuity of $p_c(\cdot, \mathbf{z})$ in (60)

We begin by calculating its functional derivative [34, Sec. 2.2]:

$$\frac{dp_c(\lambda, \mathbf{z})}{d\lambda}(\mathbf{y}) = h(\lambda, \mathbf{y}) - \int_{\mathcal{A}} \lambda(\mathbf{x}) h(\lambda, \mathbf{x}) b(\mathbf{x}, \mathbf{y}) d\mathbf{x}. \quad (13)$$

We observe from the triangle inequality that

$$\left\| \int_0^1 \nabla_{\mathbf{y}} \varphi_t(\mathbf{y}) dt \right\| \leq \int_0^1 \|\nabla_{\mathbf{y}} \varphi_t(\mathbf{y})\| \, dt \leq \sup_{t\in[0,1]} \|\nabla_{\mathbf{y}} \varphi_t(\mathbf{y})\|$$
$$\leq \sup_{t\in[0,1]} \|\nabla_{\mathbf{y}} h(\lambda_t, \mathbf{y})\|$$
$$+ \sup_{t\in[0,1]} \left\| \int_{\mathcal{A}} \lambda_t(\mathbf{x}) h(\lambda_t, \mathbf{x}) \nabla_{\mathbf{y}} b(\mathbf{x}, \mathbf{y}) d\mathbf{x} \right\| =: T_1 + T_2 \quad (14)$$

For the first term, for any $t \in [0, 1]$, the triangle inequality yields

$$\|\nabla_{\mathbf{y}} h(\lambda_t, \mathbf{y})\| = h(\lambda_t, \mathbf{y}) \left\| \nabla_{\mathbf{y}} a(\mathbf{y}) + \int_{\mathcal{A}} \lambda_t(\mathbf{u}) \nabla_{\mathbf{y}} b(\mathbf{y}, \mathbf{u}) d\mathbf{u} \right\|$$
$$\leq h(\lambda_t, \mathbf{y}) \left(\|\nabla_{\mathbf{y}} a(\mathbf{y})\| + \left\| \int_{\mathcal{A}} \lambda_t(\mathbf{u}) \nabla_{\mathbf{y}} b(\mathbf{y}, \mathbf{u}) d\mathbf{u} \right\| \right).$$

Recall that since $e^{L_\ell r} \ell(r)$ is increasing in r, we have that

$$\sup_{r\geq 0} \frac{|\ell'(r)|}{\ell(r)} \leq L_\ell. \quad (15)$$

For $\mathbf{y} \neq \mathbf{z}$, since $\|\nabla_{\mathbf{y}} d_{\mathbf{y}}\| = 1$, we can write

$$\|\nabla_{\mathbf{y}} a(\mathbf{y})\| = \frac{\kappa}{\ell^2(d_{\mathbf{y}})} |\ell'(d_{\mathbf{y}})| \leq L_\ell \frac{\kappa}{\ell(d_{\mathbf{y}})} = L_\ell a(\mathbf{y}) \quad (16)$$

$$\nabla_{\mathbf{y}} b(\mathbf{y}, \mathbf{u}) = -\frac{\tau \ell(d_{\mathbf{u}})}{(\tau \ell(d_{\mathbf{u}}) + \ell(d_{\mathbf{y}}))^2} \ell'(d_{\mathbf{y}}) (\nabla_{\mathbf{y}} d_{\mathbf{y}}) \quad (17)$$

so using the fact that $\ell(d_{\mathbf{u}}) \geq 0$, we obtain

$$\left\| \int_{\mathcal{A}} \lambda_t(\mathbf{u}) \nabla_{\mathbf{y}} b(\mathbf{y}, \mathbf{u}) d\mathbf{u} \right\| \leq L_\ell \int_{\mathcal{A}} \lambda_t(\mathbf{u}) b(\mathbf{y}, \mathbf{u}) d\mathbf{u} \quad (18)$$

