Supporting Information

Advancing vapor-deposited perovskite solar cells via machine learning

Jiazheng Wang [a], Yuchen Qi [a], Haofeng Zheng [a], Ruilong Wang [a], Siyou Bai [a], Yanan Liu [a], Qi Liu [a], Jin Xiao [a], Dechun Zou [b], and Shaocong Hou [a]*

[a.] *School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China; Wuhan Unversity Shenzhen Research Institute. E-mail: sc.hou@whu.edu.cn**

[b.] *Beijing National Laboratory for Molecular Sciences, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871*

Datasets used in this work

Article	Method	Substrate	HTL	HTL-2	Perovskite	ETL	ETL-2	Back contact	Annealing temperature/°C	Pressure/mbar	thickness/nm	Jsc/[mA cm−2]	Voc/V	FF	PCE (best) %	Structure(Normal/Inverse)
2017 IEEE Regional Symposium on Micro and Nanoelectronics (RSM). IEEE, 2017: 195-198.	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	TiO2	0	Au	70	10^-6	no	16.57	0.8	0.65	7.75	Normal
ACS Appl. Electron. Mater. 2021, 3, 3023−3033	Coevaporation	ITO	NiO	0	CsPbI3	TiO2	0	Al	335	10^-6	no	14.75	0.93	no	8.85	Normal
ACS Appl. Energy Mater. 2020, 3, 1476−1483	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	mp-TiO2	ZnAl-MMOs	Au	no	10^-6	no	22.51	1.06	0.78	18.54	Normal
ACS Appl. Energy Mater. 2020, 3, 1476−1483	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	mp-TiO2	0	Au	no	no	no	19.9	0.9	0.76	13.59	Normal
ACS Appl. Energy Mater. 2020, 3, 2350−2359	Sequential Deposition	FTO	Spiro-OMeTAD	0	MAPbI3	c-TiO2	0	Ag	180	no	no	22.4	0.877	0.595	11.7	Normal
ACS Appl. Energy Mater. 2021, 4, 4333−4343	Sequential Deposition	ITO	Spiro-OMeTAD	0	MAPbI3	C60	0	Au	100	no	230	17	0.927	0.654	10.3	Normal
ACS Appl. Energy Mater. 2022, 5, 8049−8056	Sequential Deposition	FTO	0	0	CsPbBr3	SnO2	0	Carbon	300	no	no	7.73	1.467	no	9.41	Normal
ACS Appl. Mater. Interfaces 2015, 7, 25770−25776	Sequential Deposition	FTO	Spiro-OMeTAD	0	MAPbI3	TiO2	0	Ag	110	no	100	21.26	0.962	0.66	13.5	Normal
ACS Appl. Mater. Interfaces 2018, 10, 26293−26302	Coevaporation	FTO	Spiro-OMeTAD	0	MAPbI3	PCBM	TiO2	Au	100	10^-6	500	21.3	1.007	0.75	17.1	Normal

ACS Appl. Mater. Interfaces 2018, 10, 26293−26302	Coevaporation	FTO	Spiro-OMeTAD	0	MAPbI3	TiO2	0	Au	100	10^-6	380	9.7	0.965	0.63	5.9	Normal
ACS Appl. Mater. Interfaces 2019, 11, 28851−28857	Dual source	FT0	Spiro-OMeTAD	0	MAPbI3	C60	0	Ag	no	10^-5	no	20	0.98	0.8	15.7	Normal
ACS Appl. Mater. Interfaces 2020, 12, 20456−20461	Single source	FT0	Spiro-OMeTAD	0	MA0.25FA0.75PbI2.75Br0.25	TiO2	SnO2	Au	27	no	no	23.4	1.04	0.726	17.2	Normal
ACS Appl. Mater. Interfaces 2020, 12, 20456−20461	Single source	FT0	Spiro-OMeTAD	0	MAPbI3	TiO2	SnO2	Au	27	no	500	20.3	1.04	0.764	16.2	Normal
ACS Appl. Mater. Interfaces 2020, 12, 22730−22740	Sequential deposition	ITO	NiO	0	MAPbI3	PCBM	PEI	Ag	50	no	no	20.1	1.04	0.751	15.7	Normal
ACS Appl. Mater. Interfaces 2020, 12, 22730−22740	Sequential deposition	ITO	NiO	0	FACsPbX3	PCBM	PEI	Ag	100	no	no	21.6	0.95	0.744	15.2	Normal
ACS Appl. Mater. Interfaces 2020, 12, 39261−39272	Coevaporation	ITO	MeO-2PACz	0	MAPbI3	C60	BCP	Cu	40	10^-6	750	22.43	1.15	0.796	20.5	Inverse
ACS Appl. Mater. Interfaces 2021, 13, 20034−20042	Sequential deposition	ITO	Spiro-OMeTAD	Cu2O	CsPbBr3	ZnO	0	Au	150	no	250	5.75	1.34	0.737	5.67	Normal
ACS Appl. Mater. Interfaces 2021, 13, 20034−20042	Sequential deposition	ITO	Spiro-OMeTAD	0	CsPbBr3	ZnO	0	Au	150	no	250	5.46	1.3	0.7044	5	Normal
ACS Appl. Mater. Interfaces 2015, 7, 44, 24726−24732	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	mp-TiO2	bl-TiO2	Ag	no	10^-6	280	20.17	1.03	0.72	15.6	Normal
ACS Appl. Mater. Interfaces 2020, 12, 45, 50684−50691	Coevaporation	FTO	NiOx	0	FA0.2MA0.8PbIxCl3−x	PCBM	0	Ag	no	10^-6	160	21.19	1.12	0.786	18.65	Inverse
ACS Appl. Mater. Interfaces 2020, 12, 50, 55830−55837	Sequential deposition	ITO	PC61BM	BCP	FAPbI3	NiOx	0	Ag	150	10^-6	no	23.29	1.04	0.76	18.41	Normal

