ON SPATIO-TEMPORAL TOMLINSON HARASHIMA PRECODING IN IIR CHANNELS: MMSE SOLUTION, PROPERTIES, AND FAST COMPUTATION

Sander Wahls and Holger Boche

Technische Universität München
Lehrstuhl für Theoretische Informationstechnik
Karlstraße 45-47/V, 80333 München, Germany
{sander.wahls,boche}@tum.de

ABSTRACT

We consider spatio-temporal Tomlinson Harashima Precoding where the feedforward filter is located at the transmitter and an additional scalar gain is employed as receive filter. In contrast to other works, we allow channel, feedforward, and feedbackward filters to have one-sided but infinite impulse responses. The optimal filters with respect to a minimum mean square error criterion are derived. We elaborate several interesting properties of our solution and discuss a fast implementation with only quadratic complexity in the latency time.

I. INTRODUCTION

Precoding is an attractive approach to fight typical impairments in communication systems such as intersymbol interference when low complexity receivers are sought and the channel is known at the transmitter. An especially popular precoding technique is Tomlinson Harashima Precoding (THP; see [1], [2]), which is well-known for its good compromise between performance and complexity. The basic idea of THP is to extend the modulation alphabet to an infinite lattice in the complex plane and to modify the data signal such that when the (probably unstable) inverse channel is applied to this modified data signal, a precoded signal of fixed average power results. The receiver can revert the modification of the data signal with a simple modulo operation. A typical THP basically consists of three components: modulo operators, a feedforward filter (FFF) which is used to shape the channel impulse response, and a feedbackward filter (FBF) which is used to compute the signal that neutralizes the interference in the shaped channel.

In the literature, several variants of THP can be found. The classical THP was designed for single-input single-output (SISO) channels and incorporates transmitters that only employ modulo operators and a FBF [1], [2]. Many authors chose to include an FFF at the receiver, e.g., [3]. Various works have extended this concept to multiple-input multiple-output (MIMO) channels by addition of spatial feedback, e.g., [4], [5]. A more recent variant of the THP considers systems where also the FFF has been moved to the transmitter, e.g., [6]. This has the advantage that the complexity at the receiver is reduced, but there is the expense that the FFF has to be designed subject to a transmit power constraint. Finally, Joham et. al. [7] have recently proposed to add an additional scalar receive filter. They derived the optimal THP in this scenario with respect to the minimum mean square error (MMSE) criterion (under
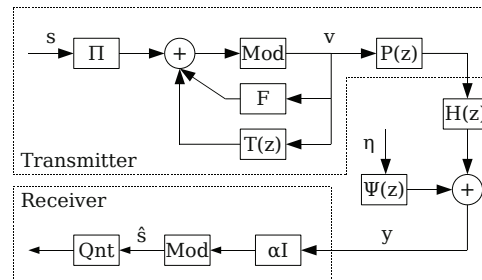
Fig. 1. System model.

some simplifying assumptions) for FIR MIMO channels. However, although FIR channel models are predominant in communications, it is well-known that the use of IIR channel models can be advantageous in some applications, e.g., DSL [8], [9]. Furthermore, also the FFF and FBF in [7] were constrained to be FIR. The goal of this paper is to address these issues. We make following contributions.

1) We derive the MMSE THP for IIR MIMO channels, where we also allow the FFF and the FBF to be IIR.[1]
2) We establish several properties of the IIR-MMSE THP. Our most striking finding is that the IIR-MMSE THP reduces to a FIR-MMSE THP if the channel is FIR.
3) We give a fast implementation of our results with only quadratic complexity in the latency time.

The paper is structured as follows. We start with the system model (Sec. II) and the problem statement (Sec. III). Then, we present our results in Sec. IV. We close the paper with a numerical example (Sec. V) and a conclusion (Sec. VI).