Combining these and substituting $w := a(\mathbf{y}) + \int_{\mathcal{A}} \lambda_t(\mathbf{u}) b(\mathbf{y}, \mathbf{u}) d\mathbf{u}$, we obtain $\|\nabla_{\mathbf{y}} h(\lambda_t, \mathbf{y})\| \leq L_\ell w e^{-w} \leq \frac{L_\ell}{e}$. Hence the bound on T_1 becomes $T_1 \leq \frac{L_\ell}{e}$. For the second term in (14), observe that

$$\nabla_{\mathbf{y}} b(\mathbf{x}, \mathbf{y}) = \frac{\tau \ell(d_{\mathbf{x}}) \ell'(d_{\mathbf{y}})}{(\tau \ell(d_{\mathbf{y}}) + \ell(d_{\mathbf{x}}))^2} (\nabla_{\mathbf{y}} d_{\mathbf{y}}) \quad (19)$$

so using $\|\nabla_{\mathbf{y}} d_{\mathbf{y}}\| = 1$ for $\mathbf{y} \neq \mathbf{z}$ and the inequality $4ab \leq (a+b)^2$, we have the bound

$$\left\| \int_{\mathcal{A}} \lambda_t(\mathbf{x}) h(\lambda_t, \mathbf{x}) \nabla_{\mathbf{y}} b(\mathbf{x}, \mathbf{y}) d\mathbf{x} \right\|$$
$$\leq \frac{|\ell'(d_{\mathbf{y}})|}{\ell(d_{\mathbf{y}})} \int_{\mathcal{A}} \lambda_t(\mathbf{x}) h(\lambda_t, \mathbf{x}) \frac{\tau \ell(d_{\mathbf{x}}) \ell(d_{\mathbf{y}})}{(\tau \ell(d_{\mathbf{y}}) + \ell(d_{\mathbf{x}}))^2} d\mathbf{x} \quad (20)$$
$$\leq \frac{L_\ell}{4} \int_{\mathcal{A}} \lambda_t(\mathbf{x}) h(\lambda_t, \mathbf{x}) d\mathbf{x} = \tfrac{1}{4} L_\ell p_c(\lambda_t, \mathbf{z}) \leq L_\ell/4 \quad (21)$$

for $t \in [0, 1]$. Therefore we have that $T_2 \leq L_\ell/4$. Adding the bounds on T_1 and T_2, we can take $L_\varphi = L_\ell(e^{-1} + 4^{-1}) < 2L_\ell/3$ and hence $p_c(\lambda, \mathbf{z})$ is $\frac{2}{3} L_\ell$-Lipschitz in λ with respect to the W_1 metric.

B. Lipschitz continuity of $p_c^a(\cdot, \mathbf{z})$ in (65)

Defining

$$p(r, \mathbf{z}) := \int_{\mathcal{B}(\mathbf{z}, r)} \lambda(\mathbf{y}) \mathrm{d}\mathbf{y} = \int_{\mathcal{A}} \lambda(\mathbf{y}) \mathbf{1}_{\{d_{\mathbf{y}} \leq r\}} \mathrm{d}\mathbf{y} \quad (22)$$

the coverage probability in (65) takes the form

$$p_c^a(\lambda, \mathbf{z}) = \int_0^\infty h^a(\lambda, r) e^{-p(r, \mathbf{z})} \frac{dp(r, \mathbf{z})}{\mathrm{d}r} \mathrm{d}r \quad (23)$$

where $d_{\mathbf{y}} := \|\mathbf{y} - \mathbf{z}\|$ and the expression for $h^a(\lambda, r)$ is provided in (70). Note that for a fixed $\mathbf{y} \in \mathcal{A}$,

$$\frac{\mathrm{d}p(r, \mathbf{z})}{\mathrm{d}\lambda}(\mathbf{y}) = \mathbf{1}_{\{d_{\mathbf{y}} \leq r\}} \quad \frac{\mathrm{d}}{\mathrm{d}\lambda(\mathbf{y})} \left[\frac{\mathrm{d}p(r, \mathbf{z})}{\mathrm{d}r} \right] = \delta(r - d_{\mathbf{y}}), \quad (24)$$

where $\delta(\cdot)$ denotes the Dirac delta function. Hence, the functional derivative of p_c^a is given by