ACS Energy Lett. 2017, 2, 2799−2804	Dual source	FT0	Spiro-OMeTAD	0	FAPbI3	C60	0	Ag	170	no	370	22.1	1.01	no	15.8	Normal
ACS Energy Lett. 2018, 3, 214−219	Coevaporation	ITO	TaTm	TaTm: F6-TCNNQ	MAPb (Br0.5I0.5)3	C60:PhIm	C60	Au	90	no	250	11.4	1.207	0.769	10.6	Normal
ACS Energy Lett. 2018, 3, 214−219	Coevaporation	ITO	TaTm	TaTm: F6-TCNNQ	MAPb (Br0.2I0.8)3	C60:PhIm	C60	Au	90	no	250	17.3	1.12	0.823	15.9	Normal
ACS Energy Lett. 2018, 3, 214−219	Coevaporation	ITO	TaTm	TaTm: F6-TCNNQ	MAPbI3	C60:PhIm	C60	Au	90	no	500	19.6	1.095	0.813	17.5	Normal
ACS Energy Lett. 2019, 4, 2748−2756	Coevaporation	ITO	PEDOT:PSS	0	(FA1−xCsx)(Sn1−yPby)I3	PCBM	BCP	Ag	450	5×10^-6	no	20.29	0.78	0.74	11.48	Inverse
ACS Energy Lett. 2020, 5, 2498−2504	Coevaporation	ITO	PTAA	0	FA0.7Cs0.3Pb (I0.9Br0.1)3	PCBM	BCP	Ag	150	no	10	23	1.06	0.746	18.2	Inverse
ACS Energy Lett. 2020, 5, 710.	Dual source	FT0	Spiro-OMeTAD	0	MAPbI3	C60	0	Au	no	no	490	21.7	1.08	0.778	18.2	Normal
ACS Energy Lett. 2021, 6, 827−836	Coevaporation	ITO	MoO3	TaTm	FA(1–n)CsnPb (I1–xBrx)3	C60	BCP	Ag	no	no	no	18	1.184	0.79	16.8	Inverse
ACS Energy Lett. 2022, 7, 1903−1911	Coevaporation	ITO	PTAA	0	FA1−yCsyPb(I1−xClx)3	C60	BCP	Ag	135	no	no	23	1.06	0.79	19.3	Inverse
ACS Omega 2017, 2, 4464−4469	Sequential Deposition	FT0	P3HT	0	CsPbI3	TiO2	0	Ag	350	4×10^-6	200	12.06	0.71	0.67	5.71	Normal
Adv. Energy Mater. 2014, 4, 1400345	Dual source	ITO	PEDOT:PSS	poly-TPD	MAPbI3	PCBM	0	Au	no	10^-6	300	18.2	1.09	0.75	14.8	Inverse

Adv. Energy Mater. 2014, 4, 1400345	Coevaporation	ITO	PEDOT:PSS	0	MAPbI3	PCBM+C60	0	Au	no	10^-6	300	19.5	1	0.52	10	Inverse
Adv. Energy Mater. 2014, 4, 1400345	Coevaporation	ITO	PEDOT:PSS	0	MAPbI3	0	0	Au	no	10^-6	300	12.4	0.6	0.58	4.6	Inverse
Adv. Energy Mater. 2016, 6, 1502087	Sequential deposition	FTO	0	0	MAPbI3	TiO2	0	carbon	100	no	500	21.27	1.04	0.65	14.38	Normal
Adv. Energy Mater. 2016, 6, 1502202	Coevaporation	FTO	0	0	CsPbIBr2	TiO2	0	Au	250	10^-6	550	8.7	0.959	0.56	4.7	Normal
Adv. Energy Mater. 2016, 6, 1502202	Dual source	FTO	0	0	CsPbIBr2	c-TiO2	0	Au	250	10^-6	97	8.7	0.956	0.56	4.7	Normal
Adv. Energy Mater. 2018, 8, 1703054	Sequential deposition	ITO	PTAA	0	MAPbI3	PCBM	0	Al	100	10^-6	500	22.97	1.04	0.7	16.7	Inverse
Adv. Energy Mater. 2018, 8, 1703054	Sequential deposition	ITO	PEDOT:PSS	0	MAPbI3	PCBM	0	Al	100	10^-6	no	20.1	0.88	0.79	14	Inverse
Adv. Energy Mater. 2018, 8, 1703506.	Coevaporation	ITO	TaTm	TaTm: F6-TCNNQ	Cs0.5FA0.4MA0.1Pb(I0.83Br0.17)3	C60:PhIm	C60	Au	no	no	430	17	1.146	0.82	16	Normal
Adv. Energy Mater. 2018, 8, 1703506.	Coevaporation	ITO	TaTm	TaTm: F6-TCNNQ	Cs0.5FA0.5Pb(I0.83Br0.17)3	C60:PhIm	C60	Au	no	no	430	18.7	0.922	0.56	9.7	Normal
Adv. Energy Mater. 2019, 9, 1802995	Dual source	ITO	NiOx	0	MAPbI3	C60	0	Au	100	10^-5	400	21.4	1.08	0.72	15.4	Inverse
Adv. Energy Mater. 2019, 9, 1900555	Dual source	ITO	PTAA	0	CsPbI3	C60	BCP	Cu	no	10^-6	500	17.8	0.96	0.73	12.5	Inverse
Adv. Energy Mater. 2021, 11, 2100299	Coevaporation	ITO	PTAA	0	CsPbI3+10%PEAI	PCBM	BCP	Ag	100	10^-6	500	17.33	1.09	0.7752	14.62	Inverse

Adv. Energy Mater. 2021, 11, 2101460	Coevaporation	ITO	MeO-2PACz	0	MAxFA1-xPbI3	C60	BCP	Cu	145	10^-6	750	24.26	1.04	0.7632	19.2	Inverse
Adv. Eng. Mater. 2020, 2000990	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	SnO2	0	Au	200	<10^-5	no	18.47	0.98	0.74	13.51	Normal
Adv. Eng. Mater. 2020, 2000990	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	SnO2	0	Au	250	<10^-5	no	17.7	0.98	0.67	11.73	Normal
Adv. Eng. Mater. 2020, 2000990	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	SnO2	0	Au	150	<10^-5	no	17.28	0.95	0.65	10.73	Normal
Adv. Eng. Mater. 2020, 2000990	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	SnO2	0	Au	300	<10^-5	no	17.04	0.95	0.61	10.09	Normal
Adv. Eng. Mater. 2020, 2000990	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	SnO2	0	Au	100	<10^-5	no	17.23	0.89	0.43	6.67	Normal
Adv. Eng. Mater. 2020, 2000990	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	SnO2	0	Au	70	<10^-5	478	8.66	0.74	0.64	4.27	Normal
Adv. Funct. Mater. 2021, 31, 2103252	Coevaporation	ITO	PTAA	0	MAPbI3	PCBM	BCP	Ag	100	10^-5	no	22.3	1.134	0.807	20.41	Inverse
Adv. Funct. Mater. 2021, 31, 2103252	Coevaporation	ITO	Spiro-TTB	0	MAPbI3	PCBM	BCP	Ag	100	10^-5	160	22.05	1.131	0.812	20.27	Inverse
Adv. Funct. Mater. 2021, 31, 2103252	Coevaporation	ITO	MeO-2PACz	0	MAPbI3	PCBM	BCP	Ag	100	10^-5	750	22.3	1.121	0.824	20.61	Inverse
Adv. Mater. 2015, 27, 7221–7228	Sequential deposition	FTO	Spiro-OMeTAD	0	CH3NH3PbIyBrx	TiO2	0	Au	no	no	380	21.8	1.025	0.74	16.9	Normal