Notation and basic definitions: We denote the space of complex $m \times n$ matrices by $\mathbb{C}^{m \times n}$ and equip it with the Frobenius norm $\|A\|_F^2 := \text{trace}\{A^*A\}$, where A^* is the hermitian and $A \in \mathbb{C}^{m \times n}$. A^T means transposition. The space of stable and causal rational $m \times n$ matrices (i.e., no poles in $|z| \geq 1$) will be denoted by $\mathcal{RH}_\infty^{m \times n}$. We equip it with the norm $\|B\|_2^2 := \int_0^{2\pi} \|B(e^{i\theta})\|_F^2 \frac{d\theta}{2\pi}$, $B \in \mathcal{RH}_\infty^{m \times n}$. The impulse response of B is the unique sequence $\{B_k\}_{k \in \mathbb{N}} \subset \mathbb{C}^{m \times n}$ such that $B(z) = \sum_{k=0}^\infty B_k z^{-k}$ for $|z| \geq 1$. $\mathcal{O}$ denotes the Landau symbol.

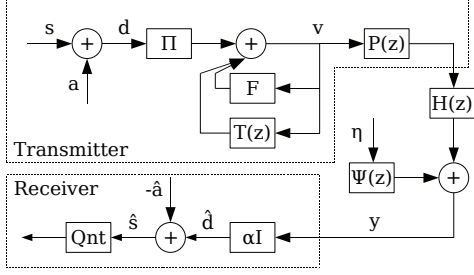[1]Note that we do so under the same simplifying assumptions as in [7].

Fig. 2. Auxiliary system model

II. SYSTEM MODEL

The system model of a communication system employing Tomlinson Harashima Precoding is depicted in Fig. 1. The transmitter consists of a permutation matrix $\Pi \in \mathbb{C}^{p \times p}$, a strictly lower triangular spatial FBF $F \in \mathbb{C}^{p \times p}$, a strictly causal temporal FBF $T \in \mathcal{RH}_\infty^{p \times p}$, and a causal FFF filter $P \in \mathcal{RH}_\infty^{p \times q}$. The Mod block denotes the modulo operator $\text{Mod} : \mathbb{C}^q \to \mathbb{C}^q, [v_1, \ldots, v_q]^T \mapsto [w_1, \ldots, w_q]^T$, where $w_k := v_k - \tau \,\text{floor}\left(\tau^{-1} \,\text{real}(v_k) + 0.5\right) - i\tau \,\text{floor}\left(\tau^{-1} \,\text{imag}(v_k) + 0.5\right)$ for all $k = 1, \ldots, q$, and some $\tau > 0$.[2] The channel $H \in \mathcal{RH}_\infty^{q \times p}$, $q \leq p$, and the noise model $0 \not\equiv \Psi \in \mathcal{RH}_\infty^{q \times l}$ are also assumed causal. At the receiver, we encounter the scalar gain $\alpha \geq 0$, another Mod block, and an quantizer Qnt which maps the estimates to the nearest element of the modulation alphabet used in the signal generation.

III. PROBLEM STATEMENT

Computation of the optimal THP in the sense that the mean square error (MSE) between $z^{-L}s$ and $\hat{s}$ in Fig. 1 is minimized is an intricate problem. Therefore, we perform some common simplifications [7]. We replace the modulo operators with two auxiliary signals a and $\hat{a}$ (which depend on s and $\hat{s}$), see Fig. 2. Note that this replacement can be made exact by proper choice of a and $\hat{a}$. Instead of trying to minimize the MSE between $z^{-L}s$ and $\hat{s}$ in Fig. 1, we will minimize the MSE between $z^{-L}d$ and $\hat{d}$ in Fig. 2. We will do so under the assumption that the output $v = \{v_k\}_{k \in \mathbb{N}}$ of the modulo operator is a temporally and spatially uncorrelated random sequence with zero mean and variance $\sigma_v^2 := \frac{\tau^2}{6}$ [7], i.e., $\mathbb{E}[v_k v_{k+j}^*] = \sigma_v^2 I_q$ if $j = 0$ and $\mathbb{E}[v_k v_{k+j}^*] = 0_q$ if $j \neq 0$. We furthermore also assume that the unfiltered noise $\eta = \{\eta_k\}_{k \in \mathbb{N}}$ is a temporally and spatially uncorrelated random sequence with zero mean and variance $\sigma_\eta^2 > 0$.