$$\frac{\mathrm{d}p_c^a(\lambda, \mathbf{z})}{\mathrm{d}\lambda}(\mathbf{y}) = -\int_0^\infty h^a(\lambda, r) e^{-p(r, \mathbf{z})} \frac{\mathbf{1}_{\{d_{\mathbf{y}} > r\}}}{1 + \frac{\ell(r)}{\tau \ell(d_{\mathbf{y}})}} \frac{\mathrm{d}p(r, \mathbf{z})}{\mathrm{d}r} \mathrm{d}r$$
$$- \int_0^\infty h^a(\lambda, r) e^{-p(r, \mathbf{z})} \mathbf{1}_{\{d_{\mathbf{y}} \leq r\}} \frac{\mathrm{d}p(r, \mathbf{z})}{\mathrm{d}r} \mathrm{d}r$$
$$+ \int_0^\infty h^a(\lambda, r) e^{-p(r, \mathbf{z})} \delta(r - d_{\mathbf{y}}) \mathrm{d}r \quad (25)$$
$$= -\int_0^{d_{\mathbf{y}}} \frac{h^a(\lambda, r)}{1 + \frac{\ell(r)}{\tau \ell(d_{\mathbf{y}})}} f_{R_{\mathbf{z}}}(r) \mathrm{d}r - \int_{d_{\mathbf{y}}}^\infty h^a(\lambda, r) f_{R_{\mathbf{z}}}(r) \mathrm{d}r$$
$$+ h^a(\lambda, d_{\mathbf{y}}) e^{-p(d_{\mathbf{y}}, \mathbf{z})} =: \varphi^a(\mathbf{y}) \quad (26)$$

Since all the terms in $\varphi^a(\mathbf{y})$ depend on $\mathbf{y}$ only through $d_{\mathbf{y}}$, let $s := d_{\mathbf{y}}$, so that it suffices to use the fact that $|d_{\mathbf{y}} - d_{\mathbf{y}'}| \leq \|\mathbf{y} - \mathbf{y}'\|$ and only bound $|\frac{d}{ds} \varphi_t^a(s)|$ where $\varphi_t^a(s)$ is defined in (26) with λ replaced by $\lambda_t = (1-t)\lambda' + t\lambda$. With a slight abuse of notation, define

$$a(s) := \frac{\kappa}{\ell(s)} \qquad b(r, s) := \frac{\tau \ell(s)}{\tau \ell(s) + \ell(r)}, \quad (27)$$

which are simply (62) when the serving and interfering BSs are at distances r and s, respectively. Then

$$\varphi_t^a(s) = h^a(\lambda_t, s) e^{-p_t(s, \mathbf{z})} - \int_0^s h^a(\lambda_t, r) b(r, s) f_{R_{\mathbf{z}}, t}(r) \mathrm{d}r$$
$$- \int_s^\infty h^a(\lambda_t, r) f_{R_{\mathbf{z}}, t}(r) \mathrm{d}r \quad (28)$$

where p_t is defined as in (22) and $f_{R_{\mathbf{z}}, t}$ in (67) with λ replaced with λ_t. Hence, we obtain

$$h^a(\lambda_t, s) e^{-p_t(s, \mathbf{z})} =$$
$$\exp\left(-a(s) - p_t(s, \mathbf{z}) - \int_{\mathcal{A} \setminus \mathcal{B}(\mathbf{z}, s)} \lambda_t(\mathbf{u}) b(s, d_{\mathbf{u}}) \mathrm{d}\mathbf{u} \right). \quad (29)$$

Using the Leibniz rule and the fact that $b(s, s) = \tau/(1+\tau)$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}s} \left[h^a(\lambda_t, s) e^{-p_t(s, \mathbf{z})} \right] = -h^a(\lambda_t, s) e^{-p_t(s, \mathbf{z})} \times \quad (30)$$
$$\left[a'(s) + \int_{\mathcal{A} \setminus \mathcal{B}(\mathbf{z}, s)} \lambda_t(\mathbf{u}) \frac{\partial b(s, d_{\mathbf{u}})}{\partial s} \mathrm{d}\mathbf{u} + \frac{1}{1+\tau} \frac{\mathrm{d}p_t(s, \mathbf{z})}{\mathrm{d}s} \right].$$