Adv. Mater. 2015, 27, 7221–7228	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	TiO2	0	Au	no	no	no	20.7	0.983	0.71	14.7	Normal
Adv. Mater. 2016, 28, 3653–3661	Sequential deposition	FTO	Spiro-OMeTAD	0	(IC2H4NH3)2(CH3NH3)n-1PbnI3n+1	TiO2	0	Au	100	no	no	14.88	0.883	0.69	9.03	Normal
Adv. Mater. 2017, 29, 1605290	Dual source	ITO	TAPC	TAPC:MoO3	CsPbI2Br	Ca	C60	Ag	260	2×10^-6	400	15.2	1.15	0.67	11.8	Normal
Adv. Mater. 2017, 29, 1605290	Dual source	ITO	TAPC	TAPC:MoO3	CsPbI3	Ca	C60	Ag	325	2×10^-6	300	12.5	1.13	0.63	9.4	Normal
Adv. Mater. 2018, 30, 1800855	Dual source	ITO	Spiro-OMeTAD	0	CsPbBr3	ZnO	0	Au	180	no	450	7.01	1.44	0.77	7.78	Normal
Adv. Mater. 2020, 1907769	Sequential deposition	ITO	Spiro-OMeTAD	0	FA1-xMAxPbI3	SnO2	0	Ag	90	no	no	24.6	1.096	0.775	20.9	Normal
Adv. Mater. 2020, 2007126	Sequential deposition	ITO	Spiro-OMeTAD	0	FAPbI3	SnO2	0	Au	145	no	400	15.1	0.947	0.506	7.2	Normal
AIP Advances 8, 095226 (2018)	Sequential deposition	ITO	PTAA	0	MAPbI3	PCBM	Ti	Au	110	6×10^-6	no	21.14	1.019	0.749	16.13	Inverse
Chem. Mater. 2020, 32, 8641−8652	Coevaporation	ITO	MoO3	TaTm	CsPbIBr2	C60	BCP	Ag	150	10^-6	250	14.3	0.958	0.731	10	Inverse
Chemical Engineering Journal 448 (2022) 137676	Sequential deposition	ITO	Spiro-OMeTAD	0	(FAPbI3)0.97(MAPbBr3)0.03	SnO2	0	Ag	no	6×10^-6	500	23.53	1.1	0.7887	20.48	Normal
Chemical Engineering Journal, 2018, 336: 732-740.	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	ZnO	0	Au	80	no	no	19.31	1.038	0.7109	14.25	Normal
Coatings 2020, 10, 1163	Coevaporation	FTO	Spiro-OMeTAD	0	MAPbI3	TiO2	0	Au	100	10^-6	no	21.4	1.108	0.747	18.3	Normal

Coatings 2020, 10, 1163.	Dual source	FT0	Spiro-OMeTAD	0	MAPbI3	TiO2	0	Au	100	no	650	22.06	1.115	0.77	19	Normal
Dalton Trans., 2015, 44,17841	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	TiO2	0	Ag	70	no	no	21.14	0.939	0.7306	14.51	Normal
Dalton Trans., 2015,44, 3967-3973	Sequential deposition	FTO	0	0	NiO(MAPbI3)	TiO2	0	carbon	no	no	450	18.2	0.89	0.71	11.4	Normal
Dalton Trans., 2015,44, 3967-3973	Sequential deposition	FTO	0	0	ZrO2 (MAPbI3)	TiO2	0	carbon	no	10^-6	443	16.4	0.818	0.6	8.2	Normal
Electronic Materials Letters (2019) 15:56–60	Coevaporation	FTO	Spiro-OMeTAD	0	CsPbI2Br	TiO2	0	Au	300	10^-6	400	10.9	1.1	0.49	5.7	Normal
Electronic Materials Letters (2019) 15:56–60	Dual source	FTO	Spiro-OMeTAD	0	CsPbIBr2	TiO2	0	Au	300	10^-6	400	10.9	1.1	0.49	5.7	Normal
Energy Environ. Sci., 2016, 9, 3456--3463	Coevaporation	ITO	TaTm	TaTm:F6-TCNNQ	MAPbI3	C60:Phlm	C60	Ag	no	no	750	22.08	1.14	0.805	20.3	Normal
Energy Environ. Sci., 2022, 15, 1144	Sequential deposition	FTO	Spiro-OMeTAD	0	M2FAn-1PbnI3n+1	TiO2	0	Au	150	no	no	25.04	1.07	0.7198	19.33	Normal
Energy Environ. Sci.2014, 7, 3989.	Dual source	FT0	Spiro-OMeTAD	0	MAPbI3-xClx	TiO2	0	Ag	no	no	330	17	1.09	0.566	9.9	Normal
Energy Technol. 2019, 7, 1800986	Single-source	FT0	Spiro-OMeTAD	0	CsPbI2Br	TiO2	PCBM	Au	280	3×10^-5	no	15.4	1.07	0.72	12.2	Normal
Energy Technol. 2020, 8, 1900878	Coevaporation	FTO	Spiro-OMeTAD	0	MAPbI3	Nb2O5	0	Au	100	<2×10^-5	210	18.9	1.028	0.73	14.2	Normal
Energy Technol. 2020, 8, 1900878	Coevaporation	FTO	Spiro-OMeTAD	0	MAPbI3	PCBM	Nb2O5	Au	100	<2×10^-5	210	21.7	1.013	0.7	15.4	Normal

Energy Technol. 2020, 8, 1900878	Coevaporation	FTO	Spiro-OMeTAD	0	MAPbI3	PCBM	0	Au	100	<2×10^-5	210	18.3	0.967	0.6	10.8	Normal
Energy Technol. 2020, 8, 1900878	Coevaporation	FTO	Spiro-OMeTAD	0	MAPbI3	PCBM	TiO2	Au	100	<2×10^-5	210	19.9	0.965	0.69	13.3	Normal
Energy Technol. 2020, 8, 1900878	Coevaporation	FTO	Spiro-OMeTAD	0	MAPbI3	TiO2	0	Au	100	<2×10^-5	425	17.1	0.856	0.56	8.3	Normal
IEEE JOURNAL OF PHOTOVOLTAICS, VOL. 7, NO. 1, JANUARY 2017	Sequential Deposition	ITO	PEDOT:PSS	0	MAPbI3	PCBM	BCP	Ag	100	5.3×10^-6	no	11.8	0.89	no	7.46	Normal
Int J Energy Res. 2020;1–11.	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	TiO2	0	Ag	no	2.67×10^-5	660	21.2	1.2	0.76	19.43	Normal
J Mater Sci (2021) 56:15205–15214	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	ZnO	0	Au	no	no	650	22.2	1.082	0.74	17.75	Normal
J Mater Sci (2022) 57:1936–1946	Sequential deposition	FTO	Spiro-OMeTAD	0	MAxFA1-xPbI3(MA0.5FA0.5PbI3)	ZnO	c-TiO2	Ag	120	<10^-6	no	21.2	0.99	0.67	14.1	Normal
J. Alloys Compd. 2020, 818, 152903.	Single source	FT0	Spiro-OMeTAD	0	CsPbBr3	TiO2	0	Au	300	10^-5	650	7.79	1.37	0.81	8.65	Normal
J. Mater. Chem. A, 2015, 3, 23888–23894	Dual source	FT0	CuPc	0	MAPbI3	C60	0	Au	no	6.65×10^-5	370	18.91	1.04	0.78	15.33	Normal
J. Mater. Chem. A, 2015, 3, 9401–9405	Sequential Deposition	FT0	Spiro-OMeTAD	0	MAPbI3	c-TiO2	0	Au	120	10^-6	300	22.27	1	0.72	16.03	Normal
J. Mater. Chem. A, 2018, 6, 21143–21148	Sequential Deposition	FT0	Spiro-OMeTAD	0	Cs0.24FA0.76PbI3-yBry	SnOx	0	Au	160	10^-6	315	22.88	1.06	0.71	17.29	Normal
J. Mater. Chem. A, 2019, 7, 6920–6929	Sequential Deposition	FT0	Spiro-OMeTAD	0	Cs0.1FA0.9PbI2.9Br0.1	SnO2	C60	Au	150	10^-6	400	20.5	0.94	0.69	13.3	Normal