Problem 1. Let the channel $H \in \mathcal{RH}_\infty^{q \times p}$, the noise model $0 \not\equiv \Psi \in \mathcal{RH}_\infty^{q \times l}$, a permutation matrix $\Pi \in \mathbb{C}^{q \times q}$, the signal and noise powers $\sigma_v^2 > 0$ and $\sigma_\eta^2 > 0$, and a latency time $L \in \mathbb{N}$ be given. Our goal is to find $P \in \mathcal{RH}_\infty^{q \times p}$, $F \in \mathbb{C}^{q \times q}$, $T \in \mathcal{RH}_\infty^{q \times q}$ and $\alpha > 0$ such that the expectation of the mean square error (MSE) in the auxiliary model, $e := \mathbb{E}[\|z^{-L}d - \hat{d}\|_2^2]$, is minimized subject to the transmit power constraint $\sigma_v^2 \|P\|_2^2 \leq E_{tr}$, the temporal causality constraint that T is strictly causal, and the spatial causality constraint that F is strictly lower triangular.

[2]The exact value of τ depends on the modulation used. See [7] for details.

IV. OPTIMAL THP

IV-A. Main result

Theorem 2. *Let $\{H_k\}_{k \in \mathbb{N}}$ denote the impulse response of H and assume that $H_{k_0} \neq 0$ for some $k_0 \in \{0, \ldots, L\}$. Furthermore, set*

$$M := \begin{bmatrix} \Pi H_0 & & & & \\ \Pi H_1 & \Pi H_0 & & & \\ \Pi H_2 & \Pi H_1 & \Pi H_0 & & \\ \vdots & \ddots & \ddots & \ddots & \\ \Pi H_L & \ldots & \Pi H_2 & \Pi H_1 & \Pi H_0 \end{bmatrix} \in \mathbb{C}^{(L+1)q \times (L+1)p}$$

(1)

as well as

$$\tilde{P}(z) := \begin{bmatrix} z^{-0}I_p & \ldots & z^{-L}I_p \end{bmatrix} \begin{bmatrix} u_1 & \ldots & u_q \end{bmatrix}$$

where, for all $k = 1, \ldots, q$, the u_k are given by

$$u_k := \left(M^* \begin{bmatrix} I_{Lq+k} & \\ & 0_{q-k} \end{bmatrix} M + \frac{\sigma_\eta^2 \|\Psi\|_2^2}{E_{tr}} I \right)^{-1} M^* e_{Lq+k}$$

and e_j denotes the jth column of the identity matrix $I_{(L+1)q}$. Finally, let $\{N_k\}_{k \in \mathbb{N}}$ denote the impulse response of $\Pi H \tilde{P}$. Then,

$$\begin{aligned} \alpha_{opt} &:= \frac{\sigma_v \|\tilde{P}\|_2}{\sqrt{E_{tr}}} & (2) \\ P_{opt} &:= \frac{1}{\alpha_{opt}} \tilde{P} & (3) \\ F_{opt} &:= \textit{strictly lower triangular part of } -N_L \\ T_{opt} &:= -\sum_{k=1}^\infty N_{k+L} z^{-k} & (4) \end{aligned}$$

are well-defined and solve Problem 1.

Proof: See the appendix. ∎

IV-B. Remarks and Discussion

FIRness: The optimal FFF P_{opt} in Theorem 2 clearly is a FIR filter of order L. Moreover, if H is a FIR channel of order N, then also the temporal FBF T_{opt} is a FIR filter of order N. Interestingly, this observation is independent of the noise model Ψ. This is a contrast to the optimal DFE where the optimal filters are FIR only if the noise model is all-pole [10]. The reason is that we cannot influence the noise at the receiver.