Likewise, differentiating the two integrals in (28) gives

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\int_0^s h^a(\lambda_t,r)b(r,s)f_{R_{\mathbf{z}},t}(r)\mathrm{d}r\right] = \frac{\tau}{1+\tau}h^a(\lambda_t,s)e^{-p_t(s,\mathbf{z})}\frac{\mathrm{d}p_t(s,\mathbf{z})}{\mathrm{d}s} + \int_0^s h^a(\lambda_t,r)\frac{\partial b(r,s)}{\partial s}f_{R_{\mathbf{z}},t}(r)\mathrm{d}r \quad (31)$$

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\int_s^\infty h^a(\lambda_t,r)f_{R_{\mathbf{z}},t}(r)\mathrm{d}r\right] = -h^a(\lambda_t,s)e^{-p_t(s,\mathbf{z})}\frac{\mathrm{d}p_t(s,\mathbf{z})}{\mathrm{d}s}. \quad (32)$$

Substituting (30)–(32) into the derivative of (28), the terms containing $\frac{\mathrm{d}p_t(s,\mathbf{z})}{\mathrm{d}s}$ cancel, and we obtain

$$(\varphi_t^a)'(s) = -h^a(\lambda_t,s)e^{-p_t(s,\mathbf{z})}\left[a'(s)+\int_{\mathcal{A}\setminus\mathcal{B}(\mathbf{z},s)}\lambda_t(\mathbf{u})\frac{\partial b(s,d_{\mathbf{u}})}{\partial s}\mathrm{d}\mathbf{u}\right] - \int_0^s h^a(\lambda_t,r)\frac{\partial b(r,s)}{\partial s}f_{R_{\mathbf{z}},t}(r)\mathrm{d}r. \quad (33)$$

whose terms can be bounded using the same arguments as those used for T_1 and T_2 in the instantaneous-association proof. Specifically, defining

$$w := a(s) + \int_{\mathcal{A}\setminus\mathcal{B}(\mathbf{z},s)}\lambda_t(\mathbf{u})b(s,d_{\mathbf{u}})\mathrm{d}\mathbf{u}, \quad (34)$$

we have $h^a(\lambda_t,s)e^{-p_t(s,\mathbf{z})} \le e^{-w}$, and therefore the magnitude of the first term in (33) is bounded as in T_1 by $L_\ell w e^{-w} \le \frac{L_\ell}{e}$. For the second term, we can likewise bound $-\frac{\partial b(r,s)}{\partial s} \le \frac{L_\ell}{4}$ so that

$$-\int_0^s h^a(\lambda_t,r)\frac{\partial b(r,s)}{\partial s}f_{R_{\mathbf{z}},t}(r)\mathrm{d}r \le \frac{L_\ell}{4}p_c^a(\lambda_t,\mathbf{z}) \le \frac{L_\ell}{4}.$$

Different from the earlier proof, the first term in (33) is nonpositive, while the second is nonnegative, and hence $-\frac{L_\ell}{e} \le \varphi_t'(s) \le \frac{L_\ell}{4}$ which implies that we can take $L_\varphi = \frac{L_\ell}{e}$. In conclusion, $p_c^a(\lambda,\mathbf{z})$ is L_ℓ/e-Lipschitz with respect to the W_1 metric.

Appendix E
Incremental Deployment

As earlier, let Φ_0 denote the set of existing BSs. Under the instantaneous-SINR association rule, a user $\mathbf{z}$ may be covered either through one of the installed BSs $\mathbf{u} \in \Phi_0$ with probability

$$p_c^1 = \sum_{\mathbf{u}\in\Phi_0} h(\lambda,\mathbf{u}) \prod_{\mathbf{v}\in\Phi_0\setminus\{\mathbf{u}\}} (1-b(\mathbf{u},\mathbf{v})) \quad (35)$$

or through one of the newly deployed BSs with probability

$$p_c^2 = \int_{\mathcal{A}} \lambda(\mathbf{x})h(\lambda,\mathbf{x}) \prod_{\mathbf{u}\in\Phi_0}(1-b(\mathbf{x},\mathbf{u}))d\mathbf{x}. \quad (36)$$

Since $\tau > 1$, at most one BS can satisfy the SINR threshold. Therefore, the total coverage probability is given by $p_c^1 + p_c^2$. Coverage probability of a specified set Φ_0 of BSs can likewise be obtained by substituting $\lambda \equiv 0$ in (35).