J. Mater. Chem. A,2016, 4, 5663.	Dual source	ITO	MoO3	NPB	MAPbI3	C60	0	Al	no	2×10^-6	550	20.1	1	0.65	14.5	Inverse
J. Mater. Chem. C, 2020, 8, 7725--7733	Coevaporation	ITO	Spiro-TTB	F6TCNNQ	Cs0.1FAxPbI2+xBr0.1	C60	0	Ag	100	10^-7	no	19.5	1.07	0.797	16.6	Inverse
J. Mater. Chem. C, 2020, 8, 7725--7733	Coevaporation	ITO	F6TCNNQ	Spiro-TTB	FA2+xCs0.1PbI2+xBr0.1	C60	0	Ag	100	10^-6	400	19.5	1.07	0.797	16.6	Inverse
J. Name., 2012, 00, 1-3 | 5	Sequential deposition	ITO	Spiro-OMeTAD	0	MAPbI3	SnO2	0	Ag	70	no	no	18.8	1.08	0.6	12.3	Normal
J. Name., 2013, 00, 1-3 | 1	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	c-TiO2	Er-TiO2	Au	120	no	no	20.08	1.03	0.6731	14.06	Normal
J. Name., 2013, 00, 1-3 | 3	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	TiO2	0	Ag	100	no	no	20.9	0.993	0.73	14.61	Normal
J. Phys. Chem. C 2017, 121, 19642−19649	Dual source	FT0	P3HT	0	CsPbI2Br	c-TiO2	0	Au	300	10^-6	230	11.5	1.01	0.67	7.7	Normal
J. Phys. Chem. C 2017, 121, 19642−19649	Coevaporation	FTO	Spiro-OMeTAD	0	CsPbI2Br	c-TiO2	0	Au	300	no	230	11.5	1.005	0.77	6.7	Normal
J. Phys. Chem. C 2021, 125, 23474−23482	Sequential deposition	ITO	PTAA	0	MAPbI3	PC61BM	BCP	Ag	no	2×10^-6	160	21.94	1.07	0.7724	19.64	Normal
J. Phys. Chem. C 2021, 125, 23474−23482	Sequential deposition	ITO	PTAA	0	MAPbI3	PC61BM	BCP	Ag	no	2×10^-6	250	20.53	1.02	0.7686	17.48	Normal
J. Phys. Chem. C 2021, 125, 23474−23482	Sequential deposition	ITO	Cs-doped VOx	0	MAPbI3	PC61BM	BCP	Ag	no	2×10^-6	660	20.71	0.91	0.7637	15.47	Normal
J. Phys. Chem. Lett. 2017, 8, 67−72	Dual source	FT0	P3HT	0	CsPbI3	c-TiO2	0	Au	320	no	no	13.8	0.1063	0.716	10.5	Normal
J. Phys. Chem. Lett. 2018, 9, 1041−1046	Dual source	ITO	TaTm	TaTm: F6-TCNNQ	MAPbI3	TiO2	C60	Au	no	10^-6	590	21.91	1.16	0.82	20.8	Normal

J. Phys. Chem. Lett. 2021, 12, 10106−10111	Sequential deposition	ITO	C60	BCP	E1G20/ PHSCN	PEDOT:PSS	0	Ag	no	6×10^-6	no	21.9	0.81	0.76	13.5	Normal
Joule 2020, 4, 1035.	Dual source	FT0	Spiro-OMeTAD	0	MAPbI3	SnO2	PCBM	Au	100	10^-5	300	23.3	1.12	0.777	20.28	Normal
Joule 4, 1035–1053, May 20, 2020	Coevaporation	FTO	Spiro-OMeTAD	0	MAPbI3	PCBM	SnO2	Au	100	10^-5	523	22.6	1.13	0.78	19.91	Normal
Joule 4, 1035–1053, May 20, 2020	Coevaporation	FTO	Spiro-OMeTAD	0	MAPbI3	PCBM	TiO2	Au	100	10^-5	no	21.8	1.11	0.7751	18.75	Normal
Journal of Alloys and Compounds 818 (2020) 152903	Single-source evaporation	FTO	Spiro-OMeTAD	0	CsPbBr3	c-TiO2	0	Au	300	no	650	7.79	1.37	0.81	8.65	Normal
Journal of Alloys and Compounds 864 (2021) 158793	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	TiO2	ZnO	Au	no	no	no	18.02	0.84	0.51	7.71	Normal
Journal of Alloys and Compounds 888 (2021) 161448	Dual source	FT0	Spiro-OMeTAD	0	MAPbI3	c-TiO2	m-TiO2	Au	70	1	375	22.2	1.07	0.66	15.6	Normal
Journal of Materials Science: Materials in Electronics (2019) 30:5487–5494	Coevaporation	FTO	PTAA	0	MAPbI3	PCBM	0	Al	30	7.3×10^-5	660	22	1.03	0.77	17.4	Inverse
Journal of Materials Science: Materials in Electronics (2019) 30:5487–5494	Coevaporation	FTO	PTAA	0	MAPbI3	PCBM	0	Al	50	1.2×10^-4	350	21.4	1	0.77	16.5	Inverse
Journal of Materials Science: Materials in Electronics (2019) 30:5487–5494	Coevaporation	FTO	PTAA	0	MAPbI3	PCBM	0	Al	75	2×10^-4	350	21.5	0.99	0.76	16	Inverse
Journal of Power Sources 443 (2019) 227269	Sequential Deposition	FT0	CuPc	0	CsPbBr3	c-TiO2	0	C	250	9×10^-6	500	7.59	1.328	0.752	7.58	Normal
Journal of Power Sources 520 (2022) 230783	Sequential deposition	FTO	NiO	0	CsxMA1-xPbI3	C60	0	Ag	100	4×10^-6	no	19.24	1.1	0.7212	15.26	Normal
Li et al., Sci. Adv. 8, eabo7422 (2022) 15 July 2022	Sequential deposition	FTO	Spiro-OMeTAD	0	Cs0.05PbI2.05-xClx(2-CF3-PEAI)	SnO2	0	Au	170	7×10−3	no	25.91	1.145	0.81	24.42	Normal