State-space realizations for IIR channels: The question remains how we can realize THP if the channel is IIR. The FFF obviously still is FIR even if the channel is IIR. However, the temporal FBF will be IIR. A simple way to realize the temporal FFF T_{opt} is via state-space realizations [11]. Let the channel have the state-space realization $H(z) = D + C(zI - A)^{-1}B$ and expand the FFF as $P_{opt}(z) = \sum_{k=0}^L P_{opt}^{(k)} z^{-k}$. Then, by [12, Lem. 5], $T_{opt}(z) = -\alpha_{opt}[CX + C(zI - A)^{-1}BX]$, where $X := \sum_{k=1}^{L+1} A^{k-1} BP_{opt}^{(L+1-k)}$ simply is a complex $q \times q$ matrix. Note that this in particular shows that the McMillan degree (see, e.g., [11]) of the temporal FBF is independent of the latency time. This is in contrast to the FIR FFF whose McMillan degree Lq obviously grows with the latency time.

The assumption $H_{k_0} \neq 0$: Similar arguments as for linear precoding [13] show that precoding requires that the first arrival delay of the channel is not larger than the latency time. Otherwise, one can show that $P_{opt} = 0_{p \times q}$, $T_{opt} = F_{opt} = 0_{q \times q}$ and $\alpha_{opt} = 0$ solve Problem 1 if $H_k = 0$ for all $k \in \{0, \ldots, L\}$.

Influence of L: Similar arguments as for linear precoding [13] show that the MSE is non-increasing in the latency time.

Choice of Π*:* In general, the optimal permutation matrix Π can only be found by exhaustive search, i.e., we have to compute the MMSE of $\Pi = \sum_{k=1}^{q} e_k e_{\pi(k)}^*$ for all possible permutations $\pi : \{1, \ldots, q\} \rightarrow \{1, \ldots, q\}$ and then use the Π which gave the best MSE. (Here, e_k denotes the kth column of I_q.) The exhaustive search becomes very costly with growing q because there are $q!$ possible permutations. However, note that suboptimal schemes with lower complexity often give close-to-optimal performance [7].

IV-C. Fast computation

The main computational complexity in Theorem 2 arises in the determination of the $u_1, \ldots, u_q$. When implemented naively, the computational cost is $\mathcal{O}(qp^3(L+1)^3)$, which can be quite expensive. In contrast, we propose Alg. 1, which, if implemented correctly, has the much lower complexity of $\mathcal{O}(p^3 L^2)$.

Let us give some remarks on algorithm and implementation.

- The correctness of the algorithm is a simple consequence of

$$RR^* = V^*V = M^* \begin{bmatrix} I_{Lq} & \\ & 0_q \end{bmatrix} M + \frac{\sigma_\eta^2 \|\Psi\|_2^2}{E_{tr}} I$$

  and the Sherman-Morrison-Woodbury formula [14, §2.1.3].
- The only computationally demanding task in Alg. 1 is the Cholesky decomposition $RR^* = V^*V$. Note that V is a full rank block Toeplitz matrix. Thus, the results in [15], [16] can be applied to compute the decomposition in $\mathcal{O}(p^3 L^2)$.
- If the channel H is FIR, we could in principle also use the approach in [7] because we have seen that the optimal filters then will also be FIR. However, this may be less efficient because no similarly fast implementation (i.e., with quadratic complexity in the latency time) of the results in [7] is known.
- The linear equation system in Step 3b) should be solved with forward and backward substitution [14, §3.1].
- The low rank of the matrix W in Step 3c) should be exploited.

V. NUMERICAL EXAMPLE

In this section, we illustrate a case where the IIR channel model has advantages over the FIR channel model. We consider 2×2 IIR channels of the form

$$H(z) = \frac{H_0 + H_1 z^{-1} + H_2 z^{-2}}{1 - z/0.95}$$

where the real and imaginary parts of the entries of each H_k are normally distributed random variables with mean zero and variance one. The pole at 0.95 ensures that the channel has a long impulse response although it has a low McMillan degree. Typical examples for such "pole-zero channels" are, e.g., DSL lines [8], [9]. The standard approach to such a channel would be to use a FIR approximation of the channel. The FIR THP filters would then be derived from this FIR approximation. We will compare this approach with our direct IIR solution.