Under average-SINR association, the serving BS is the nearest BS to the user. Let $\mathbf{u_z}$ denote the existing BS that is closest to $\mathbf{z}$ and $r_{\mathbf{u}} := \|\mathbf{u_z} - \mathbf{z}\|$. Recall that $\mathbf{x_z}$ is the newly deployed BS closest to $\mathbf{z}$ and $R_{\mathbf{z}} = \|\mathbf{x_z} - \mathbf{z}\|$. There are two distinct possibilities: $R_{\mathbf{z}} \ge r_{\mathbf{u}}$ so that the user is served by an existing BS, and $R_{\mathbf{z}} < r_{\mathbf{u}}$ so that the user is served by a new BS. The contribution to the coverage probability from the first case is

$$p_c^1 = \prod_{\mathbf{v}\in\Phi_0\setminus\{\mathbf{u_z}\}}(1-b(\mathbf{u_z},\mathbf{v})) \times \exp\left(-a(\mathbf{u_z}) - \int_{\mathcal{B}(\mathbf{z},r_{\mathbf{u}})}\lambda(\mathbf{y})\mathrm{d}\mathbf{y} - \int_{\mathcal{A}\setminus\mathcal{B}(\mathbf{z},r_{\mathbf{u}})}\lambda(\mathbf{y})b(\mathbf{u_z},\mathbf{y})\mathrm{d}\mathbf{y}\right) \quad (37)$$

where $a(\cdot)$ is defined in (62). When a newly deployed BS serves $\mathbf{z}$, the coverage probability is given by

$$p_c^2 = \int_0^{r_{\mathbf{u}}} \exp\left(-\frac{\tau\sigma^2}{P\ell(r)} - \int_{\mathcal{A}\setminus\mathcal{B}(\mathbf{z},r)}\frac{\lambda(\mathbf{y})\mathrm{d}\mathbf{y}}{1+\frac{\ell(r)}{\tau\ell(\|\mathbf{y}-\mathbf{z}\|)}}\right) \times \prod_{\mathbf{v}\in\Phi_0}\frac{\ell(r)}{\ell(r)+\tau\ell(\|\mathbf{v}-\mathbf{z}\|)}f_{R_{\mathbf{z}}}(r)\mathrm{d}r. \quad (38)$$

Finally, the total coverage probability is given by $p_c^1 + p_c^2$.

Appendix F
Convexity in the Noise-Limited Case

In this section, we show that $p_c(\lambda)$ for IPPP deployment is concave when the network is noise-limited, i.e., when the aggregate interference power $I(\mathbf{z})$ is much smaller than the noise power σ^2.

Lemma 1. *In the noise-limited case, the coverage probability expressions in* (60) *and* (65) *are both concave in* λ.

Proof: In the instantaneous-SINR association, the noise-limited coverage probability is given by [27, Sec. II]:

$$p_c(\lambda,\mathbf{z}) = 1 - \exp\left(-\int_{\mathcal{A}}\lambda(\mathbf{x})e^{-a(\mathbf{x})}\mathrm{d}\mathbf{x}\right) \quad (39)$$

which is concave since the function $1-e^{-x}$ is concave in x and the integral is linear in λ.

In the average-SINR association, the noise-limited version of (65) becomes

$$p_c^a(\lambda,\mathbf{z}) = \int_0^\infty \exp\left(-\frac{\tau\sigma^2}{P\ell(r)}\right)f_{R_{\mathbf{z}}}(r)\mathrm{d}r \quad (40)$$

$$= \int_0^\infty \mathbb{P}\left[R_{\mathbf{z}} \le r\right]\mathrm{d}\left[-\exp\left(-\frac{\tau\sigma^2}{P\ell(r)}\right)\right] \quad (41)$$

where $\mathbb{P}\left[R_{\mathbf{z}} \le r\right] = 1 - \exp\left(-\int_{\mathcal{B}(\mathbf{z},r)}\lambda(\mathbf{y})\mathrm{d}\mathbf{y}\right)$ is concave in λ and the measure $\mathrm{d}\left[-\exp\left(-\frac{\tau\sigma^2}{P\ell(r)}\right)\right]$ is nonnegative. Hence, $p_c^a(\lambda,\mathbf{z})$ is a nonnegative weighted integral of concave functions and is therefore concave in λ.