Materials 2019, 12, 1394; doi:10.3390/ma12091394	Sequential deposition	FTO	P3HT	0	MAPI	ZnO	0	Au	no	2×10^-5	300	13.2	1.22	0.6	9.4	Normal
Materials 2019, 12, 1394; doi:10.3390/ma12091394	Sequential deposition	FTO	P3HT	0	MAPI	ZnO	0	Au	140	2×10^-5	300	14	0.9	0.53	6.8	Normal
Materials Letters 169 (2016) 236–240	Sequential Deposition	FTO	Spiro-OMeTAD	0	MAPbIxCl3-x	TiO2	0	Au	80	no	no	14.64	1.01	0.6612	9.76	Normal
Materials Letters 292 (2021) 129623	Sequential deposition	FTO	Spiro-OMeTAD	0	CsFAMAPbIBr	Li-SnO2	0	Au	no	<10^-6	900	22.89	0.958	0.7303	16.01	Normal
Materials Science in Semiconductor Processing 148 (2022) 106839	Sequential Deposition	FTO	NiO	0	Cs0.14FA0.86Pb(BrxI1-x)3	C60	BCP	Ag	160	no	330	20.91	0.96	0.72	14.45	Inverse
Nano Energy (2016) 19, 88–97	Single source	ITO	PEDOT:PSS	0	BA2MA3Pb4I13	PCBM	0	Ag	no	6.65×10^-5	97	10.98	0.95	0.449	4.67	Inverse
Nano Energy 2016, 19, 88.	Sequential Deposition	FT0	Spiro-OMeTAD	0	MAPbI3	C60	0	Au	no	6.65×10^-5	no	18.9	1.1	0.754	15.7	Normal
Nano Energy 2019, 65, 104015.	Sequential Deposition	FT0	Spiro-OMeTAD	0	CsPbBr3	c-TiO2	0	Ag	350	no	300	9.78	1.498	0.7447	10.91	Normal
Nano Energy 48 (2018) 536–542	Sequential Deposition	FT0	Spiro-OMeTAD	0	Cs0.15FA0.85PbI2.85Br0.15	c-TiO2	0	Ag	160	10^-6	no	22.82	1.06	0.75	18.22	Normal
Nano Energy, 2018, 49: 109-116.	Sequential deposition	ITO	Spiro-OMeTAD	0	MAPbI3-Z057	ZnO	0	Ag	80	2.67×10^-5	220	19.96	1.09	0.7472	16.21	Normal
Nano Energy, 2018, 49: 109-116.	Sequential deposition	ITO	Spiro-OMeTAD	0	MAPbI3	ZnO	0	Ag	80	2.67×10^-5	220	19.66	1.06	0.6418	13.4	Normal
Nano-Micro Lett. (2022)14:79	Sequential deposition	ITO	Spiro-OMeTAD	0	PbI2:CsI	SnO2	0	Ag	135	8×10^-6	490	22.09	1.08	0.7601	18.13	Normal
Nanoscale, 2015, 7, 10699–10707	Sequential Deposition	ITO	PEDOT:PSS	0	MAPbI3	PCBM	BCP	Ag	50	10^-5	260	16.66	1.03	0.651	11.17	Inverse

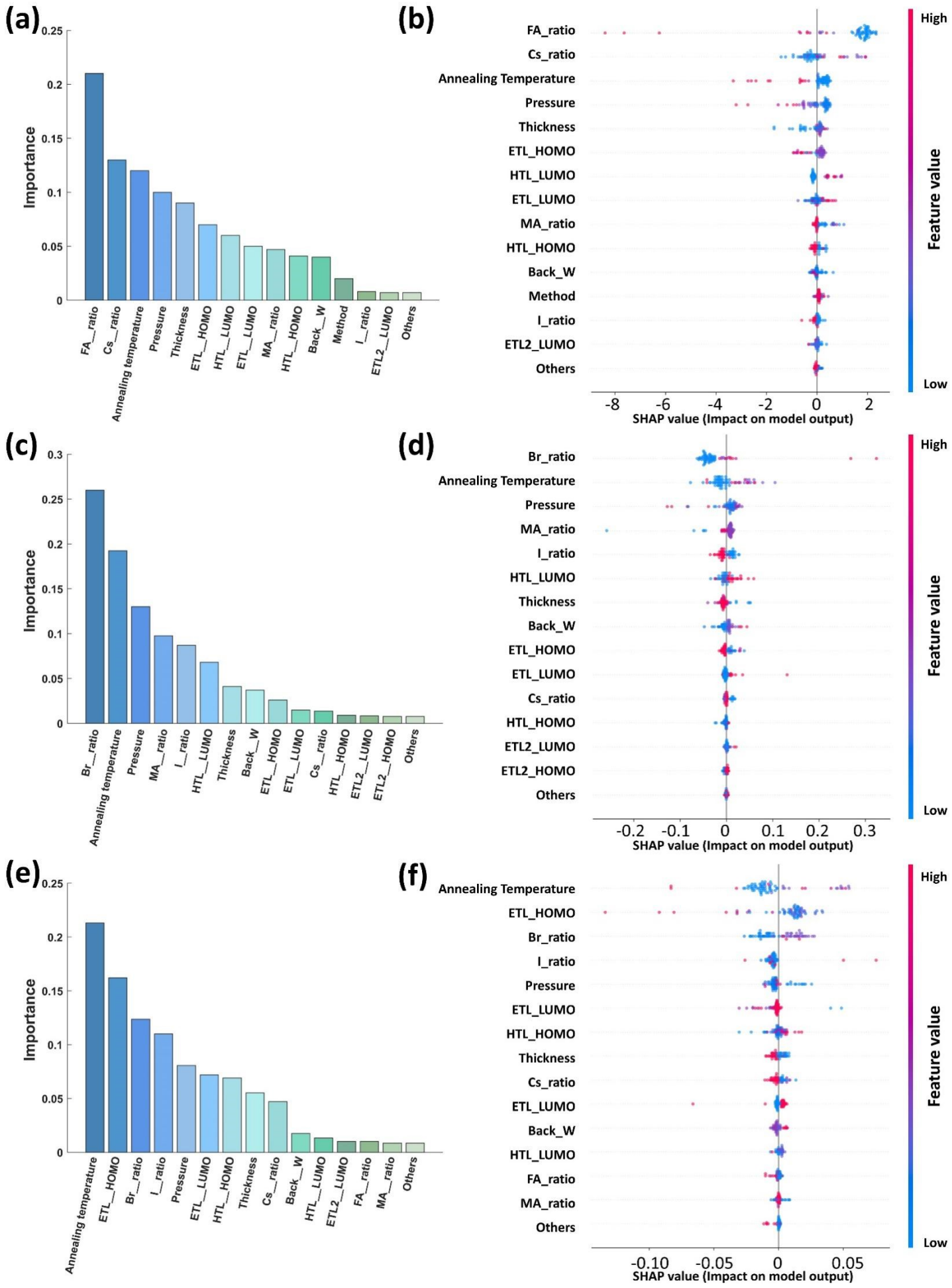
Nanoscale, 2015, 7, 10699–10707	Sequential Deposition	ITO	PEDOT:PSS	0	MAPbI3	PCBM	BCP	Ag	60	10^-5	260	17.12	1	0.621	10.63	Inverse
Nanoscale, 2015, 7, 10699–10707	Sequential Deposition	ITO	PEDOT:PSS	0	MAPbI3	PCBM	BCP	Ag	70	10^-5	260	18.58	0.98	0.603	10.97	Inverse
Nanoscale, 2017, 9, 12316–12323	Coevaporation	FT0	Spiro-OMeTAD	0	MAxCs1–xPbI3	TiO2	0	Au	150	6×10^-6	750	23.17	1.1	0.79	20.13	Normal
Nat. Photonics 2014, 8, 128.	Dual source	ITO	PEDOT:PSS	poly-TPD	MAPbI3	PCBM	0	Au	no	2×10^-6	285	16.12	1.05	0.67	12	Inverse
Nature 2013, 501, 395.	Dual source	FT0	Spiro-OMeTAD	0	MAPbI3-xClx	TiO2	0	Ag	100	10^-5	330	21.5	1.07	0.67	15.4	Normal
Organic Electronics 53 (2018) 26–34	Sequential deposition	ITO	PEDOT:PSS	0	MAPbI3(T-PbI2)	PC61BM	C70-bis	Al	100	2×10^-6	no	21.01	1.04	0.76	16.61	Inverse
Organic Electronics 53 (2018) 26–34	Sequential deposition	ITO	PEDOT:PSS	0	MAPbI3(P-PbI2)	PC61BM	C70-bis	Al	100	2×10^-6	no	15.27	1.01	0.65	10.02	Inverse
Organic Electronics 69 (2019) 208–215	Sequential deposition	FTO	Spiro-OMeTAD	0	FAxMA1–xPbI3	TiO2	0	Au	70	no	350	22.69	1.06	0.72	17.31	Normal
Organic Electronics 83 (2020) 105736	Sequential deposition	ITO	Spiro-OMeTAD	0	FAxMA(1-x)PbIyBr(3-y)	SnO2	0	Ag	150	no	330	16.8	1.09	0.64	11.9	Normal
Phys. Status Solidi RRL 2021, 15, 2000449	Sequential Deposition	ITO	PTAA	0	MAPbI3	C60	BCP	Ag	100	4×10^-6	no	23.11	1.09	0.77	19.4	Inverse
Results in Physics 17 (2020) 103122	Coevaporation	FTO	NiOx	0	MAPbI3	PCBM	BCP	Ag	no	6×10^-6	no	16.19	0.99	0.77	12.28	Inverse
RSC Adv. 2021, 11, 3380.	Single source	FT0	0	0	CsPbBr3	TiO2	0	Carbon	250	8×10^-6	no	9.27	1.36	0.71	8.95	Normal
RSC Adv., 2014, 4, 28964–28967	Sequential Deposition	ITO	0	0	MAPbI3	C60	0	Ag	100	no	350	13.6	0.8	0.5	5.4	Inverse
RSC Adv., 2016,6, 47459-47467	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	PCBM	TiO2	Ag	120	10^-6	600	19.8	1.01	0.773	15.59	Normal