In our simulations, we chose $E_{tr} = 1$, $\tau = 2\sqrt{2}$, $\sigma_v^2 = \frac{\tau^2}{6}$, $L = 6$, $\Pi = I$. We transmitted 1000 QPSK modulated data vectors per channel realization, i.e., $\frac{1}{\sqrt{2}} s_k \in \{1 \pm i, -1 \pm i\}^2$, and averaged the resulting uncoded bit error rates over 1000 realizations for various signal to noise ratios $\sigma_v^2 E_{tr} \sigma_\eta^{-2}$. We computed the optimal THP as given in Theorem 2 for the FIR approximations of the channel

---

Algorithm 1 Fast computation of the $u_1, \ldots, u_q$ in Theorem 2

Input: $\quad M$ in Theorem 2, $H_0, \ldots, H_L$, $\sigma_\eta \|\Psi\|_2 E_{tr}^{-1/2}$
Output: $\quad u_1, \ldots, u_q$ in Theorem 2

1) compute the Cholesky decomposition $RR^* = V^*V$, where

$$V := \begin{bmatrix} \frac{\sigma_\eta \|\Psi\|_2}{\sqrt{E_{tr}}} I_p & & & \\ & \ddots & & \\ & & \frac{\sigma_\eta \|\Psi\|_2}{\sqrt{E_{tr}}} I_p & \\ V_0 & & & \\ \vdots & \ddots & & \\ V_{L-1} & \cdots & & V_0 \end{bmatrix} \in \mathbb{C}^{(2L+1)p \times (L+1)p}$$

and, for $k \in \{0, \ldots, L-1\}$,

$$V_k := \begin{bmatrix} H_k \\ 0_{(p-q) \times q} \end{bmatrix} \in \mathbb{C}^{p \times p}.$$

2) $W \leftarrow 0_{(L+1)q}$
3) for $k = 1, \ldots, q$ do
   a) $v \leftarrow M^* e_{Lq+k}$
      where e_j is the jth column of $I_{(L+1)q}$
   b) solve $RR^* u = v$ for u
   c) $w \leftarrow u - Wv$
   d) $W \leftarrow W + \frac{ww^*}{1+v^*w}$
   e) $u_k \leftarrow w - \frac{ww^*}{1+v^*w} v$

H with orders $N \in \{20, 40, 60, 80, 100\}$. The results are shown in Fig. 3. Note that only the approximations $N \in \{80, 100\}$ give a reasonable performance. However, while the FFF always simply is FIR of order $L = 6$, the temporal FBFs become very complicated in these cases: they have order N. We have also computed the optimal IIR THP for the exact IIR channel using the method in Sec. IV-B. In contrast to the FIR cases we now obtain optimal performance *and* a low complexity FBF at the same time. (The FFF is the same in all cases.)

We point out that similar observations have been made for the DFE, where Crespo and Honig showed that IIR FBFs may be of much lower complexity than FIR FBFs [9].

VI. CONCLUSION

We have derived the optimal spatio-temporal IIR Tomlinson Harashima Precoder, both for FIR and IIR channels. We found that the optimal feedforward filter is always FIR, while the optimal temporal feedback feedback filter is FIR if the channel is FIR. We also derived a fast algorithm to compute the optimal filters, which is about an order of magnitude faster than the naive implementation. A numerical example illustrated potential complexity reductions compared to the conventional FIR approach.

VII. REFERENCES

[1] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electron. Lett.*, vol. 7, no. 5/6, pp. 138–139, 1971.
[2] H. Harashima and H. Miyakawa, "Matched-Transmission Technique for Channels With Intersymbol Interference," *IEEE Trans. Commun.*, vol. COM-20, pp. 774–780, Aug. 1972.

[3] J. M. Cioffi and G. P. Dudevoir, "Data transmission with mean-square partial response," in *Proc. IEEE GLOBECOM*, (Dallas, TX), pp. 1687–1691, 1989.