Finally, expectation over $\mathbf{z}$ and empirical averaging preserve concavity, and hence the coverage probability is concave in both the cases. ∎

In the noise-limited case therefore, mirror ascent converges to the global optimum and E_o can be made arbitrarily small. To obtain the bounds in closed-form however, we must

first establish that the discretized gradients of $p_c(\boldsymbol{\lambda}_\mathrm{d}, \mathbf{z})$ and $\mathrm{p}_c^a(\boldsymbol{\lambda}_\mathrm{d}, \mathbf{z})$ have bounded components.

Lemma 2. *In the noise-limited case, it holds that*

$$\|\nabla p_c(\boldsymbol{\lambda}_\mathrm{d}, \mathbf{z})\|_\infty \leq \tfrac{A}{D} \qquad \|\nabla p_c^a(\boldsymbol{\lambda}_\mathrm{d}, \mathbf{z})\|_\infty \leq \tfrac{A}{D} \tag{42}$$

where the expressions for p_c and p_c^a are provided in (60) *and* (65)*, respectively.*

Proof: Observe that in (13), both $h(\lambda, \mathbf{y})$ and $b(\mathbf{x}, \mathbf{y})$ lie in the interval $[0, 1]$. Therefore, the right-hand side is bounded as $|\frac{dp_c(\lambda, \mathbf{z})}{d\lambda}(\mathbf{y})| \leq 1$. The required bound follows as

$$\begin{aligned} \|\nabla p_c(\boldsymbol{\lambda}_\mathrm{d}, \mathbf{z})\|_\infty &= \max_i |[\nabla p_c(\boldsymbol{\lambda}_\mathrm{d}, \mathbf{z})]_i| \qquad (43) \\ &= \max_i \left| \int_{\mathcal{A}_i} \frac{dp_c(\lambda, \mathbf{z})}{d\lambda}(\mathbf{y}) \mathrm{d}\mathbf{y} \right| \leq |\mathcal{A}_i| = \tfrac{A}{D}. \end{aligned}$$

For the average-SINR association case, we can again use these bounds and the fact that $0 \leq \left(1 + \frac{\ell(r)}{\tau \ell(d_\mathbf{y})}\right)^{-1} \leq 1$ in (26), so as to conclude that $|\frac{\mathrm{d}p_c^a(\lambda, \mathbf{z})}{\mathrm{d}\lambda}(\mathbf{y})| \leq 1$. Therefore, proceeding as above, we obtain the required bound. ∎

Since the stochastic gradient $\mathbf{g}^{(t)}$ in (43) is unbiased and, from the preceding lemma, satisfies $\left\|\mathbf{g}^{(t)}\right\|_\infty \leq A/D$, we can directly apply the standard stochastic mirror-ascent bound for concave objectives [25]. In particular, using $\bar{\boldsymbol{\lambda}}_\mathrm{d}^T$ as the step-size-weighted average of the iterates and starting from uniform initialization and constant step-size, we have

$$E_o \leq N \sqrt{\frac{2 \log D}{T}}. \tag{44}$$

We can then add this bound to the other terms in Thm. 1 to obtain the full end-to-end bound.

Appendix G
Effect of noise power

In this section, we again consider the experiment with users distributed along an annulus, as in Fig. 1(a), but with different values of noise power σ^2. As expected, Fig. 1 shows that when the noise is high, the optimized intensity remains close to the users to reduce path loss. As noise decreases, a more centrally concentrated intensity can provide coverage to users around the annulus. However, as noise is reduced further, interference becomes relatively more important, and deploying multiple BSs at the center results in interference. Spreading the intensity back towards the annulus increases the likelihood that a user is close to one BS while the others are farther away, thereby improving the SINR and coverage probability.