RSC Adv.2020, 10, 8905	Sequential Deposition	FT0	Spiro-OMeTAD	0	CsPbBr3	TiO2	0	Ag	335	10^-5	600	6.49	1.42	0.79	7.22
S. Chae et al. / Materials Today Energy 14 (2019) 100341	Sequential deposition	ITO	P3HT	0	MAPbI3	PCBM	BCP	Ag	no	10^-6	no	19.198	0.91	0.7772	13.58
S. Chae et al. / Materials Today Energy 14 (2019) 100341	Sequential deposition	ITO	PEDOT:PSS	0	MAPbI3	PCBM	BCP	Ag	140	10^-6	no	16.985	0.84	0.7957	11.35
S. Prathapani et al. / Applied Materials Today 7 (2017) 112–119	Sequential deposition	FTO	Spiro-OMeTAD	0	MAPbI3	c-TiO2	mp-TiO2	Au	70	no	600	12.4	0.82	0.45	4.61
Sol. Energy Mater. Sol. Cells 2018, 187, 1.	Dual source	FT0	PCBM	0	CsPbBr3	TiO2	0	Ag	500	5×10^-5	600	6.97	1.27	0.785	6.95
Sol. Energy Mater. Sol. Cells 2020, 206, 110317.	Sequential Deposition	FT0	PCBM	0	CsPbBr3	TiO2	0	Carbon	300	no	600	7.37	1.545	0.822	9.35
Sol. RRL 2019, 3, 1900030	Sequential deposition	FTO	0	0	CsPbBr3/CsPbBr3-CsPb2Br5/CsPbBr3-Cs4PbBr6	c-TiO2	0	carbon	350	10^-6	300	9.24	1.461	0.7539	10.17
Sol. RRL 2019, 3, 1900030	Sequential deposition	FTO	0	0	CsPbBr3/CsPbBr3-CsPb2Br5	c-TiO2	0	carbon	350	10^-6	300	8.22	1.417	0.722	8.41
Sol. RRL 2019, 3, 1900030	Sequential deposition	FTO	0	0	CsPbBr3	c-TiO2	0	carbon	350	10^-6	300	7.11	1.372	0.7299	7.12
Sol. RRL 2019, 3, 1900050	Sequential Deposition	FTO	Spiro-OMeTAD	0	MAyFA1-yPbIxBr3-x	SnO2	0	Au	130	2.5×10^-3	400	19.16	1.02	0.773	15.14
Sol. RRL 2019, 3, 1900050	Sequential deposition	FTO	CuPc	0	MA0.56 FA0.44 PbI2.67 Br0.33	SnO2	0	Au	130	4×10^-6	400	19.16	1.02	0.773	15.14
Sol. RRL 2019, 3, 1900050	Sequential Deposition	FTO	Spiro-OMeTAD	0	MAPbIxBr3-x	SnO2	0	Au	130	2.5×10^-3	400	17.17	0.99	0.686	11.66
Sol. RRL 2020, 4, 1900283	Coevaporation	ITO	MoO3	PTAA	FA (Pb0.5Sn0.5)I3	C60	BCP	Ag	no	3×10^-6	385	24.3	0.73	0.733	13.98