[4] G. Ginis and J. M. Cioffi, "A multi-user precoding scheme achieving crosstalk cancellation with application to DSL systems," in *Proc. Asilomar Conf. Sig. Syst. Comp.*, (Pacific Grove, CA), pp. 1627–1631, 2000.

[5] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, "Space-Time Transmission using Tomlinson-Harashima Precoding," in *Proc. ITG Conf. Source Channel Coding*, (Berlin, Germany), 2002.

[6] J. Liu and A. Duel-Hallen, "Tomlinson-harashima transmitter precoding for synchronous multiuser communications," in *Proc. Conf. Inf. Sci. Syst.*, (Baltimore, MD), Mar. 2003.

[7] M. Joham, D. A. Schmidt, J. Brehmer, and W. Utschick, "Finite-length mmse tomlinson-harashima precoding for frequency selective vector channels," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 3073–3088, 2007.

[8] J. S. Chow, J. C. Tu, and J. M. Cioffi, "A discrete multitone transceiver system for HDSL applications," *IEEE J. Sel. Areas Commun.*, vol. 9, pp. 895–908, Aug. 1991.

[9] P. M. Crespo and M. L. Honig, "Pole-zero decision feedback equalization with a rapidly converging adaptive IIR algorithm," *IEEE J. Sel. Areas Commun.*, vol. 9, no. 6, pp. 817–829, 1991.

[10] C. Tidestav, A. Ahlen, and M. Sternad, "Realizable MIMO decision feedback equalizers: structure and design," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 121–133, 2001.

[11] P. J. Antsaklis and A. N. Michel, *Linear Systems.* Boston: Birkhäuser, 2006.

[12] S. Wahls and H. Boche, "Efficient Computation of the Realizable MIMO DFE," in *Proc. IEEE ICASSP*, (Dallas, TX), Mar. 2010.

[13] S. Wahls and H. Boche, "Linear IIR-MMSE Precoding for Frequency Selective MIMO Channels," in *Proc. IEEE ICASSP*, (Prague, Czech Republic), May 2011. accepted.

[14] G. Golub and C. F. Van Loan, *Matrix Computations.* Baltimore: Johns Hopkins University Press, 3rd ed., 1996.

[15] D. Kressner and P. Van Dooren, "Factorizations and linear system solvers for matrices with Toeplitz structure." SLICOT Working Note 2000-2 (see http://www.slicot.org), 2000.

[16] T. Kailath and A. H. Sayed, eds., *Fast Reliable Algorithms for Matrices with Structure.* Philadelphia, PA: SIAM, 1999.

[17] B. Hassibi, A. H. Sayed, and T. Kailath, *Indefinite-Quadratic Estimation and Control.* Philadelphia: SIAM, 1999.

APPENDIX

Proof: (of Theorem 2) *Ansatz*: Our approach is as follows. We first show that for any P and α the optimal T and F are trivial to find, and that when they are applied we obtain a simpler expression for the MSE which only depends on P and α. We then find the optimal P and α which minimize this semi-optimal MSE. Finally, the according T and F can be obtained from the relations in the first part of the proof. We can only sketch the proof because of space limitations.

Optimal T and F: We have $\hat{d} = \alpha HPv + \alpha\Psi\eta$ and $v = \Pi d + Fv + Tv \Leftrightarrow d = \Pi^*(I - F - T)v$. The stochastic interpretation of the energy norm (see, e.g., [17, 10.3.1]) together with $\Pi^{-1} = \Pi^*$
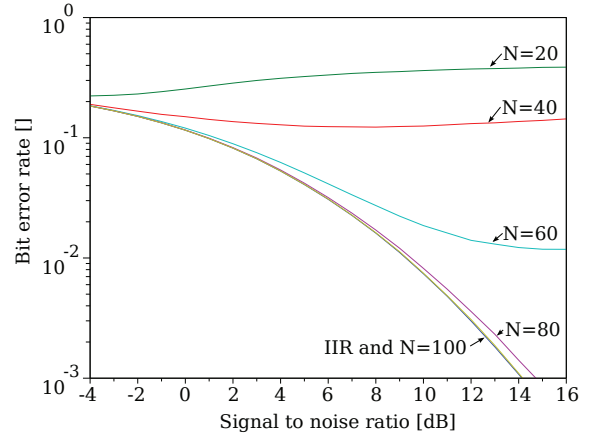