Appendix H
Effect of the Association Rule

In this section, we illustrate how the assumed association rule can affect the optimized intensity. We consider the region $\mathcal{A} = [-5, 5]^2$ with two existing BSs placed symmetrically on both sides of a corridor containing 80% of the 300 users. We also have a separate group of users at $(3, 3)$ containing the remaining users. We take $\sigma^2 = 2 \times 10^{-5}$, while the remaining propagation parameters are the same as specified in Table I. The noise parameter is chosen to make the effect of the association rule more visible.

Fig. 2 compares the intensities optimized under instantaneous-SINR association (IA) and average-SINR association (AA). The difference is concentrated in the corridor, where the users lie between the two existing BSs. Under AA, each user is associated with the nearest BS, while the other BS contributes only interference; hence, some additional intensity is allocated within the corridor. Under IA, either existing BS may serve a user depending on the instantaneous fading, reducing the need for additional deployment in this region. The user group centered at $(3, 3)$ is far from both existing BSs, and therefore the two association rules yield nearly identical intensities there. Finally, the optimized training coverage probability is $p_c = 0.71$ under instantaneous-SINR association and $p_c = 0.61$ under average-SINR association. Thus, the association rule can significantly affect both coverage and deployment priorities in regions containing multiple existing BSs.

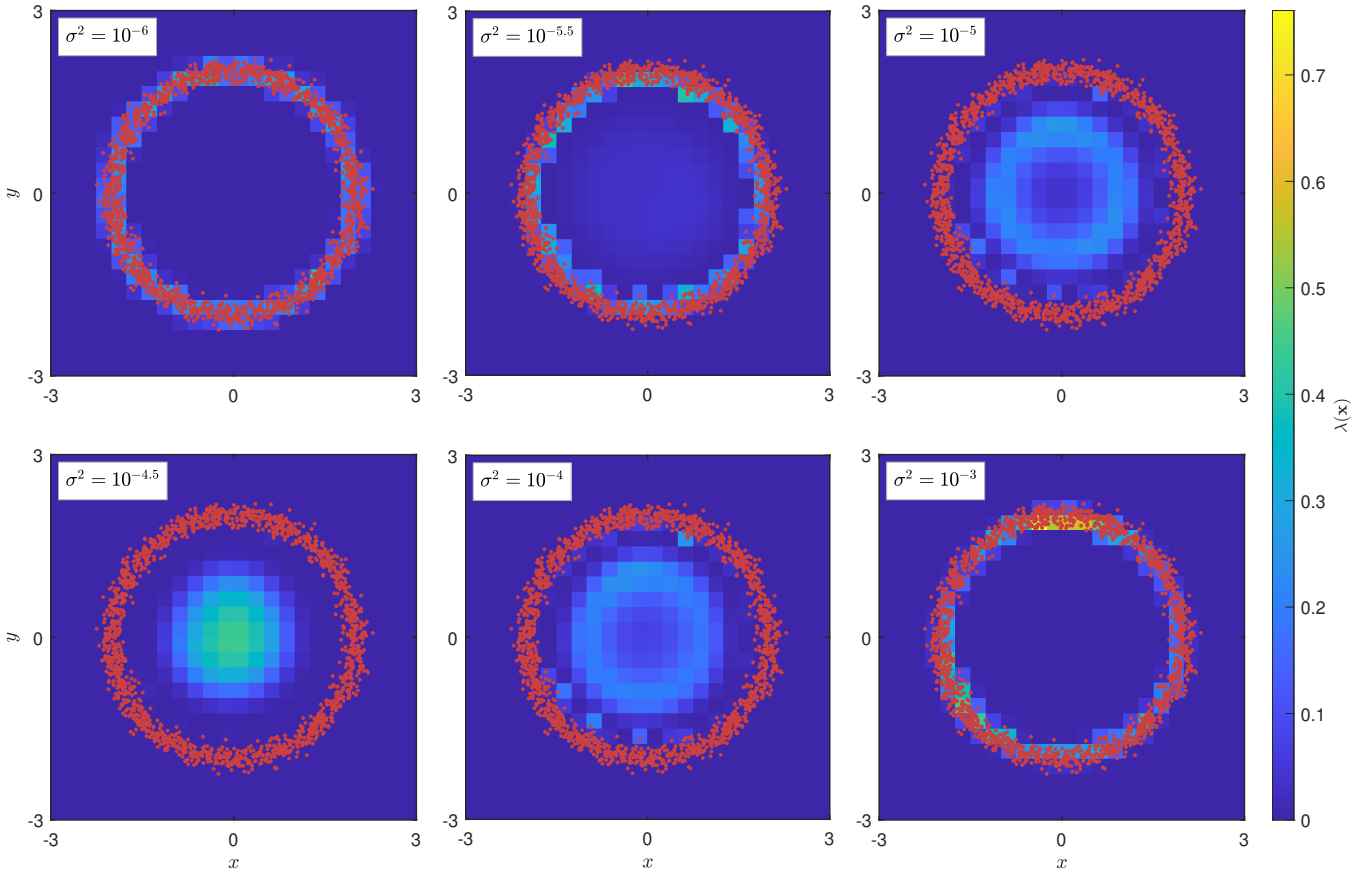


Fig. 1. Intensity maps across a range of noise powers.

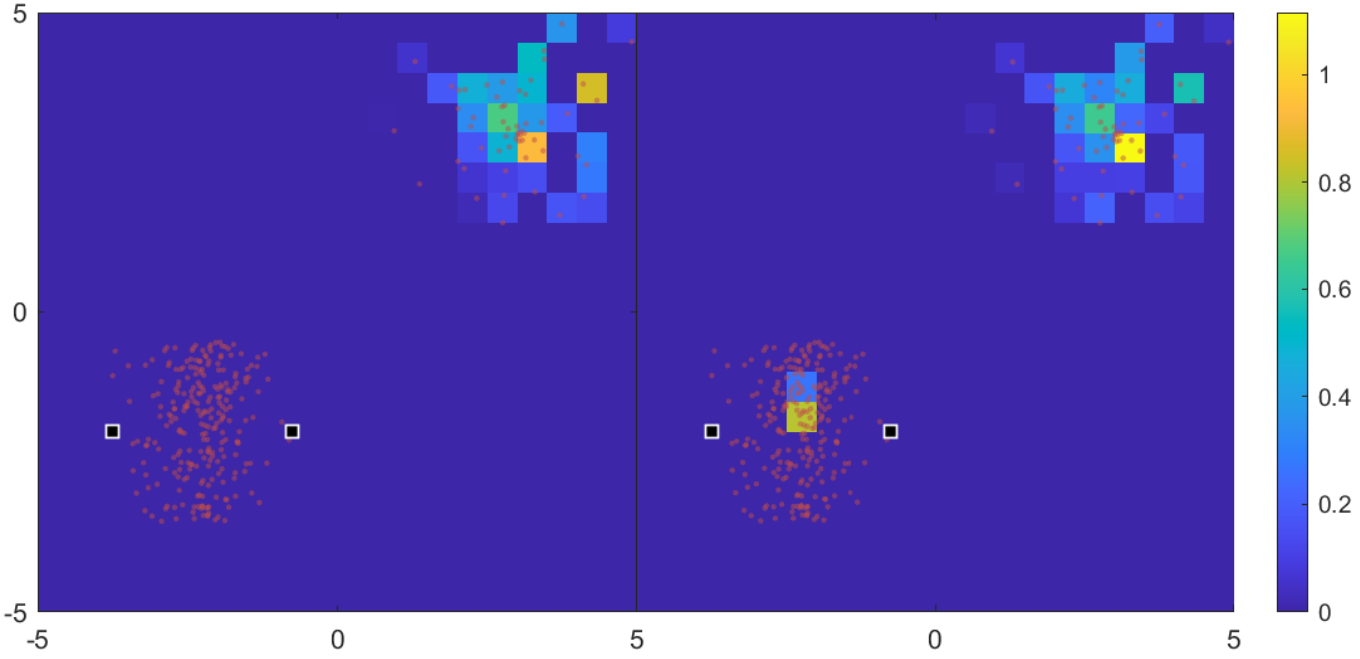


Fig. 2. Optimized intensities for the instantaneous-SINR (left) and average-SINR (right) association rules.