Sol. RRL 2021, 5, 2000552	Sequential Deposition	ITO	CuPC	0	MAPbI3	C60	BCP	Ag	100	4×10^-6	175	23.34	1.108	0.785	20.3
Sol. RRL 2021, 5, 2100102	Coevaporation	FTO	Spiro-OMeTAD	0	Cs0.14FA0.86Pb(BryI1-y)3	SnO2	0	Au	180	no	450	22.53	1.041	0.74	17.3
Sol. RRL 2022, 6, 2100842	Coevaporation	FTO	Spiro-OMeTAD	0	MAPbI3	PCBM	ALD-SnO2	Au	100	2.8×10^-4	500	22.67	1.08	0.788	19.3
Sol. RRL 2022, 6, 2200020	Sequential deposition	FTO	0	0	CsPbI2Br	TiO2	0	carbon	250	no	no	15.7	1.312	0.74	15.24
Sol. RRL 2022, 6, 2200020	Sequential deposition	FTO	0	0	CsPbI3	TiO2	0	carbon	no	no	750	20.4	1.075	no	15.35
Solar Energy 178 (2019) 56–60	Sequential Deposition	FT0	Spiro-OMeTAD	0	FAxMA1–xPbI3	c-TiO2	0	Au	100	10^-6	no	22.4	0.98	0.73	15.8
Solar Energy 215 (2021) 179–188	Sequential Deposition	FTO	Spiro-OMeTAD	0	MAPbIxCl3-x	TiO2	0	Ag	130	10^-6	150	20.55	0.907	0.545	10.14
Solar Energy 215 (2021) 179–188	Sequential Deposition	FTO	Spiro-OMeTAD	0	MAPbI3	TiO2	0	Ag	130	10^-6	150	19.65	0.862	0.517	8.76
Solar Energy Materials and Solar Cells 187 (2018) 1–8	Coevaporation	FTO	Spiro-OMeTAD	0	CsPbBr3	c-TiO2	0	Au	500	5×10^-3	600	6.97	1.27	0.78	6.95
Solar Energy Materials and Solar Cells 187 (2018) 1–8	Coevaporation	FTO	Spiro-OMeTAD	0	CsPbBr3	c-TiO2	0	Au	400	5×10^-3	600	6	1	no	5.48
Solar Energy Materials and Solar Cells 187 (2018) 1–8	Coevaporation	FTO	Spiro-OMeTAD	0	CsPbBr3	c-TiO2	0	Au	550	5×10^-3	600	5.5	0.79	no	4.84
Solar Energy Materials and Solar Cells 206 (2020) 110317	Sequential Deposition	FTO	0	0	CsPbBr3	TiO2	0	Carbon	300	10^-6	600	7.37	1.545	0.822	9.35
Solar Energy Materials and Solar Cells 206 (2020) 110317	Sequential Deposition	FTO	0	0	CsPbBr3	TiO2	0	Carbon	250	10^-6	600	6.8	1.477	0.785	7.89

Feature Importance Chart and SHAP Chart

Figure S1. J_{SC} (a), V_{OC} (c), and FF (e) in order of feature importance. The effects of each feature on J_{SC} (b), V_{OC}(d) and FF(f) were analyzed using SHAP. Each bar in the left column of the graph indicates the absolute magnitude of importance of the feature, and the ranking is from the largest to the smallest absolute value of importance. Each point in the graph in the right column represents a data sample, the darker the color represents the larger the value of the feature, the horizontal coordinate is the SHAP value, and the vertical coordinate is the different features.


(a)
Importance
0.25
0.2
0.15
0.1
0.05
0
FA_ratio
Cs_ratio
Annealing temperature
Pressure
Thickness
ETL_HOMO
HTL_LUMO
ETL_LUMO
MA_ratio
HTL_HOMO
Back_W
Method
I_ratio
ETL2_LUMO
Others
(b)
FA_ratio
Cs_ratio
Annealing Temperature
Pressure
Thickness
ETL_HOMO
HTL_LUMO
ETL_LUMO
MA_ratio
HTL_HOMO
Back_W
Method
I_ratio
ETL2_LUMO
Others
-8
-6
-4
-2
0
2
SHAP value (Impact on model output)
High
Feature value
Low
(c)
Importance
0.3
0.25
0.2
0.15
0.1
0.05
0
Br_ratio
Annealing temperature
Pressure
MA_ratio
I_ratio
HTL_LUMO
Thickness
Back_W
ETL_HOMO
ETL_LUMO
Cs_ratio
HTL_HOMO
ETL2_LUMO
ETL2_HOMO
Others
(d)
Br_ratio
Annealing Temperature
Pressure
MA_ratio
I_ratio
HTL_LUMO
Thickness
Back_W
ETL_HOMO
ETL_LUMO
Cs_ratio
HTL_HOMO
ETL2_LUMO
ETL2_HOMO
Others
-0.2
-0.1
0
0.1
0.2
0.3
SHAP value (Impact on model output)
High
Feature value
Low
(e)
Importance
0.25
0.2
0.15
0.1
0.05
0
Annealing temperature
ETL_HOMO
Br_ratio
I_ratio
Pressure
ETL_LUMO
HTL_HOMO
Thickness
Cs_ratio
Back_W
HTL_LUMO
ETL2_LUMO
FA_ratio
MA_ratio
Others
(f)
Annealing Temperature
ETL_HOMO
Br_ratio
I_ratio
Pressure
ETL_LUMO
HTL_HOMO
Thickness
Cs_ratio
ETL_LUMO
Back_W
HTL_LUMO
FA_ratio
MA_ratio
Others
-0.10
-0.05
0
0.05
SHAP value (Impact on model output)
High
Feature value
Low

Mathematical part

$$r = \frac{\sum_{i=1}^{n}(\varphi_i - \overline{\varphi})(\theta_i - \overline{\theta})}{\sqrt{\sum_{i=1}^{n}(\varphi_i - \overline{\varphi})^2}\sqrt{\sum_{i=1}^{n}(\theta_i - \overline{\theta})^2}} \quad (1)$$

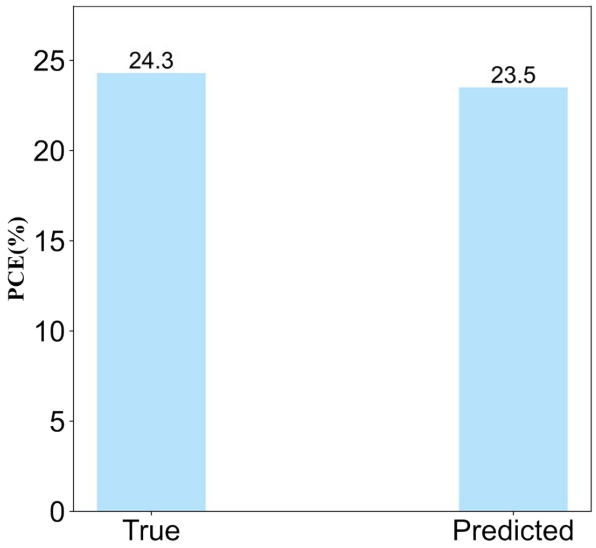
φ_i and θ_i are true value and predicted value of the i-th sample, $\overline{\varphi}$ and $\overline{\theta}$ are the mean value of true values and predicted values.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(true_i - predicted_i)^2} \quad (2)$$

N represents the volume of data.

Model Test

Figure S2. Comparative histogram of PCE values.



We verified our model with the most recent results from the just-accepted, unedited paper in Advanced Materials by L Tan, which is published during the peer-reviewing period and thus outside of our database. Our model predicts PCE value of 23.5%, very close to the experimental PCE value of 24.3%. The relative error is only 3.3%, which reflects the excellent prediction ability of our model for unknown samples and the excellent generalization performance of the model.

Machine learning model

RF is a classification algorithm, which belongs to the Bagging algorithm in ensemble learning, that is, the guided aggregation algorithm. RF consists of multiple decision trees, and each decision tree is different. When building a decision tree, we randomly select a part of the samples from the training data with replacement, and do not use all the features of the data, but randomly select parts features for training. And the final predicted value is determined by the prediction of the aggregation tree.

LR algorithm is a binary classification algorithm, and the decision function is represented by the conditional probability distribution P(Y|X). It non-linearly maps the operation result of linear regression to the [0,1] interval through the Sigmoid function, that is, the prediction category is expressed in the form of probability. LR is usually based on gradient descent algorithm to obtain training weights, and then uses decision function to make probabilistic prediction of categories.