Fig. 3. Numerical Example

allows us to rewrite the MSE in Problem 1 as $e = \sigma_v^2\|z^{-L}(I - F - T) - \alpha\Pi HP\|_2^2 + \alpha^2\sigma_\eta^2\|\Psi\|_2^2$. The fact that the energy norm $\|\cdot\|_2$ decomposes spatially and temporally can be used to show

$$e = \sigma_v^2 \left\| E - \begin{bmatrix} 0 \\ \vdots \\ 0 \\ F \end{bmatrix} \right\|_F^2 + \sigma_v^2 \sum_{k=0}^{\infty} \|-T_k - N_{k+L+1}\|_F^2 + \alpha^2\sigma_\eta^2\|\Psi\|_2^2,$$

$$E := \begin{bmatrix} 0 & \dots & 0 & I_q \end{bmatrix}^* - \alpha M \begin{bmatrix} P_0^* & \dots & P_{L-1}^* & P_L^* \end{bmatrix}^*,$$

where M was defined in (1) and $\{N_k\}_{k\in\mathbb{N}}$ is the impulse response of ΠHP. This representation implies that the choices $T(z) = -\sum_{k=L+1}^{\infty} N_k z^{-k}$ and $F = $ strictly lower triangular part of $-N_L$ are optimal for all P and α.

Semi-optimal MSE: When we insert the optimal T and F, we obtain the semi-optimal MSE $e_s := \sigma_v^2\sum_{k=1}^{q}\|R_{q-k}(v_k - \alpha M u_k)\|_F^2 + \alpha^2\sigma_\eta^2\|\Psi\|_2^2$, where we have vectorized $[u_1, \dots, u_q] := [P_0^*, \dots, P_L^*]^*$ and $[v_1, \dots, v_q] := [0, \dots, 0, I_q]^*$, and introduced $R_k := [I_{(Lq+q-k)}, 0_{(Lq+q-k)\times(L+1)q}]$ to model the optimal spatial FBF. We can rewrite the semi-optimal MSE as

$$e_s = \sigma_v^2 \left\| \begin{bmatrix} R_{q-1}v_1 \\ \vdots \\ R_0 v_q \end{bmatrix} - \alpha M_{large} \begin{bmatrix} u_1 \\ \vdots \\ u_q \end{bmatrix} \right\|_F^2 + \alpha^2\sigma_\eta^2\|\Psi\|_2^2,$$

$$M_{large} := \text{blockdiag}(R_{q-1}M, \dots, R_1 M, R_0 M).$$

Optimal P and α: Note that the semi-optimal MSE e_s is very similar to the MSE which arises in linear precoding [13]. Thus, after some modifications, the optimal P that minimizes e_s subject to the power constraint $\sigma_v^2\|P\|_2^2 \leq E_{tr}$ is found to be

$$P_{opt} = \alpha_{opt}^{-1} \begin{bmatrix} z^{-0}I_p & \dots & z^{-L}I_p \end{bmatrix} \begin{bmatrix} u_1 & \dots & u_q \end{bmatrix},$$

$$\begin{bmatrix} u_1 \\ \vdots \\ u_q \end{bmatrix} := \left(M_{large}^* M_{large} + \frac{\sigma_\eta^2\|\Psi\|_2^2}{E_{tr}}I \right)^{-1} M_{large}^* \begin{bmatrix} R_{q-1}v_1 \\ \vdots \\ R_0 v_q \end{bmatrix}$$

and $\alpha_{opt} := \sqrt{E_{tr}}\sigma_v^{-1}\|[u_1, \dots, u_q]\|_F^{-2}$. These expressions reduce to (3) and (2) because M_{large} is block diagonal. $\blacksquare$