DTree algorithm is based on the binary division strategy, and represents the result of data classification in a tree structure. Each decision point implements a test function with discrete output. DTree applies the information gain criterion on each node of the decision tree to select features, constructs the decision tree recursively, and the final fitted function is a step function between partitions.

GBoost uses the negative gradient of the loss function as an approximation of the residual in the boosting tree algorithm to train weak classifiers, thereby constructing a decision tree, and finally obtaining the predicted value according to the results of all weak classifiers. Each calculation of GBoost is to reduce the previous residual, and then build a new model in the direction of residual reduction.

ABoost is an adaptive algorithm. In the ABoost algorithm, the wrongly classified samples of the previous basic classifier will be given a higher weight, while the correctly classified samples will be given a lower weight, and the weighted samples will be used again to train the next basic classifier. At the same time, a new weak classifier is added in each round until a predetermined small enough error rate is reached or a pre-specified maximum number of iterations is reached.

XGBoost algorithm is similar to the GBoost algorithm, the difference is that GBoost reduces the residual at the fastest speed by continuously adding new trees, while XGBoost can define the loss function artificially, just need to know the first derivative and second derivative of the function for the parameters, which greatly improves the generalization ability of the model.

MLP is a deep learning algorithm in which each node is a perceptron, simulating the basic functions of neurons in biological neural networks, and these neurons are also divided into input layer, intermediate layer and output layer. A neural network trains the net with a feature vector as input, passes that vector to the hidden layer, then

computes the result through weights and activation functions, and passes the result to the next layer until the output. The weights, synapses and neurons of each layer are calculated and learned through the catenary ANN algorithm.

ETree is an algorithm very similar to RF and an improvement on Bagging. But ETree uses all of the samples, only the features are randomly selected. And ET obtains the bifurcation value completely randomly, while RF obtains the best bifurcation value within a random subset.

SVR is a dichotomous model whose purpose is to find a hyperplane to segment the samples. SVR sets a tolerance deviation on both sides of the linear function. All samples that fall into it do not calculate the loss, that is, only the support vector will affect its function model. Finally, it is obtained by minimizing the total loss and maximizing the total interval. out the optimized model.

LinearSVR is very similar to SVR, the only difference is that LinearSVR is a linear classification, does not support various low-dimensional to high-dimensional kernel functions, only supports linear kernel functions, and cannot be used for linearly inseparable data. But for linear data, LinearSVR does not need to adjust parameters to select the kernel function, and it is faster.

Genetic algorithm (GA)

Genetic algorithm (GA) is a computational model that simulates the biological evolution process of Darwinian genetic selection and natural elimination. It is a method to search for the optimal solution by simulating the natural evolution process. Starting from a population that represents the possible potential solutions of the problem, after genetic coding, the next generation is generated by operations such as replication, mutation and crossover, and individuals with low fitness function values are gradually eliminated to form a population. In this way, iteratively controls the search adaptively, and finally obtains the best population.

Data augmentation

In order to make the model get more sufficient data for training, we expanded the original dataset to generate a dataset with three times the size of the original dataset. The data expansion process is as follows:

1.Loading the original data set is recorded as L1, and creating a new storage space is recorded as L2.

2.Traverse each piece of data in L1, add random interference to the eigenvalues to form new samples, and add new samples to L2 and L3. Traverse each piece of data in L1, add random interference to the eigenvalues to form new samples, and add new samples to L2 and L3. In this step, we choose the form of Gaussian noise as the interference, and add the interference according to the features to ensure the accuracy of the model.

3.If the data in L2 and L3 space reaches the number of original data sets, go to step 4, otherwise return to step 2.

4.Randomly shuffle the order of data samples in L2 and L3 space, and store the original data to generate an expanded data set.

Code abstract

```
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import cross_val_score, GridSearchCV
from sklearn import metrics
from sklearn import datasets
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.neural_network import MLPRegressor
from sklearn.ensemble import AdaBoostRegressor
from sklearn.ensemble import ExtraTreesRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.svm import LinearSVR
from sklearn.svm import NuSVR
from sklearn.svm import SVR
from xgboost import XGBRegressor
from sklearn.metrics import accuracy_score
import numpy as np
import pandas as pd
import shap
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split

df = pd.read_csv()
features = df.iloc[:,1:].values
target= df.iloc[:,0].values
X_train, X_test, y_train, y_test
=train_test_split(features,target,test_size=0.2,random_state=0)
model=randomforest.fit(X_train, y_train)
y_pred = randomforest.predict(X_test)
yt = randomforest.predict(X_train)
importances = list(model.feature_importances_)
feature_name = list(df.columns)[1:]
```

```
feature_importances = [(feature, round(importance, 2))
for feature, importance in zip(feature_name, importances)]
feature_importances = sorted(feature_importances, key = lambda x: x[1], reverse = True)
```

```
estimator_list =[RandomForestRegressor(),
                  GradientBoostingRegressor(),
                 LinearRegression(),
                 DecisionTreeRegressor(),
                 MLPRegressor(solver='lbfgs'),
                 AdaBoostRegressor(),
                 ExtraTreesRegressor(),
                 LinearSVR(),
                 SVR(),
                 XGBRegressor(nthread = 15) ]
cv_split = ShuffleSplit(n_splits=3,test_size=0.2, random_state=18)
df_columns = ['Name', 'Parameters', 'Train Accuracy Mean', 'Test Accuracy Mean',
              'TestAccuracyStd','ComsumedTime']
df = pd.DataFrame(columns=df_columns)
```

GA:

```
if __name__ == '__main__':
        for i in range(generations):
            obj_value = cal_obj_value(pop)
            fit_value = calfitvalue(obj_value)
            [best_individual, best_fit] = best(pop, fit_value)
            temp_n_estimator, temp_max_depth = b2d(best_individual)
            results.append([best_fit,temp_n_estimator,temp_max_depth])
            selection(pop, fit_value)
            crossover(pop, pc)
             mutation(pop, pc)
        results.sort()
        print(results[-1])
```

PLOT:

```
fontsize=12
plt.figure(figsize=(3.5,3))
plt.style.use('default')
```

```
plt.rc('xtick', labelsize=fontsize)
plt.rc('ytick', labelsize=fontsize)
plt.rcParams['font.family']="Arial"
a = plt.scatter(y_train, yt, s=25,c='LightSkyBlue')
b = plt.scatter(y_test, y_pred, s=25,c='PaleGreen')
plt.tick_params(direction='in')
plt.legend((a,b),('Train','Test'),fontsize=fontsize,handletextpad=0.1,borderpad=0.1)
plt.rcParams['font.family']="Arial"
plt.plot([0.5,1.7],[0.5,1.7],"--",color="black")
plt.tight_layout()
```

SHAP:

```
model.fit(X_test,y_test)
explainer = shap.TreeExplainer(model)
shap_values = explainer.shap_values(X_test)
shap_explainer = explainer(X_test)
shap_explainer.base_values=shap_explainer.base_values[0][0]
shap_explainer.data=shap_explainer.data[]
shap_explainer.values=shap_explainer.values[]
shap.initjs()
shap.summary_plot(shap_values.values,X_test,show = False,alpha = 0.6,max_display
= 15)
```